Progress in Probability
77

Maria Eulália Vares
Roberto Fernández
Luiz Renato Fontes
Charles M. Newman
Editors

In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius

Birkhäuser

Birkhäuser

Progress in Probability

Volume 77

Progress in Probability is designed for the publication of workshops, seminars and conference proceedings on all aspects of probability theory and stochastic processes, as well as their connections with and applications to other areas such as mathematical statistics and statistical physics.

More information about this series at http://www.springer.com/series/4839

Maria Eulália Vares • Roberto Fernández •
Luiz Renato Fontes • Charles M. Newman
Editors

In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius

Birkhäuser

Editors
Maria Eulália Vares
Instituto de Matemática
Universidade Federal do Rio de Janeiro
Rio de Janeiro, RJ, Brazil

Roberto Fernández
NYU-ECNU Institute of Mathematical Sciences
at NYU Shanghai
Shanghai, China

Luiz Renato Fontes
Instituto de Matemática e Estatística
Universidade de São Paulo
São Paulo, SP, Brazil

Charles M. Newman
Courant Institute of Mathematical Sciences
New York University
New York
NY, USA

ISSN 1050-6977 ISSN 2297-0428 (electronic)
Progress in Probability
ISBN 978-3-030-60753-1 ISBN 978-3-030-60754-8 (eBook)
https://doi.org/10.1007/978-3-030-60754-8

Mathematics Subject Classification: 60-XX

This book is published under the imprint Birkhäuser, www.birkhauser-science.com by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Vlado Sidoravičiaus (1963–2019) atminimui.
Jis įkvėpė mus savo draugiškumu,
kūrybiškumu ir meile matematikai ir
gyvenimui.

In memory of Vladas Sidoravicius
(1963–2019).
He inspired us with his friendliness,
creativity, and love for mathematics and for
life.

Vladas Sidoravicius (courtesy of NYU-Shanghai; credit: Junbo Chen)

Preface

Vladas was born in Vilnius, Lithuania, on August 23, 1963, and did his undergraduate studies in Mathematics from 1982 to 1985 at Vilnius University. There, in 1986, he received a Master's degree with Honors under the supervision of Vygantas Paulauskas. While in Vilnius, his early career and interest for doing research in Mathematics benefited greatly from the mentorship of Donatas Surgailis. Pursuing research from the very beginning, Vladas moved to Moscow for the next 4 years and was awarded a Ph.D. from Moscow State University, under the supervision of Vadim Malyshev, with a dissertation on the convergence of the stochastic quantization method. At the VIth Vilnius Conference on Probability and Statistics, in 1990, Vladas gave what was probably his first presentation in an important international conference. In the meantime, he also had several other collaborators in Vilnius, already showing the vitality and initiative that were some of his characteristics. Vladas loved music and all expressions of fine art and always made clear that those years in Moscow offered him an extremely enriching experience in this aspect as well as for his mathematical development.

In 1991, Vladas had a postdoc experience at the University of Heidelberg, which he then continued for more than 1 year at the Université Paris Dauphine, working with the team of Claude Kipnis. This brought to his attention the existence of a probability research group in Brazil, which led to his arrival at IMPA, Rio de Janeiro, in February of 1993, where he held a position until 2015. While at IMPA, Vladas served as advisor to several PhD students, supervised a number of postdocs, and organized many meetings as well as remarkable conferences and schools. He always focused on offering challenging and stimulating opportunities to young researchers. As pointed out to us by Marco Isopi, this emphasis on supporting young scientists was something that Vladas and other postdocs of Claude Kipnis vowed to emulate following Kipnis' early death at age 43. Vladas made an immense contribution to the development of Probability in South America, particularly in Brazil.

A very important development in his scientific career began in 1995, when he made the first of many visits to Cornell University. It was the beginning of an extremely fruitful interaction with Harry Kesten, a towering figure in Probability Theory for six decades, who passed away shortly before Vladas. Not only did they write many joint papers, including a seminal work where a shape theorem without subadditivity was proven, but they also became very close friends. One anecdotal story has to do with the efforts made by Vladas to find a copy of Kesten's book *Percolation Theory for Mathematicians*, which was out of print. He searched by all possible methods until the day arrived when he somehow managed, with his usual soft and charmingly convincing attitude, to have one colleague make him a gift of his personal copy. That was a priceless gift, providing huge joy to Vladas.

Vladas' friendship with Harry Kesten extended into an approach to the Dutch stochastic community that resulted in a double appointment as researcher at the Centrum Wiskunde & Informatica (CWI) and visiting professor at Leiden University. During his tenure (2007–2011), Vladas developed an intensive research activity with the leaders of the main Dutch groups in probability, gave courses and seminars, and acted both as consultant and conference organizer at Eurandom. Vladas was a vocal supporter of this last institute, which he considered a model deserving emulation.

In 2015, Vladas became NYU Global Network Professor and was appointed Deputy Director of the Mathematical Institute at NYU-Shanghai. He quickly understood the immense potential of this new institution and invested in it all his energy and his capital of scientific networking. His enthusiasm and dedication helped to construct a remarkable Institute characterized by a continuous flow of distinguished visitors and an intense scientific activity. A particular achievement was the semester he organized on mathematical physics supported by the Chinese Science Foundation, which attracted most of the leading scientists in the field. He was there, in Shanghai, planning the next scientific visits, dreaming on building "the Eurandom of Asia" when his life came to an end.

Besides his great talent and creativity, Vladas had an unlimited enthusiasm for his work. He truly enjoyed it and would not be stopped by ordinary difficulties: he would make huge efforts to attend a conference or meeting that he considered important, working full day in between two long flights; he would put in full energy while organizing events and making sure that everyone felt as comfortable as possible. At IMPA, students and collaborators always knew the clock drawn on his blackboard as his daily agenda. This was always full but also always open to find some extra time. Everyone could see his immense energy and his passionate enthusiasm for the profession.

This volume contains a collection of papers by many of his collaborators and on a variety of topics in probability and statistical physics that reflects Vladas' main research interests. Among them are two projects in collaboration with him, in preparation at the time of his death.

The idea of preparing this volume grew during the XXIII Brazilian School of Probability that took place at the end of July 2019, in USP-São Carlos, also dedicated to him. We thank the scientific and organizing committees as well as all the speakers of the School for their full support.

After we wrote to Vladas' many collaborators, we received great support and excellent cooperation from them and many others who helped with this project, including series editors, authors, and anonymous referees. Most of the review process took place during a period when everyone was affected by the pandemic of Covid-19, with extra time needed for online teaching activities, but our referees were extremely generous with their help. Our sincere thanks also goes to a group of Vladas' close friends from Lithuania, for their valuable and inspiring feedback.

We acknowledge the important role played by NYU-Shanghai in the latter portion of Vladas' career and its cooperation in the preparation of this volume. The 1-day memorial event held in Shanghai, on October 22, 2019, when several of us came together to remember him, was also a source of inspiration.

As a consequence of his enormous enthusiasm and dedication to the probability community, besides organizing wonderful meetings, Vladas edited many special volumes, mostly associated to schools or conferences in probability and mathematical physics. The list includes proceedings of two editions of the Brazilian School of Probability, of which he was one of the initiators, and that he titled *In and Out of Equilibrium*. As a way to honor him and at the same time reflecting well the content of the scientific papers, we keep the title for this memorial volume.

Anyone who had the opportunity of being close to Vladas, in the profession or outside, knows his huge energy and joy for life. He also took great care of his mother, Galina, who survives him. No matter where in the world he was located, he would call her almost daily to make sure she was well and well-provided for. We all remember him in constant Celebration of Life. We miss his joyful laugh but have powerful reasons to celebrate his life and his achievements.

Rio de Janeiro, Brazil — Maria Eulália Vares
Shanghai, China — Roberto Fernández
São Paulo, Brazil — Luiz Renato Fontes
New York, NY, USA — Charles M. Newman
August 2020

With Harry Kesten (Ithaca, 2003)

Vladas, 2nd from right in top row at Vilnius University, in 1984 (courtesy of Arvydas Strumskis)

Vladas at leisure in the early 90s (courtesy of Renata Sidoraviciene)

Publications of Vladas Sidoravicius

The editors believe that this list of publications was complete at the time when this volume was prepared, but since there are a number of ongoing projects that Vladas was involved in, it is likely that there will be some future publications that include Vladas as a coauthor.

Research Articles

1. Ignatiuk, I. A., Malyshev, V. A. and Sidoravičius, V. Convergence of the stochastic quantization method. *Probability theory and mathematical statistics*, Vol. I (Vilnius, 1989) "Mokslas", Vilnius, 1990, 526–538.
2. Sidoravičius, V. Convergence of the stochastic quantization method for lattice R-gauge theories. *New trends in probability and statistics*, Vol. 1 (Bakuriani, 1990) VSP, Utrecht, 1991, 694–699.
3. Statulyavichus, V. A. and Sidoravichyus, V. Convergence to the Poisson law on algebras with canonical commutative relations and canonical anticommutative relations. *Dokl. Akad. Nauk*, 1992, Vol. 322(5), 858–860; translation in Soviet Math. Dokl. 45 (1992), no. 1, 202–205.
4. Ignatyuk, I. A., Malyshev, V. A. and Sidoravichius, V. Convergence of the stochastic quantization method. I *Teor. Veroyatnost. i Primenen.*, 1992, Vol. 37(2), 241–253; translation in Theory Probab. Appl. 37 (1992), no. 2, 209–221.
5. Ignatyuk, I. A., Malyshev, V. A. and Sidoravičius, V. Convergence of the stochastic quantization method. II. The stochastic quantization method for Grassmannian Gibbs fields. *Teor. Veroyatnost. i Primenen.*, 1992, Vol. 37(4), 621–647; translation in Theory Probab. Appl. 37 (1992), no. 4, 599–620.
6. Sidoravičius, V. and Statulevicius, V. A. Convergence to Poisson law on algebraic structures for ψ-mixing systems. *Probability theory and mathematical statistics* (Kiev, 1991) World Sci. Publ., River Edge, NJ, 1992, 354–362.
7. Sidoravičius, V. and Vares, M. E. Ergodicity of Spitzer's renewal model. *Stochastic Process. Appl.*, 1995, Vol. 55(1), 119–130.
8. Fontes, L. R., Isopi, M. and Sidoravičius, V. Analyticity of the density and exponential decay of correlations in 2-d bootstrap percolation. *Stochastic Process. Appl.*, 1996, Vol. 62(1), 169–178.
9. De Angelis, G. F., Jona-Lasinio, G. and Sidoravicius, V. Berezin integrals and Poisson processes. *J. Phys. A*, 1998, Vol. 31(1), 289–308.
10. Kesten, H., Sidoravicius, V. and Zhang, Y. Almost all words are seen in critical site percolation on the triangular lattice. Electron. J. Probab., 1998, Vol. 3, no. 10, 75 pp.

11. Sidoravicius, V., Surgailis, D. and Vares, M. E. An exclusion process with two types of particles and the hydrodynamic limit. *Markov Process. Related Fields*, 1998, Vol. 4(2), 131–174.
12. Sidoravicius, V., Triolo, L. and Vares, M. E. On the forced motion of a heavy particle in a random medium. I. Existence of dynamics. *Markov Process. Related Fields*, 1998, Vol. 4(4), 629–647.
13. Sidoravicius, V., Surgailis, D. and Vares, M. E. On the truncated anisotropic long-range percolation on $\mathbb{Z}^2$. *Stochastic Process. Appl.*, 1999, Vol. 81(2), 337–349.
14. Sidoravicius, V., Vares, M. E. and Surgailis, D. Poisson broken lines process and its application to Bernoulli first passage percolation. *Acta Appl. Math.*, 1999, Vol. 58(1–3), 311–325.
15. Brassesco, S., Presutti, E., Sidoravicius, V. and Vares, M. E. Ergodicity and exponential convergence of a Glauber + Kawasaki process. *On Dobrushin's way. From probability theory to statistical physics*, 37–49. Amer. Math. Soc. Transl. Ser. 2, 198, Adv. Math. Sci., 47, Amer. Math. Soc., Providence, RI, 2000.
16. Brassesco, S., Presutti, E., Sidoravicius, V. and Vares, M. E. Ergodicity of a Glauber + Kawasaki process with metastable states. *Markov Process. Related Fields*, 2000, Vol. 6(2), 181–203.
17. Fontes, L. R. G., Jordão Neves, E. and Sidoravicius, V. Limit velocity for a driven particle in a random medium with mass aggregation. *Ann. Inst. H. Poincaré Probab. Statist.*, 2000, Vol. 36(6), 787–805.
18. Pellegrinotti, A., Sidoravicius, V. and Vares, M. E. Stationary state and diffusion for a charged particle in a one-dimensional medium with lifetimes. *Teor. Veroyatnost. i Primenen.*, 1999, Vol. 44(4), 796–825; reprinted in Theory Probab. Appl., 2000, Vol. 44(4), 697–721.
19. Kesten, H., Sidoravicius, V. and Zhang, Y. Percolation of arbitrary words on the close-packed graph of $\mathbb{Z}^2$. *Electron. J. Probab.*, 2001, Vol. 6, no. 4, 27 pp.
20. Sidoravicius, V., Triolo, L. and Vares, M. E. Mixing properties for mechanical motion of a charged particle in a random medium. *Comm. Math. Phys.*, 2001, Vol. 219(2), 323–355.
21. Menshikov, M., Sidoravicius, V. and Vachkovskaia, M. A note on two-dimensional truncated long-range percolation. *Adv. in Appl. Probab.*, 2001, Vol. 33(4), 912–929.
22. Camia, F., Newman, C. M. and Sidoravicius, V. Approach to fixation for zero-temperature stochastic Ising models on the hexagonal lattice. *In and out of equilibrium* (Mambucaba, 2000) Birkhäuser Boston, Boston, MA, 2002, Progr. Probab. Vol. 51, 163–183.
23. Fontes, L. R., Schonmann, R. H. and Sidoravicius, V. Stretched exponential fixation in stochastic Ising models at zero temperature. *Comm. Math. Phys.*, 2002, Vol. 228(3), 495–518.
24. Camia, F., Newman, C. M. and Sidoravicius, V. Cardy's formula for some dependent percolation models. *Bull. Braz. Math. Soc. (N.S.)*, 2002, Vol. 33(2), 147–156.

25. Ramírez, A. F. and Sidoravicius, V. Asymptotic behavior of a stochastic growth process associated with a system of interacting branching random walks. *C. R. Math. Acad. Sci. Paris*, 2002, Vol. 335(10), 821–826.
26. Kesten, H. and Sidoravicius, V. Branching random walk with catalysts. *Electron. J. Probab.*, 2003, Vol. 8, no. 5, 51 pp.
27. Camia, F., Newman, C. M. and Sidoravicius, V. A particular bit of universality: scaling limits of some dependent percolation models. *Comm. Math. Phys.*, 2004, Vol. 246(2), 311–332.
28. Ramírez, A. F. and Sidoravicius, V. Asymptotic behavior of a stochastic combustion growth process. *J. Eur. Math. Soc. (JEMS)*, 2004, Vol. 6(3), 293–334.
29. Sidoravicius, V. and Sznitman, A.-S. Quenched invariance principles for walks on clusters of percolation or among random conductances. *Probab. Theory Related Fields*, 2004, Vol. 129(2), 219–244.
30. Friedli, S., de Lima, B. N. B. and Sidoravicius, V. On long range percolation with heavy tails. *Electron. Comm. Probab.*, 2004, Vol. 9, 175–177.
31. Fontes, L. R. G. and Sidoravicius, V. Percolation *School and Conference on Probability Theory*, 101–201, ICTP Lect. Notes, XVII, Abdus Salam Int. Cent. Theoret. Phys., Trieste, 2004.
32. Kesten, H. and Sidoravicius, V. The spread of a rumor or infection in a moving population. *Ann. Probab.*, 2005, Vol. 33(6), 2402–2462.
33. de Oliveira, P. M. C., Newman, C. M., Sidoravicious, V. and Stein, D. L. Ising ferromagnet: zero-temperature dynamic evolution. *J. Phys. A*, 2006, Vol. 39(22), 6841–6849.[1]
34. Alexander, K. S. and Sidoravicius, V. Pinning of polymers and interfaces by random potentials. *Ann. Appl. Probab.*, 2006, Vol. 16(2), 636–669.
35. Kesten, H. and Sidoravicius, V. A phase transition in a model for the spread of an infection. *Illinois J. Math.*, 2006, Vol. 50(1–4), 547–634.
36. Beffara, V., Sidoravicius, V., Spohn, H. and Vares, M. E. Polymer pinning in a random medium as influence percolation. *Dynamics & Stochastics*, IMS Lecture Notes Monogr. Ser., Vol. 48, 1–15. Inst. Math. Statist., Beachwood, OH, 2006.
37. Kesten, H. and Sidoravicius, V. A shape theorem for the spread of an infection. *Ann. of Math. (2)*, 2008, Vol. 167(3), 701–766.
38. Kesten, H. and Sidoravicius, V. A problem in one-dimensional diffusion-limited aggregation (DLA) and positive recurrence of Markov chains. *Ann. Probab.*, 2008, Vol. 36(5), 1838–1879.
39. van den Berg, J., Peres, Y., Sidoravicius, V. and Vares, M. E. Random spatial growth with paralyzing obstacles. *Ann. Inst. Henri Poincaré Probab. Stat.*, 2008, Vol. 44(6), 1173–1187.

[1] We have left the misspelling of Vladas' last name because that was the way it appeared in the published version of the paper.

40. Kesten, H. and Sidoravicius, V. Positive recurrence of a one-dimensional variant of diffusion limited aggregation. *In and out of equilibrium. 2* Birkhäuser, Basel, 2008, Progr. Probab., Vol. 60, 429–461.
41. Bertoin, J. and Sidoravicius, V. The structure of typical clusters in large sparse random configurations. *J. Stat. Phys.*, 2009, Vol. 135(1), 87–105.
42. Sidoravicius, V. and Sznitman, A.-S. Percolation for the vacant set of random interlacements. *Comm. Pure Appl. Math.*, 2009, Vol. 62(6), 831–858.
43. Dickman, R., Rolla, L. T. and Sidoravicius, V. Activated random walkers: facts, conjectures and challenges. *J. Stat. Phys.*, 2010, Vol. 138(1–3), 126–142.
44. Beffara, V., Sidoravicius, V. and Vares, M. E. Randomized polynuclear growth with a columnar defect. *Probab. Theory Related Fields*, 2010, Vol. 147(3–4), 565–581.
45. Kesten, H. and Sidoravicius, V. A problem in last-passage percolation. *Braz. J. Probab. Stat.*, 2010, Vol. 24(2), 300–320.
46. Marchetti, D. H. U., Sidoravicius, V. and Vares, M. E. Oriented percolation in one-dimensional $1/|x-y|^2$ percolation models. *J. Stat. Phys.*, 2010, Vol. 139(6), 941–959.
47. Hilário, M. R., Louidor, O., Newman, C. M., Rolla, L. T., Sheffield, S. and Sidoravicius, V. Fixation for distributed clustering processes. *Comm. Pure Appl. Math.*, 2010, Vol. 63(7), 926–934.
48. Rolla, L. T., Sidoravicius, V., Surgailis, D. and Vares, M. E. The discrete and continuum broken line process. *Markov Process. Related Fields*, 2010, Vol. 16(1), 79–116.
49. Bertoin, J., Sidoravicius, V. and Vares, M. E. A system of grabbing particles related to Galton-Watson trees. *Random Structures Algorithms*, 2010, Vol. 36(4), 477–487.
50. Sidoravicius, V. and Sznitman, A.-S. Connectivity bounds for the vacant set of random interlacements. *Ann. Inst. Henri Poincaré Probab. Stat.*, 2010, Vol. 46(4), 976–990.
51. Rolla, L. T. and Sidoravicius, V. Absorbing-state phase transition for driven-dissipative stochastic dynamics on $\mathbb{Z}$. *Invent. Math.*, 2012, Vol. 188(1), 127–150.
52. Kesten, H., Ramírez, A. F. and Sidoravicius, V. Asymptotic shape and propagation of fronts for growth models in dynamic random environment. *Probability in complex physical systems.* Springer Proc. Math., Vol. 11, 195–223, Springer, Heidelberg, 2012.
53. den Hollander, F., dos Santos, R. and Sidoravicius, V. Law of large numbers for non-elliptic random walks in dynamic random environments. *Stochastic Process. Appl.*, 2013, Vol. 123(1), 156–190.
54. Marchetti, D. H. U., Sidoravicius, V. and Vares, M. E. Commentary to: Oriented percolation in one-dimension $1/|x-y|^2$ percolation models. *J. Stat. Phys.*, 2013, Vol. 150(4), 804–805.
55. Damron, M., Kogan, H., Newman, C. M. and Sidoravicius, V. Fixation for coarsening dynamics in 2D slabs. *Electron. J. Probab.*, 2013, Vol. 18, No. 105, 20 pp.

56. Ben Arous, G., Fribergh, A. and Sidoravicius, V. Lyons-Pemantle-Peres monotonicity problem for high biases. *Comm. Pure Appl. Math.*, 2014, Vol. 67(4), 519–530.
57. Kesten, H., de Lima, B. N. B., Sidoravicius, V. and Vares, M. E. On the compatibility of binary sequences. *Comm. Pure Appl. Math.*, 2014, Vol. 67(6), 871–905.
58. Cabezas, M., Rolla, L. T. and Sidoravicius, V. Non-equilibrium phase transitions: activated random walks at criticality. *J. Stat. Phys.*, 2014, Vol. 155(6), 1112–1125.
59. den Hollander, F., Kesten, H. and Sidoravicius, V. Random walk in a high density dynamic random environment. *Indag. Math. (N.S.)*, 2014, Vol. 25(4), 785–799
60. Damron, M., Kogan, H., Newman, C. M. and Sidoravicius, V. Coarsening in 2D slabs. *Topics in percolative and disordered systems.* Springer Proc. Math. Stat., Vol. 69, 15–22. Springer, New York, 2014.
61. Rolla, L. T., Sidoravicius, V. and Tournier, L. Greedy clearing of persistent Poissonian dust. *Stochastic Process. Appl.*, 2014, Vol. 124(10), 3496–3506.
62. Ahlberg, D., Sidoravicius, V. and Tykesson, J. Bernoulli and self-destructive percolation on non-amenable graphs *Electron. Commun. Probab.*, 2014, Vol. 19, no. 40, 6 pp.
63. Hilário, M. R., de Lima, B. N. B., Nolin, P. and Sidoravicius, V. Embedding binary sequences into Bernoulli site percolation on $\mathbb{Z}^3$. *Stochastic Process. Appl.*, 2014, Vol. 124(12), pp. 4171–4181.
64. Dembo, A., Huang, R. and Sidoravicius, V. Walking within growing domains: recurrence versus transience. *Electron. J. Probab.*, 2014, Vol. 19, no. 106, 20 pp.
65. Dembo, A., Huang, R. and Sidoravicius, V. Monotone interaction of walk and graph: recurrence versus transience. *Electron. Commun. Probab.*, 2014, Vol. 19, no. 76, 12 pp.
66. Sidoravicius, V. Criticality and phase transitions: five favorite pieces. *Proceedings of the International Congress of Mathematicians—Seoul 2014*. Vol. IV Kyung Moon Sa, Seoul, 2014, 199–224.
67. Sidoravicius, V. and Stauffer, A. Phase transition for finite-speed detection among moving particles. *Stochastic Process. Appl.*, 2015, Vol. 125(1), 362–370.
68. Aizenman, M., Duminil-Copin, H. and Sidoravicius, V. Random currents and continuity of Ising model's spontaneous magnetization. *Comm. Math. Phys.*, 2015, Vol. 334(2), 719–742.
69. Foss, S., Rolla, L. T. and Sidoravicius, V. Greedy walk on the real line. *Ann. Probab.*, 2015, Vol. 43(3), 1399–1418.
70. Damron, M., Eckner, S. M., Kogan, H., Newman, C. M. and Sidoravicius, V. Coarsening dynamics on $\mathbb{Z}^d$ with frozen vertices. *J. Stat. Phys.*, 2015, Vol. 160(1), 60–72.

71. Ahlberg, D., Duminil-Copin, H., Kozma, G. and Sidoravicius, V. Seven-dimensional forest fires. *Ann. Inst. Henri Poincaré Probab. Stat.*, 2015, Vol. 51(3), 862–866.
72. Damron, M., Newman, C. M. and Sidoravicius, V. Absence of site percolation at criticality in $\mathbb{Z}^2 \times \{0, 1\}$. *Random Structures Algorithms*, 2015, Vol. 47(2), 328–340.
73. Hilário, M. R., den Hollander, F., dos Santos, R. S., Sidoravicius, V. and Teixeira, A. Random walk on random walks. *Electron. J. Probab.*, 2015, Vol. 20, no. 95, 35 pp.
74. Hilário, M. R., Sidoravicius, V. and Teixeira, A. Cylinders' percolation in three dimensions. Probab. Theory Related Fields, 2015, Vol. 163(3–4), 613–642.
75. Kozma, G. and Sidoravicius, V. Lower bound for the escape probability in the Lorentz mirror model on $\mathbb{Z}^2$. *Israel J. Math.*, 2015, Vol. 209(2), 683–685.
76. Kiss, D., Manolescu, I. and Sidoravicius, V. Planar lattices do not recover from forest fires. *Ann. Probab.*, 2015, Vol. 43(6), 3216–3238.
77. Markarian, R., Rolla, L. T., Sidoravicius, V., Tal, F. A. and Vares, M. E. Stochastic perturbations of convex billiards. *Nonlinearity*, 2015, Vol. 28(12), 4425–4434.
78. Ahlberg, D., Damron, M. and Sidoravicius, V. Inhomogeneous first-passage percolation. *Electron. J. Probab.*, 2016, Vol. 21, Paper No. 4, 19 pp.
79. Damron, M., Kogan, H., Newman, C. M. and Sidoravicius, V. Coarsening with a frozen vertex. *Electron. Commun. Probab.*, 2016, Vol. 21, Paper No. 9, 4 pp.
80. Duminil-Copin, H., Sidoravicius, V. and Tassion, V. Absence of infinite cluster for critical Bernoulli percolation on slabs. *Comm. Pure Appl. Math.*, 2016, Vol. 69(7), 1397–1411.
81. Aymone, M. and Sidoravicius, V. Partial sums of biased random multiplicative functions *J. Number Theory*, 2017, Vol. 172, 343–382.
82. Duminil-Copin, H., Sidoravicius, V. and Tassion, V. Continuity of the phase transition for planar random-cluster and Potts models with $1 \leq q \leq 4$. *Comm. Math. Phys.*, 2017, Vol. 349(1), 47–107.
83. Sidoravicius, V. and Teixeira, A. Absorbing-state transition for stochastic sandpiles and activated random walks. *Electron. J. Probab.*, 2017, Vol. 22, Paper No. 33, 35 pp.
84. Grassberger, P., Hilário, M. R. and Sidoravicius, V. Percolation in media with columnar disorder. *J. Stat. Phys.*, 2017, Vol. 168(4), 731–745.
85. Rolla, L. T. and Sidoravicius, V. Stability of the greedy algorithm on the circle. *Comm. Pure Appl. Math.*, 2017, Vol. 70(10), 1961–1986.
86. Sidoravicius, V. and Tournier, L. Note on a one-dimensional system of annihilating particles. *Electron. Commun. Probab.*, 2017, Vol. 22, Paper No. 59, 9 pp.
87. Cabezas, M., Rolla, L. T. and Sidoravicius, V. Recurrence and density decay for diffusion-limited annihilating systems. *Probab. Theory Related Fields*, 2018, Vol. 170(3–4), 587–615.
88. Berger, N., Hoffman, C. and Sidoravicius, V. Non-uniqueness for specifications in $\ell^{2+\epsilon}$. Ergodic Theory Dynam. Systems, 2018, Vol. 38(4), 1342–1352.

89. Duminil-Copin, H., Hilário, M. R., Kozma, G. and Sidoravicius, V. Brochette percolation. *Israel J. Math.*, 2018, Vol. 225(1), 479–501.
90. Kious, D. and Sidoravicius, V. Phase transition for the once-reinforced random walk on $\mathbb{Z}^d$-like trees. *Ann. Probab.*, 2018, Vol. 46(4), 2121–2133.
91. Basu, R., Sidoravicius, V. and Sly, A. Lipschitz embeddings of random fields. *Probab. Theory Related Fields*, 2018, Vol. 172(3–4), 1121–1179.
92. Huang, R., Kious, D., Sidoravicius, V. and Tarrès, P. Explicit formula for the density of local times of Markov jump processes. *Electron. Commun. Probab.*, 2018, Vol. 23, Paper No. 90, 7 pp.
93. Curien, N., Kozma, G., Sidoravicius, V. and Tournier, L. Uniqueness of the infinite noodle. *Ann. Inst. Henri Poincaré D*, 2019, Vol. 6(2), 221–238.
94. Rolla, L. T., Sidoravicius, V. and Zindy, O. Universality and Sharpness in Activated Random Walks. *Ann. Henri Poincaré*, 2019, Vol. 20(6), 1823–1835.
95. Cabezas, M., Dembo, A., Sarantsev, A., Sidoravicius, V. Brownian particles with rank-dependent drifts: out-of-equilibrium behavior. *Comm. Pure Appl. Math.*, 2019, Vol. 72(7), 1424–1458.
96. Sidoravicius, V., Stauffer, A. Multi-particle diffusion limited aggregation. *Invent. Math.*, 2019, Vol. 218(2), 491–571.
97. Hilário, M. R., Sidoravicius, V. Bernoulli line percolation. *Stochastic Process. Appl.*, 2019, Vol. 129(12), 5037–5072.
98. Blondel, O., Hilário, M. R., dos Santos, R. S., Sidoravicius, V., Teixeira, A. Random walk on random walks: higher dimensions. *Electron. J. Probab.*, 2019, Vol. 24, Paper No. 80, 33 pp.
99. Collevecchio, A., Kious, D., Sidoravicius, V. The branching-ruin number and the critical parameter of once-reinforced random walk on trees. *Comm. Pure Appl. Math.*, 2020, Vol. 73(1), 210–236.
100. Liu, Y., Sidoravicius, V., Wang, L., Xiang, K. An invariance principle and a large deviation principle for the biased random walk on $\mathbb{Z}^d$. *J. Appl. Probab.*, 2020, Vol. 57(1), 295–313.
101. Bock, B., Damron, M., Newman, C. M., Sidoravicius, V. Percolation of finite clusters and shielded paths. *J. Stat. Phys.*, 2020, Vol. 179(3), 789–807.
102. de Lima, B. N. B., Sanchis, R., dos Santos, D. C., Sidoravicius, V., Teodoro, R. The constrained-degree percolation model. *Stochastic Process. Appl.*, 2020, Vol. 130(9), 5492–5509.
103. Blondel, O., Hilário, M. R., dos Santos, R. S., Sidoravicius, V., Teixeira, A. Random walk on random walks: Low densities. Ann. Appl. Probab., 2020, Vol. 30(4), 1614–1641.
104. Duminil-Copin, H., Kesten, H., Nazarov, F., Peres, Y., Sidoravicius, V. On the number of maximal paths in directed last-passage percolation. *Ann. Probab.*, 2020, Vol. 48(5), 2176–2188.
105. Aymone, M., de Lima, B. N. B., Hilário, M, Sidoravicius, V. Bernoulli hiperplane percolation. (this volume)
106. Fribergh, A., Kious, D., Sidoravicius, V., Stauffer, A. Random memory walk. (this volume)

Vladas at NYU-Shanghai in 2016 (courtesy of NYU-Shanghai; credit: Junbo Chen)

Contents

Existence and Coexistence in First-Passage Percolation

Daniel Ahlberg

Abstract We consider first-passage percolation with i.i.d. non-negative weights coming from some continuous distribution under a moment condition. We review recent results in the study of geodesics in first-passage percolation and study their implications for the multi-type Richardson model. In two dimensions this establishes a dual relation between the existence of infinite geodesics and coexistence among competing types. The argument amounts to making precise the heuristic that infinite geodesics can be thought of as 'highways to infinity'. We explain the limitations of the current techniques by presenting a partial result in dimensions $d > 2$.

Keywords First-passage percolation · Competing growth · Geodesics · Busemann functions

1 Introduction

In first-passage percolation the edges of the $\mathbb{Z}^d$ nearest neighbour lattice, for some $d \geq 2$, are equipped with non-negative i.i.d. random weights ω_e, inducing a random metric T on $\mathbb{Z}^2$ as follows: For $x, y \in \mathbb{Z}^d$, let

$$T(x, y) := \inf\left\{ \textstyle\sum_{e \in \pi} \omega_e : \pi \text{ is a self-avoiding path from } x \text{ to } y \right\}. \qquad (1)$$

Since its introduction in the 1960s, by Hammersley and Welsh [18], a vast body of literature has been generated seeking to understand the large scale behaviour of distances, balls and geodesics in this random metric space. The state of the art has been summarized in various volumes over the years, including [4, 21, 23, 32]. We will here address questions related to geodesics, and shall for this reason make

D. Ahlberg (✉)
Department of Mathematics, Stockholm University, Stockholm, Sweden
e-mail: daniel.ahlberg@math.su.se

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_1

the common assumption that the edge weights are sampled from a continuous distribution. Since many of the results we shall rely on require a moment condition for their conclusions to hold, we shall assume in what follows that $\mathbb{E}[Y^d] < \infty$, where Y denotes the minimum weight among the $2d$ edges connected to the origin.

In the 1960s, the study of first-passage percolation led to the development of an ergodic theory for subadditive ergodic sequences, culminating with the ergodic theorem due to Kingman [24]. As a consequence thereof, one obtains the existence of a norm $\mu : \mathbb{R}^d \to [0, \infty)$, simply referred to as the *time constant*, such that for every $z \in \mathbb{Z}^d$, almost surely,

$$\lim_{n\to\infty} \frac{1}{n} T(0, nz) = \mu(z).$$

Richardson [30], and later work of Cox and Durrett [9], extended the above *radial* convergence to *simultaneous* convergence in all directions. Their results show that the ball $\{z \in \mathbb{Z}^d : T(0, z) \leq t\}$ in the metric T once rescaled by $1/t$ approaches the unit ball in the norm μ. The unit ball in μ, henceforth denoted by $\text{Ball} := \{x \in \mathbb{R}^d : \mu(x) \leq 1\}$, is therefore commonly referred to as the *asymptotic shape*, and known to be compact and convex with non-empty interior. In addition, the shape retains the symmetries of $\mathbb{Z}^d$. However, little else is known regarding the properties of the shape in general. This, we shall see, is a major obstacle for our understanding of several other features of the model.

Although questions regarding geodesics were considered in the early work of Hammersley and Welsh, it took until the mid 1990s before Newman [28] together with co-authors [25, 26, 29] initiated a systematic study of the geometry of geodesics in first-passage percolation. Under the assumption of continuous weights there is almost surely a unique path attaining the minimum in (1); we shall denote this path $\text{geo}(x, y)$ and refer to it as the *geodesic* between x and y. The graph consisting of all edges on $\text{geo}(0, y)$ for some $y \in \mathbb{Z}^d$ is a tree spanning the lattice. Understanding the properties of this object, such as the number of topological ends, leads one to the study of *infinite* geodesics, i.e. infinite paths of which every finite segment is a geodesic. We shall write $\mathscr{T}_0$ for the collection of infinite geodesics starting at the origin. A simple compactness argument shows that the cardinality $|\mathscr{T}_0|$ of $\mathscr{T}_0$ is always at least one.[1] In two dimensions, Newman [28] predicted that $|\mathscr{T}_0| = \infty$ almost surely, and proved this under an additional assumption of uniform curvature of the asymptotic shape, which remains unverified to this day.

As a means to make rigorous progress on Newman's prediction, Häggström and Pemantle [17] introduced a model for competing growth on $\mathbb{Z}^d$, for $d \geq 2$, known as the *two-type Richardson model*. In this model, two sites x and y are initially coloured red and blue respectively. As time evolves an uncoloured site turns red

[1] Consider the sequence of finite geodesics between the origin and $n\mathbf{e}_1$, where $\mathbf{e}_1$ denotes the first coordinate vector. Since the number of edges that connect to the origin is finite, one of them must be traversed for infinitely many n. Repeating the argument results in an infinite path which by construction is a geodesic.

at rate 1 times the number of red neighbours, and blue at rate λ times the number or blue neighbours. A central question of interest is for which values of λ there is positive probability for both colours to coexist, in the sense that they both are responsible for the colouring of infinitely many sites.

There is an intimate relation between the existence of infinite geodesics and coexistence in the Richardson model that we shall pay special interest in. In the case of equal strength competitors ($\lambda = 1$), one way to construct the two-type Richardson model is to equip the edges of the $\mathbb{Z}^d$ lattice with independent exponential weights, thus exhibiting a direct connection to first-passage percolation. The set of sites eventually coloured red in the two-type Richardson model is then equivalent to the set of sites closer to x than y in the first-passage metric. That is, an analogous way to phrase the question of coexistence is whether there are infinitely many points closer to x than y as well as infinitely many points closer to y than x in the first-passage metric. As before, a compactness argument will show that on the event of coexistence there are disjoint infinite geodesics g and g' that respectively originate from x and y. Häggström and Pemantle [17] showed that, for $d = 2$, coexistence of the two types occurs with positive probability, and deduced as a corollary that

$$\mathbb{P}(|\mathscr{T}_0| \geq 2) > 0.$$

Their results were later extended to higher dimensions and more general edge weight distributions in parallel by Garet and Marchand [13] and Hoffman [19]. In a later paper, Hoffman [20] showed that in two dimensions coexistence of four different types has positive probability, and that $\mathbb{P}(|\mathscr{T}_0| \geq 4) > 0$. The best currently known general lower bound on the number of geodesics is a strengthening of Hoffman's result due to Damron and Hanson [10], showing that

$$\mathbb{P}(|\mathscr{T}_0| \geq 4) = 1.$$

In this paper we shall take a closer look at the relation between existence of infinite geodesics and coexistence in competing first-passage percolation. We saw above that on the event of coexistence of various types, a compactness argument gives the existence of equally many infinite geodesics. It is furthermore conceivable that it is possible to locally modify the edge weight in such a way that these geodesics are re-routed through the origin. Conversely, interpreting infinite geodesics as 'highways to infinity', along which the different types should be able to escape their competitors, it seems that the existence of a given number of geodesics should accommodate an equal number of surviving types. These heuristic arguments suggest a duality between existence and coexistence, and it is this dual relation we shall make precise.

Given sites $x_1, x_2, \ldots, x_k$ in $\mathbb{Z}^d$, we let $\mathrm{Coex}(x_1, x_2, \ldots, x_k)$ denote the event that for every $i = 1, 2, \ldots, k$ there are infinitely many sites $z \in \mathbb{Z}^d$ for which the distance $T(x_j, z)$ is minimized by $j = i$. (The continuous weight distribution assures that there are almost surely no ties.) In two dimensions the duality between existence and coexistence that we prove takes the form:

$$\exists\, x_1, x_2, \ldots, x_k \text{ such that } \mathbb{P}\big(\mathrm{Coex}(x_1, x_2, \ldots, x_k)\big) > 0 \quad \Leftrightarrow \quad \mathbb{P}(|\mathscr{T}_0| \geq k) > 0. \tag{2}$$

Turning the above heuristic into a proof is more demanding that it may seem. In order to derive the relation in (2) we shall rely on the recently developed ergodic theory for infinite geodesics. This theory has its origins in the work of Hoffman [19, 20], and was developed further by Damron and Hanson [10, 11], before it reached its current status in work of Ahlberg and Hoffman [1]. The full force of this theory is currently restricted to two dimensions, which prevents us from obtaining an analogue to (2) in higher dimensions. In higher dimensions we deduce a partial result based on results of Damron and Hanson [10] and Nakajima [27].

1.1 The Dual Relation

Before we state our results formally, we remind the reader that Y denotes the minimum weight among the $2d$ edges connected to the origin. We recall (from [9]) that $\mathbb{E}[Y^d] < \infty$ is both necessary and sufficient in order for the shape theorem to hold in dimension $d \geq 2$.

Theorem 1 *Consider first-passage percolation on $\mathbb{Z}^2$ with continuous edge weights satisfying $\mathbb{E}[Y^2] < \infty$. For any $k \geq 1$, including $k = \infty$, and $\varepsilon > 0$ we have:*

(i) If $\mathbb{P}\big(\mathrm{Coex}(x_1, \ldots, x_k)\big) > 0$ for some $x_1, \ldots, x_k$ in $\mathbb{Z}^2$, then $\mathbb{P}(|\mathscr{T}_0| \geq k) = 1$.
(ii) If $\mathbb{P}(|\mathscr{T}_0| \geq k) > 0$, then $\mathbb{P}\big(\mathrm{Coex}(x_1, \ldots, x_k)\big) > 1 - \varepsilon$ for some $x_1, \ldots, x_k$ in $\mathbb{Z}^2$.

In dimensions higher than two we shall establish parts of the above dual relation, and recall next some basic geometric concepts in order to state this result precisely. A hyperplane in the d-dimensional Euclidean space divides $\mathbb{R}^d$ into two open half-spaces. A *supporting hyperplane* to a convex set $S \subset \mathbb{R}^d$ is a hyperplane that contains some boundary point of S and contains all interior points of S in one of the two half-spaces associated to the hyperplane. It is well-known that for every boundary point of a convex set S there exists a supporting hyperplane that contains that point. A supporting hyperplane to S is called a *tangent hyperplane* if it is the unique supporting hyperplane containing some boundary point of S. Finally, we define the number of *sides* of a compact convex set S as the number of (distinct) tangent hyperplanes to S. Hence, the number of sides is finite if and only if S is a (finite) convex polygon ($d = 2$) or convex polytope ($d \geq 3$). A deeper account on convex analysis can be found in [31].

Theorem 2 *Consider first-passage percolation on $\mathbb{Z}^d$, for $d \geq 2$, with continuous edge weights. For any $k \geq 1$, including $k = \infty$, and $\varepsilon > 0$ we have*

(i) If $\mathbb{E}[\exp(\alpha\omega_e)] < \infty$ and $\mathbb{P}\big(\mathrm{Coex}(x_1, \ldots, x_k)\big) > 0$ for some $\alpha > 0$ and $x_1, \ldots, x_k$ in $\mathbb{Z}^d$, then $\mathbb{P}(|\mathscr{T}_0| \geq k) = 1$.
(ii) If $\mathbb{E}[Y^d] < \infty$ and Ball has at least k sides, then $\mathbb{P}\big(\mathrm{Coex}(x_1, \ldots, x_k)\big) > 1 - \varepsilon$ for some $x_1, \ldots, x_k$ in $\mathbb{Z}^d$.

In Sect. 2 we shall review the recent development in the study of infinite geodesics that will be essential for the deduction, in Sect. 3, of the announced dual result. Finally, in Sect. 4, we prove the partial result in higher dimensions.

1.2 A Mention of Our Methods

One aspect of the connection between existence and coexistence is an easy observation, and was hinted at already above. Namely, if $\mathrm{Geos}(x_1, x_2, \ldots, x_k)$ denotes the event that there exist k pairwise disjoint infinite geodesics, each originating from one of the points $x_1, x_2, \ldots, x_k$, then

$$\mathrm{Coex}(x_1, x_2, \ldots, x_k) \subseteq \mathrm{Geos}(x_1, x_2, \ldots, x_k). \tag{3}$$

To see this, let V_i denote the set of sites closer to x_i than to any other x_j, for $j \neq i$, in the first-passage metric. (Note that $T(x, y) \neq T(z, y)$ for all $x, y, z \in \mathbb{Z}^2$ almost surely, due to the assumptions of continuous weights.[2]) On the event $\mathrm{Coex}(x_1, x_2, \ldots, x_k)$ each set V_i is infinite, and for each i a compactness argument gives the existence of an infinite path contained in V_i, which by construction is a geodesic. Since $V_1, V_2, \ldots, V_k$ are pairwise disjoint, due to uniqueness of geodesics, so are the resulting infinite geodesics.

Let $\mathscr{N}$ denote the maximal number of pairwise disjoint infinite geodesics. Since $\mathscr{N}$ is invariant with respect to translations (and measurable) it follows from the ergodic theorem that $\mathscr{N}$ is almost surely constant. Hence, positive probability for coexistence of k types implies the almost sure existence of k pairwise disjoint geodesics. That $|\mathscr{T}_0| \leq \mathscr{N}$ is trivial, given the tree structure of $\mathscr{T}_0$. The inequality is in fact an equality, which was established by different means in [1, 27]. Together with (3), this resolves the first part of Theorems 1 and 2.

Above it was suggested that infinite geodesics should, at least heuristically, be thought of as 'highways to infinity' along which the different types may escape the competition. The concept of Busemann functions, and their properties, will be central in order to make this heuristic precise. These functions have their origin in the work of Herbert Busemann [7] on metric spaces. In first-passage percolation, Busemann-related limits first appeared in the work of Newman [28] as a means to

[2]This will be referred to as having *unique passage times*.

describe the microscopic structure of the boundary (or surface) of a growing ball $\{z \in \mathbb{Z}^2 : T(0, z) \leq t\}$ in the first-passage metric. Later work of Hoffman [19, 20] developed a method to describe asymptotic properties of geodesics via the study of Busemann functions. Hoffman's approach has since become indispensable in the study of various models for spatial growth, including first-passage percolation [1, 10, 11], the corner growth model [15, 16] and random polymers [2, 14]. In a tangential direction, Bakhtin et al. [5] used Busemann functions to construct stationary space-time solutions to the one-dimensional Burgers equation, inspired by earlier work of Cator and Pimentel [8].

Finally, we remark that (for $d = 2$) it is widely believed that the asymptotic shape is not a polygon, in which case it follows from [20] that both $\mathbb{P}(|\mathscr{T}_0| = \infty) = 1$ and for every $k \geq 1$ there are $x_1, x_2, \ldots, x_k$ such that $\mathbb{P}(\text{Coex}(x_1, x_2, \ldots, x_k)) > 0$. The latter was extended to infinite coexistence by Damron and Hochman [12]. Thus, proving that the asymptotic shape is non-polygonal would make our main theorem obsolete. However, understanding the asymptotic shape is a notoriously hard problem, which is the reason an approach sidestepping Newman's curvature assumption has been developed in the first place.

2 Geodesics and Busemann Functions

In this section we review the recent developments in the study of infinite geodesics in first-passage percolation. We shall focus on the two-dimensional setting, and remark on higher dimensions only at the end. We make no claim in providing a complete account of previous work, and instead prefer to focus on the results that will be of significance for the purposes of this paper. A more complete description of these results, save those reported in the more recent studies [1, 27], can be found in [4].

2.1 Geodesics in Newman's Contribution to the 1994 ICM Proceedings

The study of geodesics in first-passage percolation was pioneered by Newman and co-authors [25, 26, 28, 29] in the mid 1990s. Their work gave rise to a precise set of predictions for the structure of infinite geodesics. In order to describe these predictions we shall need some notation. First, we say that an infinite geodesic $g = (v_1, v_2, \ldots)$ has *asymptotic direction* θ, in the unit circle $S^1 := \{x \in \mathbb{R}^2 : |x| = 1\}$, if the limit $\lim_{k\to\infty} v_k/|v_k|$ exists and equals θ. Second, two infinite geodesics g and g' are said to *coalesce* if their symmetrical difference $g \Delta g'$ is finite. The predictions originating from the work of Newman and his collaborators can be summarized as, under mild conditions on the weight distribution, the following should hold:

(a) with probability one, every infinite geodesic has an asymptotic direction;
(b) for every direction θ, there is an almost surely unique geodesic in $\mathscr{T}_0$ with direction θ;
(c) for every direction θ, any two geodesics with direction θ coalesce almost surely.

In particular, these statements would imply that $|\mathscr{T}_0| = \infty$ almost surely.

Licea and Newman [25, 28] proved conditional versions of these statements under an additional curvature assumption of the asymptotic shape. While this assumption seems plausible for a large family of edge weight distributions, there is no known example for which it has been verified. Rigorous proofs of the corresponding statements for a rotation invariant first-passage-like model, where the asymptotic shape is known to be a Euclidean disc, has been obtained by Howard and Newman [22]. Since proving properties like strict convexity and differentiability of the boundary of the asymptotic shape in standard first-passage percolation appears to be a major challenge, later work has focused on obtaining results without assumptions on the shape.

2.2 Busemann Functions

Limits reminiscent of Busemann functions first appeared in the first-passage literature in the work of Newman [28], as a means of describing the microscopic structure of the boundary of a growing ball in the first passage metric. The method for describing properties of geodesics via Busemann functions developed in later work of Hoffman [19, 20].

Given an infinite geodesic $g = (v_1, v_2, \dots)$ in $\mathscr{T}_0$ we define the *Busemann function* $B_g : \mathbb{Z}^2 \times \mathbb{Z}^2 \to \mathbb{R}$ of g as the limit

$$B_g(x, y) := \lim_{k\to\infty} \big[T(x, v_k) - T(y, v_k)\big]. \tag{4}$$

As observed by Hoffman [19], with probability one the limit in (4) exists for every $g \in \mathscr{T}_0$ and all $x, y \in \mathbb{Z}^2$, and satisfies the following properties:

- $B_g(x, y) = B_g(x, z) + B_g(z, y)$ for all $x, y, z \in \mathbb{Z}^2$;
- $|B_g(x, y)| \le T(x, y)$;
- $B_g(x, y) = T(x, y)$ for all $x, y \in g$ such that $x \in \text{geo}(0, y)$.

In [19] Hoffman used Busemann functions to establish that there are at least two disjoint infinite geodesics almost surely. In [20] he used Busemann functions to associate certain infinite geodesics with sides (tangent lines) of the asymptotic shape. The approach involving Busemann functions in order to study infinite geodesics was later developed further in work by Damron and Hanson [10, 11] and Ahlberg and Hoffman [1]. Studying Busemann functions of geodesics, as opposed to the geodesics themselves, has allowed these authors to establish rigorous versions of Newman's predictions regarding the structure of geodesics. Describing parts of these results in detail will be essential in order to understand the duality between existence of geodesics and coexistence in competing first-passage percolation.

2.3 Linearity of Busemann Functions

We shall call a linear functional $\rho : \mathbb{R}^2 \to \mathbb{R}$ *supporting* if the line $\{x \in \mathbb{R}^2 : \rho(x) = 1\}$ is a supporting line to ∂Ball through some point, and *tangent* if $\{x \in \mathbb{R}^2 : \rho(x) = 1\}$ is the unique supporting line (i.e. the tangent line) through some point of ∂Ball. Given a supporting functional ρ and a geodesic $g \in \mathscr{T}_0$ we say that the Busemann function of g is *asymptotically linear* to ρ if

$$\limsup_{|y|\to\infty} \frac{1}{|y|}\big|B_g(0, y) - \rho(y)\big| = 0. \tag{5}$$

Asymptotic linearity of Busemann functions is closely related to asymptotic directions of geodesics in the sense that (5), together with the third of the properties of Busemann functions exhibited by Hoffman, provides information on the direction of $g = (v_1, v_2, \dots)$: The set of limit points of the sequence $(v_k/|v_k|)_{k\geq 1}$ is contained in the arc $\{x \in S^1 : \mu(x) = \rho(x)\}$, corresponding to a point or a flat edge of ∂Ball.

Building on the work of Hoffman [20], Damron and Hanson [10] showed that for every tangent line of the asymptotic shape there exists a geodesic whose Busemann function is described by the corresponding linear functional. In a simplified form their result reads as follows:

Theorem 3 *For every tangent functional $\rho : \mathbb{R}^2 \to \mathbb{R}^2$ there exists, almost surely, a geodesic in $\mathscr{T}_0$ whose Busemann function is asymptotically linear to ρ.*

While the work of Damron and Hanson proves *existence* of geodesics with linear Busemann functions, later work of Ahlberg and Hoffman [1] has established that *every* geodesic has a linear Busemann function, and that the associated linear functionals are *unique*. We summarize these results in the next couple of theorems.

Theorem 4 *With probability one, for every geodesic $g \in \mathscr{T}_0$ there exists a supporting functional $\rho : \mathbb{R}^2 \to \mathbb{R}$ such that the Busemann function of g is asymptotically linear to ρ.*

To address uniqueness, note that the set of supporting functionals is naturally parametrized by the direction of their gradients. Due to convexity of the shape, these functionals stand in 1-1 correspondence with the unit circle S^1. We shall from now on identify the set of supporting functionals with S^1.

Theorem 5 *There exists a closed (deterministic) set $\mathscr{C} \subseteq S^1$ such that, with probability one, the (random) set of supporting functionals ρ for which there exists* *a geodesic in $\mathscr{T}_0$ with Busemann function asymptotically linear to ρ equals $\mathscr{C}$. Moreover, for every $\rho \in \mathscr{C}$ we have*

$$\mathbb{P}\big(\exists \text{ two geodesics in } \mathscr{T}_0 \text{ with Busemann function linear to } \rho\big) = 0.$$

From Theorem 3 it follows that $\mathscr{C}$ contains all tangent functionals. As a consequence, if Ball has at least k sides (i.e. tangent lines), then we have $|\mathscr{T}_0| \geq k$ almost surely. On the other hand, it follows from Theorem 4 that every geodesic has a linear Busemann function, and by Theorem 5 that the set of linear functionals describing these Busemann functions is deterministic. Consequently, if with positive probability $\mathscr{T}_0$ has size at least k, then by the uniqueness part of Theorem 5 the set $\mathscr{C}$ has cardinality at least k, so that there exist k geodesics described by distinct linear functionals almost surely. All these observations will be essential in proving part *(ii)* of Theorem 1.

Due to the connection between asymptotic directions and linearity of Busemann functions mentioned above, Theorems 3–5 may be seen as rigorous, although somewhat weaker, versions of Newman's predictions *(a)–(b)*. The rigorous results are weaker in the sense that we do not know whether $\mathscr{C}$ equals S^1 or not. Note, however, that Theorem 5 provides an 'ergodic theorem' in this direction. As we shall describe next, the cited papers provide a rigorous version also of *(c)*.

2.4 Coalescence

An aspect of the above development that we have ignored so far is that of coalescence. For instance, Theorem 3 is a simplified version of a stronger statement proved in [10], namely that for every tangent functional $\rho : \mathbb{R}^2 \to \mathbb{R}$ there exists, almost surely, a family of geodesics $\Gamma = \{\gamma_z : z \in \mathbb{Z}^2\}$, where $\gamma_z \in \mathscr{T}_z$, such that any one geodesic in Γ has Busemann function linear to ρ and any two geodesics in Γ coalesce. (The latter of course implies that the Busemann functions of all geodesics in Γ coincide.) In a similar spirit, we have the following from [1]:

Theorem 6 *For every supporting functional $\rho \in \mathscr{C}$, with probability one, any two geodesics $g \in \mathscr{T}_y$ and $g' \in \mathscr{T}_z$ with Busemann function asymptotically linear to ρ coalesce.*

We remark that coalescence was irrelevant for the proof of Theorem 3 in [10], but instrumental for the deduction of Theorems 4 and 5 in [1]. In short, the importance of coalescence lies in the possibility to apply the ergodic theorem to asymptotic properties of shift invariant families of coalescing geodesics, resulting in the ergodic properties of Theorem 5.

The results described above together address the cardinality of the set $\mathscr{T}_0$. Recall that $\mathscr{N}$ denotes the maximal number of pairwise disjoint infinite geodesics and is almost surely constant. The following was first established in [1], and can be derived as a corollary to Theorems 4–6. A more direct argument, assuming a stronger moment condition, was later given by Nakajima [27].

Corollary 1 *With probability one $|\mathscr{T}_0|$ is constant and equal to $\mathscr{N}$.*

To see how the corollary follows, first note that clearly $|\mathscr{T}_0| \le \mathscr{N}$. In addition, $|\mathscr{C}| \le |\mathscr{T}_0|$ almost surely due to the ergodic part of Theorem 5, and in the case that $\mathscr{C}$ is finite, equality follows from Theorem 4 and the uniqueness part of Theorem 5. Consequently, also $|\mathscr{T}_0|$ is almost surely constant. Finally, it follows from the coalescence property in Theorem 6 that either $|\mathscr{T}_0|$ (and therefore also $\mathscr{N}$) is almost surely infinite, or $|\mathscr{T}_0| = \mathscr{N} = k$ holds almost surely for some finite k, leading to the claimed result.

2.5 Geodesics in Higher Dimensions

Whether the description of geodesics detailed above remains correct also in higher dimensions is at this point unknown. Although it has been suggested that coalescence should fail for large d, it seems plausible that results analogous to Theorems 3–5 should hold for all $d \ge 2$, and that an analogue to Theorem 6 could hold for small d. See recent work of Alexander [3] for a further discussion of these claims. Indeed, establishing the existence of coalescing families of geodesics in the spirit of [10] also in three dimensions should be considered a major open problem.

What is known is that the argument behind Theorem 3 can be extended to all dimensions $d \ge 2$ under minor adjustments; see [6]. However, the proofs of Theorems 4–6 exploit planarity in a much more fundamental way, and are not known to extend to higher dimensions. On the other hand, an argument of Nakajima [27] shows that Corollary 1 remains valid in all dimensions under the additional condition that $\mathbb{E}[\exp(\alpha\omega_e)] < \infty$ for some $\alpha > 0$. These properties will be sufficient in order to prove Theorem 2.

3 The Dual Relation in Two Dimensions

With the background outlined in the previous section we are now ready to prove Theorem 1. We recall that, with probability one, by Theorem 4 every geodesic has an asymptotically linear Busemann function, and by Theorem 5 there is a deterministic set $\mathscr{C}$ of linear functionals that correspond to these Busemann functions. Moreover, for each $\rho \in \mathscr{C}$, by Theorem 5 there is for every $z \in \mathbb{Z}^2$ an almost surely unique geodesic in $\mathscr{T}_z$ with Busemann function asymptotically linear to ρ, and by Theorem 6 these geodesics coalesce almost surely. In particular $|\mathscr{T}_0| = |\mathscr{C}|$ almost surely, and we shall in the sequel write B_ρ for the Busemann function of the almost surely unique geodesic (in $\mathscr{T}_0$) corresponding to ρ.

3.1 Part (i): Coexistence Implies Existence

The short proof of part *(i)* is an easy consequence of Corollary 1. Suppose that for some choice of $x_1, x_2, \dots, x_k$ in $\mathbb{Z}^2$ we have $\mathbb{P}(\mathrm{Coex}(x_1, x_2, \dots, x_k)) > 0$. By (3) we have $\mathbb{P}(\mathscr{N} \geq k) > 0$, and since $\mathscr{N}$ is almost surely constant it follows from Corollary 1 that

$$\mathbb{P}(|\mathscr{T}_0| \geq k) = 1.$$

While the above argument is short, it hides much of the intuition for why the implication holds. We shall therefore give a second argument based on coalescence that may be more instructive, even if no more elementary. This argument will make explicit the heuristic that geodesics are 'highways to infinity' along which the different types will have to move in order to escape the competition.

Before attending to the proof, we claim that for any $\rho \in \mathscr{C}$ we have

$$\mathbb{P}\big(B_\rho(x, y) \neq 0 \text{ for all } x \neq y\big) = 1. \tag{6}$$

To see this, let A_ρ denote the event that for each z in $\mathbb{Z}^2$ there is a unique geodesic g_z in $\mathscr{T}_z$ corresponding to ρ, and that all these geodesics coalesce, so that A_ρ has measure one. We note that on the event A_ρ coalescence of the geodesics $\{g_z : z \in \mathbb{Z}^2\}$ implies that for any $x, y \in \mathbb{Z}^2$ the limit $B_\rho(x, y)$ (which is defined through (4) for $g = g_0$) is attained after a finite number of steps. More precisely, on the event A_ρ, for any $x, y \in \mathbb{Z}^2$ and v contained in $g_x \cap g_y$ we have

$$B_\rho(x, y) = T(x, v) - T(y, v).$$

Hence, (6) follows due to unique passage times.

We now proceed with the second proof. Again by Corollary 1, either $\mathscr{T}_0$ is almost surely infinite, in which case there is nothing to prove, or $\mathbb{P}(|\mathscr{T}_0| = k) = 1$ for some integer $k \geq 1$. We shall suppose the latter, and argue that for any choice of $x_1, x_2, \dots, x_{k+1}$ in $\mathbb{Z}^2$ we have $\mathbb{P}(\mathrm{Coex}(x_1, x_2, \dots, x_{k+1})) = 0$.

On the event that $\mathscr{T}_0$ is almost surely finite, $\mathscr{C}$ is in one-to-one correspondence with the elements of $\mathscr{T}_0$. It follows from (6) that for any $g \in \mathscr{T}_0$ the Busemann function $B_g(0, x)$ has a unique minimizer over finite subsets of $\mathbb{Z}^2$ almost surely. The last statement can be rephrased in terms of competition between a finite number of types as follows: For each geodesic g in $\mathscr{T}_0$ there will be precisely one type that reaches infinitely many sites along g almost surely; it is the one whose starting position minimizes $B_g(0, x_i)$. Hence, if $|\mathscr{T}_0| = k$ almost surely, but there are $k + 1$ competing types, then at least one of them will not reach infinitely many sites along any geodesic in $\mathscr{T}_0$. Suppose that the type left out starts at a site x. Since for each geodesic in $\mathscr{T}_x$ there is a geodesic in $\mathscr{T}_0$ with which it coalesces (as of Theorem 6), it follows that for each geodesic $g \in \mathscr{T}_x$ the type starting at x will be closer than the other types to at most finitely many sites along g. Choose n so that these sites are all within distance n from x. Finally, note that for at most finitely many sites z in $\mathbb{Z}^2$ the (finite) geodesic from x to z will diverge from all geodesics in $\mathscr{T}_x$ within distance n from x. Consequently, all but finitely many sites in $\mathbb{Z}^2$ will lie closer to the starting point of some other type, implying that the $k+1$ types do not coexist.

3.2 Part (ii): Existence Implies Coexistence

Central in the proof of part *(ii)* is the linearity of Busemann functions. The argument that follows is a modern take on an argument originally due to Hoffman [20].

Let k be an integer and suppose that $|\mathscr{T}_0| \geq k$ with positive probability. Then, indeed, $|\mathscr{T}_0| = |\mathscr{C}| \geq k$ almost surely. Fix $\varepsilon > 0$ and let $\rho_1, \rho_2, \ldots, \rho_k$ be distinct elements of $\mathscr{C}$. In order to show that $\mathbb{P}(\mathrm{Coex}(x_1, x_2, \ldots, x_k)) > 1 - \varepsilon$ for some choice of $x_1, x_2, \ldots, x_k$, we shall choose these points so that with probability $1 - \varepsilon$ we have $B_{\rho_i}(x_i, x_j) < 0$ for all $i = 1, 2, \ldots, k$ and $j \neq i$. On this event, for each i, the site x_i is closer to all points along the geodesic in $\mathscr{T}_{x_i}$ corresponding to ρ_i than any of the x_j for $j \neq i$, implying that $\mathrm{Coex}(x_1, x_2, \ldots, x_k)$ occurs.

Given $\rho \in \mathscr{C}$, $z \in \mathbb{Z}^2$, $\delta > 0$ and $M \geq 1$ we let $A_\rho(z, \delta, M)$ denote the event that

$$\big|B_\rho(z, y) - \rho(y - z)\big| < \delta|y - z| \quad \text{for all } |y - z| \geq M.$$

Due to linearity of Busemann functions (Theorems 4 and 5) there exists for every $\rho \in \mathscr{C}$ and $\delta, \gamma > 0$ an $M < \infty$ such that

$$\mathbb{P}\big(A_\rho(z, \delta, M)\big) > 1 - \gamma \quad \text{for every } z \in \mathbb{Z}^2. \tag{7}$$

We further introduce the following notation for plane regions related to ρ:

$$H_\rho(z, \delta) := \big\{y \in \mathbb{R}^2 : \rho(y - z) \leq -\delta|y - z|\big\};$$
$$C_\rho(z, \delta) := \big\{y \in \mathbb{R}^2 : |\rho(y - z)| \leq \delta|y - z|\big\}.$$

Note that on the event $A_\rho(z, \delta, M)$ we have for all $y \in H_\rho(z, \delta)$ such that $|y - z| \geq M$ that $B_\rho(z, y) < 0$. Hence, $H_\rho(z, \delta)$ corresponds to sites that are likely to be at a further distance to far out vertices along the geodesic corresponding to ρ as compared to z.

Given $\rho_1, \rho_2, \ldots, \rho_k$ we now choose $\delta > 0$ so that the cones $C_{\rho_i}(0, \delta)$, for $i = 1, 2, \ldots, k$, intersect only at the origin. Next, we choose M large so that for all i

$$\mathbb{P}\big(A_{\rho_i}(z, \delta, M)\big) > 1 - \varepsilon/k.$$

Finally, due to the choice of δ we may choose $x_1, x_2, \ldots, x_k$ so that $|x_i - x_j| \geq M$ for all $i \neq j$ and such that for each i the set $H_{\rho_i}(x_i, \delta)$ contains x_j for all $j \neq i$. (For instance, position the sites on a circle of large radius, in positions roughly corresponding to the directions of $\rho_1, \rho_2, \ldots, \rho_k$.) Due to these choices we will on the event $\bigcap_{i=1,2,\ldots,k} A_{\rho_i}(x_i, \delta, M)$, which occurs with probability at least $1-\varepsilon$, have for all $i = 1, 2, \ldots, k$ that $B_{\rho_i}(x_i, x_j) < 0$ for all $j \neq i$, as required.

It remains to show that if $|\mathscr{T}_0| = \infty$ with positive probability, then it is possible to find a sequence $(x_i)_{i\geq 1}$ for which $\mathrm{Coex}(x_1, x_2, \ldots)$ occurs with probability close to one. If $|\mathscr{T}_0| = \infty$ with positive probability, then it does with probability one, and $|\mathscr{C}| = \infty$ almost surely. Let $(\rho_i)_{i\geq 1}$ be an increasing sequence in $\mathscr{C}$ (considered as a sequence in $[0, 2\pi)$). By symmetry of $\mathbb{Z}^2$ we may assume that each ρ_i corresponds to an angle in $(0, \pi/2)$. Fix $\varepsilon > 0$ and set $\varepsilon_i = \varepsilon/2^i$. We choose δ_1 so that $C_{\rho_1}(0, \delta_1)$ intersect each of the lines $C_{\rho_j}(0, 0)$, for $j \geq 2$, only at the origin, and M_1 so that $\mathbb{P}(A_{\rho_1}(z, \delta_1, M_1)) > 1 - \varepsilon_1$. Inductively we choose δ_i so that $C_{\rho_i}(0, \delta_i)$ intersects each cone $C_{\rho_j}(0, \delta_j)$ for $j < i$ and each line $C_{\rho_j}(0, 0)$ for $j > i$ only at the origin, and M_i so that $\mathbb{P}(A_{\rho_i}(z, \delta_i, M_i)) > 1 - \varepsilon_i$. For any sequence $(x_i)_{i\geq 1}$ we have

$$\mathbb{P}\Big(\bigcap_{i\geq 1} A_{\rho_i}(x_i, \delta_i, M_i)\Big) > 1 - \varepsilon.$$

It remains only to verify that we may choose the sequence $(x_i)_{i\geq 1}$ so that for each $i \geq 1$ we have $|x_i - x_j| \geq M_i$ and $x_j \in H_{\rho_i}(x_i, \delta_i)$ for all $j \neq i$. For $i \geq 1$ we take $v_{i+1} \in \mathbb{Z}^2$ such that $|v_{i+1}| > \max\{M_1, M_2, \ldots, M_{i+1}\}$, $\rho_{i+1}(v_{i+1}) > \delta_{i+1}|v_{i+1}|$ and $\rho_j(v_{i+1}) < -\delta_j|v_{i+1}|$ for all $j \leq i$. We note that this is possible since the sequence $(\rho_i)_{i\geq 1}$ is increasing and the cone-shaped regions $C_{\rho_i}(0, \delta_i)$ and $C_{\rho_j}(0, \delta_j)$ for $i \neq j$ intersect only at the origin. Finally, take $x_1 = (0, 0)$, and for $i \geq 1$ set $x_{i+1} = x_i + v_{i+1}$.

4 Partial Duality in Higher Dimensions

The proof of Theorem 2 is similar to that of Theorem 1. So, instead of repeating all details we shall only outline the proof and indicate at what instances our current understanding of the higher dimensional case inhibits us from deriving the full duality. In the sequel we assume $d \geq 2$.

The proof of the first part of the theorem is completely analogous. Suppose that

$$\mathbb{P}\big(\mathrm{Coex}(x_1, x_2, \ldots, x_k)\big) > 0$$

for some choice of $x_1, x_2, \ldots, x_k$ in $\mathbb{Z}^d$, possibly infinitely many. Then $\mathscr{N} \geq k$ almost surely, and by (Nakajima's version, which requires an exponential moment assumption, of) Corollary 1 we have $|\mathscr{T}_0| \geq k$ almost surely.

For the second part of the argument we will need to modify slightly the approach from the two dimensional case. In the general case we do not know that every geodesic has an asymptotically linear Busemann function. However, from (the higher dimensional version of) Theorem 3 we know that if the shape has at least

k sides (that is, tangent hyperplanes), then, almost surely, there are k geodesics in $\mathscr{T}_0$ which all have asymptotically linear Busemann functions described by different linear functionals. Based on this we may repeat the proof of part *(ii)* of Theorem 1 to obtain coexistence of k types with probability arbitrarily close to one.

In the case the shape has infinitely many sides, then with probability one there are infinitely many geodesics in $\mathscr{T}_0$ with asymptotically linear Busemann functions, all described by different linear functionals. Let $(\rho_i)_{i\geq 1}$ be a sequence of such linear functionals. Denote by L_i the intersection of the hyperplane $\{x \in \mathbb{R}^d : \rho_i(x) = 0\}$ and the x_1x_2-plane, i.e., the plane spanned by the first two coordinate vectors. Each L_i has dimension zero, one or two, and by exploiting the symmetries of $\mathbb{Z}^d$ we may assume that sequence $(\rho_i)_{i\geq 1}$ is chosen so that they all have dimension one. Each L_i is then a line through the origin in the x_1x_2-plane, and by restricting to a subsequence we may assume that the sequence $(\nu_i)_{i\geq 1}$ of normal vectors of these lines is monotone (considered as elements in $[0, 2\pi)$). We may now proceed and select a sequence of points $(x_i)_{i\geq 1}$ in the x_1x_2-plane in an analogous manner as in the two-dimensional case, leading to coexistence of infinitely many types with probability arbitrarily close to one.

Acknowledgments The author would like to express his gratitude for the years of encouragement and support received from Vladas Sidoravicius, to whom the volume in which this text appears is dedicated to. The author is also grateful for the detailed comments received from an anonymous referee. This work was in part supported by the Swedish Research Council (VR) through grant 2016-04442.

References

1. Ahlberg, D., Hoffman, C.: Random coalescing geodesics in first-passage percolation. Preprint, see arXiv:1609.02447
2. Alberts, T., Rassoul-Agha, F., Simper, M.: Busemann functions and semi-infinite O'Connell–Yor polymers. Bernoulli **26**(3), 1927–1955 (2020)
3. Alexander, K.S.: Geodesics, bigeodesics, and coalescence in first passage percolation in general dimension. Preprint, see arXiv:2001.08736
4. Auffinger, A., Damron, M., Hanson, J.: 50 Years of First-Passage Percolation, Volume 68 of University Lecture Series. American Mathematical Society, Providence (2017)
5. Bakhtin, Y., Cator, E., Khanin, K.: Space-time stationary solutions for the Burgers equation. J. Am. Math. Soc. **27**(1), 193–238 (2014)
6. Brito, G., Damron, M., Hanson, J.: Absence of backward infinite paths for first-passage percolation in arbitrary dimension. Preprint, see arXiv:2003.03367
7. Busemann, H.: The Geometry of Geodesics. Academic, New York (1955)
8. Cator, E., Pimentel, L.P.R.: Busemann functions and equilibrium measures in last passage percolation models. Probab. Theory Relat. Fields **154**(1–2), 89–125 (2012)
9. Cox, J.T., Durrett, R.: Some limit theorems for percolation processes with necessary and sufficient conditions. Ann. Probab. **9**(4), 583–603 (1981)
10. Damron, M., Hanson, J.: Busemann functions and infinite geodesics in two-dimensional first-passage percolation. Commun. Math. Phys. **325**(3), 917–963 (2014)
11. Damron, M., Hanson, J.: Bigeodesics in first-passage percolation. Commun. Math. Phys. **349**(2), 753–776 (2017)

12. Damron, M., Hochman, M.: Examples of nonpolygonal limit shapes in i.i.d. first-passage percolation and infinite coexistence in spatial growth models. Ann. Appl. Probab. **23**(3), 1074–1085 (2013)
13. Garet, O., Marchand, R.: Coexistence in two-type first-passage percolation models. Ann. Appl. Probab. **15**(1A), 298–330 (2005)
14. Georgiou, N., Rassoul-Agha, F., Seppäläinen, T.: Variational formulas and cocycle solutions for directed polymer and percolation models. Commun. Math. Phys. **346**(2), 741–779 (2016)
15. Georgiou, N., Rassoul-Agha, F., Seppäläinen, T.: Geodesics and the competition interface for the corner growth model. Probab. Theory Relat. Fields **169**(1–2), 223–255 (2017)
16. Georgiou, N., Rassoul-Agha, F., Seppäläinen, T.: Stationary cocycles and Busemann functions for the corner growth model. Probab. Theory Relat. Fields **169**(1–2), 177–222 (2017)
17. Häggström, O., Pemantle, R.: First passage percolation and a model for competing spatial growth. J. Appl. Probab. **35**(3), 683–692 (1998)
18. Hammersley, J.M., Welsh, D.J.A.: First-passage percolation, subadditive processes, stochastic networks, and generalized renewal theory. In: Proc. Internat. Res. Semin., Statist. Lab., Univ. California, Berkeley, Calif, pp. 61–110. Springer, New York (1965)
19. Hoffman, C.: Coexistence for Richardson type competing spatial growth models. Ann. Appl. Probab. **15**(1B), 739–747 (2005)
20. Hoffman, C.: Geodesics in first passage percolation. Ann. Appl. Probab. **18**(5), 1944–1969 (2008)
21. Howard, C.D.: Models of first-passage percolation. In: Probability on Discrete Structures, Volume 110 of Encyclopaedia Math. Sci., pp. 125–173. Springer, Berlin (2004)
22. Howard, C.D., Newman, C.M.: Geodesics and spanning trees for Euclidean first-passage percolation. Ann. Probab. **29**(2), 577–623 (2001)
23. Kesten, H.: Aspects of first passage percolation. In: École d'été de probabilités de Saint-Flour, XIV—1984, Volume 1180 of Lecture Notes in Math., pp. 125–264. Springer, Berlin (1986)
24. Kingman, J.F.C.: The ergodic theory of subadditive stochastic processes. J. R. Stat. Soc. Ser. B **30**, 499–510 (1968)
25. Licea, C., Newman, C.M.: Geodesics in two-dimensional first-passage percolation. Ann. Probab. **24**(1), 399–410 (1996)
26. Licea, C., Newman, C.M., Piza, M.S.T.: Superdiffusivity in first-passage percolation. Probab. Theory Relat. Fields **106**(4), 559–591 (1996)
27. Nakajima, S.: Ergodicity of the number of infinite geodesics originating from zero. Preprint, see arXiv:1807.05900
28. Newman, C.M.: A surface view of first-passage percolation. In: Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994), pp. 1017–1023. Birkhäuser, Basel (1995)
29. Newman, C.M., Piza, M.S.T.: Divergence of shape fluctuations in two dimensions. Ann. Probab. **23**(3), 977–1005 (1995)
30. Richardson, D.: Random growth in a tessellation. Proc. Camb. Philos. Soc. **74**, 515–528 (1973)
31. Rockafellar, R.T.: Convex Analysis. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton (1970)
32. Smythe, R.T., Wierman, J.C.: First-Passage Percolation on the Square Lattice, Volume 671 of Lecture Notes in Mathematics. Springer, Berlin (1978)

Ground State Stability in Two Spin Glass Models

L.-P. Arguin, C. M. Newman, and D. L. Stein

Abstract An important but little-studied property of spin glasses is the stability of their ground states to changes in one or a finite number of couplings. It was shown in earlier work that, if multiple ground states are assumed to exist, then fluctuations in their energy differences—and therefore the possibility of multiple ground states—are closely related to the stability of their ground states. Here we examine the stability of ground states in two models, one of which is presumed to have a ground state structure that is qualitatively similar to other realistic short-range spin glasses in finite dimensions.

Keywords Spin glass · Highly disordered model · Strongly disordered model · Critical droplets

L.-P. Arguin
Department of Mathematics, City University of New York, Baruch College and Graduate Center, New York, NY, USA
e-mail: louis-pierre.arguin@baruch.cuny.edu

C. M. Newman
Courant Institute of Mathematical Sciences, New York, NY, USA

NYU-ECNU Institute of Mathematical Sciences at NYU Shanghai, Shanghai, China
e-mail: newman@cims.nyu.edu

D. L. Stein (✉)
Department of Physics and Courant Institute of Mathematical Sciences, New York University, New York, NY, USA

NYU-ECNU Institutes of Physics and Mathematical Sciences at NYU Shanghai, Shanghai, China
Santa Fe Institute, Santa Fe, NM, USA
e-mail: daniel.stein@nyu.edu

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_2

1 Introduction and Definitions

Vladas was a remarkable mathematician, collaborator, colleague and friend: often exciting, always interesting, sometimes frustrating but never boring. We will miss him greatly, but are confident that his memory will survive for a very long time.

Although he never worked directly on spin glasses himself, Vladas maintained a longstanding interest in the problem, and we enjoyed numerous discussions with him about possible ways of proving nonuniqueness of Gibbs states, energy fluctuation bounds, overlap properties, and many other open problems. In this paper we discuss another aspect of spin glasses, namely ground state stability and its consequences, a topic we think Vladas would have enjoyed.

The stability of a spin glass ground state can be defined in different ways; here we will adopt the notion introduced in [14, 15] and further developed and exploited in [3, 5]. For specificity consider the Edwards-Anderson (EA) Ising model [8] in a finite volume $\Lambda_L = [-L, L]^d \cap \mathbb{Z}^d$ centered at the origin, with Hamiltonian

$$H_{\Lambda,J}(\sigma) = -\sum_{\langle xy\rangle \in E(\Lambda)} J_{xy}\sigma_x\sigma_y, \qquad \sigma \in \{-1, 1\}^{\Lambda}, \tag{1}$$

where $E(\Lambda)$ denotes the set of nearest-neighbor edges $\langle xy\rangle$ with both endpoints in Λ. The couplings J_{xy} are i.i.d. random variables sampled from a continuous distribution $\nu(dJ_{xy})$, which for specificity we take to be $\mathcal{N}(0, 1)$. If periodic or free boundary conditions are imposed, ground states appear as spin-reversed pairs.

For any fixed Λ_L and accompanying boundary condition, the ground state configuration (or the ground state pair if the boundary condition has spin-flip symmetry) is denoted by α. One may now ask the question, how does the lowest-energy spin configuration α change when one selects an arbitrary edge b_0 and varies its associated coupling J_0 from $-\infty$ to $+\infty$? If J_0 is satisfied, increasing its magnitude will only increase the stability of α and so the lowest-energy spin configuration pair is unchanged. However, if its magnitude is decreased, α becomes less stable, and there exists a specific value J_c for which a cluster of connected spins (which we shall refer to as the "critical droplet") will flip, leading to a new ground state pair α'. The same result follows if J_0 is unsatisfied and its magnitude is then increased.

More precisely, note that a ground state pair (hereafter GSP) is a spin configuration such that the energy $E_{\partial\mathcal{D}}$ of any closed surface $\partial\mathcal{D}$ in the dual lattice satisfies the condition

$$E_{\partial\mathcal{D}} = \sum_{\langle xy\rangle \in \partial\mathcal{D}} J_{xy}\sigma_x\sigma_y > 0\,. \tag{2}$$

The critical value J_c corresponds to the coupling value at which $\sum_{\langle xy\rangle \in \partial\mathcal{D}} J_{xy}\sigma_x\sigma_y = 0$ in α for a *single* closed surface whose boundary passes through b_0, while all other such closed surfaces satisfy (2). The cluster of spins enclosed by the zero-energy surface $\partial\mathcal{D}_c(b_0, \alpha)$ is denoted the "critical droplet" of b_0 in the GSP α. Because the couplings are i.i.d., J_c depends on α and all coupling values *except* that associated with b_0; that is, the critical value J_c is independent of J_0. For a fixed coupling realization in which $J(b_0) = J_0$, we can therefore define the *flexibility* $\mathscr{F}_{b_0,\alpha}$ of b_0 in α as

$$\mathscr{F}_{b_0,\alpha} = |E_{\partial\mathcal{D}_c(b_0,\alpha)}(J_c) - E_{\partial\mathcal{D}_c(b_0,\alpha)}(J_0)| \,. \tag{3}$$

Because the couplings are i.i.d. and drawn from a continuous distribution, all flexibilities are strictly positive with probability one.

The presentation just given is informal; a complete discussion requires use of the *excitation metastate* [2–4, 16] which we omit here for the sake of brevity. A precise definition of the above concepts and quantities can be found in [5].

The concepts of critical droplets and flexibilities for a particular GSP in a fixed coupling realization provide a foundation for quantifying (at least one version of) the stability of a given ground state. From an energetic standpoint, one can consider, e.g., the distribution of flexibilities over all bonds. One can also approach the problem from a geometric perspective, by considering the sizes and geometries of the critical droplets associated with each of the bonds. This latter approach has recently proved to be useful, in that the distribution of critical droplet sizes has been shown [5] to be closely related to the energy fluctuations associated with collections of incongruent GSP's, i.e., GSP's whose mutual interfaces comprise a positive fraction of all edges in the infinite-volume limit [9, 10].

The problem with this approach, for now at least, is that there currently exist no tools or insights into determining ground state stability properties in ordinary EA models. In this paper we discuss two models, one of which should belong in the same universality class as the ordinary EA model, in which some information on these properties *can* be determined.

2 The Highly Disordered Model

2.1 Definition and Properties

The highly disordered model was introduced in [12, 13] (see also [6]). It is an EA-type model defined on the lattice $\mathbb{Z}^d$ whose Hamiltonian in any $\Lambda \subset \mathbb{Z}^d$ is still given by (1); the difference is that now the coupling distribution is volume-dependent even though the coupling values remain i.i.d. for each Λ. The idea is to "stretch out" the coupling distribution so that, with probability one, in sufficiently large volumes each coupling magnitude occurs on its own scale. More precisely, each coupling magnitude is at least twice as large as the next smaller one and no more than half as large as the next larger one.

While there are many possibilities for the volume-dependent distribution of couplings, we have found it convenient to work with the following choice. First, we associate two new i.i.d. random variables with each edge $\langle xy \rangle$: $\epsilon_{xy} = \pm 1$ with equal probability and K_{xy} which is uniformly distributed in the closed interval [0, 1]. We then define the set of couplings $J_{xy}^{(L)}$ within Λ_L as follows:

$$J_{xy}^{(L)} = c_L \epsilon_{xy} e^{-\lambda^{(L)} K_{xy}} , \tag{4}$$

where c_L is a scaling factor chosen to ensure a sensible thermodynamic limit (but which plays no role in ground state selection), and $\lambda^{(L)}$ is a scaling parameter that grows quickly enough with L to ensure that the condition described at the end of the previous paragraph holds. It was shown in [13] that $\lambda^{(L)} \geq L^{2d+1+\delta}$ for any $\delta > 0$ is a sufficient condition.

We should emphasize that although the couplings $J_{xy}^{(L)}$ depend on L, the K_{xy}'s and ϵ_{xy}'s do not; hence there is a well-defined infinite-volume notion of ground states for the highly disordered model on all of $\mathbb{Z}^d$. This is the subject of the theorem in the next subsection.

When the highly disordered condition is satisfied, the problem of finding ground states becomes tractable; in fact, a simple greedy algorithm provides a fast and efficient way to find the exact ground state in a fixed volume with given boundary conditions [12, 13]. Moreover, the ground state problem can be mapped onto invasion percolation [12, 13] which facilitates analytic study. It was further shown in [12, 13] that in the limit of infinite volume the highly disordered model has a single pair of ground states in low dimension, and uncountably many pairs in high dimension. The crossover dimension was found to be six in [11]. It should be noted that this result, related to the minimal spanning tree, is rigorous only in dimension two (or in quasi-planar lattices [18]).

The details of ground state structure in the highly disordered model have been described at length in [12, 13] (see also [11, 17]) and are not recounted here. In this contribution we present a new result, concerning the ground state stability of the highly disordered model, where it turns out that this model is tractable as well. The result we prove below is twofold: first, that with probability one *all* couplings have finite critical droplets in any ground state, and moreover this result is dimension-independent, and therefore independent of ground state pair multiplicity. We caution, however, that (as with all other results pertaining to this model) these results may be confined to the highly disordered model alone and have not been shown to carry over to the Edwards-Anderson or other realistic spin glass models. We will address this question more in the following section.

Before proceeding, we need to introduce some relevant properties and nomenclature pertaining to the highly disordered model. One of its distinguishing features—and the central one for our purposes—is the separation of all bonds into two distinct classes [12, 13]. The first class, which we denote as *S1 bonds* are those that are satisfied in any ground state *regardless* of the sign of the coupling, i.e., that of ϵ_{xy}. These are bonds that are always satisfied, in every ground state. The remaining bonds, which we call *S2*, are those in which a change of sign of their ϵ_{xy} value changes their status in any ground state from satisfied to unsatisfied or vice-versa. (Obviously, any unsatisfied bond in any ground state is automatically S2, but a satisfied bond could *a priori* be of either type.)

To make this distinction formal, we introduce the concept of *rank*: In a given Λ_L, the coupling with largest magnitude (regardless of sign) has rank one (this is the coupling with highest rank and the smallest value of K_{xy}); the coupling with the next largest magnitude has rank two; and so on. We then define an S1 bond as follows:

A bond $\langle xy \rangle$ is S1 in Λ_L if it has greater rank than at least one coupling in any path (excluding the bond itself) that connects its two endpoints **x** and **y**.

In the above definition, we need to specify what is meant by a path if each endpoint connects to a point on the boundary. For fixed boundary conditions of the spins on $\partial\Lambda_L$, all points on the boundary are considered connected (often called wired boundary conditions), so disjoint paths from **x** to $\partial\Lambda_L$ and **y** to $\partial\Lambda_L$ are considered as connecting **x** and **y**. It follows from the definition that for wired boundary conditions an S1 bond in Λ_L remains S1 in all larger volumes. These bonds completely determine the ground state configurations, while the S2 bonds play no role.[1] For free boundary conditions, a path connects **x** and **y** only if it stays entirely within Λ_L, never touching the boundary; i.e., points on the boundary are no longer considered connected. For periodic or antiperiodic boundary conditions, boundary points are considered connected to their image points but to no others. The reasons for these distinctions are provided in [13], but are not relevant to the present discussion and are presented only for completeness.

It was proved in [12] and [13] that the set of all S1 bonds forms a union of trees, that every site belongs to some S1 tree, and that every S1 tree touches the boundary of Λ_L. The S1 bonds in a given Λ_L in some fixed dimension form either a single tree or else a union of disjoint trees. Although not immediately obvious, it was proved in [12, 13] that the tree structure has a natural infinite volume limit, and moreover every tree is infinite. Moreover, a result from Alexander [1], adapted to the current context, states that if the corresponding independent percolation model has no infinite cluster at p_c, then from every site there is a single path to infinity along S1 edges; i.e., there are no doubly-infinite paths. It is widely believed that in independent percolation there is no infinite cluster at p_c in any dimension, but this has not yet been proven rigorously for $3 \le d \le 10$.

Finally, combined with results of Jackson and Read [11], we have that below six dimensions there is a single S1 tree spanning the sites of $\mathbb{Z}^d$ (corresponding to a

[1] One can define S1 and S2 bonds for the EA model as well, in the sense that the EA model also possesses bonds that are satisfied in every ground state (though the precise definition used above no longer applies). There are of course far fewer of these in the EA model than in the highly disordered model, and there is no evidence that these "always satisfied" bonds play any special role in ground state selection in that model. (One possibly relevant result, that unsatisfied edges don't percolate in the ground state, was proved in [7].)

single pair of ground states), while above six dimensions that are infinitely many trees (corresponding to an uncountable infinity of ground states).

2.2 Ground State Stability in the Highly Disordered Model

Unlike in realistic spin glass models, the ground state structure in the highly disordered model can be analyzed and understood in great detail. This allows us to solve other, related properties of the model, in particular some of the critical droplet properties that have so far been inaccessible in most other spin glass models. In particular we can prove the following result:

Theorem 1 *In the highly disordered model on the infinite lattice $\mathbb{Z}^d$ in any d, if there is no percolation at p_c in the corresponding independent bond percolation model, then for a.e. realization of the couplings, any ground state α, and any bond b_0, the critical droplet boundary $\partial\mathscr{D}_c(b_0, \alpha)$ is finite. Correspondingly, in finite volumes Λ_L with sufficiently large L, the size of the droplet is independent of L.*

Remark As noted above, it has been proved that there is no percolation at p_c in the corresponding independent bond percolation model in all dimensions except $3 \leq d \leq 10$, but it is widely believed to be true in all finite dimensions. Theorem 2 does not specify the distribution of critical droplet boundary sizes, which is potentially relevant especially for larger critical droplets, although integrability of the distribution requires a weak upper bound falloff such as $O(L^{-(1+\epsilon)})$, $\epsilon > 0$, for large L.

Proof Choose an arbitrary S1 bond and a volume sufficiently large so that the tree it belongs to has the following property: The branch emanating from one of its endpoints (call it x_1) touches the boundary (on which we apply fixed boundary conditions) and the branch emanating from the other endpoint (x_2) does not. This remains the case as the boundary moves out to infinity: for any S1 bond and a sufficiently large volume, this is guaranteed to be the case by the result of Alexander mentioned above [1].

We use the fact, noted in Sect. 1, that as the coupling value of any bond varies from $-\infty$ to $+\infty$ while all other couplings are held fixed, there is a single, well-defined critical point at which a unique cluster of spins, i.e., the critical droplet, flips, changing the ground state. (This is true regardless of whether one is considering a finite volume with specified boundary condition or the infinite system.) Now keep the magnitude of the S1 bond fixed but change its sign. Because the S1 bond must still be satisfied, this must cause a droplet flip, which as noted above must be the critical droplet.

Now consider the state of the spins at either endpoint of the bond. Suppose that originally the bond was ferromagnetic, and the spins at x_1 and x_2 were both $+1$. After changing the sign of the coupling, the spin at x_1, remains $+1$ (because it is connected to the boundary, as explained in the first paragraph of this proof) while the spin at x_2 is now -1. This must simultaneously flip all the spins on the branch of the tree connected to x_2. This is a finite droplet and as the chosen $S1$ bond was arbitrary, the critical droplet of any $S1$ bond likewise must be finite.

Consider now an S2 bond. Without changing its sign, make its coupling magnitude sufficiently large (or equivalently, its K_{xy} value sufficiently small) so that it becomes S1. (This will cause a rearrangement of one or more trees, but it can be seen that any corresponding droplet flip must also be finite.) Now change the sign of the coupling. The same argument as before shows that the corresponding droplet flip is again finite. But given that the critical droplet corresponding to a given bond is unique, this was also the critical droplet of the original S2 bond. □

3 The Strongly Disordered Model

Although the highly disordered model is useful because of its tractability, it is clearly an unrealistic model for laboratory spin glasses. This leads us to propose a related model that, while retaining some of the simplifying features of the highly disordered model, can shed light on the ground state properties of realistic spin glass models. We will refer to this new model as the *strongly disordered model* of spin glasses.

The main difference between the two models is that in the strongly disordered model the couplings have the same distribution for all volumes. This is implemented by removing the volume dependence of the parameter λ:

The strongly disordered model is identical to the highly disordered model but with Eq. (4) replaced by

$$J_{xy} = \epsilon_{xy} e^{-\lambda K_{xy}} \tag{5}$$

with the constant $\lambda \gg 1$ independent of L.

In the strongly disordered model, the condition that every coupling value is no more than half the next larger one and no less than twice the next smaller one breaks down in sufficiently large volumes. This can be quantified: let $g(\lambda) = \text{Prob}(1/2 \leq e^{-\lambda K_{xy}}/e^{-\lambda K_{x'y'}} \leq 2)$. That is, $g(\lambda)$ is the probability that any two arbitrarily chosen bonds have coupling values that do *not* satisfy the highly disordered condition. A straightforward calculation gives $g(\lambda) = 2\ln 2/\lambda$.

The strongly disordered model carries two advantages. On the one hand, its critical droplet properties are analytically somewhat tractable given its similarity to the highly disordered model. On the other hand, since its coupling distribution is i.i.d. with mean zero and finite variance, and not varying with L, we expect global properties such as ground state multiplicity to be the same as in other versions of the EA spin glass with more conventional coupling distributions.

Theorem 2 *If there is no percolation at p_c in the corresponding independent bond percolation model, then in the strongly disordered model, the critical droplet of an arbitrary but fixed bond is finite with probability approaching one as $\lambda \to \infty$.*

Proof Consider a fixed, infinite-volume ground state on $\mathbb{Z}^d$; this induces a (coupling-dependent and ground-state-specific) spin configuration on the boundary $\partial\Lambda_L$ of any finite volume $\Lambda_L \subset \mathbb{Z}^d$.

Consider an arbitrary edge $\{x_0, y_0\}$. Let R denote the (random) smallest value in the invasion/minimal spanning forest model on $\mathbb{Z}^d$, defined by the i.i.d. K_{xy} (but with $K_{x_0y_0}$ set to zero, for convenience of the argument) such that one of the branches from x_0 or y_0 is contained within a cube of side length $2R$ centered at $\{x_0, y_0\}$. By the result of Alexander [1] mentioned earlier, R is a finite random variable (depending on the K_{xy}'s) if there is no percolation at p_c in the corresponding independent bond percolation model.

Now choose a deterministic Λ_L and consider the two events: (a) $A_L = \{R < L/2\}$ and (b) $B_L = \{$the highly disordered condition is valid in the cube of side L centered at $\{x_0, y_0\}\}$. Because R is a finite random variable, Prob(A_L) can be made arbitrarily close to one for L large. Moreover, from the definition of the highly disordered model Prob(B_L) can also be made close to one by choosing λ large (for the given L). Specifically, let P_0 denote the probability that the critical droplet of $\{x_0, y_0\}$ is finite. Then $P_0 \geq 1 - \epsilon$ if Prob$(R > L/2) + CL^{2d}/\lambda \leq \epsilon$ for some fixed $C > 0$. But for any $\epsilon > 0$ one can choose a sufficiently large L so that Prob$(R > L/2) \leq \epsilon/2$, and then choose λ such that $CL^{2d}/\lambda \leq \epsilon/2$. The result then follows. □

Theorem 2 sets a strong upper bound $O(\lambda^{-1})$ on the fraction of bonds that might *not* have a finite critical droplet. We do not yet know whether this gap can be closed in the sense that the strongly disordered model might share the property that *all* bonds have finite critical droplets. It could be that this is not the case, but that if the number of bonds with infinite critical droplets is sufficiently small, theorems analogous to those in [5] can be applied. Work on these questions is currently in progress.

Acknowledgments The research of L.-P. A. was supported in part by NSF CAREER DMS-1653602. The research of CMN was supported in part by NSF Grant DMS-1507019.

References

1. Alexander, K.S.: Percolation and minimal spanning forests in infinite graphs. Ann. Prob. **23**, 87–104 (1995)
2. Arguin, L.-P., Damron, M.: Short-range spin glasses and random overlap structures. J. Stat. Phys. **143**, 226–250 (2011)
3. Arguin, L.-P., Damron, M., Newman, C.M., Stein, D.L.: Uniqueness of ground states for short-range spin glasses in the half-plane. Commun. Math. Phys. **300**, 641–657 (2010)
4. Arguin, L.-P., Newman, C.M., Stein, D.L., Wehr, J.: Fluctuation bounds for interface free energies in spin glasses. J. Stat. Phys. **156**, 221–238 (2014)
5. Arguin, L.-P., Newman, C.M., Stein, D.L.: A relation between disorder chaos and incongruent states in spin glasses on $\mathbb{Z}^d$. Commun. Math. Phys. **367**, 1019–1043 (2019)

6. Banavar, J.R., Cieplak, M., Maritan, A.: Optimal paths and domain walls in the strong disorder limit. Phys. Rev. Lett. **72**, 2320–2323 (1994)
7. Berger, N., Tessler, R. J.: No percolation in low temperature spin glass. Electron. J. Probab. **22**, Paper no. 88 (2017)
8. Edwards, S., Anderson, P.W.: Theory of spin glasses. J. Phys. F **5**, 965–974 (1975)
9. Fisher, D.S., Huse, D.A.: Absence of many states in realistic spin glasses. J. Phys. A **20**, L1005–L1010 (1987)
10. Huse, D.A., Fisher, D.S.: Pure states in spin glasses. J. Phys. A **20**, L997–L1003 (1987)
11. Jackson, T.S., Read, N.: Theory of minimum spanning trees. I. Mean-field theory and strongly disordered spin-glass model. Phys. Rev. E **81**, 021130 (2010)
12. Newman, C.M., Stein, D.L.: Spin-glass model with dimension-dependent ground state multiplicity. Phys. Rev. Lett. **72**, 2286–2289 (1994)
13. Newman, C.M., Stein, D.L.: Ground state structure in a highly disordered spin glass model. J. Stat. Phys. **82**, 1113–1132 (1996)
14. Newman, C.M., Stein, D.L.: Nature of ground state incongruence in two-dimensional spin glasses. Phys. Rev. Lett. **84**, 3966–3969 (2000)
15. Newman, C.M., Stein, D.L.: Are there incongruent ground states in $2D$ Edwards-Anderson spin glasses? Commun. Math. Phys. **224**, 205–218 (2001)
16. Newman, C.M., Stein, D.L.: Interfaces and the question of regional congruence in spin glasses. Phys. Rev. Lett. **87**, 077201 (2001)
17. Newman, C.M., Stein, D.L.: Realistic spin glasses below eight dimensions: a highly disordered view. Phys. Rev. E **63**, 16101 (2001)
18. Newman, C.M., Tassion, V., Wu, W.: Critical percolation and the minimal spanning tree in slabs. Commun. Pure Appl. Math. **70**, 2084–2120 (2017)

Approximate and Exact Solutions of Intertwining Equations Through Random Spanning Forests

Luca Avena, Fabienne Castell, Alexandre Gaudillière, and Clothilde Mélot

Abstract For different reversible Markov kernels on finite state spaces, we look for families of probability measures for which the time evolution almost remains in their convex hull. Motivated by signal processing problems and metastability studies we are interested in the case when the size of such families is *smaller* than the size of the state space, and we want such distributions to be with "small overlap" among them. To this aim we introduce a *squeezing* function to measure the common overlap of such families, and we use random forests to build random approximate solutions of the associated intertwining equations for which we can bound from above the expected values of both squeezing and total variation errors. We also explain how to modify some of these approximate solutions into exact solutions by using those eigenvalues of the associated Laplacian with the largest size.

Keywords Intertwining · Markov process · Finite networks · Multiresolution analysis · Metastability · Random spanning forests

2010 Mathematical Subject Classification 05C81, 15A15, 60J28

1 Main Results, Motivations and Heuristic

The aim of this work is to build exact and approximate solutions of certain intertwining equations between Markov kernels on finite state spaces. The intertwining equations we look at are related to the two following problems. First, we want

---

L. Avena
Universiteit Leiden, Leiden, Netherlands
e-mail: l.avena@math.leidenuniv.nl

F. Castell · A. Gaudillière (✉) · C. Mélot
Aix-Marseille Université, CNRS, Centrale Marseille, Marseille, France
e-mail: fabienne.castell@univ-amu.fr; alexandre.gaudilliere@math.cnrs.fr; clothilde.melot@univ-amu.fr

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_3

to build wavelet-like multiresolution schemes for signal processing on arbitrary weighted graphs. Second, we want to make sense of the notion of metastability without asymptotics, in a finite setup where no large-volume or low-temperature limits are in place. We will partially address these problems by giving "good approximate solutions" of the intertwining equations, making use of random spanning forests.

1.1 Intertwining Equations

The basic object in this paper is an irreducible stochastic matrix P on a finite state space $\mathscr{X}$. P is associated (see Sect. 1.3.1 for precise definitions) with the generator $\mathscr{L}$ of a continuous time process X on $\mathscr{X}$ defined by

$$\mathscr{L}f(x) = \sum_{y\in\mathscr{X}} w(x,y)\big[f(y) - f(x)\big], \qquad f : \mathscr{X} \to \mathbb{R}, \qquad x \in \mathscr{X},$$

or, equivalently, with a connected edge weighted graph $G = (\mathscr{X}, \mathscr{E}, w)$, or $G = (\mathscr{X}, w)$, with $\mathscr{X}$ as vertex set, and

$$\mathscr{E} = \Big\{(x, y) \in \mathscr{X} \times \mathscr{X} : w(x, y) > 0\Big\}$$

as edges set. We will assume throughout the paper that P (or $\mathscr{L}$) is reversible with respect to some probability measure μ on $\mathscr{X}$:

$$\forall x, y \in \mathscr{X},\ \ \mu(x)w(x, y) = \mu(y)w(y, x)\,. \tag{1}$$

We look at solutions $(\Lambda, \bar{P})$ of the intertwining equations

$$\Lambda P = \bar{P}\Lambda, \tag{2}$$

and, for $q' > 0$,

$$\Lambda K_{q'} = \bar{P}\Lambda, \tag{3}$$

where

- $\bar{P}$ is a stochastic matrix defined on some finite state space $\bar{\mathscr{X}}$;
- $\Lambda : \bar{\mathscr{X}} \times \mathscr{X} \to [0, 1]$ is a rectangular stochastic matrix;

and $K_{q'}$ is the transition kernel on $\mathscr{X}$ given by

$$K_{q'}(x, y) := P_x(X(T_{q'}) = y) = q'(q'\,\mathrm{Id} - \mathscr{L})^{-1}(x, y)\,, \tag{4}$$

with $T_{q'}$ an exponential random variable with parameter q' that is independent of X.

Solving Eq. (2) amounts to find a family of probability measures $\nu_{\bar{x}} = \Lambda(\bar{x}, \cdot)$ on $\mathscr{X}$ such that, for some stochastic matrix $\bar{P}$,

$$\nu_{\bar{x}} P = \Lambda P(\bar{x}, \cdot) = \bar{P}\Lambda(\bar{x}, \cdot) = \sum_{\bar{y} \in \bar{\mathscr{X}}} \bar{P}(\bar{x}, \bar{y})\nu_{\bar{y}}, \qquad \bar{x} \in \bar{\mathscr{X}}. \tag{5}$$

In other words the one step evolution of the $\nu_{\bar{x}}$'s have to remain in their convex hull. Solving Eq. (3) is the same, except that the "one step evolution" has now to be considered in continuous time and on time scale $1/q'$. In both cases a trivial solution is always given by taking all the $\nu_{\bar{x}}$ equal to the equilibrium measure μ.

Related Literature

Intertwining relations, restricted to measures $\nu_{\bar{x}}$ with disjoint support, appeared in the context of diffusion processes in the paper by Rogers and Pitman [20], as a tool to state identities in laws. This method was later successfully applied to many other examples (see for instance [5, 11, 15]). In the context of Markov chains, intertwining was used by Diaconis and Fill [6] without the disjoint support restriction to build strong stationary times and to control convergence rates to equilibrium. This approach initiated in [6] is intimately related with metastability, as will be made clearer in Sects. 1.2.2 and 1.2.3, and it has been recently developed in different directions, see e.g. [16] and [14]. However, contrary to our setup, in these references intertwining relations have mainly been considered with an absorbing point for $\bar{P}$ in $\bar{\mathscr{X}}$ and with size m of $\bar{\mathscr{X}}$ being (much) larger than or equal to the size n of $\mathscr{X}$. At present, applications of intertwining include random matrices [7], particle systems [24], spectral clustering [1] ...

Our Contribution

Motivated by signal processing and metastability problems (see Sect. 1.2), in this paper we are instead interested in the case where

(R1) the size m of $\bar{\mathscr{X}}$ is smaller than the size n of $\mathscr{X}$,
(R2) $\bar{P}$ is irreducible,
(R3) the probability measures $\left(\nu_{\bar{x}} : \bar{x} \in \bar{\mathscr{X}}\right)$ are linearly independent and have small "joint overlap".

We will define the *squeezing* of a collection of probability measures to control this overlap (see Sect. 1.3.2) and a small "joint overlap" will correspond to little squeezed probability measures. We will see in Sect. 2.2 that, for any reversible stochastic kernel P with non-negative eigenvalues and for any positive $m < n$, non-degenerate solutions of Eq. (2) with $|\bar{\mathscr{X}}| = m$ always exist. By "non-degenerate solutions" we mean linearly independent probability measures such that Eq. (5) holds for some irreducible $\bar{P}$. But we will argue that exact solutions tend to be squeezed solutions. Then, rather than looking at the less squeezed solutions in the large space of all solutions for a given m, we will first consider approximate solutions with small squeezing. To this aim we will make use of random spanning forests to build random approximate solutions for which we will be able to bound both the expected value of an error term in intertwining Eq. (2) and the expected value of the squeezing (Theorem 1). Then we will use the same random forests to build random approximate solutions of Eq. (3) with no overlap, i.e., with disjoint support (Theorem 2). Assuming knowledge of the $n - m$ largest eigenvalues of $-\mathscr{L}$, we will finally see how to modify such an approximate solution of (3) with m probability measures $\nu_{\bar{x}}$ into exact solutions for q' small enough (Theorem 3).

Structure of the Paper

In the rest of this section, we detail our motivations, linking signal processing and metastability studies, and we give some heuristics in Sect. 1.2. After having fixed some notation in Sect. 1.3.1, we define the squeezing of a probability measure family in Sect. 1.3.2, we introduce random forests in Sect. 1.3.3, and state our main results in Sects. 1.3.4–1.3.6. In Sect. 2 we prove some preliminary results, and we give the proofs of our three main theorems in the three last sections. We conclude with an appendix that contains the proof of the main statement that links metastability studies with Eq. (5).

1.2 Motivations and Heuristics

Before stating precise results, we would like to explain why we are interested in solutions to (2) and (3) satisfying requirements (R1–3). These come from two motivating problems we describe now, the first one being the construction of a multiresolution analysis for signals on graphs, the second one being a proposal of metastability results without asymptotics.

1.2.1 Pyramidal Algorithms in Signal Processing

First we are interested in extending classical pyramidal algorithms of signal processing on the discrete torus

$$\mathscr{X} = \mathscr{X}_0 = \mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}$$

to the case of signals on generic edge-weighted graphs. Such algorithms are used for example to analyze or compress a given signal

$$f = f_0 : \mathscr{X}_0 \to \mathbb{R}$$

through *filtering* and *subsampling* operations. A *filter* is a linear operator which is diagonal in the same base as the discrete Laplacian $\mathscr{L}$. A *low-pass filter* K has eigenvalues of order 1 for low frequency modes, i.e., eigenvectors that are associated with small eigenvalues of $-\mathscr{L}$, and it has small eigenvalues for high frequency modes, i.e., eigenvectors that are associated with large eigenvalues of $-\mathscr{L}$. Assuming that n is an even number, a pyramidal algorithm first computes $m = n/2$ *approximation coefficients* by

- computing a low-pass filtered version Kf of the original signal f,
- subsampling Kf by keeping one in each two of its n values, those in some $\mathscr{X}_1 = \bar{\mathscr{X}} \subset \mathscr{X}$, for example the $n/2$ values in the even sites of $\mathbb{Z}_n$.

In doing so it defines a function

$$\bar{f} : \bar{x} \in \bar{\mathscr{X}} \mapsto Kf(\bar{x}) \in \mathbb{R}$$

that can naturally be seen as a signal $f_1 : \mathbb{Z}_{n/2} \to \mathbb{R}$ on a twice smaller torus. It then computes an *approximation* $\tilde{f}$ of f on $\mathscr{X}$ as a function of the approximation coefficients, and a *detail function* $\tilde{g} = f - \tilde{f}$, which in turn can be encoded into $n - m$ *detail coefficients*. Wavelet decomposition algorithms are of this kind. It then applies a similar treatment to f_1, to define f_2, then $f_3, \ldots$ up to reaching a simple signal defined on a small torus made of a few points only. The reason why this can be useful for compression is that, for well chosen filters, many of the detail coefficients obtained at the different levels are very small or negligible for a large class of smooth signals f. And one just has to store the few non-negligible detail coefficients together with the coarsest approximation's coefficients to reconstruct a good approximation of the original signal f. The point is then to find "good" filters, i.e. "good" $\varphi_{\bar{x}}$ in $\ell_2(\mu)$ (in this case μ is the uniform measure on $\mathscr{X}$, the reversible measure of the simple random walk associated with the discrete Laplacian) so that, for all $f \in \ell_2(\mu)$,

$$\bar{f}(\bar{x}) = \langle \varphi_{\bar{x}}, f \rangle = Kf(\bar{x}).$$

And a basic requirement for good filters is that, for each $\bar{x}$, $\varphi_{\bar{x}}$ is *localized* around $\bar{x}$. Even though the measures $d\nu_{\bar{x}} = \varphi_{\bar{x}} d\mu$ (so that $\bar{f}(\bar{x}) = \langle \nu_{\bar{x}} | f \rangle$) are usually signed measures and not measures, this is the reason why we want to think of the computation of the approximation coefficients $\bar{f}(\bar{x})$ as computation of *local means*. K being a low-pass filter, $\varphi_{\bar{x}}$ needs also to be "localized" in Fourier space (written in the diagonalizing basis of $\mathscr{L}$, it must have small coefficients on high-frequency modes). Thus the difficulty comes from Heisenberg principle, which roughly says that no function $\varphi_{\bar{x}}$ can be well localized both in Fourier space and around $\bar{x}$. Part of the art of wavelet design lies in the ability to make a good compromise with Heisenberg principle (see for example Chapter 7 in [23] for more details on this point).

When moving to the case of signal processing for generic edge-weighted graph, there are three main issues one has immediately to address to build pyramidal algorithms:

(Q1) What kind of subsampling should one use? What could "one every second node" mean?

(Q2) Which kind of filter should one use? How to compute local means?
(Q3) On which (weighted) graph should the approximation coefficients $\bar{f}(\bar{x})$ be defined to iterate the procedure?

On a general weighted finite graph $G = (\mathscr{X}, \mathscr{E}, w)$, none of these questions has a canonical answer. Several attempts to tackle these issues and to generalize wavelet constructions have been proposed: see [19, 22] for recent reviews on this subject and [9] for one of the most popular method. A good starting point to partially answer questions (Q2) and (Q3), is to look for a solution $(\Lambda, \bar{P})$ to intertwining Eq. (2), since any row $\nu_{\bar{x}}$ of Λ automatically belongs to an eigenspace of P, and is therefore frequency localized. Moreover, $\bar{P}$ is a candidate to define the graph structure on $\bar{\mathscr{X}}$. Requirements (R1) (R3) on $(\Lambda, \bar{P})$ reflect then the need of a subsampling procedure (with $m = |\bar{\mathscr{X}}|$ and n of the same order), and of space localization of the $\nu_{\bar{x}}$. (R2) is more technical, and essentially ensures that we can deal with $\bar{P}$ at the next level in the pyramidal algorithm in the same way we deal with P. We could however continue the pyramidal algorithm with a signal defined on unconnected graphs. Question (Q1) is left apart for the time being. This is where the random forest comes into the play, and we will come back to this question in Sect. 1.2.4.

Based on Theorem 1, we developed in [3] a novel wavelet transform. To our knowledge, our approach is the first one based on the solution of intertwining equations.

1.2.2 Metastability and Intertwining

Our second motivation stems from metastability studies, where it is common to build a coarse-grained version $\bar{X}$ of a Markov process X, possibly by seeing $\bar{X}$ as a measure-valued process on a small state space, these measures being probability measures on the large state space $\mathscr{X}$, on which X is defined. For example, when we want to describe the crystallisation of a slightly supersaturated vapor, we can do it in the following way. Vapor and crystal are defined by probability measures concentrated on very different parts of a very large state space. On this space a Markov process describing the temporal evolution of a microscopic configuration evolves, and this Markovian evolution has to be "macroscopically captured" by a new two-state Markov process evolving from gas (a probability measure on the large state space) to crystal (another probability measure on the same space almost non-overlapping with the previous one). And this evolution is such that the gas should appear as a *local equilibrium* left only to reach a more stable crystalline equilibrium. This is usually done in some asymptotic regime (e.g. large volume or low temperature asymptotic) and we refer to [17] and [4] for mathematical accounts on the subject.

But we are here outside any asymptotic regime: we are given a finite graph $(\mathscr{X}, w)$ or a Markov process X and we want to define a finite coarse-grained version of this graph and Markov process, $(\bar{\mathscr{X}}, \bar{w})$ and $\bar{X}$. Solving the intertwining equation $\Lambda P = \bar{P}\Lambda$, with the size of $\bar{P}$ smaller than the size of P, provides a clean way to do so. In this equation P is given and stands for the transition kernel of a discrete time skeleton $\hat{X}$ of X (see Sect. 1.3.1 for a precise definition) and we look for an $m \times m$ stochastic matrix $\bar{P}$ together with a collection of m probability measures $\nu_{\bar{x}}$ on $\mathscr{X}$ that defines the rectangular matrix Λ by

$$\Lambda(\bar{x}, x) = \nu_{\bar{x}}(x), \qquad \bar{x} \in \bar{\mathscr{X}}, \qquad x \in \mathscr{X}.$$

This equation reads

$$\nu_{\bar{x}} P = \sum_{\bar{y} \in \bar{\mathscr{X}}} \bar{P}(\bar{x}, \bar{y}) \nu_{\bar{y}} \tag{6}$$

for all $\bar{x}$ in $\bar{\mathscr{X}}$ and it suggests that the evolution of X can be roughly described through that of $\bar{X}$, associated with the transition kernel $\bar{P}$: from state or local equilibrium $\nu_{\bar{x}}$ the process X evolves towards a new state or local equilibrium $\nu_{\bar{y}}$ which is chosen according to the Markovian kernel $\bar{P}$. This can be turned into a rigorous and powerful mathematical statement by the following proposition, which is a partial rewriting of Section 2.4 of [6] in the spirit of [13], and whose proof is given in appendix.

Proposition 1 *If Eq. (6) is in force for some $\bar{x}$ in $\bar{\mathscr{X}}$, then there are a filtration $\mathscr{F}$ for which $\hat{X}$ is $\mathscr{F}$-adapted, a $\mathscr{F}$-stopping time $T_{\bar{x}}$ and a $\mathscr{F}_{T_{\bar{x}}}$-measurable random variable $\bar{Y}_{\bar{x}}$ with value in $\bar{\mathscr{X}} \setminus \{\bar{x}\}$ such that, for $\hat{X}$ started in $\nu_{\bar{x}}$:*

1. *$T_{\bar{x}}$ is geometric with parameter $1 - \bar{P}(\bar{x}, \bar{x})$;*
2. *$\nu_{\bar{x}}$ is stationary up to $T_{\bar{x}}$, i.e., for all $t \geq 0$,*

$$P_{\nu_{\bar{x}}}\left(\hat{X}(t) = \cdot \;\middle|\; t < T_{\bar{x}}\right) = \nu_{\bar{x}}\,; \tag{7}$$

3. *$P_{\nu_{\bar{x}}}\left(\bar{Y}_{\bar{x}} = \bar{y}\right) = \frac{\bar{P}(\bar{x}, \bar{y})}{1 - \bar{P}(\bar{x}, \bar{x})}$ for all $\bar{y}$ in $\bar{\mathscr{X}} \setminus \{\bar{x}\}$;*
4. *$P_{\nu_{\bar{x}}}\left(\hat{X}(T_{\bar{x}}) = \cdot \;\middle|\; \bar{Y}_{\bar{x}} = \bar{y}\right) = \nu_{\bar{y}}(\cdot)$;*
5. *$\left(\bar{Y}_{\bar{x}}, \hat{X}(T_{\bar{x}})\right)$ and $T_{\bar{x}}$ are independent.*

Notice the slight abuse of notation. In fact, in the above statement, $P_{\nu_{\bar{x}}}$ captures also the extra-randomness of the random variables $T_{\bar{x}}$ and $\bar{Y}_{\bar{x}}$.

As far as metastability is concerned, a possibly more natural approach is, instead of $\Lambda P = \bar{P}\Lambda$, to look for a solution of $\Lambda K_{q'} = \bar{P}\Lambda$ for a small q' and with $K_{q'}$ the transition kernel associated with our process X looked along a Poisson process of intensity q' (see Eq. (4) of Sect. 1.1). It is indeed on a "long" time scale $1/q'$ that one is usually looking at a coarse-grained Markovian version of X. But whatever the equation we are looking at, $\Lambda P = \bar{P}\Lambda$ or $\Lambda K_{q'} = \bar{P}\Lambda$, again we want solutions $\nu_{\bar{x}}$ that are localized in well distinct part of the state space, that is solutions satisfying (R3). Concerning (R1), in metastability studies, we are often interested in cases where m is very small with respect to n. However if one implements a complete pyramidal algorithm, one will solve at the same time intertwining equations with very different m and n by transitivity of the coarse-graining procedure.

1.2.3 Heisenberg Principle, Approximate Solutions and Related Work

There is actually at least a fourth question without canonical answer that arises when going from classical pyramidal or wavelet algorithms to signal processing for generic weighted graphs: what is a "Heisenberg principle" limiting the localization of our $\nu_{\bar{x}}$? We do not have an answer to this question, but, although we explained why we are interested in localized, non-overlapping, little squeezed solutions of the intertwining equations, we will see in Sect. 2 that exact solutions of intertwining equations are strongly localized in Fourier domain, then, a priori, poorly localized in space. This is the main difficulty faced by the present approach and this is one of the two reasons why we turned to approximate solutions of intertwining equations. We will also see in the next section that one needs a detailed knowledge of the spectrum and the eigenvectors of the Laplacian $\mathscr{L}$ to build exact solutions of intertwining equations. From an algorithmic point of view this can be very costly, and this is the other reason why we turned to approximate solutions.

In [3] we analyse the full pyramidal algorithm, including a wavelet basis construction, rather than simply focusing on intertwining equations of a one-step reduction. But we are still looking for a generalized Heisenberg principle that could serve as a guideline for similar constructions. And our results suggest that such a Heisenberg principle should degenerate in presence of a gap in the spectrum (see Sect. 1.3.4).

Before concluding this introductory part on intertwining equations, let us note that Proposition 1 can still be used to make sense of *approximate* intertwining. We will show in Sect. 2, denoting by d_{TV} the total variation distance:

Proposition 2 *If $\hat{X}$ and $\bar{X}$ are discrete time Markov chains on finite spaces $\mathscr{X}$ and $\bar{\mathscr{X}}$ with transitions kernels P and $\bar{P}$, if, for each $\bar{x}$ in $\bar{\mathscr{X}}$, $\nu_{\bar{x}}$ is a probability measure on $\mathscr{X}$, then, setting*

$$\epsilon = \max_{\bar{x}\in\bar{\mathscr{X}}} d_{\mathrm{TV}}\left(\nu_{\bar{x}} P, \sum_{\bar{y}\in\bar{\mathscr{X}}} \bar{P}(\bar{x},\bar{y})\nu_{\bar{y}}\right)$$

and for any convex combination

$$\nu = \sum_{\bar{x}\in\bar{\mathscr{X}}} \bar{\nu}(\bar{x})\nu_{\bar{x}},$$

there is a coupling between $\hat{X}^{\nu}$ and $\bar{X}^{\bar{\nu}}$, i.e., $\hat{X}$ and $\bar{X}$ started from ν and $\bar{\nu}$, a Markov chain $(Z,\bar{Z})$ on some product space $\mathscr{X}\times\tilde{\mathscr{X}}$, with $\bar{\mathscr{X}}\subset\tilde{\mathscr{X}}$, and two geometric *random times T and $\bar{T}$ with mean value $1/\epsilon$ such that for all $k \geq 0$,*

$$P\big(Z_k = x \mid \bar{Z}_k = \bar{x}\big) = \nu_{\bar{x}}(x), \qquad x \in \mathscr{X}, \qquad \bar{x} \in \bar{\mathscr{X}},$$

$\hat{X}_k^\nu = Z_k$ conditionally to $\{T > k\}$ and $\bar{X}_k^{\bar{\nu}} = \bar{Z}_k$ conditionally to $\{\bar{T} > k\}$. In particular it holds, for all $k \geq 0$,

$$d_{\mathrm{TV}}\left(P\big(\hat{X}_k^\nu = \cdot\big), \sum_{\bar{x} \in \bar{\mathscr{X}}} P\big(\bar{X}_k^{\bar{\nu}} = \bar{x}\big)\nu_{\bar{x}}\right) \leq P\big(T \leq k\big) + P\big(\bar{T} \leq k\big)$$
$$= 2\big(1 - (1-\epsilon)^k\big) \leq 2k\epsilon. \tag{8}$$

Comment: It is then possible to use approximate intertwining and the coarse-grained version $\bar{X}$ of $\hat{X}$ to control, for example, the mixing time of $\hat{X}$ from that of $\bar{X}$: if $1/\epsilon$ is large with respect to the latter, one can upper bound the mixing time of $\hat{X}$ by adding that of $\bar{X}$ to the time k needed for all the $\delta_x P^k$—distribution of $\hat{X}_k$ when $\hat{X}$ is started in x—to be close to the convex hull of the $\nu_{\bar{x}}$. Note that the latter will be related with the squeezing of the $\nu_{\bar{x}}$ in the sense of Sect. 1.3.2.

1.2.4 Some Heuristics on the Subsampling Question: Well Distributed Points, Renormalization and Determinantal Processes

We now go back to the subsampling question (Q1), i.e. the issue of finding m points, a fraction of n, that are in some sense well distributed in $\mathscr{X}$. This question turns out to be much simpler than (Q2) and (Q3), and a random solution is proposed in [2]. This solution is based on a random spanning forest Φ (i.e. a random collection of oriented rooted trees on the graph $G = (\mathscr{X}, w)$ exhausting $\mathscr{X}$), whose law depends on a real parameter $q > 0$. We denote by $\rho(\Phi)$ the set of tree roots of Φ. This forest will be precisely described in Sect. 1.3.3, but we review at once some of its features related to question (Q1). Let us denote, for any subset A of $\mathscr{X}$, by H_A and H_A^+ the hitting time of and the return time to A for the process X:

$$H_A := \inf\{t \geq 0, X(t) \in A\},$$

$$H_A^+ := \inf\{t \geq \sigma_1, X(t) \in A\},$$

with σ_1 the first time of the Poisson process that links $\hat{X}$ with X (see Sect. 1.3.1). For each x in $\mathscr{X}$ the mean hitting time $E_x[H_{\rho(\Phi)}]$ is a random variable, since so are Φ and $\rho(\Phi)$ (E_x being the expectation w.r.t the law of X starting from x). And it turns out that its expected value, with or without conditioning on the size of $\rho(\Phi)$, does not depend on x. In this sense the roots of the random forest are "well spread" on $\mathscr{X}$. More precisely, denoting by $\mathbb{E}_{x,q}$ the expectation w.r.t to the joint law of the

Markov process X and of the random forest Φ, and by $\mathbb{E}_q$ expectation w.r.t to the law $\mathbb{P}_q$ of Φ, we have (see [2]):

Proposition 3 *For any $x \in \mathscr{X}$ and $m \in \{1, \cdots, n\}$ it holds*

$$\mathbb{E}_{x,q}\left[H_{\rho(\Phi)}\right] = \frac{\mathbb{P}_q\left[|\rho(\Phi)| > 1\right]}{q}; \tag{9}$$

$$\mathbb{E}_{x,q}\left[H_{\rho(\Phi)} \;\middle|\; |\rho(\Phi)| = m\right] = \frac{\mathbb{P}_q\left[|\rho(\Phi)| = m+1\right]}{q\mathbb{P}_q\left[|\rho(\Phi)| = m\right]}; \tag{10}$$

$$\mathbb{E}_q\left[\frac{1}{m}\sum_{\bar{x}\in\rho(\Phi)} E_{\bar{x}}\left[H^+_{\rho(\Phi)} \;\middle|\; |\rho(\Phi)| = m\right]\right] = \frac{n}{\alpha m}. \tag{11}$$

This suggests to take $\bar{\mathscr{X}} = \rho(\Phi)$.

This is in line, in the context of very low temperature metastability systems, with Scoppola's renormalization introduced in [21] and with Freidlin and Wentzell's W-graphs [8]. Renormalization consists in individuating a sequence of smaller and smaller subsets of $\mathscr{X}$ with strong recurrence properties on longer and longer time scales. The coarse-grained Markov processes of this approach are the traces of the original process on these subsets. These subsets are naturally built as the roots of forests, or W-graphs in [8], made of bigger and bigger trees. These forests arise in this context of very low temperature systems as almost deterministic limits of our random forests Φ, and local equilibria reduce to (unsqueezed) Dirac masses. In moving away from this asymptotic regime through intertwining equations we consider dealing with more squeezed local equilibria.

As a consequence of Burton and Pemantle's transfer current Theorem, $\rho(\Phi)$ is a determinantal process on $\mathscr{X}$, and its kernel is $K_q = q(q\mathrm{Id} - \mathscr{L})^{-1}$ (see [2]):

Proposition 4 *For any subset A of $\mathscr{X}$,*

$$\mathbb{P}_q(A \subset \rho(\Phi)) = det_A(K_q),$$

where det_A applied to some matrix is the minor defined by the rows and columns corresponding to A.

By using reversibility, one can see that the determinant of $(K_q(x, y))_{x,y\in A}$ is, up to a multiplicative factor $\prod_{x\in A}\mu(x)$, the Gram matrix of the distributions $P_x\big(X(\tilde{T}_q) = \cdot\big)$, $x \in A$, with $\tilde{T}_q$ the square of an independent centered Gaussian variable with variance $1/(2q)$ (in such a way that the sum of two independent copies of $\tilde{T}_q$ has the same law as T_q). This means that a family of nodes $\bar{x}$ is unlikely to be part of $\rho(\Phi)$ if the volume of the parallelepiped formed by these distributions is small. It suggests that the distributions $\big(P_{\bar{x}}\big(X(\tilde{T}_q) = \cdot\big), \bar{x} \in \rho(\Phi)\big)$ are typically little squeezed (i.e. well space-localized) and so should be the distributions $\big(K_q(\bar{x}, \cdot), \bar{x} \in \rho(\Phi)\big)$, which are easier to deal with. To have a trade-off between squeezing and approximation error in intertwining equations, it will be convenient to introduce a second parameter $q' > 0$ and set $\nu_{\bar{x}} = K_{q'}(\bar{x}, \cdot)$ for $\bar{x}$ in $\rho(\Phi)$. At this point the choice made for $\bar{P}$ in Sect. 1.3.4 may be the most natural one.

Finally, when dealing with metastability issues, building local equilibria $\nu_{\bar{x}}$ from single "microscopic configurations" $\bar{x}$ in $\rho(\Phi)$ seems rather unnatural. In our previous example, no special microscopic configuration should play a role in defining what a metastable vapor should be. One should better look for larger structures associated with Φ, like the partition $\mathscr{A}(\Phi)$ of $\mathscr{X}$ defined by the trees of Φ, rather than $\rho(\Phi)$. Then, in view of the following proposition from [2], the unsqueezed measures $\mu_{A(\bar{x})}$ appear to be natural candidates for giving approximate solutions of (3):

Proposition 5 *Conditional law of the roots, given the partition.*
Let m be fixed, and $A_1, \ldots, A_m$ be a partition of $\mathscr{X}$. For any $x_1 \in A_1, \cdots, x_m \in A_m$,

$$\mathbb{P}_q \left[\rho(\Phi) = \{x_1, \cdots, x_m\} \,\middle|\, \mathscr{A}(\phi) = (A_1, \cdots, A_m)\right] = \prod_{i=1}^{m} \mu_{A_i}(x_i)\,, \tag{12}$$

where μ_A is the invariant measure μ conditioned to A ($\mu_A(B) = \mu(A \cap B)/\mu(A)$). Hence, given the partition, the roots are independent, and distributed according to the invariant measure.

Again, in the context of very low temperature metastable systems, this is in line with the so-called cycle decomposition ([8, 17]).

1.2.5 About the Reversibility Assumption

Concerning signal processing issues, the reversibility assumption (1) is rather standard. Actually, the classical multiresolution analysis of signals defined on the regular grid assumes the "reversibility of the grid", even in the case of audio signal where $\mathscr{X}$ is a time interval. When considering oriented and non-reversible graphs, the question of building a suitable Fourier analysis is already a delicate one, beyond the scope of the present paper.

In metastability studies, both reversible and non-reversible settings have been considered. Common approaches are usually initiated in the former context, where a richer palette of techniques is available. In our case, looking at metastable issues through intertwining equations and random forests does not rely on reversibility hypotheses, but our squeezing analysis is based on the description of determinantal processes associated with self-adjoint kernels: these are mixture of determinantal processes with a deterministic size and associated with a projector. Without such a reversibility hypotheses the full description of determinantal process kernels is still an open question. It is worth noting that the root process we use for subsampling is an example of such a non-reversible determinantal process, and that our total variation estimates in Theorem 1 still hold in this context.

1.3 Notations and Main Results

We describe now our main results, and for this purpose, introduce notations used throughout the paper.

1.3.1 Functions, Measures, Markov Kernel and Generator

Let $\mathscr{X}$ be a finite space with cardinality $|\mathscr{X}| = n$. We consider an irreducible continuous time Markov process $(X(t), t \geq 0)$ on $\mathscr{X}$, with generator $\mathscr{L}$:

$$\mathscr{L} f(x) := \sum_{y \in \mathscr{X}} w(x, y)(f(y) - f(x)), \tag{13}$$

where $f : \mathscr{X} \to \mathbb{R}$ is an arbitrary function, and $w : \mathscr{X} \times \mathscr{X} \to [0, +\infty[$ gives the transition rates. For $x \in \mathscr{X}$, let

$$w(x) := \sum_{y \in \mathscr{X} \setminus \{x\}} w(x, y)\,.$$

Note that $\mathscr{L}$ acts on functions as the matrix, still denoted by $\mathscr{L}$, the entries of which are:

$$\mathscr{L}(x, y) = w(x, y) \text{ for } x \neq y\,;\;\; \mathscr{L}(x, x) = -w(x)\,.$$

Let $\alpha > 0$ be defined by

$$\alpha = \max_{x \in \mathscr{X}} w(x)\,. \tag{14}$$

Hence, $P := \mathscr{L}/\alpha + \mathrm{Id}$ is an irreducible stochastic matrix, and we denote by $(\hat{X}_k, k \in \mathbb{N})$ a discrete time Markov chain with transition matrix P. The process $(X(t), t \geq 0)$ can be constructed from $(\hat{X}_k, k \in \mathbb{N})$ and an independent Poisson point process $(\sigma_i, i > 0)$ on $\mathbb{R}^+$ with rate α. At each point, or time, in the Poisson process, X moves according to the trajectory of $\hat{X}$, i.e., with $\sigma_0 = 0$:

$$X(t) = \sum_{i=0}^{+\infty} \hat{X}_i \mathbb{1}_{\sigma_i \leq t < \sigma_{i+1}}\,.$$

We assume that X is reversible with respect to the probability measure μ on $\mathscr{X}$, (i.e. (1)). The process X being irreducible, μ is strictly positive. The operator $-\mathscr{L}$ is self-adjoint and positive; we denote by $(\lambda_i; i = 0, \cdots, n-1)$ the real eigenvalues of $-\mathscr{L}$ in increasing order. It follows from the fact that P is irreducible that

$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \cdots \leq \lambda_{n-1} \leq 2\alpha \,. \tag{15}$$

A function f on $\mathscr{X}$ will be seen as a column vector, whereas a signed measure on $\mathscr{X}$ will be seen as a row vector. For $p \geq 1$, $\ell_p(\mu)$ is the space of functions endowed with the norm

$$\|f\|_p = \left(\sum_{x \in \mathscr{X}} |f(x)|^p \, \mu(x) \right)^{1/p} .$$

The scalar product of two functions f and g in $\ell_2(\mu)$ is

$$\langle f, g \rangle = \sum_{x \in \mathscr{X}} f(x) g(x) \mu(x)$$

The corresponding norm is denoted by $\|\cdot\| = \|\cdot\|_2$. When f is a function and ν is a signed measure, the duality bracket between ν and f is

$$\langle \nu | f \rangle = \sum_{x \in \mathscr{X}} \nu(x) f(x) \,.$$

$\ell_p^*(\mu)$ denotes the dual space of $\ell_p(\mu)$ with respect to $\langle \cdot | \cdot \rangle$. It is the space of signed measures endowed with the norm:

$$\|\nu\|_p^* = \left(\sum_{x \in \mathscr{X}} \left| \frac{\nu(x)}{\mu(x)} \right|^{p^*} \mu(x) \right)^{1/p^*}$$

where p^* is the conjugate exponent of p: $1/p + 1/p^* = 1$. $\ell_p^*(\mu)$ is identified with $\ell_{p^*}(\mu)$ through the isometry: $\nu \in \ell_p^*(\mu) \mapsto \nu^* \in \ell_{p^*}(\mu)$, where $\nu^*(x) = \nu(x)/\mu(x)$ is the density of ν with respect to μ. The inverse of this isometry is still denoted by *. It associates to a function $f \in \ell_p(\mu)$, the signed measure $f^* \in \ell_{p^*}^*(\mu)$ whose density with respect to μ is f: $f^*(A) = \sum_{x \in A} \mu(x) f(x)$ for all subset A of $\mathscr{X}$. $\ell_2^*(\mu)$ is an Euclidean space whose scalar product is denoted by:

$$\langle \nu, \rho \rangle^* := \left\langle \nu^*, \rho^* \right\rangle = \sum_{x \in \mathscr{X}} \nu(x) \rho(x) \frac{1}{\mu(x)} \,.$$

The corresponding norm is denoted by $\|\cdot\|^*$. For $\nu \in \ell_2^*(\mu)$ and $f \in \ell_2(\mu)$, one gets

$$\langle \nu | f \rangle = \left\langle \nu, f^* \right\rangle^* = \left\langle \nu^*, f \right\rangle \,.$$

1.3.2 Squeezing of a Collection of Probability Measures

For some finite space $\bar{\mathscr{X}}$ of size $m \le n$, let $(\nu_{\bar{x}} : \bar{x} \in \bar{\mathscr{X}})$ be a collection of m probability measures on $\mathscr{X}$ which is identified with the matrix Λ, the row vectors of which are the $\nu_{\bar{x}}$'s: $\Lambda(\bar{x}, \cdot) = \nu_{\bar{x}}$ for each $\bar{x}$ in $\bar{\mathscr{X}}$. Since these measures form acute angles between them ($\langle \nu_{\bar{x}}, \nu_{\bar{y}} \rangle^* \ge 0$ for all $\bar{x}$ and $\bar{y}$ in $\bar{\mathscr{X}}$) and have disjoint supports if and only if they are orthogonal, one could use the volume of the parallelepiped they form to measure their "joint overlap". The square of this volume is given by the determinant of the Gram matrix:

$$\mathrm{Vol}(\Lambda) = \sqrt{\det(\Gamma)},$$

with Γ the square matrix on $\bar{\mathscr{X}}$ with entries $\Gamma(\bar{x}, \bar{y}) = \left\langle \nu_{\bar{x}}, \nu_{\bar{y}} \right\rangle^*$, that is

$$\Gamma := \Lambda D(1/\mu) \Lambda^t \,, \tag{16}$$

where $D(1/\mu)$ is the diagonal matrix with entries given by $(1/\mu(x), x \in \mathscr{X})$, and Λ^t is the transpose of Λ. Loosely speaking, the less overlap, the largest the volume.

We will instead use the *squeezing* of Λ, that we define by

$$\mathscr{S}(\Lambda) := \begin{cases} +\infty & \text{if } \det(\Gamma) = 0, \\ \sqrt{\mathrm{Trace}\left(\Gamma^{-1}\right)} \in]0, +\infty[& \text{otherwise,} \end{cases} \tag{17}$$

to measure this "joint overlap". We call it "squeezing" because the $\nu_{\bar{x}}$ and the parallelepiped they form are squeezed when $\mathscr{S}(\Lambda)$ is large. This is also the half diameter of the rectangular parallelepiped that circumscribes the ellipsoid defined by the Gram matrix Γ : this ellipsoid is squeezed too when $\mathscr{S}(\Lambda)$ is large. We note finally that our squeezing controls the volume of Λ. Indeed, by comparison between harmonic and geometric mean applied to the eigenvalues of the Gram matrix, small squeezing implies large volume: $\mathrm{Vol}(\Lambda)^{1/n} \mathscr{S}(\Lambda) \ge \sqrt{n}$. We will also show in Sect. 2:

Proposition 6 *Let $(\nu_{\bar{x}}, \bar{x} \in \bar{\mathscr{X}})$ be a collection of m probability measures on $\mathscr{X}$.*

1. *We have*

$$\mathscr{S}(\Lambda) \ge \sqrt{\sum_{\bar{x} \in \bar{\mathscr{X}}} \frac{1}{\|\nu_{\bar{x}}\|^{*2}}} \,. \tag{18}$$

   Equality holds if and only if the $(\nu_{\bar{x}}, \bar{x} \in \bar{\mathscr{X}})$ are orthogonal.
2. *Assume that μ is a convex combination of the $(\nu_{\bar{x}}, \bar{x} \in \bar{\mathscr{X}})$. Then,*

$$\mathscr{S}(\Lambda) \ge 1 \,.$$

   Equality holds if and only if the $(\nu_{\bar{x}}, \bar{x} \in \bar{\mathscr{X}})$ are orthogonal.

Comment: $\mathscr{S}(\Lambda)$ is thus maximal when the $\nu_{\bar{x}}, \bar{x} \in \bar{\mathscr{X}}$, are linearly dependent, and minimal when they are orthogonal. Moreover, we know the minimal value of $\mathscr{S}(\Lambda)$, when μ is a convex combination of the $(\nu_{\bar{x}}, \bar{x} \in \bar{\mathscr{X}})$. Note that this is necessarily the case if the convex hull of the $\nu_{\bar{x}}$ is stable under P, i.e. when $\Lambda P = \bar{P}\Lambda$ for some stochastic $\bar{P}$. Indeed it is then stable under $e^{t\mathscr{L}}$ for any $t > 0$ and the rows of $\Lambda e^{t\mathscr{L}}$ converge to μ when t goes to infinity. Note also that we are using "$\ell_2(\mu)$ computations" (through the Gram matrix) to define the squeezing of measures that are normalized in $\ell_1(\mu) \sim \ell_\infty^*(\mu)$ (these are *probability* measures). This proposition shows that such a mixture of norms is not meaningless.

1.3.3 Random Forests

Note that the weight function w induces a structure of oriented graph on $\mathscr{X}$, $e = (x, y)$ being an oriented edge if and only if $w(e) := w(x, y) > 0$. Let $\mathscr{E}$ be the set of oriented edges, and $G = (\mathscr{X}, \mathscr{E})$ the oriented graph just defined. An oriented forest ϕ on $\mathscr{X}$ is a collection of rooted trees that are subgraphs of G, oriented from their leaves towards their root. A spanning oriented forest (s.o.f.) on $\mathscr{X}$ is an oriented forest which exhausts the points in $\mathscr{X}$. The set of roots of a spanning oriented forest ϕ is denoted by $\rho(\phi)$.

We introduce now a real parameter $q > 0$, and associate to each oriented forest a weight

$$w_q(\phi) := q^{|\rho(\phi)|} \prod_{e \in \phi} w(e) . \tag{19}$$

These weights can be renormalized to define a probability measure on the set of spanning oriented forest,

$$\pi_q(\phi) := \frac{w_q(\phi)}{Z(q)} , \tag{20}$$

where the partition function $Z(q)$ is given by

$$Z(q) := \sum_{\phi \text{ s.o.f.}} w_q(\phi) . \tag{21}$$

We can sample from π_q by using Wilson's algorithm [18, 25] which can be described as follows. Let Φ_c be the current state, an oriented forest, of the spanning oriented forest being constructed. At the beginning, Φ_c has no nodes or edges. While Φ_c is not spanning, i.e., while there is a vertex in $\mathscr{X}$ which is not in the vertex set

$V(\Phi_c)$ of Φ_c, perform the following steps:

- Choose a point x in $\mathscr{X} \setminus V(\Phi_c)$, in any deterministic or random way.
- Let evolve the Markov process $(X(t), t \geq 0)$ from x, and stop it at $T_q \wedge H_{V(\Phi_c)}$ with T_q an independent exponential time of parameter q and $H_{V(\Phi_c)}$ the hitting time of $V(\Phi_c)$.
- Erase the loops, in order of appearance, of the trajectory drawn by X to obtain a self-avoiding path C starting from x and oriented towards its end-point.
- Add C to Φ_c.

Each iteration of the "while loop" stopped by the exponential time, gives birth to another tree. Wilson's algorithm is not only a way to sample π_q, it is also a powerful tool to study it. The main strength of this algorithm is the freedom one has in choosing the starting points x's of X.

In the sequel, Φ will denote a random variable defined on some probability space $(\Omega_f, \mathscr{A}_f, \mathbb{P}_q)$, having distribution π_q. The corresponding expectation will be denoted by $\mathbb{E}_q$. We will often work with two independent sources of randomness: the Markov process X, and the random forest Φ. Integration with respect to X starting from x will be denoted by P_x and E_x. When X is started with an initial measure π, we will use the notations P_π and E_π. When we integrate over both randomness, we will use the notations $\mathbb{E}_{x,q}$, $\mathbb{E}_{\pi,q}$ and $\mathbb{P}_{x,q}$, $\mathbb{P}_{\pi,q}$. The random forest Φ defines a partition of $\mathscr{X}$, two points being in the same set of the partition if they belong to the same tree. This partition will be denoted by $\mathscr{A}(\Phi)$. A point $x \in \mathscr{X}$ being fixed, τ_x is the tree of Φ containing x, ρ_x its root, and $A(x)$ the unique element of $\mathscr{A}(\Phi)$ containing x.

A theorem of Kirchhoff [10] gives in this context that

$$Z(q) = \det(q\,\mathrm{Id} - \mathscr{L}) = \prod_{j<n} (q + \lambda_j), \tag{22}$$

and this implies (see for example [2] for more details, a proof of (22) and the following proposition):

Proposition 7 *For all $k \in \{0, \cdots, n\}$,*

$$\mathbb{P}_q\left[|\rho(\Phi)| = k\right] = \sum_{\substack{J \subset \{0,\cdots,n-1\} \\ |J|=k}} \prod_{j\in J} \frac{q}{q+\lambda_j} \prod_{j\notin J} \frac{\lambda_j}{q+\lambda_j}.$$

Otherwise stated, the number of roots has the same law as $\sum_{j=0}^{n-1} B_j$ where $B_0, \ldots, B_{n-1}$ are independent, B_j having Bernoulli distribution with parameter $\frac{q}{q+\lambda_j}$.

1.3.4 Approximate Solution of $\Lambda P = \bar{P}\Lambda$

Assume that we sampled Φ from π_q for some parameter $q > 0$. For $q' > 0$ we then set

- $\bar{\mathscr{X}} := \rho(\Phi)$;
- For any $\bar{x} \in \bar{\mathscr{X}}$, $\nu_{\bar{x}}(\cdot) := K_{q'}(\bar{x}, \cdot)$ (cf. Eq. (4)), i.e. $\Lambda = K_{q'}|_{\bar{\mathscr{X}} \times \mathscr{X}}$;
- $\bar{P}(\bar{x}, \bar{y}) := P_{\bar{x}}\left[X\left(H^+_{\bar{\mathscr{X}}}\right) = \bar{y}\right]$ with, for any $A \subset \mathscr{X}$,

$$H^+_A := \inf\{t \geq \sigma_1, X(t) \in A\} . \tag{23}$$

H^+_A is in other words the return time in A, and $\bar{P}$ is the (irreducible and reversible) Markovian kernel associated with the trace chain of X on $\bar{\mathscr{X}}$.

Here $\bar{\mathscr{X}}$ is a random subset of $\mathscr{X}$, and so is its cardinality. If we want to keep approximately m points from $\mathscr{X}$, we have to ensure that

$$\mathbb{E}_q\left[|\bar{\mathscr{X}}|\right] = \sum_{i=0}^{n-1} \frac{q}{q+\lambda_i} \approx m . \tag{24}$$

This can be obtained, starting from any q to sample Φ, by updating q according to $q \leftarrow qm/|\rho(\Phi)|$ before re-sampling Φ and going so up to getting a satisfactory number of roots (see [2] for more details).

Let us remind the definition (14) of α, and let us define

$$p_j := \frac{q}{q+\lambda_j}, \qquad p'_j := \frac{q'}{q'+\lambda_j}, \qquad j < n,$$

and denote by d_{TV} the total variation distance: if ν and ν' are two probability measures on $\mathscr{X}$,

$$d_{TV}(\nu, \nu') = \frac{1}{2}\sum_{x \in \mathscr{X}} \left|\nu(x) - \nu'(x)\right| .$$

Theorem 1 *For all $m \in \{1, \cdots, n\}$,*

$$\mathbb{E}_q\left[\sum_{\bar{x} \in \bar{\mathscr{X}}} d_{TV}(\Lambda P(\bar{x}, \cdot), \bar{P}\Lambda(\bar{x}, \cdot)) \;\Big|\; |\bar{\mathscr{X}}| = m\right] \leq \frac{q'(n-m)}{\alpha}, \tag{25}$$

and

$$\mathbb{E}_q\left[\sum_{\bar{x} \in \bar{\mathscr{X}}} d_{TV}(\Lambda P(\bar{x}, \cdot), \bar{P}\Lambda(\bar{x}, \cdot))\right] \leq \frac{q'}{\alpha}\sum_{i=1}^{n-1} \frac{\lambda_i}{q+\lambda_i} . \tag{26}$$

In addition, with

$$S_n := \sum_{j=1}^{n-1} p_j'^2 (1-p_j)^2 \ ; \ T_n := \sum_{j=1}^{n-1} \frac{p_j^2}{p_j'^2} \ ; \ V_n = \sum_{j=1}^{n-1} p_j(1-p_j)\,,$$

it holds

$$\mathbb{E}_q\left[\mathscr{S}(\Lambda)\,\middle|\, |\bar{\mathscr{X}}| = m\right] \leq \frac{\min\left\{\sqrt{1+\sqrt{\frac{T_n}{S_n}}}\ \exp\left(\sqrt{S_n T_n} - V_n\right);\ \sqrt{1+T_n}\exp\left(\frac{(1+S_n T_n))}{2} - V_n\right)\right\}}{\mathbb{P}_q\left[|\bar{\mathscr{X}}| = m\right]} \tag{27}$$

for any $m \in \{1, \cdots, n\}$.

Proof See Sect. 3.

Comment: Our upper bounds depend on $\mathscr{L}$ through its spectrum only. They show that if there is a gap in this spectrum—that is if for some $1 < m < n$ it holds $\lambda_{m-1} \ll \lambda_m$—then we can have asymptotically exact solutions with small squeezing by choosing $\lambda_{m-1} \ll q \ll q' \ll \lambda_m$. We then have indeed $q' \ll \alpha$ since $\lambda_m \leq 2\alpha$ and $p_j \sim p_j' \sim 1$ for $j < m$, while $p_j \ll p_j' \ll 1$ for $j \geq m$. We can then have a vanishing error in the approximation, see (26). In addition we can have $V_n \ll 1$, $S_n \ll 1$, $T_n \sim m-1$, $\mathbb{P}_q[|\bar{\mathscr{X}}| = m] \sim 1$ (recall Proposition 7) and an upper bound on the mean value of $\mathscr{S}(\Lambda)$ that goes like $\sqrt{m}$. This upper bound has to be compared with the lower bounds of Proposition 6, i.e. with 1 if we have asymptotic solutions of intertwining equations. For some simple low temperature metastable systems as quickly mentioned in Sect. 1.2.4, there is such a gap in the spectrum and this construction will give indeed asymptotic solutions with $\mathscr{S}(\Lambda)$ going to 1. There is room for improvement in the sense that our approximate solutions can be even less squeezed that what is ensured by the theorem.

1.3.5 Approximate Solutions of $\Lambda K_{q'} = \bar{P}\Lambda$

Assume once again that we sampled Φ from π_q for some parameter $q > 0$. But let us modify our choices for $\bar{\mathscr{X}}$, Λ and $\bar{P}$, by using this time the partition $\mathscr{A}(\Phi)$. Set:

- $\bar{\mathscr{X}} := \rho(\Phi)$ (one could rather think that $\bar{\mathscr{X}}$ is the set of the different pieces forming the partition $\mathscr{A}(\Phi)$ but the notation will be simpler by using the set of roots, which obviously is in one to one correspondence through the map $A : \bar{x} \in \rho(\Phi) \mapsto A(\bar{x})$);

- for any $\bar{x} \in \bar{\mathscr{X}}$, $\nu_{\bar{x}}(\cdot) := \mu_{A(\bar{x})}(\cdot)$, with, for any $A \subset \mathscr{X}$, μ_A being defined by the probability μ conditioned to A: $\mu_A := \mu(\cdot|A)$;
- for any $\bar{x}, \bar{y} \in \bar{\mathscr{X}}$, $\bar{P}(\bar{x}, \bar{y}) := P_{\mu_{A(\bar{x})}}\left[X(T_{q'}) \in A(\bar{y})\right]$, with $T_{q'}$ being as previously an exponential random variable of parameter q' that is independent from X. Irreducibility and reversibility of $\bar{P}$ are then inherited from those of P.

It follows from Proposition 6 that the squeezing of $\left\{\nu_{\bar{x}}, \bar{x} \in \bar{\mathscr{X}}\right\}$ is minimal and equal to one.

To bound the distance between $\Lambda K_{q'}$ and $\bar{P}\Lambda$, we introduce another random forest Φ' distributed as $\pi_{q'}$ and independent of Φ and X. For any $x \in \mathscr{X}$, t'_x is the tree containing x in Φ', ρ'_x its root, $A'(x)$ the unique element of $\mathscr{A}(\Phi')$ containing x, and Γ'_x is the path going from x to ρ'_x in Φ'. By Wilson algorithm started at x, Γ'_x is the trajectory of a loop-erased random walk started from x and stopped at an exponential time $T_{q'}$. We denote by $|\Gamma'_x|$ its length, that is the number of edges to be crossed in Φ' to go from x to ρ'_x.

Theorem 2 *Let $p \geq 1$, and p^* its conjugate exponent, so that $\frac{1}{p} + \frac{1}{p^*} = 1$. Then,*

$$\mathbb{E}_q\left[\sum_{\bar{x}\in\bar{\mathscr{X}}} d_{TV}(\Lambda K_{q'}(\bar{x},\cdot), \bar{P}\Lambda(\bar{x},\cdot))\right] \leq \left(\mathbb{E}_q\left[|\rho(\Phi)|\right]\right)^{1/p}\left(\frac{q'}{q}\sum_{x\in\mathscr{X}} \mathbb{E}_{q'}\left[|\Gamma'_x|\right]\right)^{1/p^*}.$$

Proof See Sect. 4.

Comment: Note that

$$q'\mathbb{E}_{q'}\left[|\Gamma'_x|\right] = \alpha\frac{\mathbb{E}_{q'}\left[|\Gamma'_x|\right]}{\alpha/q'}$$

is, up to the factor α, the ratio between the mean number of steps of the loop-erased random walk and the mean number of steps of the simple random walk up to time $T_{q'}$, that is the time fraction spent outside loops up to time $T_{q'}$. As a consequence "the more recurrent is X on time scale $1/q'$", the smaller is this ratio.

1.3.6 Exact Solutions of $\Lambda K_{q'} = \bar{P}\Lambda$

We finally modify the previous random measures $\mu_{A(\bar{x})}$ to build exact solution of Eq. (3) for q' small enough. We will use to this end a result due to Micchelli and Willoughby [12]: for any $m > 0$

$$MW_m := \prod_{j\geq m}\frac{1}{\lambda_j}\left(\mathscr{L} + \lambda_j\mathrm{Id}\right)$$

is a Markovian kernel (one can see [2] for a probabilistic insight into the proof of this result). Assume then that we sampled Φ from π_q for some parameter $q > 0$, let us keep $\bar{\mathscr{X}} = \rho(\Phi)$, but let us now set

$$\nu_{\bar{x}} = \mu_{A(\bar{x})} M W_m, \qquad \bar{x} \in \bar{\mathscr{X}},$$

with $m = |\bar{\mathscr{X}}|$.

Theorem 3 *If the $\nu_{\bar{x}}$ have finite squeezing, then for q' small enough, the $\nu_{\bar{x}} K_{q'}$ are in the convex hull of the $\nu_{\bar{x}}$.*

Proof See Sect. 5.

Comment: Since we do not give quantitative bounds on how small q' has to be for the thesis to hold, and we do not bound the squeezing of these $\nu_{\bar{x}}$, Theorem 3 is at first not a very insightful result. However the proof we will give suggests that the $\nu_{\bar{x}}$ are natural candidates for not too squeezed solution associated with some non-very small q'. It will also give further motivation to use squeezing to measure joint overlap. We actually got to our squeezing definition by looking for quantitative bounds for this theorem.

2 Preliminary Results

2.1 Proof of Proposition 6

If Γ is not invertible, points (1) and (2) are obviously true. We assume therefore that Γ is invertible. Let $\tilde{\Lambda} := \Gamma^{-1}\Lambda$, and let $(\tilde{\nu}_{\bar{x}}, \bar{x} \in \bar{\mathscr{X}})$ be the row vectors of $\tilde{\Lambda}$. Note that

$$\tilde{\Lambda} D(1/\mu) \Lambda^t = \Gamma^{-1} \Lambda D(1/\mu) \Lambda^t = \Gamma^{-1}\Gamma = \mathrm{Id}\,.$$

$$\tilde{\Lambda} D(1/\mu) \tilde{\Lambda}^t = \Gamma^{-1} \Lambda D(1/\mu) \Lambda^t \Gamma^{-1} = \Gamma^{-1}\,.$$

Hence, for all $\bar{x}, \bar{y} \in \bar{\mathscr{X}}$, $\langle \tilde{\nu}_{\bar{x}}, \nu_{\bar{y}} \rangle^* = \delta_{\bar{x}\bar{y}}$ and $\|\tilde{\nu}_{\bar{x}}\|^{*2} = (\Gamma^{-1})(\bar{x}, \bar{x})$.

1. We have $\mathscr{S}(\Lambda)^2 = \sum_{\bar{x} \in \bar{\mathscr{X}}} \|\tilde{\nu}_{\bar{x}}\|^{*2} \geq \sum_{\bar{x} \in \bar{\mathscr{X}}} \dfrac{\langle \tilde{\nu}_{\bar{x}}, \nu_{\bar{x}} \rangle^{*2}}{\|\nu_{\bar{x}}\|^{*2}} = \sum_{\bar{x} \in \bar{\mathscr{X}}} \dfrac{1}{\|\nu_{\bar{x}}\|^{*2}}$.

   Assume now that the $\nu_{\bar{x}}$'s, $\bar{x} \in \bar{\mathscr{X}}$ are orthogonal. $\Gamma = \mathrm{diag}(\|\nu_{\bar{x}}\|^{*2})$, so that $\mathrm{Trace}(\Gamma^{-1}) = \sum_{\bar{x} \in \bar{\mathscr{X}}} \dfrac{1}{\|\nu_{\bar{x}}\|^{*2}}$.

   In the opposite direction, assume instead that $\mathrm{Trace}(\Gamma^{-1}) = \sum_{\bar{x} \in \bar{\mathscr{X}}} \dfrac{1}{\|\nu_{\bar{x}}\|^{*2}}$.

   Then for any $\bar{x} \in \bar{\mathscr{X}}$, $\left|\langle \tilde{\nu}_{\bar{x}}, \nu_{\bar{x}} \rangle^*\right| = \|\tilde{\nu}_{\bar{x}}\|^* \, \|\nu_{\bar{x}}\|^*$. This implies that for all $\bar{x} \in \bar{\mathscr{X}}$, there exists a real number $\alpha(\bar{x}) \neq 0$ such that $\tilde{\nu}_{\bar{x}} = \alpha(\bar{x}) \nu_{\bar{x}}$. Taking the scalar product with $\nu_{\bar{y}}$ leads to $\delta_{\bar{x}\bar{y}} = \langle \tilde{\nu}_{\bar{x}}, \nu_{\bar{y}} \rangle^* = \alpha(\bar{x}) \langle \nu_{\bar{x}}, \nu_{\bar{y}} \rangle^*$. Hence $(\nu_{\bar{x}}, \bar{x} \in \bar{\mathscr{X}})$ are orthogonal.

2. Let us write μ as a convex combination of the $(\nu_{\bar{x}}, \bar{x} \in \bar{\mathscr{X}})$:

$$\mu = \sum_{\bar{x}\in\bar{\mathscr{X}}} \alpha(\bar{x})\nu_{\bar{x}}, \qquad \alpha(\bar{x}) \geq 0, \qquad \sum_{\bar{x}\in\bar{\mathscr{X}}} \alpha(\bar{x}) = 1.$$

Note that for any probability measure ν, $\langle \mu, \nu\rangle^* = \sum_{x\in\mathscr{X}} \mu(x)\nu(x)/\mu(x) = 1$. As a special case, for any $\bar{y} \in \bar{\mathscr{X}}$,

$$1 = \langle \mu, \nu_{\bar{y}}\rangle^* = \sum_{\bar{x}\in\bar{\mathscr{X}}} \alpha(\bar{x}) \langle \nu_{\bar{x}}, \nu_{\bar{y}}\rangle^* \geq \alpha(\bar{y}) \left\| \nu_{\bar{y}} \right\|^{*2} . \tag{28}$$

By point (1), we deduce that

$$\mathscr{S}(\Lambda)^2 \geq \sum_{\bar{x}\in\bar{\mathscr{X}}} \frac{1}{\|\nu_{\bar{x}}\|^{*2}} \geq \sum_{\bar{x}\in\bar{\mathscr{X}}} \alpha(\bar{x}) = 1 .$$

Equality holds if and only if (28) and (18) are equalities. By point (1), this implies that the $(\nu_{\bar{x}}, \bar{x} \in \bar{\mathscr{X}})$ are orthogonal. In the opposite direction, when the $\nu_{\bar{x}}$, for $\bar{x} \in \bar{\mathscr{X}}$, are orthogonal, (28) and (18) are equalities, and $\mathscr{S}(\Lambda) = 1$.

2.2 Elementary Observations on Intertwining Equations

Consider Eq. (2) for any reversible and irreducible stochastic kernel P, and assume an $m \times n$ rectangular stochastic matrix $\Lambda = (\Lambda(\bar{x}, x))_{\bar{x}\in\bar{\mathscr{X}}, x\in\mathscr{X}}$ to be a solution for some $\bar{P}$ with $m \leq n$. Let us write $(\theta_j)_{j<n} = (1 - \lambda_j/\alpha)_{j<n}$ for the n eigenvalues of P in decreasing order:

$$1 = \theta_0 > \theta_1 \geq \cdots \geq \theta_{n-1} \geq -1.$$

We also set $[n] = \{0, 1, 2, \ldots, n-1\}$, call μ the reversible measure of P, and write $\nu_{\bar{x}} = \Lambda(\bar{x}, \cdot)$ for the rows of Λ.

Lemma 1 *Assume Eq. (2) is in force. If Λ is non-degenerate, i.e., if Λ is of rank m, then there is an orthonormal basis of left eigenvectors $(\mu_j : 0 \leq j < n)$ of P such that*

$$\mu_j P = \theta_j \mu_j, \qquad j < n,$$

there is a subset J of $[n]$ such that $0 \in J$ and $|J| = m$ and there is an invertible matrix $C = (C(\bar{x}, j))_{\bar{x}\in\bar{\mathscr{X}}, j\in J}$ with $C(\bar{x}, 0) = 1$ for all $\bar{x}$ in $\bar{\mathscr{X}}$, such that

$$\nu_{\bar{x}} = \sum_{j\in J} C(\bar{x}, j)\mu_j, \qquad \bar{x} \in \bar{\mathscr{X}}, \tag{29}$$

and

$$\bar{P}C(\cdot, j) = \theta_j C(\cdot, j), \qquad j \in J. \tag{30}$$

In particular, the spectrum of $\bar{P}$ is contained in that of P, with eigenvalue multiplicities that do not exceed the corresponding ones for P.

Proof Let V be the subspace of $\ell_2^*(\mu)$ spanned by the $\nu_{\bar{x}}$. Since Λ is non-degenerate, V is of dimension m. Since $\Lambda P = \bar{P}\Lambda$, the $\nu_{\bar{x}}P$ are convex combinations of the $\nu_{\bar{x}}$ and V is stable by the self-adjoint operator P. It follows that there is an orthonormal basis of left eigenvectors μ_j, with $\mu_0 = \mu$, a subset $J \subset [n]$ of size m, and an invertible matrix C such that (29) holds. Since for $j > 0$ one has $\langle \mu, \mu_j \rangle^* = 0$, by computing the scalar product with μ of both sides of Eqs. (29), it follows that 0 belongs to J and $C(\bar{x}, 0) = 1$ for each $\bar{x}$.

Now, applying P on both sides of (29) we obtain

$$\sum_{j \in J} \sum_{\bar{y} \in \bar{\mathscr{X}}} \bar{P}(\bar{x}, \bar{y}) C(\bar{y}, j) \mu_j = \sum_{j \in J} \theta_j C(\bar{x}, j) \mu_j, \qquad \bar{x} \in \bar{\mathscr{X}}.$$

By identifying the decomposition coefficients in the basis of the μ_j's, this gives (30). Since the m column vectors $C(\cdot, j)$ are linearly independent, they form a basis of the functions on $\bar{\mathscr{X}}$. This is why Eqs. (30) completely describe the spectrum of $\bar{P}$ and we can conclude that the spectrum of $\bar{P}$ is contained in that of P with the multiplicity constraint. □

The previous lemma shows on the one hand a localisation property in Fourier space of exact solutions of intertwining equations: the $\nu_{\bar{x}}$ have to be with no component on $n - m$ eigenvectors of the Laplacian $\mathscr{L}$ (see Eqs. (29)). On the other hand, it shows that finding exact solutions of intertwining equation implies to have a detailed knowledge of the eigenvectors of the Laplacian.

Conversely, it is now possible to describe all the non-degenerate solutions of the intertwining equations in terms of, on the one hand, the eigenvectors and eigenvalues of P and, on the other hand, the set of diagonalizable stochastic matrices $\bar{P}$ with a given spectrum contained in that of P, and satisfying the multiplicity constraint. Any right eigenvector basis $(C(\cdot, j) : j \in J)$—satisfying (30) and with $C(\cdot, 0) \equiv 1$—of such a $\bar{P}$ will provide, through Eqs. (29) and possibly after rescaling, a non-degenerate solution of the intertwining equations. The only delicate point to check is indeed the non-negativity of the $\nu_{\bar{x}}$. But if this fails, and since $\mu = \mu_0$ charges all points in $\mathscr{X}$, one just has to replace the $C(\cdot, j)$ for positive j in J, by some $\delta_j C(\cdot, j)$ for some small enough δ_j.

At this point we just have to give sufficient conditions for the set of diagonalisable stochastic matrices with a given spectrum to ensure that our intertwining equations do have solutions. The next lemma shows that, if P has non-negative eigenvalues, then we will find solutions with $\bar{\mathscr{X}}$ of any size $m < n$. We further note that this hypothesis will always be fulfilled if instead of considering P we consider its lazy version $(P + Id)/2$.

Lemma 2 *For any*

$$1 = \theta_0 > \theta_1 \geq \theta_2 \geq \cdots \geq \theta_{m-1} \geq 0$$

there always exists a reversible and irreducible *stochastic matrix $\bar{P}$ with such a spectrum. In particular, if P is a reversible and irreducible stochastic matrix that admits $(\theta_j : j < m)$ as a subsequence of its ordered spectrum with multiplicities, then the Markov chains associated with P and $\bar{P}$ are intertwined.*

Proof Let us set

$$A = \begin{pmatrix} 1 & -1 & 0 & \cdots & \cdots & 0 \\ 1 & 1 & -2 & \ddots & & \vdots \\ 1 & 1 & 1 & -3 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & 0 \\ \vdots & \vdots & \vdots & & \ddots & -(m-1) \\ 1 & 1 & 1 & \cdots & \cdots & 1 \end{pmatrix},$$

a matrix with orthogonal rows, and introduce the diagonal matrices

$$D_\theta = \begin{pmatrix} \theta_0 & & & & \\ & \theta_1 & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \theta_{m-1} \end{pmatrix}, \qquad D_{\bar{\mu}} = \begin{pmatrix} \frac{1}{1\times 2} & & & & \\ & \frac{1}{2\times 3} & & & \\ & & \ddots & & \\ & & & \frac{1}{(m-1)m} & \\ & & & & \frac{1}{m} \end{pmatrix},$$

the second one being such that $Q = D_{\bar{\mu}}^{1/2} A$ is orthogonal. We compute

$$\bar{P} = D_{\bar{\mu}}^{-1/2} Q D_\theta Q^t D_{\bar{\mu}}^{1/2} = A D_\theta A^t D_{\bar{\mu}}$$

to find

$$\bar{P} = \begin{pmatrix} \frac{\Sigma_1+\theta_1}{1\times 2} & \frac{\Sigma_1-\theta_1}{2\times 3} & \frac{\Sigma_1-\theta_1}{3\times 4} & \cdots & \frac{\Sigma_1-\theta_1}{(m-1)m} & \frac{\Sigma_1-\theta_1}{m} \\ \frac{\Sigma_1-\theta_1}{1\times 2} & \frac{\Sigma_2+2^2\theta_2}{2\times 3} & \frac{\Sigma_2-2\theta_2}{3\times 4} & \cdots & \frac{\Sigma_2-2\theta_2}{(m-1)m} & \frac{\Sigma_2-2\theta_2}{m} \\ \frac{\Sigma_1-\theta_1}{1\times 2} & \frac{\Sigma_2-2\theta_2}{2\times 3} & \frac{\Sigma_3+3^2\theta_3}{3\times 4} & \cdots & \frac{\Sigma_3-3\theta_3}{(m-1)m} & \frac{\Sigma_3-3\theta_3}{m} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\Sigma_1-\theta_1}{1\times 2} & \frac{\Sigma_2-2\theta_2}{2\times 3} & \frac{\Sigma_3-3\theta_3}{3\times 4} & \cdots & \frac{\Sigma_{m-1}+(m-1)^2\theta_{m-1}}{(m-1)m} & \frac{\Sigma_{m-1}-(m-1)\theta_{m-1}}{m} \\ \frac{\Sigma_1-\theta_1}{1\times 2} & \frac{\Sigma_2-2\theta_2}{2\times 3} & \frac{\Sigma_3-3\theta_3}{3\times 4} & \cdots & \frac{\Sigma_{m-1}-(m-1)\theta_{m-1}}{(m-1)m} & \frac{\Sigma_m}{m} \end{pmatrix} \tag{31}$$

with, for all $1 \le k \le m$, $\Sigma_k = \sum_{j<k} \theta_j$. $\bar{P}$ is stochastic, irreducible and reversible with respect to $\bar{\mu}$ defined by

$$\bar{\mu}(k) = \begin{cases} \frac{1}{k(k+1)} & \text{if } k < m, \\ \frac{1}{m} & \text{if } k = m. \end{cases}$$

It also has the desired spectrum. □

Comment: The proof actually shows that the positivity hypothesis on the θ_j's can be slightly relaxed: we only have to require the numerators of the diagonal coefficients in (31) to be non-negative.

We conclude this section by observing that the universal solution we just provided is not fully satisfactory. First, it requires a detailed knowledge of the spectrum that can be practically unavailable. Second, we can expect such a universal solution to produce very squeezed solutions. Indeed, the coefficients $C(\bar{x}, j)$ in (30) will be given by the matrix $C = D_{\bar{\mu}}^{-1/2} Q = A$ or by $C = A D_\delta$ with D_δ a rescaling diagonal matrix

$$D_\delta = \begin{pmatrix} 1 & & & \\ & \delta_1 & & \\ & & \ddots & \\ & & & \delta_{m-1} \end{pmatrix}$$

ensuring the non-negativity of the $\nu_{\bar{x}}$. The fact that the δ_i's may have to be chosen very small can be the source of very strong squeezing.

2.3 Proof of Proposition 2

Let $\xi_{\bar{x}}^+$ and $\xi_{\bar{x}}^-$ be, for each $\bar{x}$ in $\bar{\mathscr{X}}$, the positive and negative part of the signed measure $\xi_{\bar{x}} = \xi_{\bar{x}}^+ - \xi_{\bar{x}}^-$ such that

$$\nu_{\bar{x}} P = \xi_{\bar{x}} + \sum_{\bar{y} \in \bar{\mathscr{X}}} \bar{P}(\bar{x}, \bar{y}) \nu_{\bar{y}}. \tag{32}$$

Since $\nu_{\bar{x}} P$ and the convex combination of the $\nu_{\bar{y}}$ are both probability measures, $\xi_{\bar{x}}^+$ and $\xi_{\bar{x}}^-$ have the same mass

$$\epsilon_{\bar{x}} = d_{\mathrm{TV}}\left(\nu_{\bar{x}} P, \sum_{\bar{y} \in \bar{\mathscr{X}}} \bar{P}(\bar{x}, \bar{y}) \nu_{\bar{y}} \right).$$

Adding $\xi_{\bar{x}}^-$ on both sides of Eq. (32) and dividing by $1+\epsilon_{\bar{x}}$, we get

$$\frac{1}{1+\epsilon_{\bar{x}}}\nu_{\bar{x}}P + \frac{\epsilon_{\bar{x}}}{1+\epsilon_{\bar{x}}}\pi_{\bar{x}}^- = \frac{\epsilon_{\bar{x}}}{1+\epsilon_{\bar{x}}}\pi_{\bar{x}}^+ + \frac{1}{1+\epsilon_{\bar{x}}}\sum_{\bar{y}\in\bar{\mathscr{X}}}\bar{P}(\bar{x},\bar{y})\nu_{\bar{y}}, \tag{33}$$

where $\pi_{\bar{x}}^+$ and $\pi_{\bar{x}}^-$ are the probability measures obtained by normalization from $\xi_{\bar{x}}^+$ and $\xi_{\bar{x}}^-$.

Let us build a new set $\tilde{\mathscr{X}}$ by associating some $\bar{x}'$ with each $\bar{x}$ in $\bar{\mathscr{X}}$, calling $\bar{\mathscr{X}}'$ the set of these associated $\bar{x}'$ and setting $\tilde{\mathscr{X}} = \bar{\mathscr{X}} \cup \bar{\mathscr{X}}'$ to get a twice as large set. We can then read Eq. (33) as an exact intertwining equation at site $\bar{x}$, between a Markov chain $Z_{\bar{x}}$ with values in $\mathscr{X}$ and transition probabilities

$$P_{\bar{x}}(x,y) = \frac{1}{1+\epsilon_{\bar{x}}}P(x,y) + \frac{\epsilon_{\bar{x}}}{1+\epsilon_{\bar{x}}}\pi_{\bar{x}}^-(y), \qquad x,y\in\mathscr{X},$$

on the one hand, and on the other hand a Markov chain $\bar{Z}$ with values in the augmented set $\tilde{\mathscr{X}}$ and transition probabilities

$$\tilde{P}(\bar{x},\bar{y}) = \frac{1}{1+\epsilon_{\bar{x}}}\bar{P}(\bar{x},\bar{y}), \qquad \bar{x}\in\bar{\mathscr{X}}, \qquad \bar{y}\in\bar{\mathscr{X}},$$

$$\tilde{P}(\bar{x},\bar{y}') = \frac{\epsilon_{\bar{x}}}{1+\epsilon_{\bar{x}}}\mathbb{1}_{\{\bar{y}'=\bar{x}'\}}, \qquad \bar{x}\in\bar{\mathscr{X}}, \qquad \bar{y}'\in\bar{\mathscr{X}}',$$

$$\tilde{P}(\bar{x}',\tilde{y}) = \mathbb{1}_{\{\tilde{y}=\bar{x}'\}}, \qquad \bar{x}'\in\bar{\mathscr{X}}', \qquad \tilde{y}\in\tilde{\mathscr{X}}.$$

The linking probabilities at our $\bar{x}$ in Eq. (33) are the $\nu_{\bar{y}}$ and $\nu_{\bar{x}'} = \pi_{\bar{x}}^+$. The process $Z_{\bar{x}}$ can be constructed with a sequence of independent uniform random variables $(U_k, k\ge 1)$. Assuming that at time k, $Z_{\bar{x}}(k)=z$ and that $U_{k+1} > \frac{\epsilon_{\bar{x}}}{1+\epsilon_{\bar{x}}}$, $Z_{\bar{x}}(k+1)$ is sampled with $P(z,\cdot)$, while if $U_{k+1}\le\frac{\epsilon_{\bar{x}}}{1+\epsilon_{\bar{x}}}$, $Z_{\bar{x}}(k+1)$ is sampled with $\pi_{\bar{x}}^-$.

By Proposition 1, we get a stopping time $T_{\bar{x}}$ and a random variable $\tilde{Y}_{\bar{x}}$, with values in $\{\bar{x}'\}\cup\bar{\mathscr{X}}\setminus\{\bar{x}\}$, such that, conditionally to $\tilde{Y}_{\bar{x}}$, the law of $Z_{\bar{x}}(T_{\bar{x}})$ is $\nu_{\tilde{Y}_{\bar{x}}}$. For $\nu=\nu_{\bar{x}}$ we define then the Markov chain $(Z,\bar{Z})$ on $\mathscr{X}\times\bar{\mathscr{X}}$ in the following way:

- the law of $(Z,\bar{Z})(0)$ is $\nu_{\bar{x}}\otimes\delta_{\bar{x}}$;
- for $k<T_{\bar{x}}$, $(Z,\bar{Z})(k) = (Z_{\bar{x}}(k),\bar{x})$;
- $(Z,\bar{Z})(T_{\bar{x}}) = (Z_{\bar{x}}(T_{\bar{x}}),\tilde{Y}_{\bar{x}})$.
- If $\tilde{Y}_{\bar{x}}=\bar{x}'$ then we set $(Z,\bar{Z})(k) = (Z,\bar{Z})(T_{\bar{x}})$ for all $k\ge T_{\bar{x}}$. Otherwise $\tilde{Y}_{\bar{x}}$ plays now the previous role of $\bar{x}$.

This construction is naturally adapted to $\bar{Z}$ started in $\bar{\nu}$ and Z in $\nu_{\bar{Z}}$. We get the desired equality on the law of Z_k conditioned on $\bar{Z}_k$ as a consequence of properties (2) and (4) of Proposition 1.

By properties (1), (3) and (5) of Proposition 1, $\bar{Z}$ and $\bar{X}^{\bar{\nu}}$ have the same law before the absorbing time

$$\bar{K} = \min\left\{k > 0 : \bar{Z}_k \notin \bar{\mathscr{X}}\right\}.$$

We can then set $\bar{X}^{\bar{\nu}}_k = \bar{Z}_k$ for $k < \bar{K}$, and build $\bar{X}^{\bar{\nu}}$ independently from $(Z, \bar{Z})$ for $k \geq \bar{K}$. Setting

$$K = \min\left\{k < \bar{K} : U_k \leq \frac{\epsilon_{\bar{Z}_k}}{1 + \epsilon_{\bar{Z}_k}}\right\},$$

with the usual convention that the minimum of the empty set is $+\infty$, we can also set $\hat{X}^{\nu}_k = Z_k$ for $k < K$, and build $\hat{X}^{\nu}$ independently of $(Z, \bar{Z})$ for $k \geq K$. We simply conclude this coupling construction by observing that K and $\bar{K}$ stochastically dominate two (correlated) geometric random variables T and $\bar{T}$ with success probability

$$\epsilon \geq \max_{\bar{x} \in \bar{\mathscr{X}}} \frac{\epsilon_{\bar{x}}}{1 + \epsilon_{\bar{x}}}.$$

Let us finally explain why this implies our claimed upper bound on the total variation distance between the law of $\hat{X}^{\nu}_k$ and

$$\xi_k = \sum_{\bar{x} \in \bar{\mathscr{X}}} P\big(\bar{X}^{\bar{\nu}}_k = \bar{x}\big)\nu_{\bar{x}}$$

for any $k \geq 0$. For any $A \subset \mathscr{X}$ it holds

$$\begin{aligned}
\xi_k(A) &= \sum_{\bar{x} \in \bar{\mathscr{X}}} P\big(\bar{X}^{\bar{\nu}}_k = \bar{x}, \bar{T} \leq k\big)\nu_{\bar{x}}(A) + \sum_{\bar{x} \in \bar{\mathscr{X}}} P\big(\bar{X}^{\bar{\nu}}_k = \bar{x}, \bar{T} > k\big)\nu_{\bar{x}}(A) \\
&\leq \sum_{\bar{x} \in \bar{\mathscr{X}}} P\big(\bar{X}^{\bar{\nu}}_k = \bar{x}, \bar{T} \leq k\big) + \sum_{\bar{x} \in \bar{\mathscr{X}}} P\big(\bar{Z}_k = \bar{x}\big)\nu_{\bar{x}}(A) \\
&= P\big(\bar{T} \leq k\big) + \sum_{\bar{x} \in \bar{\mathscr{X}}} P\big(\bar{Z}_k = \bar{x}, Z_k \in A\big) \\
&\leq P\big(\bar{T} \leq k\big) + P\big(Z_k \in A\big) \\
&\leq P\big(\bar{T} \leq k\big) + P\big(T \leq k\big) + P\big(\hat{X}^{\nu}_k \in A\big).
\end{aligned}$$

The same inequality holds with the complementary of A and this concludes the proof.

3 Proof of Theorem 1

3.1 Total Variation Estimates

Inequality (26) is a direct consequence of Inequality (25) and Proposition 7. Indeed,

$$\begin{aligned}
&\mathbb{E}_q\left[\sum_{\bar{x}\in\bar{\mathscr{X}}} d_{TV}(\Lambda P(\bar{x},\cdot),\bar{P}\Lambda(\bar{x},\cdot))\right]\\
&\quad=\sum_{i=1}^{n}\mathbb{E}_q\left[\sum_{\bar{x}\in\bar{\mathscr{X}}} d_{TV}(\Lambda P(\bar{x},\cdot),\bar{P}\Lambda(\bar{x},\cdot))\,\middle|\,|\bar{\mathscr{X}}|=i\right]\mathbb{P}_q\left[|\bar{\mathscr{X}}|=i\right]\\
&\quad\le\sum_{i=1}^{n}\frac{q'(n-i)}{\alpha}\mathbb{P}_q\left[|\bar{\mathscr{X}}|=i\right]\\
&\quad=\frac{q'}{\alpha}\mathbb{E}_q\left[n-|\bar{\mathscr{X}}|\right].
\end{aligned}$$

It remains thus to prove (25). Applying Markov property at time σ_1, we get

$$\bar{P}(\bar{x},\cdot)=\sum_{y\in\mathscr{X}}P(\bar{x},y)P_y\left[X(H_{\bar{\mathscr{X}}})=\cdot\right].$$

Moreover, set $\delta_{\bar{x}}$ the Dirac measure at $\bar{x}$, seen both as a probability measure on $\mathscr{X}$ and as a row vector of dimension n. Then, we can rewrite

$$\begin{aligned}
\Lambda P(\bar{x},\cdot)&=\delta_{\bar{x}}K_{q'}P(\cdot)=\delta_{\bar{x}}PK_{q'}(\cdot)=\sum_{y\in\mathscr{X}}P(\bar{x},y)P_y\left[X(T_{q'})=\cdot\right]\\
&=\sum_{y\in\mathscr{X}}P(\bar{x},y)P_y\left[H_{\bar{\mathscr{X}}}<T_{q'};X(T_{q'})=\cdot\right]\\
&\qquad+\sum_{y\in\mathscr{X}}P(\bar{x},y)P_y\left[H_{\bar{\mathscr{X}}}\ge T_{q'};X(T_{q'})=\cdot\right]\\
&=\sum_{y\in\mathscr{X},\bar{z}\in\bar{\mathscr{X}}}P(\bar{x},y)P_y\left[H_{\bar{\mathscr{X}}}<T_{q'};X(H_{\bar{\mathscr{X}}})=\bar{z}\right]P_{\bar{z}}\left[X(T_{q'})=\cdot\right]\\
&\qquad+\sum_{y\in\mathscr{X}}P(\bar{x},y)P_y\left[H_{\bar{\mathscr{X}}}\ge T_{q'};X(T_{q'})=\cdot\right]
\end{aligned}$$

$$
= \bar{P}\Lambda(\bar{x},\cdot) - \sum_{y\in\mathscr{X},\bar{z}\in\bar{\mathscr{X}}} P(\bar{x},y)P_y\left[H_{\bar{\mathscr{X}}} \geq T_{q'}; X(H_{\bar{\mathscr{X}}}) = \bar{z}\right] P_{\bar{z}}\left[X(T_{q'}) = \cdot\right]
+ \sum_{y\in\mathscr{X}} P(\bar{x},y)P_y\left[H_{\bar{\mathscr{X}}} \geq T_{q'}; X(T_{q'}) = \cdot\right].
$$

Therefore,

$$
\begin{aligned}
d_{TV}&(\Lambda P(\bar{x},\cdot), \bar{P}\Lambda(\bar{x},\cdot)) \\
&= \frac{1}{2}\sum_{x\in\mathscr{X}} \left|\Lambda P(\bar{x},x) - \bar{P}\Lambda(\bar{x},x)\right| \\
&\leq \frac{1}{2}\sum_{x\in\mathscr{X},y\in\mathscr{X},\bar{z}\in\bar{\mathscr{X}}} P(\bar{x},y)P_y\left[H_{\bar{\mathscr{X}}} \geq T_{q'}; X(H_{\bar{\mathscr{X}}}) = \bar{z}\right] P_{\bar{z}}\left[X(T_{q'}) = x\right] \\
&\quad + \frac{1}{2}\sum_{x\in\mathscr{X},y\in\mathscr{X}} P(\bar{x},y)P_y\left[H_{\bar{\mathscr{X}}} \geq T_{q'}; X(T_{q'}) = x\right] \\
&= \sum_{y\in\mathscr{X}} P(\bar{x},y)P_y\left[H_{\bar{\mathscr{X}}} \geq T_{q'}\right] \\
&= \sum_{y\in\mathscr{X}} P(\bar{x},y)E_y\left[1 - e^{-q'H_{\bar{\mathscr{X}}}}\right] \\
&\leq \sum_{y\in\mathscr{X}} P(\bar{x},y)E_y\left[q'H_{\bar{\mathscr{X}}}\right] = q'E_{\bar{x}}\left[H^+_{\bar{\mathscr{X}}} - \sigma_1\right].
\end{aligned}
$$

We now take the expectation with respect to $\mathbb{E}_q$.

$$
\mathbb{E}_q\left[\sum_{\bar{x}\in\bar{\mathscr{X}}} d_{TV}(\Lambda P(\bar{x},\cdot), \bar{P}\Lambda(\bar{x},\cdot)) \,\middle|\, |\bar{\mathscr{X}}| = m\right]
\leq q'\mathbb{E}_q\left[\sum_{\bar{x}\in\bar{\mathscr{X}}} E_{\bar{x}}\left[H^+_{\bar{\mathscr{X}}} - \sigma_1\right] \,\middle|\, |\bar{\mathscr{X}}| = m\right].
$$

Formula (11) gives then the desired result.

3.2 Squeezing Estimates

We now prove the quantitative upper bounds on the squeezing of Λ stated in (27). We begin with the following lemma:

Lemma 3 *For any* $m \in \{1, \cdots, n\}$,

$$\mathbb{E}_q\left[\mathscr{S}(\Lambda)\,\Big|\,|\bar{\mathscr{X}}|=m\right] \leq \frac{\sqrt{\sum_{|J|=m-1}\prod_{j\in J}p_j'^2}\sqrt{\sum_{|J|=m}\prod_{j\in J}p_j'^{-2}\prod_{j\in J}p_j^2\prod_{j\notin J}(1-p_j)^2}}{\sum_{|J|=m}\prod_{j\in J}p_j\prod_{j\notin J}(1-p_j)}. \tag{34}$$

Proof Note first that

$$\mathscr{S}(\Lambda)^2 = \sum_{\bar{x}\in\bar{\mathscr{X}}}\Gamma^{-1}(\bar{x},\bar{x}) = \sum_{\bar{x}\in\bar{\mathscr{X}}}\frac{\det_{\bar{\mathscr{X}}\setminus\{\bar{x}\}}(\Gamma)}{\det(\Gamma)} = \sum_{\bar{x}\in\bar{\mathscr{X}}}\frac{\mathrm{Vol}^2(\nu_{\bar{y}};\,\bar{y}\in\bar{\mathscr{X}},\,\bar{y}\neq\bar{x})}{\mathrm{Vol}^2(\nu_{\bar{y}};\,\bar{y}\in\bar{\mathscr{X}})}.$$

Hence,

$$\mathbb{E}_q\left[\mathscr{S}(\Lambda)\,\Big|\,|\bar{\mathscr{X}}|=m\right] = \sum_{|R|=m}\mathbb{P}_q\left[\bar{\mathscr{X}}=R\,\Big|\,|\bar{\mathscr{X}}|=m\right]\frac{\sqrt{\sum_{\bar{x}\in R}\mathrm{Vol}^2(\nu_{\bar{y}};\,\bar{y}\in R,\,\bar{y}\neq\bar{x})}}{\sqrt{\mathrm{Vol}^2(\nu_{\bar{y}};\,\bar{y}\in R)}}. \tag{35}$$

From Proposition 4, $\bar{\mathscr{X}} = \rho(\Phi)$ is a determinantal process associated to the kernel K_q. Remind that for all $j \in \{0, \cdots, n-1\}$, $\mu_j(-\mathscr{L}) = \lambda_j\mu_j$. The μ_j are orthogonal by symmetry of $-\mathscr{L}$, and we assume that for all $j \in \{0, \cdots, n-1\}$, $\|\mu_j\|^* = 1$, so that $\mu_0 = \mu$. Hence, we get $\mu_j K_q = \frac{q}{q+\lambda_j}\mu_j$. One way to construct $\rho(\Phi)$, the number of roots being fixed equal to m, is to choose m eigenvectors of K_q, according to Bernoulli random variables with parameters p_j, and then to choose $\bar{\mathscr{X}}$ according to the determinantal process associated to the projector operator onto the m chosen eigenvectors. More formally,

$$\mathbb{P}_q\left[\bar{\mathscr{X}}=R\,\Big|\,|\bar{\mathscr{X}}|=m\right] = \frac{1}{Z_{m,q}}\sum_{|J|=m}\prod_{j\in J}\frac{q}{q+\lambda_j}\prod_{j\notin J}\frac{\lambda_j}{q+\lambda_j}\det{}^2\left(\left\langle\frac{\delta_{\bar{x}}}{\|\delta_{\bar{x}}\|^*};\mu_j\right\rangle^*_{\bar{x}\in R,\,j\in J}\right), \tag{36}$$

where $Z_{m,q}$ is a normalizing constant ($Z_{m,q} = \mathbb{P}_q\left[\left|\bar{\mathscr{X}}\right| = m\right]$). We go back to (35) and turn to the term $\mathrm{Vol}^2(\nu_{\bar{y}};\, \bar{y} \in R)$. It follows from Cauchy-Binet formula that

$$\mathrm{Vol}^2(\nu_{\bar{y}};\, \bar{y} \in R) = \sum_{|J|=m} \det{}^2\left(\left\langle \nu_{\bar{y}}, \mu_j \right\rangle^*, \bar{y} \in R, j \in J\right) .$$

Note that

$$\nu_{\bar{y}} = \delta_{\bar{y}} K_{q'} = \sum_{j=0}^{n-1} \left\langle \delta_{\bar{y}}; \mu_j \right\rangle^* \mu_j K_{q'} = \sum_{j=0}^{n-1} p'_j \left\langle \delta_{\bar{y}}; \mu_j \right\rangle^* \mu_j .$$

Thus $\left\langle \nu_{\bar{y}}, \mu_j \right\rangle^* = p'_j \left\langle \delta_{\bar{y}}; \mu_j \right\rangle^*$. We obtain then

$$\mathrm{Vol}^2(\nu_{\bar{y}};\, \bar{y} \in R) = \sum_{|J|=m} \prod_{j\in J} p'^2_j \det{}^2\left(\left\langle \delta_{\bar{y}}; \mu_j \right\rangle^*, \bar{y} \in R, j \in J\right) . \tag{37}$$

Putting (36) and (37) into (35), we are led to

$$\mathbb{E}_q\left[\mathscr{S}(\Lambda) \,\middle|\, \left|\bar{\mathscr{X}}\right| = m\right] = \frac{1}{Z_{m,q}} \sum_{|R|=m} \frac{\sqrt{\sum_{\bar{x}\in R} \mathrm{Vol}^2(\nu_{\bar{y}};\, \bar{y} \in R, \bar{y} \neq \bar{x})}}{\prod_{\bar{x}\in R} \|\delta_{\bar{x}}\|^{*2}}$$
$$\times \frac{\sum_{|J|=m} \prod_{j\in J} p_j \prod_{j\notin J} (1-p_j) \det{}^2\left(\left\langle \delta_{\bar{x}}; \mu_j \right\rangle^*, \bar{x} \in R, j \in J\right)}{\sqrt{\sum_{|J|=m} \prod_{j\in J} p'^2_j \det{}^2\left(\left\langle \delta_{\bar{y}}; \mu_j \right\rangle^*, \bar{y} \in R, j \in J\right)}} .$$

Cauchy-Schwartz inequality then yields

$$\frac{\sum_{|J|=m} \prod_{j\in J} p_j \prod_{j\notin J} (1-p_j) \det{}^2\left(\left\langle \delta_{\bar{x}}; \mu_j \right\rangle^*, \bar{x} \in R, j \in J\right)}{\sqrt{\sum_{|J|=m} \prod_{j\in J} p'^2_j \det{}^2\left(\left\langle \delta_{\bar{y}}; \mu_j \right\rangle^*, \bar{y} \in R, j \in J\right)}}$$
$$\leq \sqrt{\sum_{|J|=m} \prod_{j\in J} \frac{p_j^2}{p'^2_j} \prod_{j\notin J} (1-p_j)^2 \det{}^2\left(\left\langle \delta_{\bar{x}}; \mu_j \right\rangle^*, \bar{x} \in R, j \in J\right)}$$

and

$$\mathbb{E}_q\left[\mathscr{S}(\Lambda)\,\middle|\,|\bar{\mathscr{X}}|=m\right]$$
$$\le \frac{1}{Z_{m,q}}\sqrt{\sum_{|R|=m}\frac{\sum_{\bar{x}\in R}\mathrm{Vol}^2(\nu_{\bar{y}};\,\bar{y}\in R,\,\bar{y}\neq\bar{x})}{\prod_{\bar{x}\in R}\|\delta_{\bar{x}}\|^{*2}}}$$
$$\times\sqrt{\sum_{\substack{|R|=m\\|J|=m}}\prod_{j\in J}\frac{p_j^2}{p_j'^2}\prod_{j\notin J}(1-p_j)^2\,\det{}^2\left(\left\langle\frac{\delta_{\bar{x}}}{\|\delta_{\bar{x}}\|^*};\mu_j\right\rangle^*;\bar{x}\in R,\,j\in J\right)}\,.$$

Using again Cauchy-Binet formula, we get

$$\sum_{|R|=m}\det{}^2\left(\left\langle\frac{\delta_{\bar{x}}}{\|\delta_{\bar{x}}\|^*};\mu_j\right\rangle^*,\bar{x}\in R,\,j\in J\right)=\mathrm{Vol}^2(\mu_j,\,j\in J)=1\,,$$

so that the term in the second square root is equal to

$$\sum_{|J|=m}\prod_{j\in J}\frac{p_j^2}{p_j'^2}\prod_{j\notin J}(1-p_j)^2\,.$$

We turn now to the term in the first square root, which can be rewritten, by using twice the Cauchy-Binet formula, as

$$\sum_{\bar{x}\in\mathscr{X}}\frac{1}{\|\delta_{\bar{x}}\|^{*2}}\sum_{|R|=m,\bar{x}\in R}\frac{\mathrm{Vol}^2(\nu_{\bar{y}};\,\bar{y}\in R,\,\bar{y}\neq\bar{x})}{\prod_{\bar{x}\in R\setminus\{\bar{x}\}}\|\delta_{\bar{x}}\|^{*2}}$$
$$=\sum_{\bar{x}\in\mathscr{X}}\mu(\bar{x})\sum_{R\subset\mathscr{X}\setminus\{\bar{x}\},|R|=m-1}\frac{\mathrm{Vol}^2(\nu_{\bar{y}};\,\bar{y}\in R)}{\prod_{\bar{y}\in R}\|\delta_{\bar{y}}\|^{*2}}$$
$$=\sum_{\bar{x}\in\mathscr{X}}\mu(\bar{x})\sum_{R\subset\mathscr{X}\setminus\{\bar{x}\},|R|=m-1}\sum_{|J|=m-1}\prod_{j\in J}p_j'^2\,\det{}^2\left(\left\langle\frac{\delta_{\bar{y}}}{\|\delta_{\bar{y}}\|^*};\mu_j\right\rangle^*;\bar{y}\in R,\,j\in J\right)$$
$$\le\sum_{\bar{x}\in\mathscr{X}}\mu(\bar{x})\sum_{|J|=m-1}\prod_{j\in J}p_j'^2\sum_{|R|=m-1}\det{}^2\left(\left\langle\frac{\delta_{\bar{y}}}{\|\delta_{\bar{y}}\|^*};\mu_j\right\rangle^*,\bar{y}\in R,\,j\in J\right)$$

$$= \sum_{|J|=m-1} \prod_{j \in J} p_j'^2 \operatorname{Vol}^2(\mu_j, j \in J)$$

$$= \sum_{|J|=m-1} \prod_{j \in J} p_j'^2 .$$

To end the proof of the lemma, it is sufficient to note that

$$Z_{m,q} = \mathbb{P}_q \left(\left| \bar{\mathscr{X}} \right| = m \right) = \sum_{|J|=m} \prod_{j \in J} p_j \prod_{j \notin J} (1 - p_j). \qquad \square$$

We can now conclude the proof of (27) and of Theorem 1. For any $t > 0$ it holds

$$\sum_{|J|=m-1} \prod_{j \in J} p_j'^2 \leq \frac{1}{t^{m-1}} \prod_{j=0}^{n-1} (1 + t p_j'^2) = \frac{1+t}{t^{m-1}} \prod_{j=1}^{n-1} (1 + t p_j'^2) .$$

since the left-hand is the coefficient of t^{m-1} in the product $\prod_{j=0}^{n-1}(1 + t p_j'^2)$. In the same way, for any $x > 0$,

$$\begin{aligned}
&\sum_{|J|=m-1, J \subset \{1,\cdots,n-1\}} \prod_{j \in J} \frac{p_j^2}{p_j'^2} \prod_{j \in \{1,\cdots,n-1\} \setminus J} (1 - p_j)^2 \\
&\qquad = \prod_{j=1}^{n-1} (1 - p_j)^2 \sum_{|J|=m-1, J \subset \{1,\cdots,n-1\}} \prod_{j \in J} \frac{p_j^2}{p_j'^2 (1 - p_j)^2} \\
&\qquad \leq \frac{\prod_{j=1}^{n-1} (1 - p_j)^2}{x^{m-1}} \prod_{j=1}^{n-1} \left(1 + x \frac{p_j^2}{p_j'^2 (1 - p_j)^2} \right) \\
&\qquad = \frac{1}{x^{m-1}} \prod_{j=1}^{n-1} \left((1 - p_j)^2 + x \frac{p_j^2}{p_j'^2} \right) .
\end{aligned}$$

Hence, for any $x, t > 0$, for any $m \in \{1, \cdots, n\}$,

$$\begin{aligned}
&\mathbb{E}_q \left[\mathscr{S}(\Lambda) \,\middle|\, \left| \bar{\mathscr{X}} \right| = m \right] \\
&\qquad \leq \frac{1}{\mathbb{P}_q \left[\left| \bar{\mathscr{X}} \right| = m \right]} \frac{\sqrt{1+t}}{(tx)^{\frac{m-1}{2}}} \sqrt{\prod_{j=1}^{n-1} (1 + t p_j'^2) \left((1 - p_j)^2 + x \frac{p_j^2}{p_j'^2} \right)} .
\end{aligned}$$

One can check that

$$(1+tp_j'^2)\left((1-p_j)^2+x\frac{p_j^2}{p_j'^2}\right)=\left(1+(\sqrt{xt}-1)p_j\right)^2+\left(\sqrt{t}p_j'(1-p_j)-\sqrt{x}\frac{p_j}{p_j'}\right)^2.$$

Take now $xt=1$. We obtain that for any $t>0$, for any $m\in\{1,\cdots,n\}$,

$$\mathbb{E}_q\left[\mathscr{S}(\Lambda)\mathbb{1}_{|\bar{\mathscr{X}}|=m}\right]\leq\sqrt{1+t}\sqrt{\prod_{j=1}^{n-1}\left(1+\left(\sqrt{t}p_j'(1-p_j)-\frac{1}{\sqrt{t}}\frac{p_j}{p_j'}\right)^2\right)}$$

$$\leq\sqrt{1+t}\,\exp\left(\frac{1}{2}\sum_{j=1}^{n-1}\left(\sqrt{t}p_j'(1-p_j)-\frac{1}{\sqrt{t}}\frac{p_j}{p_j'}\right)^2\right)$$

$$=\sqrt{1+t}\,\exp\left(\frac{t}{2}\sum_{j=1}^{n-1}p_j'^2(1-p_j)^2+\frac{1}{2t}\sum_{j=1}^{n-1}\frac{p_j^2}{p_j'^2}-\sum_{j=1}^{n-1}p_j(1-p_j)\right).$$

Optimizing the exponential term in t and choosing $t=T_n$ lead to (27).

4 Proof of Theorem 2

Let us first rewrite $K_{q'}$ in terms of μ.

Lemma 4 *For $x\in\mathscr{X}$,*

$$K_{q'}(x,\cdot)=\mathbb{E}_{q'}\left[\mu_{A'(x)\cap A(\rho_x')}(\cdot)\right],\ \mathbb{P}_q\ a.s..$$

Proof Starting Wilson's algorithm from x to construct Φ', we get

$$\begin{aligned}K_{q'}(x,y)&=P_x\left[X(T_{q'})=y\right]\\&=\mathbb{P}_{q'}\left[\rho_x'=y\right]\\&=\mathbb{E}_{q'}\left[\mathbb{P}_{q'}\left[\rho_x'=y\mid\mathscr{A}(\Phi')\right]\right]\\&=\mathbb{E}_{q'}\left[\mu_{A'(x)}(y)\right],\end{aligned}$$

where the last equality comes from Proposition 5. Hence, $\mathbb{P}_q$ a.s.,

$$\begin{aligned}K_{q'}(x,y)&=\sum_{\bar{x}\in\bar{\mathscr{X}}}\mathbb{E}_{q'}\left[\mu_{A'(x)\cap A(\bar{x})}(y)\,\mu_{A'(x)}(A(\bar{x}))\right]\\&=\sum_{\bar{x}\in\bar{\mathscr{X}}}\mathbb{E}_{q'}\left[\mu_{A'(x)\cap A(\bar{x})}(y)\mathbb{P}_{q'}\left[\rho_x'\in A(\bar{x})\mid\mathscr{A}(\Phi')\right]\right]\end{aligned}$$

$$= \sum_{\bar{x} \in \bar{\mathscr{X}}} \mathbb{E}_{q'} \left[\mu_{A'(x) \cap A(\bar{x})}(y) \mathbb{1}_{A(\bar{x})}(\rho'_x) \right]$$

$$= \mathbb{E}_{q'} \left[\mu_{A'(x) \cap A(\rho'_x)}(y) \right] . \qquad \square$$

Lemma 5 *For any $x \in \mathscr{X}$, set $\tilde{K}_{q'}(x, \cdot) = \mathbb{E}_{q'} \left[\mu_{A(\rho'_x)}(\cdot) \right]$. Then, $\mathbb{P}_q$ a.s.,*

$$\Lambda \tilde{K}_{q'} = \bar{P} \Lambda .$$

Proof $\mathbb{P}_q$ a.s., for any $x, y \in \mathscr{X}$,

$$\tilde{K}_{q'}(x, y) = \mathbb{E}_{q'} \left[\mu_{A(\rho'_x)}(y) \right] = \sum_{\bar{y} \in \bar{\mathscr{X}}} \mu_{A(\bar{y})}(y) \mathbb{P}_{q'} \left[\rho'_x \in A(\bar{y}) \right]$$

$$= \sum_{\bar{y} \in \bar{\mathscr{X}}} \nu_{\bar{y}}(y) P_x \left[X(T_{q'}) \in A(\bar{y}) \right] .$$

Hence, $\mathbb{P}_q$ a.s., for any $\bar{x} \in \bar{\mathscr{X}}$, and $y \in \mathscr{X}$,

$$\nu_{\bar{x}} \tilde{K}_{q'}(y) = \sum_{x \in \mathscr{X}} \sum_{\bar{y} \in \bar{\mathscr{X}}} \nu_{\bar{x}}(x) \nu_{\bar{y}}(y) P_x \left[X(T_{q'}) \in A(\bar{y}) \right]$$

$$= \sum_{\bar{y} \in \bar{\mathscr{X}}} \nu_{\bar{y}}(y) P_{\nu_{\bar{x}}} \left[X(T_{q'}) \in A(\bar{y}) \right]$$

$$= \bar{P} \Lambda(\bar{x}, y) . \qquad \square$$

Therefore, $\mathbb{P}_q$ a.s., for any $\bar{x} \in \bar{\mathscr{X}}$,

$$\begin{aligned} d_{TV}(\Lambda K_{q'}(\bar{x}, \cdot), \bar{P} \Lambda(\bar{x}, \cdot)) &= d_{TV}(\Lambda K_{q'}(\bar{x}, \cdot), \Lambda \tilde{K}_{q'}(\bar{x}, \cdot)) \\ &\leq \sum_{x \in \mathscr{X}} \nu_{\bar{x}}(x)\, d_{TV}(K_{q'}(x, \cdot), \tilde{K}_{q'}(x, \cdot)) \\ &\leq \sum_{x \in \mathscr{X}} \nu_{\bar{x}}(x) \mathbb{E}_{q'} \left[d_{TV}(\mu_{A'(x) \cap A(\rho'_x)}, \mu_{A(\rho'_x)}) \right] . \end{aligned}$$

When B is a subset of C, one has $d_{TV}(\mu_B, \mu_C) = \mu_C(B^c)$. This yields

$$\begin{aligned} d_{TV}(\Lambda K_{q'}(\bar{x}, \cdot), \bar{P} \Lambda(\bar{x}, \cdot)) &\leq \sum_{x \in \mathscr{X}} \nu_{\bar{x}}(x) \mathbb{E}_{q'} \left[\mu_{A(\rho'_x)}(A'(x)^c) \right] \\ &= \sum_{x \in \mathscr{X}} \nu_{\bar{x}}(x) \mathbb{E}_{q'} \left[\mathbb{P}_q \left[\rho_{\rho'_x} \notin A'(x) | \mathscr{A}(\Phi) \right] \right] \end{aligned}$$

Note that

$$\sum_{\bar{x}\in\bar{\mathscr{X}}} \nu_{\bar{x}}(x) = \sum_{\bar{x}\in\bar{\mathscr{X}}} \frac{\mu(x)}{\mu(A(\bar{x}))} \mathbb{1}_{A(\bar{x})}(x) = \sum_{\bar{x}\in\bar{\mathscr{X}}} \frac{\mu(x)}{\mu(A(x))} \mathbb{1}_{A(\bar{x})}(x)$$
$$= \frac{\mu(x)}{\mu(A(x))} \sum_{\bar{x}\in\bar{\mathscr{X}}} \mathbb{1}_{A(\bar{x})}(x) = \frac{\mu(x)}{\mu(A(x))} \,.$$

Summing on $\bar{x}$ and integrating w.r.t. $\mathbb{E}_q$, leads to

$$\mathbb{E}_q \left[\sum_{\bar{x}\in\bar{\mathscr{X}}} d_{TV}(\Lambda K_{q'}(\bar{x},\cdot), \bar{P}\Lambda(\bar{x},\cdot)) \right] \leq \sum_{x\in\mathscr{X}} \mathbb{E}_{q,q'} \left[\mu_{A(x)}(x) \mathbb{1}_{A'(x)^c}(\rho_{\rho'_x}) \right] .$$

Let $p \geq 1$ and p^* its conjugate exponent. Using Hölder's inequality, we get

$$\mathbb{E}_q \left[\sum_{\bar{x}\in\bar{\mathscr{X}}} d_{TV}(\Lambda K_{q'}(\bar{x},\cdot), \bar{P}\Lambda(\bar{x},\cdot)) \right]$$
$$\leq \left(\sum_{x\in\mathscr{X}} \mathbb{E}_{q,q'} \left[\mu_{A(x)}(x)^p \right] \right)^{1/p} \left(\sum_{x\in\mathscr{X}} \mathbb{P}_{q,q'} \left[\rho_{\rho'_x} \notin A'(x) \right] \right)^{1/p^*}$$
$$\leq \left(\sum_{x\in\mathscr{X}} \mathbb{E}_{q} \left[\mu_{A(x)}(x) \right] \right)^{1/p} \left(\sum_{x\in\mathscr{X}} \mathbb{P}_{q,q'} \left[\rho_{\rho'_x} \notin A'(x) \right] \right)^{1/p^*} .$$

Note that

$$\sum_{x\in\mathscr{X}} \mathbb{E}_q \left[\mu_{A(x)}(x) \right] = \sum_{x\in\mathscr{X}} \mathbb{P}_q \left[\rho_x = x \right] = \sum_{x\in\mathscr{X}} \mathbb{P}_q \left[x \in \rho(\Phi) \right] = \mathbb{E}_q \left[|\rho(\Phi)| \right] .$$

Therefore,

$$\mathbb{E}_q \left[\sum_{\bar{x}\in\bar{\mathscr{X}}} d_{TV}(\Lambda K_{q'}(\bar{x},\cdot), \bar{P}\Lambda(\bar{x},\cdot)) \right]$$
$$\leq \left(\mathbb{E}_q \left[|\rho(\Phi)| \right] \right)^{1/p} \left(\sum_{x\in\mathscr{X}} \mathbb{P}_{q,q'} \left[\rho_{\rho'_x} \notin A'(x) \right] \right)^{1/p^*} . \tag{38}$$

To conclude the proof of our theorem we evaluate $\mathbb{P}_{q,q'} \left[\rho_{\rho'_x} \notin A'(x) \right]$ for x any given point in $\mathscr{X}$.

Lemma 6 *For any $x \in \mathscr{X}$, let Γ'_x be the path going from x to ρ'_x in Φ'. Then,*

$$\mathbb{P}_{q,q'}\left[\rho_{\rho'_x} \notin A'(x)\right] \leq \frac{q'}{q}\mathbb{E}_{q'}\left[\left|\Gamma'_x\right|\right] .$$

Proof To decide whether $\rho_{\rho'_x}$ is in $A'(x)$ or not, we do the following construction:

1. We begin the construction of Φ' using Wilson's algorithm starting from x. Thus, we let evolve the Markov process starting from x until an exponential time of parameter q', and erase the loop. The result is an oriented path γ' ($= \Gamma'_x$) without loops from x to a point y ($= \rho'_x$).
2. We go on with the construction of Φ with Wilson's algorithm starting from y. We let evolve the Markov process starting from y until an exponential time T_q of parameter q. The Markov process stops at a point v ($= \rho_{\rho'_x}$).
3. Finally, we continue the construction of Φ' using Wilson's algorithm starting from v. We let evolve the Markov process starting from v, and we stop it after an exponential time $T_{q'}$ of parameter q', or when it reaches the already constructed path γ'. At this point, we are able to decide whether $\rho_{\rho'_x}$ is in $A'(x)$ or not, since $\rho_{\rho'_x} \in A'(x)$ if and only if $T_{q'}$ is bigger than the hitting time of γ'.

Using this construction, we get that for any self-avoiding path γ' from x to y,

$$\mathbb{P}_{q,q'}\left[\rho_{\rho'_x} \notin A'(x) | \Gamma'_x = \gamma'; \rho'_x = y\right] = P_y\left[T_{q'} < H_{\gamma'} \circ \theta_{T_q}\right] ,$$

where θ_t denotes the time shift. Recall that σ_1 is the first time of the clock process on which X is build from $\hat{X}$, and let S_i be the successive return times to γ':

$$S_0 = 0\,,\;\; S_1 = \inf\left\{t \geq \sigma_1; X(t) \in \gamma'\right\} = H^+_{\gamma'}\,,\;\; S_{i+1} = S_i + S_1 \circ \theta_{S_i}\,.$$

Then,

$$P_y\left[T_{q'} < H_{\gamma'} \circ \theta_{T_q}\right] = \sum_{i=0}^{\infty} P_y\left[S_i \leq T_q < S_{i+1}; T_{q'} < H_{\gamma'} \circ \theta_{T_q}\right] .$$

Now, if $S_i \leq T_q < S_i + \sigma_1 \circ \theta_{S_i}$, $X(T_q) \in \gamma'$ and $H_{\gamma'} \circ \theta_{T_q} = 0 < T_{q'}$. If $T_q \geq S_i + \sigma_1 \circ \theta_{S_i}$ and $T_q < S_{i+1}$, $X(T_q) \notin \gamma'$ and $H_{\gamma'} \circ \theta_{T_q} = S_{i+1} - T_q$. Therefore,

$$\begin{aligned} P_y\left[T_{q'} < H_{\gamma'} \circ \theta_{T_q}\right] &= \sum_{i=0}^{\infty} P_y\left[S_i + \sigma_1 \circ \theta_{S_i} \leq T_q < T_{q'} + T_q < S_{i+1}\right] \\ &= \sum_{i=0}^{\infty}\sum_{z\in\gamma'} P_y\left[S_i \leq T_q; X(S_i) = z\right] P_z\left[\sigma_1 \leq T_q < T_{q'} + T_q < H^+_{\gamma'}\right], \end{aligned}$$

using Markov property at time S_i. Set $\tilde{G}_q(y,z,\gamma') = E_y\left[\sum_{i=0}^{+\infty} \mathbb{1}_{S_i \le T_q; X(S_i)=z}\right]$. Since $z \in \gamma'$, $\tilde{G}_q(y,z,\gamma')$ is the mean number of visits to the point z up to time T_q. We have obtained that

$$P_y\left[T_{q'} < H_{\gamma'} \circ \theta_{T_q}\right] = \sum_{z \in \gamma'} \tilde{G}_q(y,z,\gamma') P_z\left[\sigma_1 \le T_q < T_q + T_{q'} < H^+_{\gamma'}\right].$$

We now use Markov property at time σ_1 to write

$$\begin{aligned} &P_z\left[\sigma_1 \le T_q < T_q + T_{q'} < H^+_{\gamma'}\right] \\ &\quad = \sum_{u \notin \gamma'} P_z\left[\sigma_1 \le T_q, X(\sigma_1) = u\right] P_u(T_q < T_{q'} + T_q < H_{\gamma'}) \\ &\quad \le \sum_{u \notin \gamma'} \frac{\alpha}{q+\alpha} P(z,u) P_u(T_{q'} < H_{\gamma'}) \\ &\quad = \sum_{u \notin \gamma'} \frac{1}{q+\alpha} w(z,u) \mathbb{P}_{q'}\left[\rho'_u \ne y \mid \Gamma'_x = \gamma'\right], \end{aligned}$$

using that $\alpha P(z,u) = \mathscr{L}(z,u) = w(z,u)$ for $z \ne u$. Integrating over γ' and y, we are led to

$$\mathbb{P}_{q,q'}\left[\rho_{\rho'_x} \notin A'(x)\right] \le \sum_{y \in \mathscr{X}} \sum_{\gamma : x \rightsquigarrow y} \sum_{z \in \gamma} \sum_{u \notin \gamma} \frac{\tilde{G}_q(y,z,\gamma)}{q+\alpha} w(z,u) \mathbb{P}_{q'}\left[\rho'_u \ne y; \Gamma'_x = \gamma; \rho'_x = y\right]$$

where the sum over γ' is the sum on all self-avoiding paths going from x to y. Now, introducing for any such path γ

$$\mathscr{F}_1(y,\gamma,u) := \{\phi\ s.o.f.;\ y \in \rho(\phi), \gamma \subset \phi, \rho_u \ne y\},$$

this can be rewritten, with $w(\phi) = \prod_{e \in \phi} w(e)$, as

$$\mathbb{P}_{q,q'}\left[\rho_{\rho'_x} \notin A'(x)\right] = \sum_{y \in \mathscr{X}} \sum_{\gamma : x \rightsquigarrow y} \sum_{z \in \gamma} \sum_{u \notin \gamma} \sum_{\phi \in \mathscr{F}_1(y,\gamma,u)} \frac{\tilde{G}_q(y,z,\gamma)}{q+\alpha} w(z,u) \frac{(q')^{|\rho(\phi)|} w(\phi)}{Z(q')}.$$

Lemma 7 *Let* $G_q(y,z) = E_y\left[\int_0^{T_q} \mathbb{1}_{X(s)=z}\, ds\right]$. *Then* $G_q(y,z) = \tilde{G}_q(y,z,\gamma)/(q+\alpha)$ *for any self-avoiding path* γ *that contains* z *and goes from* x *to* y.

Proof Let V_i be the successive return times to z:

$$V_0 = 0,\ \ V_1 = \inf\{t \ge \sigma_1; X(t) = z\},\ \ V_{i+1} = V_i + V_1 \circ \theta_{V_i}.$$

Then $\tilde{G}_q(y,z,\gamma) = \delta_y(z) + \sum_{i=1}^{+\infty} E_y[\mathbb{1}_{V_i \le T_q}]$. Moreover, using Markov's property at time V_i,

$$\begin{aligned}
G_q(y,z) &= \sum_{i=0}^{\infty} E_y\left[\int_{V_i}^{V_{i+1}} \mathbb{1}_{T_q \ge s}\mathbb{1}_{X(s)=z}\, ds\right] \\
&= \sum_{i=0}^{\infty} E_y\left[\mathbb{1}_{V_i \le T_q} E_{X(V_i)}\left[\int_0^{V_1} \mathbb{1}_{T_q \ge s}\mathbb{1}_{X(s)=z}\, ds\right]\right] \\
&= E_y\left[\int_0^{V_1} \mathbb{1}_{T_q \ge s}\mathbb{1}_{X(s)=z}\, ds\right] + \sum_{i=1}^{\infty} E_y\left[\mathbb{1}_{V_i \le T_q}\right] E_z\left[\int_0^{V_1} \mathbb{1}_{T_q \ge s}\mathbb{1}_{X(s)=z}\, ds\right] \\
&= \left(\delta_y(z) + \sum_{i=1}^{\infty} E_y\left[\mathbb{1}_{V_i \le T_q}\right]\right) E_z\left[\int_0^{V_1} \mathbb{1}_{T_q \ge s}\mathbb{1}_{X(s)=z}\, ds\right] \\
&= \tilde{G}_q(y,z,\gamma) E_z\left[\int_0^{V_1} \mathbb{1}_{T_q \ge s}\mathbb{1}_{X(s)=z}\, ds\right].
\end{aligned}$$

Now, $E_z\left[\int_0^{V_1} \mathbb{1}_{T_q \ge s}\mathbb{1}_{X(s)=z}\, ds\right] = E_z\left[\int_0^{\sigma_1} \mathbb{1}_{T_q \ge s}\, ds\right] = E\left[\sigma_1 \wedge T_q\right] = \frac{1}{q+\alpha}$. □

Hence,

$$\mathbb{P}_{q,q'}\left[\rho_{\rho'_x} \notin A'(x)\right] \le \sum_{y\in\mathcal{X}} \sum_{\gamma: x\rightsquigarrow y} \sum_{z\in\gamma} \sum_{u\notin\gamma} \sum_{\phi\in\mathcal{F}_1(y,\gamma,u)} G_q(y,z) w(z,u) \frac{(q')^{|\rho(\phi)|} w(\phi)}{Z(q')}.$$

We fix y, γ and z and want to perform the summations over u and ϕ. With any pair (u,ϕ), with $u \notin \gamma$ and $\phi \in \mathcal{F}_1(y,\gamma,u)$, we associate a new forest $\tilde{\phi} = \tilde{\phi}(u,\phi)$ in the following way:

1. we reverse the edges from z to y along γ;
2. we add the edge (z,u).

The forest $\tilde{\phi}$ is such that:

- $\left|\rho(\tilde{\phi})\right| = |\rho(\phi)| - 1$;
- $z \notin \rho(\tilde{\phi})$.
- the piece $\gamma_{x\rightsquigarrow z}$ of the path γ going from x to z belongs to $\tilde{\phi}$;
- the path $\overleftarrow{\gamma}_{y\rightsquigarrow z}$ consisting of the reversed path γ from z to y, belongs to $\tilde{\phi}$.

Using reversibility, one has $\mu(z)\prod_{e\in\gamma_{z\rightsquigarrow y}} w(e) = \mu(y)\prod_{e\in\overleftarrow{\gamma}_{y\rightsquigarrow z}} w(e)$, and

$$w(z,u)w(\phi) = w(\tilde{\phi})\mu(y)/\mu(z).$$

Set $\mathscr{F}_2(y,z,\gamma) = \left\{\phi\, s.o.f.\,;\; z \notin \rho(\phi), \gamma_{x\rightsquigarrow z} \subset \phi,\, \overleftarrow{\gamma}_{y\rightsquigarrow z} \subset \phi\right\}$. Note that the function

$$(u,\phi) \in \{(u,\phi), u \notin \gamma, \phi \in \mathscr{F}_1(y,\gamma,u)\} \mapsto \tilde{\phi} \in \mathscr{F}_2(y,z,\gamma)$$

is one to one. Indeed, given $\tilde{\phi}$ in $\mathscr{F}_2(y,z,\gamma)$, u is the "ancestor" of z in $\tilde{\phi}$, and once we know u, ϕ is obtained by cutting the edge (z,u), and by reversing the path $\overleftarrow{\gamma}_{y\rightsquigarrow z}$. Therefore, we obtain

$$\begin{aligned}
&\sum_{u\notin\gamma}\sum_{\phi\in\mathscr{F}_1(y,\gamma,u)} G_q(y,z)w(z,u)\frac{(q')^{|\rho(\phi)|}w(\phi)}{Z(q')}\\
&\quad= \sum_{\phi\in\mathscr{F}_2(y,z,\gamma)} G_q(y,z)\frac{\mu(y)}{\mu(z)}\frac{(q')^{|\rho(\phi)|+1}w(\phi)}{Z(q')} = \sum_{\phi\in\mathscr{F}_2(y,z,\gamma)} G_q(z,y)\frac{(q')^{|\rho(\phi)|+1}w(\phi)}{Z(q')}
\end{aligned}$$

by reversibility. At this point, we are led to

$$\mathbb{P}_{q,q'}\left[\rho_{\rho'_x} \notin A'(x)\right] \le \sum_{y\in\mathscr{X}}\sum_{\gamma:x\rightsquigarrow y}\sum_{z\in\gamma}\sum_{\phi\in\mathscr{F}_2(y,z,\gamma)} G_q(z,y)\frac{(q')^{|\rho(\phi)|+1}w(\phi)}{Z(q')}\,.$$

We now perform the summations over z and γ and ϕ, y being fixed. Note that if $\phi \in \mathscr{F}_2(y,z,\gamma)$ for some z and γ, x and y are in the same tree ($\tau_x = \tau_y$ using the notations of Sect. 1.3.3), and z is their first common ancestor $a(x,y)$ in that tree. Let us then denote

$$\mathscr{F}_3(y,x) = \left\{\phi\, s.o.f.\,;\; \tau_x = \tau_y, a(x,y) \notin \rho(\phi)\right\}\,.$$

Then,

$$\cup_{\gamma:x\rightsquigarrow y}\cup_{z\in\gamma}\,\mathscr{F}_2(y,z,\gamma) \subset \mathscr{F}_3(y,x)\,.$$

In addition, given a forest $\phi \in \mathscr{F}_3(y,x)$, there is a unique $\gamma : x \rightsquigarrow y$, and $z \in \gamma$ such that $\phi \in \mathscr{F}_2(y,z,\gamma)$: z is the first common ancestor $a(x,y)$ of x and y, whereas γ is the concatenation of the path going from x to $a(x,y)$ and the reversed path from y to $a(x,y)$. Therefore,

$$\begin{aligned}
&\sum_{\gamma:x\rightsquigarrow y}\sum_{z\in\gamma}\sum_{\phi\in\mathscr{F}_2(y,z,\gamma)} G_q(z,y)\frac{(q')^{|\rho(\phi)|+1}w(\phi)}{Z(q')}\\
&\quad= \sum_{\phi\in\mathscr{F}_3(y,x)} G_q(a(x,y),y)\frac{(q')^{|\rho(\phi)|+1}w(\phi)}{Z(q')}\,.
\end{aligned}$$

It remains to sum over y. When moving y in τ_x, $a(x, y)$ moves along the path γ_x going from x to the root of τ_x. Hence,

$$\sum_{y\in\mathscr{X}} \sum_{\phi\in\mathscr{F}_3(y,x)} G_q(a(x, y), y)\frac{(q')^{|\rho(\phi)|+1}w(\phi)}{Z(q')}$$
$$= \sum_{\phi\, s.o.f.} \sum_{z\in\gamma_x, z\neq\rho_x} \sum_{y\in\tau_x; a(x,y)=z} G_q(z, y)\frac{(q')^{|\rho(\phi)|+1}w(\phi)}{Z(q')}$$
$$\leq \frac{q'}{q} \sum_{\phi\, s.o.f.} \sum_{z\in\gamma_x, z\neq\rho_x} \pi_{q'}(\phi) \leq \frac{q'}{q}\mathbb{E}_{q'}\left[|\Gamma'_x|\right]. \qquad \square$$

5 Proof of Theorem 3

Let us first rewrite our approximate solutions of Eq. (3) with error terms. There are signed measures $\epsilon_{\bar{x},q'}$ such that, for all $\bar{x}$ in $\bar{\mathscr{X}}$,

$$\mu_{A(\bar{x})}K_{q'} = \sum_{\bar{y}\in\bar{\mathscr{X}}} P_{\mu_{A(\bar{x})}}\big(X(T_{q'}) \in A(\bar{y})\big)\, \mu_{A(\bar{y})} + \epsilon_{\bar{x},q'}.$$

Let us now apply the "low-pass filter" MW_m on both sides of the equations. On the one hand, $K_{q'}$ and MW_m commute. On the other hand, our linear independence (i.e. finite squeezing) hypothesis implies that the $\epsilon_{\bar{x},q'}MW_m$ are linear combinations of the $\mu_{A(\bar{x})}MW_m$. Indeed, since the image $\mathrm{im}(MW_m)$ of MW_m is a vector space of dimension m that contains the m linearly independent $\nu_{\bar{x}}$, the latter should span $\mathrm{im}(MW_m)$. We then get, by using the notation of the proof of Proposition 6,

$$\nu_{\bar{x}}K_{q'} = \sum_{\bar{y}\in\bar{\mathscr{X}}} \left(P_{\mu_{A(\bar{x})}}\big(X(T_{q'}) \in A(\bar{y})\big) + \langle\tilde{\nu}_{\bar{y}}, \epsilon_{\bar{x},q'}MW_m\rangle^*\right) \nu_{\bar{y}}.$$

Now, when q' goes to 0, $P_{\mu_{A(\bar{x})}}\big(X(T_{q'}) \in A(\bar{y})\big)$ converges to $\mu(A(\bar{y})) > 0$, and, by Theorem 2, $\epsilon_{\bar{x},q'}$ goes to zero. Since our $\nu_{\bar{x}}$ do not depend on q', this concludes the proof of the theorem.

Let us list what would be needed to give quantitative bounds on q' to ensure that we can build in this way exact solutions of (3). We would need:

1. upper bounds on the $\epsilon_{\bar{x},q'}$;
2. upper bounds on the $\|\tilde{\nu}_{\bar{x}}\|$;
3. lower bounds on the $P_{\mu_{A(\bar{x})}}\big(X(T_{q'}) \in A(\bar{y})\big)$.

The latter are out of reach in such a general framework, the first ones are provided by Theorem 2, the second ones would be a consequence of upper bounds on the squeezing. This is the reason why we introduce the squeezing to measure joint overlap. We note that given Proposition 5 and Eq. (27) in Theorem 1, we are not so far of getting such bounds. But no convexity inequality leads here to the conclusion.

Appendix: Proof of Proposition 1

If such random variables exist then, for all $\bar{x}$, $\bar{y} \neq \bar{x}$ and y,

$$P_{\nu_{\bar{x}}}\left(T_{\bar{x}} = 1, \bar{Y}_{\bar{x}} = \bar{y} \;\middle|\; \hat{X}(1) = y\right) = \frac{P_{\nu_{\bar{x}}}\left(T_{\bar{x}} = 1, \bar{Y}_{\bar{x}} = \bar{y}, \hat{X}(1) = y\right)}{(\nu_{\bar{x}} P)(y)}$$

$$= \frac{(1 - \bar{P}(\bar{x}, \bar{x}))\frac{\bar{P}(\bar{x},\bar{y})}{1-\bar{P}(\bar{x},\bar{x})}\nu_{\bar{y}}(y)}{(\nu_{\bar{x}} P)(y)}$$

$$= \frac{\bar{P}(\bar{x}, \bar{y})\nu_{\bar{y}}(y)}{(\nu_{\bar{x}} P)(y)}\,.$$

By summing on $\bar{y}$ we get

$$P_{\nu_{\bar{x}}}\left(T_{\bar{x}} = 1 \;\middle|\; \hat{X}(1) = y\right) = \frac{(\nu_{\bar{x}} P)(y) - \bar{P}(\bar{x}, \bar{x})\nu_{\bar{x}}(y)}{(\nu_{\bar{x}} P)(y)} = 1 - \frac{\bar{P}(\bar{x}, \bar{x})\nu_{\bar{x}}(y)}{(\nu_{\bar{x}} P)(y)}\,.$$

We also have

$$P_{\nu_{\bar{x}}}\left(\bar{Y}_{\bar{x}} = \bar{y} \;\middle|\; \hat{X}(1) = y, T_{\bar{x}} = 1\right) = \frac{P_{\nu_{\bar{x}}}\left(\bar{Y}_{\bar{x}} = \bar{y}, T_{\bar{x}} = 1 \;\middle|\; \hat{X}(1) = y\right)}{P_{\nu_{\bar{x}}}\left(T_{\bar{x}} = 1 \;\middle|\; \hat{X}(1) = y\right)}$$

$$= \frac{\bar{P}(\bar{x}, \bar{y})\nu_{\bar{y}}(y)}{(\nu_{\bar{x}} P)(y) - \bar{P}(\bar{x}, \bar{x})\nu_{\bar{x}}(y)}\,.$$

We are then led to build $T_{\bar{x}} \geq 1$ and $\bar{Y}_{\bar{x}}$ in the following way.

1. At $t = 1$ we set $T_{\bar{x}} = 1$ with probability $1 - \bar{P}(\bar{x}, \bar{x})\nu_{\bar{x}}(\hat{X}(1))/(\nu_{\bar{x}} P)(\hat{X}(1))$ by using a uniform random variable U_1 which is independent of $\hat{X}$—it holds

$$\bar{P}(\bar{x}, \bar{y})\nu_{\bar{x}}(y)/(\nu_{\bar{x}} P)(y) \leq 1$$

   for all y in $\mathscr{X}$, as a consequence of Eq. (6).
2. If we just set $T_{\bar{x}} = 1$ we then set $\bar{Y}_{\bar{x}} = \bar{y} \neq \bar{x}$ with a probability given by the ratio $\bar{P}(\bar{x}, \bar{y})\nu_{\bar{y}}(\hat{X}(1))/[(\nu_{\bar{x}} P)(\hat{X}(1)) - \bar{P}(\bar{x}, \bar{x})\nu_{\bar{x}}(\hat{X}(1))]$ by using a uniform random variable U_1' that is independent of U_1 and $\hat{X}$. (Once again (6) ensures that these are positive quantities summing to one.)

3. If for all $s < t$ we did not decide to set $T_{\bar{x}} = s$ then we set in the same way $T_{\bar{x}} = t$ with probability $1 - \bar{P}(\bar{x}, \bar{x})\nu_{\bar{x}}(\hat{X}(t))/(\nu_{\bar{x}} P)(\hat{X}(t))$, in which case we set $\bar{Y}_{\bar{x}} = \bar{y} \neq \bar{x}$ with probability $\bar{P}(\bar{x}, \bar{y})\nu_{\bar{y}}(\hat{X}(t))/[(\nu_{\bar{x}} P)(\hat{X}(t)) - \bar{P}(\bar{x}, \bar{x})\nu_{\bar{x}}(\hat{X}(t))]$. This is naturally done by using uniform random variable that are independent of $\hat{X}$ and $U_1, U_1', U_2, U_2', \ldots, U_{t-1}, U_{t-1}'$.

At this point, the key property to check is the stationarity of $\nu_{\bar{x}}$ up to $T_{\bar{x}}$. To this end it suffices to check Eq. (7) with $t = 1$. And one has

$$
\begin{aligned}
P_{\nu_{\bar{x}}}\left(\hat{X}(1) = y \,\middle|\, T_{\bar{x}} > 1\right) &= \frac{P_{\nu_{\bar{x}}}\left(\hat{X}(1) = y, T_{\bar{x}} > 1\right)}{P_{\nu_{\bar{x}}}\left(T_{\bar{x}} > 1\right)} \\
&= \frac{P_{\nu_{\bar{x}}}\left(\hat{X}(1) = y\right) - P_{\nu_{\bar{x}}}\left(\hat{X}(1) = y, T_{\bar{x}} = 1\right)}{1 - P_{\nu_{\bar{x}}}\left(T_{\bar{x}} = 1\right)} \\
&= \frac{\nu_{\bar{x}} P(y) - \nu_{\bar{x}} P(y)\left(1 - \frac{\bar{P}(\bar{x},\bar{x})\nu_{\bar{x}}(y)}{\nu_{\bar{x}} P(y)}\right)}{1 - \sum_z \nu_{\bar{x}} P(z)\left(1 - \frac{\bar{P}(\bar{x},\bar{x})\nu_{\bar{x}}(z)}{\nu_{\bar{x}} P(z)}\right)} \\
&= \frac{\bar{P}(\bar{x}, \bar{x})\nu_{\bar{x}}(y)}{1 - \sum_z \nu_{\bar{x}} P(z) + \sum_z \bar{P}(\bar{x}, \bar{x})\nu_{\bar{x}}(z)} = \nu_{\bar{x}}(y).
\end{aligned}
$$

Points (1)–(5) immediately follow.

References

1. Andersen, R., Gharan, S.O., Peres, Y., Trevisan, L.: Almost optimal local graph clustering using evolving sets. J. ACM **63**, 1–31 (2016)
2. Avena, L., Gaudillière, A.: Two applications of random spanning forests. J. Theor. Probab. **31**, 1975–2004 (2018)
3. Avena, L., Castell, F., Gaudillière, A., Mélot, C.: Intertwining wavelets or multiresolution analysis through random forests on graphs. Appl. Comput. Harmon. Anal. **48**, 949–992 (2020)
4. Bovier, A., den Hollander, F.: Metastability: A Potential-Theoretic Approach. Springer, Cham (2015)
5. Carmona, P., Petit, F., Yor, M.: Beta-gamma random variables and intertwining relations between certain Markov processes. Rev. Mat. Iberoamericana **14**, 311–367 (1998)
6. Diaconis, P., Fill, J.A.: Strong stationary times via a new form of duality. Ann. Probab. **18**, 1483–1522 (1990)
7. Donati-Martin, C., Doumerc, Y., Matsumoto, H., Yor, M.: Some properties of the Wishart processes and a matrix extension of the Hartman-Watson laws. Publ. Res. Inst. Math. Sci. **40**, 1385–1412 (2004)
8. Freidlin, M.I., Wentzell, A.D.: Random Pertubations of Dynamical Systems. Springer, Berlin (1984)
9. Hammond, D.K., Vandergheynst, P., Gribonval, R.: Wavelets on graphs via spectral graph theory. Appl. Comput. Harmon. Anal. **30**, 129–150 (2009)

10. Kirchhoff, G.: Über die Auflösung der Gleichungen, auf welche man bei der Untersuchung der lineraren Verteilung galvanisher Ströme geführt wird. Ann. Phys. Chem. **72**, 497–508 (1847)
11. Matsumoto, H., Yor, M.: An analogue of Pitman's $2M - X$ theorem for exponential Wiener functionals, I. A time-inversion approach. Nagoya Math. J. **159**, 125–166 (2000)
12. Micchelli, C.A., Willoughby, R.A.: On functions which preserve the class of Stieltjes matrices. Lin. Alg. Appl. **23**, 141–156 (1979)
13. Miclo, L.: On absorption times and Dirichlet eigenvalues. ESAIM Prob. Stat. **14**, 117–150 (2010)
14. Miclo, L.: On the construction of set-valued dual processes. Electron. J. Probab. **25**, 1–64 (2020)
15. Miclo, L., Patie, P.: On intertwining relations between Ehrenfest, Yule and Ornstein-Uhlenbeck processes. ArXiv:1904.10693 (2019)
16. Morris, B., Peres, Y.: Evolving sets, mixing and heat kernel bounds. Probab. Theory Relat. Fields **133**, 245–266 (2005)
17. Olivieri, E., Vares, M.E.: Large Deviations and Metastability. Cambridge University Press, Cambridge (2005)
18. Propp, J., Wilson, D.: How to get a perfectly random sample from a generic Markov chain and generate a random spanning tree of a directed graph. J. Algorithms **27**, 179–217 (1998)
19. Ricaud, B., Borgnat, P., Tremblay, N., Goncalves, P., Vandergheynst, P.: Fourier could be a data scientist: from graph Fourier transform to signal processing on graphs. C. R. Phys. **20**, 474–488 (2019)
20. Rogers, L.C.G., Pitman, J.W.: Markov functions. Ann. Probab. **9**, 573–582 (1981)
21. Scoppola, E.: Renormalization group for Markov chains and application to metastability. J. Stat. Phys. **73**, 83–121 (1993)
22. Shuman, D.I., Narang, S.K., Frossard, P., Ortega, A., Vandergheynst, P.: The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains. IEEE Signal Process. Mag. **30**, 83–98 (2013)
23. Vetterli, M., Kovacevic, J., Goyal, V.K.: Foundations of Signal Processing. Cambridge University Press, Cambridge (2014)
24. Warren, J.: Dyson's Brownian motions, intertwining and interlacing. Electron. J. Probab. **12**, 573–590 (2007)
25. Wilson, D.: Generating random spanning trees more quickly than the cover time. In: Proceedings of the Twenty-Eighth Annual ACM Symposium on the Theory of Computing, pp. 296–303 (1996)

Bernoulli Hyperplane Percolation

Marco Aymone, Marcelo R. Hilário, Bernardo N. B. de Lima, and Vladas Sidoravicius

Abstract We study a dependent site percolation model on the n-dimensional Euclidean lattice where, instead of single sites, entire hyperplanes are removed independently at random. We extend the results about Bernoulli line percolation showing that the model undergoes a non-trivial phase transition and proving the existence of a transition from exponential to power-law decay within some regions of the subcritical phase.

Keywords Dependent percolation · Phase transition · Connectivity decay

1 Introduction

In Bernoulli site percolation on the $\mathbb{Z}^n$-lattice, vertices are removed independently with probability $1 - p$. For $n \geq 2$, the model undergoes a phase transition at $p_c = p_c(\mathbb{Z}^n) \in (0, 1)$: For $p < p_c$ all the connected components are finite almost surely whereas, for $p > p_c$, there exists an infinite connected component almost surely [2]. In a different percolation model on $\mathbb{Z}^n$, $n \geq 3$, called Bernoulli line percolation, instead of single sites, bi-infinite lines (or columns) of sites that are parallel to the coordinate axes are removed independently. This model, that was introduced in the physics literature by Kantor [8] and later studied both from the numerical [5, 12] and mathematical [7] points of view, also exhibits a phase transition as the probability of removal of single lines is varied. However the geometric properties of the resulting connected components differ substantially in these two models. In fact, while for Bernoulli site percolation the connectivity decay is exponential [1, 3, 10] except exactly at the critical point, for Bernoulli line percolation, transitions from

M. Aymone · M. R. Hilário (✉) · B. N. B. de Lima
Universidade Federal de Minas Gerais, Departamento de Matemática, Belo Horizonte, MG, Brazil
e-mail: aymone@ufmg.br; mhilario@mat.ufmg.br; bnblima@mat.ufmg.br

V. Sidoravicius
Shanghai New York University, Pudong New District, Shanghai, China

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_4

exponential to power-law decay occur within the subcritical phase [7, Theorem 1.2]. In the present paper we study a higher dimensional version of the Bernoulli line percolation model that we call *Bernoulli hyperplane percolation*. In this model, for fixed $n \geq 3$ and k with $1 \leq k \leq n$ we remove from $\mathbb{Z}^n$ entire $(n-k)$-dimensional 'affine hyperplanes'. We introduce the model precisely in the following section.

1.1 Definition of the Model and Main Results

In this section we define the Bernoulli hyperplane percolation model.

It can be formulated in terms of orthogonal projections onto "the coordinate hyperplanes" as follows: For $n \geq 2$ and $1 \leq k \leq n$ we write

$$\mathscr{I} = \mathscr{I}(k; n) := \{I \subset [n] : \#I = k\}, \tag{1}$$

where $[n] := \{1, \ldots, n\}$. For a fixed $I \in \mathscr{I}(k, n)$ we denote $\mathbb{Z}_I^k$ the set of all the linear combinations of the canonical vectors $(e_i)_{i \in I}$ with integer coefficients, that is,

$$\mathbb{Z}_I^k := \Big\{ \sum_{i \in I} x_i e_i \in \mathbb{Z}^n : x_i \in \mathbb{Z} \text{ for all } i \in I \Big\}. \tag{2}$$

Since each one of the $\binom{n}{k}$ sets $\mathbb{Z}_I^k$ is isomorphic to the $\mathbb{Z}^k$-lattice they will be called the k-dimensional coordinate hyperplanes of $\mathbb{Z}^n$. Let us define, independently on each $\mathbb{Z}_I^k$, a Bernoulli site percolation $\omega_I \in \{0, 1\}^{\mathbb{Z}_I^k}$ with parameter $p_I \in [0, 1]$ that is, a process in which $(\omega_I(u))_{u \in \mathbb{Z}_I^k}$ are independent Bernoulli random variables with mean $p_I \in [0, 1]$. We may interpret these processes by considering that each $u \in \mathbb{Z}_I^k$ is removed (that is $\omega_I(u) = 0$) independently with probability $1 - p_I$.

Let $\pi_I : \mathbb{Z}^n \to \mathbb{Z}_I^k$ stand for the orthogonal projection from $\mathbb{Z}^n$ onto $\mathbb{Z}_I^k$

$$\pi_I \Big(\sum_{i=1}^{n} x_i e_i \Big) := \sum_{i \in I} x_i e_i. \tag{3}$$

The Bernoulli (n, k)-hyperplane percolation on $\mathbb{Z}^n$ is the process $\omega = (\omega(v))_{v \in \mathbb{Z}^n} \in \{0, 1\}^{\mathbb{Z}^n}$, where

$$\omega(v) = \prod_{I \in \mathscr{I}} \omega_I(\pi_I(v)). \tag{4}$$

We denote $\mathbf{p} = (p_I)_{I \in \mathscr{I}}$. Each entry $p_I \in [0, 1]$ is called a parameter of $\mathbf{p}$. We write $\mathbb{P}_{\mathbf{p}}$ for the law in $\{0, 1\}^{\mathbb{Z}^n}$ of the random element ω defined in (4).

Since $\omega(v) = 1$ if and only if $\omega_I(\pi_I(v)) = 1$ for all $I \in \mathscr{I}(k, n)$, we may interpret the process ω in terms of removal of sites in $\mathbb{Z}^n$ as follows: v is removed (that is $\omega(v) = 0$) if and only if, for at least one of the $I \in \mathscr{I}(k; n)$ its orthogonal projection into $\mathbb{Z}_I^k$ has been removed in ω_I. One can easily check that this is equivalent to perform independent removal (or drilling) of $(n-k)$-dimensional hyperplanes that are parallel to the coordinate hyperplanes.

Let $[o \leftrightarrow \infty]$ denote the event that the origin belongs to an infinite connected component of sites v such that $\omega(v) = 1$ and let $[o \nleftrightarrow \infty]$ be the complementary event. Also denote by $[o \leftrightarrow \partial B(K)]$ the event that the origin is connected to some vertex lying at l_∞-distance K from it via a path of sites v such that $\omega(v) = 1$.

Our first result generalizes Theorem 1.1 in [7].

Theorem 1 *Let $n \geq 3$ and $2 \leq k \leq n-1$. The Bernoulli (n,k)-hyperplane percolation model undergoes a non-trivial phase transition, that is: If all the parameters of* $\mathbf{p}$ *are sufficiently close to* 1 *then* $\mathbb{P}_{\mathbf{p}}(o \leftrightarrow \infty) > 0$. *On the other hand, when all the parameters of* $\mathbf{p}$ *are sufficiently close to* 0 *then* $\mathbb{P}_{\mathbf{p}}(o \leftrightarrow \infty) = 0$.

The proof of Theorem 1 will be divided into two parts. The second assertion which concerns the regime in which all the parameters are small is proved in Sect. 2 (see Remark 1 therein). In Sect. 3 we prove the first assertion which concerns the regime in which all the parameters are large (see Remark 3 therein).

Our next result states that for some range of the parameter vector $\mathbf{p}$ the connectivity cannot decay faster than a power law.

Theorem 2 *Let $n \geq 3$ and $2 \leq k \leq n-1$. If, for all $I \in \mathscr{I}(k;n)$, the parameters $p_I < 1$ are sufficiently close to 1, then there exists $c = c(\mathbf{p}) > 0$ and $\alpha = \alpha(\mathbf{p}) > 0$ such that*

$$\mathbb{P}_{\mathbf{p}}(o \leftrightarrow \partial B(K),\ o \nleftrightarrow \infty) \geq cK^{-\alpha} \tag{5}$$

for every integer $K > 0$.

In the special case $k = 2$ we can determine more precisely some regions of the parameter space for which power-law decay holds:

Theorem 3 *Let $n \geq 3$ and $k = 2$ and assume that $p_I > 0$ for every $I \in \mathscr{I}(k;n)$. Denote $I_j := \{1, j\}$ and assume that $p_{I_j} > p_c(\mathbb{Z}^2)$ for every $2 \leq j \leq n$. Assume further that $p_I < 1$ for some $I \in \mathscr{I} \setminus \{I_2, \dots, I_n\}$. Then there exist $c = c(\mathbf{p}) > 0$ and $\alpha = \alpha(\mathbf{p}) > 0$ such that (5) holds for every integer $K > 0$.*

Having stated our main results, we now provide some remarks about the contribution of this paper.

The case $k = 1$ does not admit a (non-trivial) phase transition. In fact, as soon as $p_I > 0$ for every I, drilling $(n-1)$-dimensional hyperplanes has the effect of splitting the lattice into finite rectangles. The case $k = n$ corresponds to Bernoulli site percolation (here we interpret 0-dimensional hyperplanes as being just single sites). For these reasons in the above statements we have $2 \leq k \leq n-1$. Moreover, the case $k = n-1$ corresponds to the Bernoulli line percolation model studied in [7] so the results are not novel in this specific case.

Theorem 3, that only concerns the case $k = 2$, states that when the Bernoulli site percolation processes ω_I defined on the $n-1$ coordinate planes that contain e_1 are all supercritical, then power-law decay holds regardless of the values of the parameters fixed for the Bernoulli percolation processes in the remaining $\binom{n-1}{2}$ hyperplanes (provided that at least one of these is smaller than 1). It generalizes the first statement in [7, Theorem 1.2]. Of course, the arbitrary choice of fixing the 'direction' 1 is made purely for convenience; any other choice would result in an analogous result.

Still for $k = 2$, the same argument used to prove Equation (1.3) in [7] can be employed to show that if $p_I < p_c(\mathbb{Z}^2)$ for at least $\binom{n-1}{2} + 1$ parameters, then $\mathbb{P}_{\mathbf{p}}(o \leftrightarrow \partial B(K))$ is exponentially small in K, see Remark 2 for a sketch of the argument. Hence, like in Bernoulli line percolation, there is a transition from exponential to power-law decay within the subcritical phase. This contrasts with the classical Bernoulli site percolation in which exponential decay holds everywhere outside the critical point [1, 3, 10] and raises the question whether the phase transition for Bernoulli hyperplane percolation is sharp in the sense that the expected size of the cluster containing a vertex is finite in the whole subcritical regime. For Bernoulli site percolation sharpness is an immediate consequence of exponential decay.

Let us now briefly comment on some related results obtained for percolation models presenting infinite-range correlations along columns. One of these models is the so called Winkler's percolation [13] for which a power-law decay as in (5) has been proved by Gács [4] whenever the model is supercritical. For another model called corner percolation, although all the connected components are finite almost surely, Pete [11] has obtained a power-law lower bound for $\mathbb{P}_{\mathbf{p}}(o \leftrightarrow \partial B(K))$. A variation of Bernoulli line percolation was studied in [6]. In this paper, only columns that extend along a single direction are removed and Bernoulli line percolation is performed on the remaining graph. Here (5) holds in some parts of the subcritical phase and throughout the whole supercritical phase.

We finish this section presenting a brief overview of the remainder of the paper. In Sect. 1.2 we introduce some of the notation that we will need. Section 2 is devoted to the study of the subcritical phase, that is, the regime in which infinite clusters occur with null probability. Lemma 2 identifies values of the parameters which fall inside the subcritical phase and thus implies the second assertion in Theorem 1. We also present other results that add more information about the subcritical phase including bounds on the parameters that guarantee exponential decay of correlations (Remark 2). In Sect. 3 we prove the existence of the supercritical phase, that is, the regime in which there exists at least one infinite open cluster with probability one. This corresponds to the first assertion in Theorem 1. In Sect. 4 we present the proof of Theorem 3 and show how to modify it in order to obtain a proof for Theorem 2. These are perhaps the most interesting results in our work since they highlight the presence of power-law decay of connectivity in some regimes and show that the transition from the subcritical to the supercritical phase is more delicate than that exhibited by ordinary percolation models with finite range dependencies.

1.2 Notation

In this section we make precise the notation and definitions used in the previous section and introduce some further notation that will be used in the remainder of the paper.

The n-dimensional Euclidean lattice (here called simply the $\mathbb{Z}^n$-lattice) is the pair $\mathbb{Z}^n = (V(\mathbb{Z}^n), E(\mathbb{Z}^n))$ whose vertex set $V(\mathbb{Z}^n)$ is composed of vectors $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ having integer coordinates x_i and $E(\mathbb{Z}^n)$ is the set of pairs of vertices in $V(\mathbb{Z}^n)$ lying at Euclidean distance one from each other, called edges (or bonds). Vertices $x \in V(\mathbb{Z}^n)$ will also be called sites. We abuse notation using $\mathbb{Z}^n$ to refer both to the $\mathbb{Z}^n$-lattice and to its set of vertices. We denote $\|x\| = \sum_{i=1}^n |x_i|$ the l_1-norm of $x = (x_1, \dots, x_n) \in \mathbb{Z}^n$.

The vertex $o = (0, \dots, 0) \in \mathbb{Z}^n$ will be called the origin. Note that it also belongs to each one of the $\mathbb{Z}_I^k$. We write $B(K) := [-K, K]^n \cap \mathbb{Z}^n$ for the l_∞-ball of radius K centered at o. For a given $I \in \mathscr{I}(k; n)$, we will also use $B(K)$ instead of $\pi_I(B(K))$ for the corresponding box contained in the hyperplanes $\mathbb{Z}_I^k$ (recall the definitions of the index set $\mathscr{I}(k; n)$ in (1) and of the k-dimensional coordinate hyperplanes $\mathbb{Z}_I^k$ in (2)).

Consider $\Omega = \{0, 1\}^{\mathbb{Z}^n}$ endowed with the canonical sigma-field $\mathscr{F}$ generated by the cylinder sets. A probability measure μ on $(\Omega, \mathscr{F})$ is called a site percolation on $\mathbb{Z}^n$. Any random element $(\omega(v))_{v\in\mathbb{Z}^n}$ which is distributed as μ is also called a percolation process in $\mathbb{Z}^n$. For $p \in [0, 1]$, we denote $\mathbb{P}_p$ the probability measure in $(\Omega, \mathscr{F})$ under which the projections $(\omega(v))_{v\in\mathbb{Z}^n}$ are i.i.d. Bernoulli random variables of mean p. This is the so-called Bernoulli site percolation on $\mathbb{Z}^n$ with parameter p.

Fix integers $n \geq 3$ and $2 \leq k \leq n - 1$ and let $p_I \in [0, 1]$ for each $I \in \mathscr{I}(k; n)$. Consider $\Omega_I = \{0, 1\}^{\mathbb{Z}_I^k}$ endowed with the canonical sigma-field $\mathscr{F}_I$. The definition of Bernoulli site percolation with parameter p_I extends to $\mathbb{Z}_I^k$ naturally yielding the measures $\mathbb{P}_{p_I}$ on $(\Omega_I, \mathscr{F}_I)$. The probability measure $\mathbb{P}_{\mathbf{p}}$ which was defined below (4), is the unique measure in $(\Omega, \mathscr{F})$ satisfying

$$\mathbb{P}_{\mathbf{p}} = \left(\otimes_{I\in\mathscr{I}_{k,n}} \mathbb{P}_{p_I} \right) \circ \omega^{-1} \tag{6}$$

where ω is defined in (4). We will denote $\mathbb{E}_{\mathbf{p}}(\cdot)$ the expectation with respect to $\mathbb{P}_{\mathbf{P}}$.

Let G be either $\mathbb{Z}^n$ or $\mathbb{Z}_I^k$ for some $I \in \mathscr{I}$. Given $\eta = (\eta(x))_G \in \{0, 1\}^G$, we say that a site $x \in G$ is η-open when $\eta(x) = 1$. Otherwise x is said η-closed. A site $x \in \mathbb{Z}^n$ is said ω_I-open if $\pi_I(x) \in \mathbb{Z}_I^k$ is ω_I-open, i.e., if $\omega_I(\pi_I(x)) = 1$. Otherwise, x is said ω_I-closed. Since x is ω-open if and only if it is ω_I-open for all $I \in \mathscr{I}(k, n)$ and since the percolation processes ω_I are independent we have

$$\mathbb{E}_{\mathbf{p}}(\omega(o)) = \mathbb{P}_{\mathbf{p}}(\omega(x) = 1) = \prod_{I\in\mathscr{I}} p_I. \tag{7}$$

Therefore one can show that

$$\mathbb{P}_{\mathbf{p}}\big(\text{all sites in } B(R) \text{ are open}\big) = \big[\mathbb{E}_{\mathbf{p}}(\omega(o))\big]^{(2R+1)^k} \tag{8}$$

which is similar to the Bernoulli site percolation case where the exponent $(2R+1)^k$ has to be replaced by $(2R+1)^n$.

Let G be either $\mathbb{Z}^n$ or $\mathbb{Z}^k_I$ for some $I \in \mathscr{I}$. A path in G is either a finite set $\Gamma = \{x_0, x_1, \dots, x_m\}$ or an infinite set $\Gamma = \{x_0, x_1, x_2 \dots\}$ such that $x_i \in G$ for all i, $x_i \neq x_j$ whenever $i \neq j$ and $||x_i - x_{i-1}|| = 1$ for all $i = 1, \dots, m$. For $\eta = (\eta(x))_{x\in G} \in \{0,1\}^G$ we denote $\mathscr{V}_\eta = \mathscr{V}_\eta(G) = \{x \in G : \eta(x) = 1\}$. For such η we say that $x, y \in \mathscr{V}_\eta$ are connected and write $x \leftrightarrow y$ if there exists a path composed exclusively of η-open sites that starts at x and finishes at y. Otherwise, we write $x \not\leftrightarrow y$. For $A \subset \mathbb{Z}^n$, we write $x \leftrightarrow A$ if $x \leftrightarrow y$ for some $y \in A$. We say that $\mathscr{G} \subset \mathscr{V}_\eta$ is a connected component (or a cluster) of $\mathscr{V}_\eta$ when every pair of sites $x, y \in \mathscr{G}$ is such that $x \leftrightarrow y$. In addition, for $x \in G$ we denote by $\mathscr{V}_\eta(x) = \mathscr{V}_\eta(x; G)$ the maximal connected component in $\mathscr{V}_\eta$ containing x, that is, $\mathscr{V}_\eta(x; G) = \{y \in G \colon x \leftrightarrow y \text{ in } \eta\}$. We say that a site x belongs to an infinite connected component in $\mathscr{V}_\eta$ and denote it $x \leftrightarrow \infty$ if $\#\mathscr{V}_\eta(x) = \infty$.

We say that the Bernoulli (n,k)-hyperplane percolation exhibits a non-trivial phase transition if there exists $\mathbf{p} = (p_I)_{I\in\mathscr{I}}$ with $p_I > 0$ for all $I \in \mathscr{I}$ and $\mathbf{q} = (q_I)_{I\in\mathscr{I}}$ with $q_I < 1$ for all $I \in \mathscr{I}$, such that $\mathbb{P}_{\mathbf{p}}(o \leftrightarrow \infty) = 0$ and $\mathbb{P}_{\mathbf{q}}(o \leftrightarrow \infty) > 0$. The set of all $\mathbf{p}$ for which $\mathbb{P}_{\mathbf{q}}(o \leftrightarrow \infty) > 0$ is called the supercritical phase whereas the set of all $\mathbf{p}$ for which $\mathbb{P}_{\mathbf{p}}(o \leftrightarrow \infty) = 0$ is the subcritical phase.

For $x \in \mathbb{Z}^k_I$ we denote by

$$\mathscr{P}_I(x) := \pi_I^{-1}(x) = \{z \in \mathbb{Z}^n \colon \pi_I(z) = x\} \tag{9}$$

the pre-image of x under π_I, so that $\{\mathscr{P}_I(x) \colon x \in \mathbb{Z}^k_I\}$ foliates $\mathbb{Z}^n$ into disjoint 'parallel $(n-k)$-dimensional afine hyperplanes'. Observe that

$$\inf\big\{\|v - w\| : v \in \mathscr{P}_I(x), w \in \mathscr{P}_I(y)\big\} = 1 \text{if and only if } \|x - y\| = 1.$$

Let $I \in \mathscr{I}(k;n)$. The graph $\mathscr{H} = \mathscr{H}(I)$ with vertices $V(\mathscr{H}) := \{\mathscr{P}_I(x) : x \in \mathbb{Z}^k_I\}$, and with edges linking pairs of vertices $\mathscr{P}_I(x)$ and $\mathscr{P}_I(y)$ satisfying

$$\inf\{||v - w|| : v \in \mathscr{P}_I(x), w \in \mathscr{P}_I(y)\} = 1$$

is called a $\mathbb{Z}^k$-decomposition of $\mathbb{Z}^n$. Notice that $\mathscr{H}$ is isomorphic to $\mathbb{Z}^k$.

For a fixed $x \in \mathbb{Z}^k_I$, the projection $\pi_{[n]\setminus I} : \mathbb{Z}^n \to \mathbb{Z}^{n-k}_{[n]\setminus I}$ maps $\mathscr{P}_I(x)$ isomorphically to $\mathbb{Z}^{n-k}_{[n]\setminus I}$ which is, in turn, isomorphic to $\mathbb{Z}^{n-k}$. Thus for each $v \in \mathbb{Z}^{n-k}_{[n]\setminus I}$, there exists a unique $u \in \mathscr{P}_I(x)$ such that $\pi_{[n]\setminus I}(u) = v$. We say that v is $\mathscr{P}_I(x)$-closed if there exists $J \in \mathscr{I}(n;k) \setminus \{I\}$ for which u is ω_J-closed.

Otherwise we say that v is $\mathscr{P}_I(x)$-open. Observe that if x and y are different vertices in $\mathbb{Z}_I^k$ and $v \in \mathbb{Z}_{[n]\setminus I}^{n-k}$ we might have that v is $\mathscr{P}_I(x)$-open and $\mathscr{P}_I(y)$-closed.

We say that $T \subset \mathbb{Z}^n$ surrounds the origin if there exists a partition of $\mathbb{Z}^n \setminus T = A \cup B$ such that:

$$A \text{ is connected, } o \in A \text{ and } \#A < \infty; \tag{10}$$

$$\inf\{\|a - b\| : a \in A, b \in B\} \geq 2. \tag{11}$$

Similar definitions can be made replacing $\mathbb{Z}^n$ by any of the $\mathbb{Z}_I^k$. A useful fact that we will use below it that $\mathscr{V}_\omega(o)$ is finite if and only if there exists $T \subset \mathbb{Z}^n$ that surrounds the origin and whose sites are all ω-closed (and similarly for $\mathscr{V}_{\omega_I}$).

2 The Existence of a Subcritical Phase

This section is dedicated to the existence of a subcritical phase. Indeed we show that Bernoulli hyperplane percolation does not present infinite connected components a.s. when some of the parameters of **p** are sufficiently small. This corresponds to the second assertion in Theorem 1 which is a consequence of Lemma 2 below. Roughly speaking, Lemma 2 asserts that the probability that a given site belongs to an infinite open cluster vanishes as soon as a single parameter p_I is taken subcritical and at least other (well-chosen) $n - k$ parameters do not equal 1. In order to get the same conclusion, Lemma 3 requires that n/k parameters are subcritical regardless of the fact that the other parameters can even be equal to 1. Remark 2 contains the sketch of an argument showing that if we get sufficiently many subcritical parameters then actually exponential decay holds (hence infinite connected components cannot exist *a.s.*). We begin with the following deterministic result which will also be useful in Sect. 4 when we present a proof of Theorem 3.

Lemma 1 *Let ω be as in (4) and fix $I \in \mathscr{I}(k; n)$. Assume that the two following conditions hold:*

(i) The cluster $\mathscr{V}_{\omega_I}(o; \mathbb{Z}_I^k)$ is finite;

(ii) There exists $T \subset \mathbb{Z}_{[n]\setminus I}^{n-k}$ that surrounds the origin in $\mathbb{Z}_{[n]\setminus I}^{n-k}$ and such that every $v \in T$ is $\mathscr{P}_I(x)$-closed for every $x \in \mathscr{V}_{\omega_I}(o; \mathbb{Z}_I^k)$.

Then $\mathscr{V}_\omega(o; \mathbb{Z}^n)$ is finite.

Proof Assume that there exists an infinite path $\{o = z_1, z_2, \dots\} \subset \mathbb{Z}^n$ starting at the origin and composed of ω−open sites only. Let T be as in Condition (ii) and $A \subset \mathbb{Z}_{[n]\setminus I}^{n-k}$ be the corresponding set given as (10) and (11). We claim that z_i satisfies

$$\pi_I(z_i) \in \mathscr{V}_{\omega_I}(o, \mathbb{Z}_I^k) \quad \text{and} \quad \pi_{[n]\setminus I}(z_i) \in A \quad \text{for all } i = 1, 2, \dots \tag{12}$$

Since $\mathscr{V}_{\omega_I}(o; \mathbb{Z}^k_I)$ and A are finite this would contradict the fact that all the z_i's are distinct.

Since $z_1 = o$, (12) holds for $i = 1$. Now assume that (12) holds for some $i \geq 1$. Since $\|z_i - z_{i+1}\| = 1$, either we have $\|\pi_{[n]\setminus I}(z_{i+1}) - \pi_{[n]\setminus I}(z_i)\| = 1$ and $\|\pi_I(z_{i+1}) - \pi_I(z_i)\| = 0$ or else $\|\pi_{[n]\setminus I}(z_{i+1}) - \pi_{[n]\setminus I}(z_i)\| = 0$ and $\|\pi_I(z_{i+1}) - \pi_I(z_i)\| = 1$. In the first case, let $x = \pi_I(z_i) = \pi_I(z_{i+1})$. We have $\pi_I(z_{i+1}) \in \mathscr{V}_{\omega_I}(o, \mathbb{Z}^k_I)$. Moreover, since $\|\pi_{[n]\setminus I}(z_{i+1}) - \pi_{[n]\setminus I}(z_i)\| = 1$, we have $\pi_{[n]\setminus I}(z_{i+1}) \in A \cup T$. But $\pi_{[n]\setminus I}(z_{i+1})$ is not $\mathscr{P}_I(x)$-closed, so $\pi_{[n]\setminus I}(z_{i+1}) \notin T$, therefore we must have $\pi_{[n]\setminus I}(z_{i+1}) \in A$. In the second case, $z_i \in \mathscr{P}_I(x)$ and $z_{i+1} \in \mathscr{P}_I(y)$, where $\|x - y\| = 1$ and $x \in \mathscr{V}_{\omega_I}(o; \mathbb{Z}^k_I)$ thus, since $\omega_I(y) = 1$ we must have that $y \in \mathscr{V}_{\omega_I}(o; \mathbb{Z}^k_I)$. Also $\pi_{[n]\setminus I}(z_i) = \pi_{[n]\setminus I}(z_{i+1})$ and hence $\pi_{[n]\setminus I}(z_{i+1}) \in A$. Therefore, (12) follows by induction. □

We use Lemma 1 in order to prove the following result that settles the existence of a subcritical phase and hence proves the second assertion in Theorem 1.

Lemma 2 *Assume that $p_I < p_c(\mathbb{Z}^k)$ for some $I \in \mathscr{I}(k; n)$ and that $p_{J_i} < 1$ for the $n - k$ distinct $J_i \in \mathscr{I}(k; n)$ such that $\#(I \cap J_1 \cap \ldots \cap J_{n-k}) = k - 1$. Then $\mathscr{V}_\omega(o; \mathbb{Z}^n)$ is finite $\mathbb{P}_{\mathbf{p}}$-a.s.*

Proof The proof is divided into 2 cases:

First Case: $k = n - 2$, $k \geq 2$
Let us assume for simplicity that $I = \{1, \ldots, k\}$, $J_1 = \{1, \ldots, k-1, k+1\}$ and $J_2 = \{1, \ldots, k-1, k+2\}$ (thus $p_I < p_c(\mathbb{Z}^k)$ and $p_{J_1}, p_{J_2} < 1$). Let $\mathscr{H} = \mathscr{H}(I)$ be the $\mathbb{Z}^k$-decomposition of $\mathbb{Z}^n$ associated to I. Then $\mathscr{H}$ is isomorphic to $\mathbb{Z}^k$ and each site of $\mathscr{H}$ is isomorphic to $\mathbb{Z}^2$. Since $p_I < p_c(\mathbb{Z}^k)$, there exists $a.s.$ a (random) non-negative integer N such that $\mathscr{V}_{\omega_I}(o; \mathbb{Z}^k_I) \subset B(N)$. In particular, Condition (i) in Lemma 1 holds $a.s.$ and all we need to show is that Condition (ii) holds a.s. on the event $[\mathscr{V}_{\omega_I}(o; \mathbb{Z}^k_I) \subset B(N)]$ for each fixed N.

To this end, first recall that for each $x \in \mathbb{Z}^k_I$, $\mathscr{P}_I(x) = \{z \in \mathbb{Z}^n : \pi_I(z) = x\}$. Since $p_{J_1} < 1$, the Borel-Cantelli Lemma guarantees that, almost surely, there exists $x^*_{k+1} \in \mathbb{N}$ such that $\omega_{J_1}(x_1, \ldots, x_{k-1}, x^*_{k+1}) = \omega_{J_1}(x_1, \ldots, x_{k-1}, -x^*_{k+1}) = 0$, for all $(x_1, \ldots, x_{k-1}) \in [-N, N]^{k-1} \cap \mathbb{Z}^{k-1}$. This implies that for each $x = (x_1, \ldots, x_k) \in [-N, N]^k \cap \mathbb{Z}^k$

$$\omega(x_1, \ldots, x_k, x^*_{k+1}, x_{k+2}) = \omega(x_1, \ldots, x_k, -x^*_{k+1}, x_{k+2}) = 0, \ \forall\, x_{k+2} \in \mathbb{Z}.$$

In other words, for each $x \in B(N) \subset \mathbb{Z}^k_I$ the set of $\mathscr{P}_I(x)$-closed sites of $\pi_{[n]\setminus I}(\mathscr{P}_I(x)) \subset \mathbb{Z}^2_{[n]\setminus I}$ contains the lines $\{(0, \ldots, 0, x^*_{k+1}, s) : s \in \mathbb{Z}\}$ and $\{(0, \ldots, 0, -x^*_{k+1}, s) : s \in \mathbb{Z}\}$. Similarly, we find x^*_{k+2} such that for each $x = (x_1, \ldots, x_k) \in [-N, N]^k \cap \mathbb{Z}^k$

$$\omega(x_1, \ldots, x_k, x_{k+1}, x^*_{k+2}) = \omega(x_1, \ldots, x_k, x_{k+1}, -x^*_{k+2}) = 0, \ \forall\, x_{k+1} \in \mathbb{Z}.$$

Hence for each $x \in B(N) \subset \mathbb{Z}_I^k$, the set of $\mathscr{P}_I(x)$-closed sites of $\pi_{[n]\setminus I}(\mathscr{P}_I(x)) \subset \mathbb{Z}^2$ contains the lines $\{(0,\ldots,0,t,x^*_{k+2}) : t \in \mathbb{Z}\}$ and $\{(0,\ldots,0,-t^*_{k+2}) : t \in \mathbb{Z}\}$. Therefore, for each fixed N, Condition (ii) in Lemma 1 holds $a.s.$ in the event $\mathscr{V}_{\omega_I}(o;\mathbb{Z}_I^k) \subset [-N,N]^k$ with $T \subset \mathbb{Z}^2_{[n]\setminus I}$ being the rectangle delimited by the lines $\{(0,\ldots,0,x^*_{k+1},s) : s \in \mathbb{Z}\}$, $\{(0,\ldots,0,-x^*_{k+1},s) : s \in \mathbb{Z}\}$, $\{(0,\ldots,0,t,x^*_{k+2}) : t \in \mathbb{Z}\}$ and $\{(0,\ldots,0,t,-x^*_{k+2}) : x \in \mathbb{Z}\}$. This completes the proof in the case $n = k+2$.

The General Case: $n = k + l$**,** $k, l \geq 2$
Let us assume for simplicity that $I = \{1,\ldots,k\}$ and $J_i = \{1,\ldots,k-1,k+i\}$ for $1 \leq i \leq l$. Let $\mathscr{H} = \mathscr{H}(I)$ be the $\mathbb{Z}^k$-decomposition of $\mathbb{Z}^n$ associated to I. Then $\mathscr{H}$ is isomorphic to $\mathbb{Z}^k$ and each site in $\mathscr{H}$ is isomorphic to $\mathbb{Z}^l$. Similar to above, condition $p_I < p_c(\mathbb{Z}^k)$ implies that Condition (i) in Lemma 1 holds $a.s.$ Hence we only need to show that $p_{J_i} < 1$ for $1 \leq i \leq l$ implies that Condition (ii) in the same Lemma holds $a.s.$ Similarly as above, for each $1 \leq i \leq l$, the family of random variables $\{\omega_{J_i}(x_1,\ldots,x_{k-1},x_{k+i}) : -N \leq x_1,\ldots,x_{k-1} \leq N, x_{k+i} \in \mathbb{Z}\}$ are independent, and hence by the Borel-Cantelli Lemma there exists $a.s.$ a set of nonnegative integers $\{x^*_{k+i}\}_{i=1}^l$ such that $\omega_{J_i}(x_1,\ldots,x_{k-1},x^*_{k+i}) = \omega_{J_i}(x_1,\ldots,x_{k-1},-x^*_{k+i}) = 0$, for all $-N \leq x_1,\ldots,x_{k-1} \leq N$ and all $i = 1,\ldots,l$. This implies that, for every $i = 1,\ldots,l$, the set of $\mathscr{P}_I(x)$-closed sites of $\pi_{[n]\setminus I}(\mathscr{P}_I(x))$ contains the hyperplane

$$T_i := \left\{z = (0,\ldots,0,z_{k+1},\ldots,z_{k+l}) \in \mathbb{Z}^l_{[n]\setminus I} \colon\, z_{k+i} = \pm x^*_{k+1}\right\}.$$

This shows that Condition (ii) holds $a.s.$ with

$$T := \{0\} \times \cdots \times \{0\} \times \partial\big([-x^*_{k+1},x^*_{k+1}] \times \ldots \times [-x^*_{k+l},x^*_{k+l}]\big) \cap \mathbb{Z}^l_{[n]\setminus I}.$$

□

Remark 1 (Proof of the Second Assertion in Theorem 1) It follows directly from the statement of Lemma 2 that when the parameters p_I are sufficiently small (e.g. if they all belong to the interval $(0, p_c(Z^k))$) then percolation does not occur. Notice, however, that Lemma 2 provides much more detail on the location of the subcritical phase in the space of parameters.

The next result also implies the existence of the subcritical phase. Strictly speaking, it only holds in the particular setting when k divides n and, although it will not be used in the remainder of the paper, we decided to include it here because it adds some further information to the phase diagram in this specific setting. Its proof also uses Lemma 1.

Lemma 3 *Assume that k divides n, and let ω be a Bernoulli (n,k)-hyperplane percolation process. Let $I_1,\ldots,I_{n/k} \in \mathscr{I}(k;n)$ be a partition of $[n]$. If for each $1 \leq j \leq n/k$ we have $p_{I_j} < p_c(\mathbb{Z}^k)$, then $\mathscr{V}_\omega(o;\mathbb{Z}^n)$ is finite $\mathbb{P}_{\mathbf{p}}$-$a.s.$*

Proof Write $n = lk$. Without loss of generality, assume that for each $1 \leq j \leq l$, $I_j = \{(j-1)k+1,\ldots,jk\}$. We will argue by induction on $l \geq 2$.

We begin fixing $l = 2$. Let $\mathscr{H} = \mathscr{H}(I_1)$ be the $\mathbb{Z}^k$-decomposition of $\mathbb{Z}^{2k}$ corresponding to I_1. Then each site $\mathscr{P}_{I_1}(x)$ of $\mathscr{H}$ is isomorphic to $\mathbb{Z}^k$. We want to verify that Conditions (i) and (ii) in Lemma 1 hold $a.s.$ with $I = I_1$. Since $p_{I_1} < p_c(\mathbb{Z}^k)$, the cluster $\mathscr{V}_{\omega_{I_1}}(o, \mathbb{Z}^{k_{I_1}})$ is finite $a.s.$, hence Condition (i) holds $a.s.$ Let us condition on $\mathscr{V}_{\omega_{I_1}}(o, \mathbb{Z}^{k_{I_1}})$ and show that Condition (ii) also holds $a.s.$

Since $l = 2$ we have $\mathbb{Z}^{n-k}_{[n]\setminus I} = \mathbb{Z}^k_{I_2}$. Now, since $p_{I_2} < p_c(\mathbb{Z}^k)$, the cluster $\mathscr{V}_{\omega_{I_2}}(o, \mathbb{Z}^k_{I_2})$ is finite a.s. Hence, almost surely, there exists $T \subset \mathbb{Z}^k_{I_2}$ that surrounds the origin and whose sites are ω_{I_2}-closed. In particular, they are $\mathscr{P}_{I_1}(x)$-closed for every $x \in \mathscr{V}_{\omega_1}(o, \mathbb{Z}^k_{I_1})$. This shows that Condition (ii) in Lemma 1 holds $a.s.$ with $I = I_1$. This concludes the proof for the case $l = 2$.

Now, assume that the result holds for some $l \in \mathbb{N}$ and let $n = (l+1)k$. Since $p_{I_1} < p_c(\mathbb{Z}^k)$, Condition (i) in the same lemma holds $a.s.$ with $I = I_1$. Let us condition on the cluster $\mathscr{V}_{\tilde{\omega}_1}(o; \mathbb{Z}^l_{I_1})$ and show that Condition (ii) holds.

Since $\mathbb{Z}^{lk}_{[n]\setminus I_1}$ is isomorphic to $\mathbb{Z}^{lk}$ we can define naturally Bernoulli (lk, k)-hyperplane percolation processes on it. In fact, one can show that $(\tilde{\omega}(v))_{v \in \mathbb{Z}^{lk}_{[n]\setminus I_1}}$ defined as

$$\tilde{\omega}(v) = \prod_{\substack{J \in \mathscr{I}(k;n) \\ J \cap I_1 = \varnothing}} \omega_I(\pi_J(v))$$

is indeed an instance of such a percolation process. Since for each $2 \le j \le l+1$ we have $p_{I_j} < p_c(\mathbb{Z}^k)$, one can use the induction hypothesis to obtain that $\mathscr{V}_{\tilde{\omega}}(o; \mathbb{Z}^{lk}_{[n]\setminus I_1})$ is finite almost surely. Thus, almost surely, there exists a set $T \subset \mathbb{Z}^{lk}_{[n]\setminus I_1}$ that surrounds the origin in $\mathbb{Z}^{lk}_{[n]\setminus I_1}$ and whose sites are all $\tilde{\omega}$-closed. Therefore, for each $v \in T$, there exists $J \in \mathscr{I}(k; n) \setminus I_1$ such that $\omega_J(v) = 0$ which means that v is $\mathscr{P}_{\omega_{I_1}}(x)$-closed for every $x \in \mathbb{Z}^k_{I_1}$, in particular, for every $x \in \mathscr{V}_{\omega_{I_1}}(o; \mathbb{Z}^k_{I_1})$. This establishes Condition (ii). The result follows by induction.

□

We close this section presenting a sketch to a proof for the existence of regimes in which the connectivity decay is exponential.

Remark 2 (Exponential Decay) Assume that at least $\binom{n-1}{k} + 1$ of the parameters $I \in \mathscr{I}(k; n)$ satisfy $p_I < p_c(\mathbb{Z}^k)$. If the event $\big[o \leftrightarrow \partial B(K)\big]$ holds for some integer $K > 1$ then there must be at least one site $x = (x_1, \ldots, x_n) \in \partial B(K)$ for which $[o \leftrightarrow x]$. Such a site x has at least one coordinate, say x_{i_o}, with $x_{i_o} = K$. By continuity of the projections into the coordinate planes, the events $\big[o \leftrightarrow \pi_J\big(\partial B(K)\big)\big]$ must occur for all the indices $J \in \mathscr{I}(k; n)$ containing i_o. This amounts for exactly $\binom{n-1}{k-1}$ indices, hence by our assumption, there must be at least one of these indices J for which $p_J < p_c(\mathbb{Z}^k)$. This implies that $\mathbb{P}_{p_J}\big(o \leftrightarrow \pi_J\big(\partial B(K)\big)\big)$ decays exponentially fast [1, 10] (see also [3] for a more elementary proof). Therefore, $\mathbb{P}_{\mathbf{p}}\big(o \leftrightarrow \partial B(K)\big)$ must also decay exponentially fast.

3 The Existence of a Supercritical Phase

Our aim in this section is to prove that Bernoulli hyperplane percolation contain infinite connected components almost surely as soon as the parameters of $\mathbf{p}$ are large enough. According to (7) this is equivalent to $\mathbb{E}_{\mathbf{p}}\big(\omega(o)\big)$ being sufficiently close to 1. This is the content of the next result which readily implies the first assertion in Theorem 1.

Theorem 4 *Let $2 \le k \le n-2$ and ω be as in (4). If $\mathbb{E}_{\mathbf{p}}(\omega(o))$ is sufficiently close to 1, then $\mathbb{P}_{\mathbf{p}}(o \longleftrightarrow \infty) > 0$.*

Remark 3 (Proof of the First Assertion in Theorem 1) By (7), $E_{\mathbf{p}}(\omega(o))$ can be arbitrarily close to 1 provided that all the p_I are sufficiently large (all of them still smaller than 1). Therefore, the first assertion in Theorem 1 follows readily from Theorem 4.

In percolation, such a result is usually obtained with the help of Peierls-type arguments, that is, by restricting the process to the plane $\mathbb{Z}^2$ and showing that, as long as the control parameter are made large enough, large closed sets surrounding the origin in $\mathbb{Z}^2$ are very unlikely. In our case, restricting the model to $\mathbb{Z}^2$ is not useful since the plane will be disconnected into finite rectangles. We therefore replace $\mathbb{Z}^2$ with an subgraph resembling a plane that is inclined with respect to the coordinate axis in order to gain some independence.

For that we will use the following auxiliary result whose proof relies on elementary arguments and is presented in the Appendix.

Lemma 4 *Let $2 \le k \le n-1$. There exist orthogonal vectors w_1 and w_2 in $\mathbb{Z}^n$ with $\|w_1\| = \|w_2\|$ such that the linear application $A : \mathbb{R}^2 \to \mathbb{R}^n$ given by $A(x, y) = xw_1 + yw_2$ satisfies the following properties:*

i. For every $I \in \mathscr{I}(k; n)$ the mapping $\pi_I \circ A : \mathbb{Z}^2 \to \mathbb{Z}^{\#I}$ is injective;
ii. There exists a constant $c = c(w_1, w_2) > 0$ such that for every $I \in \mathscr{I}(k; n)$, and every u and v in $\mathbb{R}^2$, $\|\pi_I(Au - Av)\| \ge c\|u - v\|$.

For the rest of this section we fix the dimension n, the vectors and w_1 and w_2, and the corresponding linear application A as in Lemma 4. We define $\mathscr{G}_0 = A(\mathbb{Z}^2)$. By Condition i. in Lemma (4) for every $2 \le k \le n-1$, the Bernoulli (n, k)-hyperplane percolation process ω restricted to $\mathscr{G}_0$ has i.i.d. states i.e., the process $\eta_0 := \{\omega(Av)\}_{v \in \mathbb{Z}^2}$ is a standard Bernoulli site percolation process in $\mathbb{Z}^2$ with parameter $p = \mathbb{E}_{\mathbf{p}}(\omega(o))$. Thus, for p close to 1, η_0 has an infinite cluster $a.s.$ whose image under A is also an infinite set composed of ω-open sites in $\mathbb{Z}^n$. However, $\mathscr{G}_0$ is not necessarily a connected subgraph of $\mathbb{Z}^n$ and thus we did not prove that $A(\eta_0)$ has an infinite open cluster. To fix this issue we will add sites to $\mathscr{G}_0$ in such a way to guarantee that we get a connected subgraph $\mathscr{G} \subset \mathbb{Z}^n$. Now the family of random variables $\{\omega(x)\}_{x \in \mathscr{G}}$ may no longer be independent. However it will still dominate an independent family as long as the parameters p_I are large enough. This will allow us to find an infinite cluster in $\mathscr{G}$ a.s.

3.1 Construction of the Graph $\mathscr{G}$

Let w_1 and w_2 be as above and denote $w_1 = (\alpha_1, \ldots, \alpha_n)$ and $w_2 = (\beta_1, \ldots, \beta_n)$. Let $p_0 = q_0 = o$ and define inductively for $1 \le j \le n$:

$$p_j = p_{j-1} + \alpha_j e_j,$$
$$q_j = q_{j-1} + \beta_j e_j.$$

Given $u, v \in \mathbb{Z}^n$ such that $u - v = ze_j$ for some $z \in \mathbb{Z}$, denote $[u, v] = \{w \in \mathbb{Z}^n : w = u + l(z/|z|)e_j,\ l = 0, \ldots, |z|\}$ (if $z = 0$, then $u = v$ so set $[u, v] = \{u\}$). Let $\Gamma(0,0) := \bigcup_{j=1}^{n} [p_{j-1}, p_j] \cup [q_{j-1}, q_j]$ and $\Gamma(x, y) = \{A(x, y) + v : v \in \Gamma(0, 0)\}$ which contains a path that starts at $A(x, y)$ and ends at $A(x + 1, y)$ and another that starts at $A(x, y)$ and ends at $A(x, y + 1)$. Therefore, if we denote

$$\mathscr{G} := \bigcup_{(x,y)\in\mathbb{Z}^2} \Gamma(x, y),$$

then, when regarded as a subgraph of the $\mathbb{Z}^n$ lattice, $\mathscr{G}$ is connected.

As mentioned above, we will study the percolation process restricted to $\mathscr{G}$ which we hope will dominate a supercritical percolation process. In implementing these ideas, an standard result due to Liggett, Schonmann and Stacey [9] is very useful. Before we state it precisely, let us give the relevant definitions.

A random element $\big(f(x)\big)_{x\in\mathbb{Z}^n} \in \{0, 1\}^{\mathbb{Z}^n}$ is said of class $C(n, \chi, p)$ if for every $x \in \mathbb{Z}^n$ and $S \subset \mathbb{Z}^n$ such that $\inf\{\|a - x\| : a \in S\} \ge \chi$, we have $\mathbb{P}\big(f(x) = 1 | (f(a))_{a\in S}\big) \ge p$. Such elements appear naturally when performing one-step renormalization arguments. We are ready to state a result that will help to control the process restricted to $\mathscr{G}$ and will also be used in Sect. 4. It consists of a rephrasing of the part of the statement of Theorem 0.0 in [9] that serves our purposes.

Theorem 5 (Theorem 0.0 in [9]) *For every $\rho > 0$ and χ there exists p_0 such that every random element $(f(x))_{x\in\mathbb{Z}^n}$ of class $C(n, \chi, p)$ with $p > p_0$ dominates stochastically an i.i.d. family of Bernoulli random variables $(g(x))_{x\in\mathbb{Z}^n}$ such that $\mathbb{P}(g(x) = 1) = \rho$. Moreover, ρ can be taken arbitrarily close to* 1 *provided that p_0 is also made sufficiently close to* 1.

Let $\{\eta(x, y)\}_{(x,y)\in\mathbb{Z}^2}$ be such that

$$\eta(x, y) = \begin{cases} 1, & \text{if all sites in } \Gamma(x, y) \text{ are open,} \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

Lemma 5 *There exists $\chi \in \mathbb{N}$ and $s = s(\mathbb{E}_{\mathbf{p}}(\omega(o)))$ such that, under $\mathbb{P}_{\mathbf{p}}$, the process η given by* (13) *is of class $C(2, \chi, s)$. Furthermore, s can be made arbitrarily close to* 1 *provided that all the parameters p_I are made close enough to* 1.

Proof For $x \in \mathbb{Z}^n$ and $R > 0$, let $B(x; R)$ be the set of sites $y \in \mathbb{Z}^n$ such that $\|x - y\| \leq R$. Let $R = \|w_1\| = \|w_2\|$. Observe that for each $v \in \mathbb{Z}^2$ we have that $\Gamma(v) \subset B(Av; R)$. In particular, for all $I \in \mathscr{I}(k; n)$, $\pi_I(\Gamma(v)) \subset \pi_I(B(\pi_I \circ Av; R))$. By Lemma 4-*ii*. there exists $c > 0$ such that

$$\|\pi_I(Av - Au)\| \geq c\|u - v\|.$$

Thus, the choice $\chi = 3R/c$ gives that $\pi_I(B(\pi_I \circ Au; R))$ and $\pi_I(B(\pi_I \circ Av; R))$ are disjoint provided that $\|v - u\| \geq \chi$. In particular $\eta(v)$ is independent of $\{\eta(u) : u \in \mathbb{Z}^2, \|v - u\| \geq \chi\}$.

Since $\Gamma(v) \subset B(Av; R)$ we can use (8) to obtain

$$\mathbb{P}_{\mathbf{p}}(\eta(v) = 1) \geq \mathbb{P}_{\mathbf{p}}\big(\text{all sites in } B(Av, R) \text{ are open}\big) \geq \big[\mathbb{E}_{\mathbf{p}}(\omega(o))\big]^{(2R+1)^k}.$$

Hence η is of class $C\big(2, 3R/c, [\mathbb{E}_{\mathbf{p}}(\omega(o))]^{(2R+1)^k}\big)$. □

We are now ready to present the proof of 4.

Proof of Theorem 4 In light of Theorem 5, we can choose all $p_I < 1$ sufficiently close to 1 so that $\mathbb{E}\omega(o) = \prod_{I \in \mathscr{I}} p_I$ is large enough to guarantee that the process η defined in (13) dominates stochastically a standard supercritical Bernoulli site percolation process in $\mathbb{Z}^2$. In particular, with positive probability we have that $\mathscr{V}_\eta(o)$ is infinite. We conclude by observing that each path $\{o, v_1, v_2 \ldots, \} \in \mathbb{Z}^2$ of η-open sites such that $\lim_{j\to\infty} \|v_j\| = \infty$ can be mapped into a path $\{o, x_1, x_2, \ldots\} \subset \mathscr{G}$ with $\lim_{j\to\infty} \|x_j\| = \infty$ and whose sites are ω-open. □

4 Polynomial Decay of Connectivity

In this section we prove Theorem 3 and indicate the few modifications that lead to the proof of Theorem 2. Our method follows essentially the ideas presented in [7] for Bernoulli line percolation. However, there is a complication and we need to adapt Lemma 4.7 therein to the higher dimension setting. The main problem is that the proof presented in [7] only works in 3-dimensions. We replace that result by our Proposition 1 whose proof relies on Lemma 6.

4.1 Crossing Events

Given integers $a < b$, $c < d$ and a configuration $\{\eta(x)\}_{x\in\mathbb{Z}^2} \in \{0, 1\}^{\mathbb{Z}^2}$ we say that there is a bottom to top crossing in the rectangle $R := [a, b] \times [c, d] \cap \mathbb{Z}^2$ if there is a path $\{(x_0, y_0), \ldots, (x_T, y_T)\} \subset R$ of η-open sites such that $y_0 = c$ and $y_T = d$. We denote $\mathscr{BT}(R)$ the event that such a crossing occurs that is, the set of all the configurations η for which there is a bottom to top crossing in R. Similarly, we say that there is a left to right crossing in the rectangle R if there exists a path of open sites $\{(w_0, z_0), \ldots, (w_H, z_H)\} \subset R$ with $w_0 = a$ and $w_H = b$, and similarly we denote this event by $\mathscr{L}\mathscr{R}(R)$.

Let N and $\{m_j\}_{j=2}^n$ be non-negative integers and denote

$$B = B(N, m_2, \ldots, m_n) := [0, N] \times [0, m_2] \times \ldots \times [0, m_n] \cap \mathbb{Z}^n.$$

Throughout this section we will regard the first coordinate as measuring the height of the rectangle B. Thus, for a random element $\{\eta(x)\}_{x \in \mathbb{Z}^n} \in \{0, 1\}^{\mathbb{Z}^n}$, we can refer to bottom to top crossings in B: We denote by $\mathscr{B}\mathscr{T}(B)$ the set of all the configurations for which there exists a path of η-open sites $\{x_0, \ldots, x_T\} \subset B$ such that $\pi_{\{1\}}(x_0) = 0$ and $\pi_{\{1\}}(x_T) = N$.

Let $k = 2$ and $I_j = \{1, j\}$, $2 \leq j \leq n$. Notice that the set $\pi_{I_j}(B) \subset \mathbb{Z}^2_{I_j}$ is isomorphic to a rectangle in $\mathbb{Z}^2$ with side lengths m_j and N corresponding to the jth and first coordinate respectively. Define

$$\xi(x) = \omega_{I_2}(\pi_{I_2}(x)) \cdots \omega_{I_n}(\pi_{I_n}(x))$$

Notice that $\xi \leq \omega$ (see (4)). If there is a bottom to top crossing in B of sites $x \in \mathbb{Z}^n$ that are ω_{I_j}-open for all $2 \leq j \leq n$ (which is to say, $\xi \in \mathscr{B}\mathscr{T}(B)$) then a simple projection onto the coordinate planes $\mathbb{Z}^2_{I_j}$ shows that $\omega_{I_j} \in \mathscr{B}\mathscr{T}(\pi_{I_j}(B))$ for all $2 \leq j \leq n$. Our next result states that the converse is also true. For that, given paths that cross the projections $\pi_{I_j}(B)$ from top to bottom we will need to construct a path inside B which is projected under π_{I_j} to the given crossing in $\pi_{I_j}(B)$. Although it may sound somewhat intuitive that it is possible to do so, we did not find any existing proof for this fact. Therefore, we have produced a combinatorial proof that may be interesting on its own.

Proposition 1 *Let $k = 2$. For each $i = 2, \ldots, n$, let $\gamma_{I_j} : \{0\} \cup [H_j] \to \pi_{I_j}(B)$ be a path composed of ω_j-open sites such that: $\pi_{\{1\}} \circ \gamma_{I_j}(0) = 0$, $\pi_{\{1\}} \circ \gamma_{I_j}(H_j) = N$ and for all $0 \leq t < H_j$ we have $\pi_{\{1\}} \circ \gamma_{I_j}(t) < N$. Then there exists a path $\lambda \colon \{0\} \cup [T] \to B$ whose sites are ω_{I_j}-open and satisfying that for every $j = 2, \ldots, n$, $\pi_{I_j}(\lambda(0)) = \gamma_{I_j}(0)$ and that $\pi_{I_j}(\lambda(T)) = \gamma_{I_j}(H_j)$. In particular, if $\omega_{I_j} \in \mathscr{B}\mathscr{T}(\pi_{I_j}(B))$ for all $2 \leq j \leq n$, then there exists a bottom to top crossing in B whose sites are ω_{I_j}-open for each $2 \leq j \leq n$.*

Proposition 1 implies that $[\xi \in \mathscr{B}\mathscr{T}(B)] = \bigcap_{j=2}^n \mathscr{B}\mathscr{T}(\pi_{I_j}(B))$ hence, by independence:

$$\mathbb{P}(\xi \in \mathscr{B}\mathscr{T}(B)) = \prod_{j=2}^{n} \mathbb{P}_{p_{I_j}}(\mathscr{B}\mathscr{T}(\pi_{I_j}(B)), \tag{14}$$

where $\mathbb{P}$ stands for $\otimes_{I\in\mathscr{I}(k;n)}\mathbb{P}_{p_I}$. For $n = 3$ and $k = 2$, this result has already been proved in [7, Lemma 4.7]. Here we extend this result for any $n \geq 3$ and $k = 2$.

For the proof of Proposition 1 we use the following lemma that is inspired by the Two Cautious Hikers Algorithm discussed by G. Pete in [11, page 1722].

Lemma 6 *Let N and $\{T_i\}_{i=1}^n$ be non-negative integers. Let, for each $i \in [n]$, $S_i : \{0\} \cup [T_i] \to \{0\} \cup [N]$ be functions satisfying:*

i. $|S_i(t) - S_i(t-1)| = 1, \forall t \in [T_i]$ *;*
ii. $0 \leq S_i(t) < N$, *for all* $0 \leq t < T_i$*;*
iii. $S_i(0) = 0$ *and* $S_i(T_i) = N$.

Then there exists $T \in \mathbb{N}$ and $f_i : \{0\} \cup [T] \to \{0\} \cup [T_i]$, $1 \leq i \leq n$, that satisfy:

(a) $|f_i(t) - f_i(t-1)| = 1$, *for all* $t \in [T]$*;*
(b) $S_1 \circ f_1(t) = S_j \circ f_j(t)$, *for all* $t \in \{0\} \cup [T]$, *for each* $1 \leq j \leq n$*;*
(c) $S_1(f_1(0)) = 0$ *and* $S_1(f_1(T)) = N$.

Before we give a proof for this lemma let us clarify its statement. The functions S_i can be thought of as n different random walks parametrized by $t \in \{0, \ldots T_i\}$ and that can, at each step, jump one unit up, jump one unit down or remain put. They are required to start at height 0, to remain above 0 and finish at height N. The conclusion is that it is possible to introduce delays to the individual random walks or even require them to backtrack (by means of composing them the f_i's) so that they will all be parametrized by the same interval $\{0, \ldots, T\}$ and always share the same height for any time inside this interval. The arguments in [11] can be modified in order to obtain a proof for $n = 2$. Below we present a proof that works for general n.

Proof of Lemma 6 Let $G = (V(G), E(G))$ be a graph with vertex set $V(G) = \left\{v = (t_1, \ldots, t_n) \in \mathbb{N}^n :\ t_i \in \{0, \ldots, T_i\} \text{ and } S_1(t_1) = S_j(t_j) \text{ for every } 1 \leq j \leq n\right\}$ and whose edge set $E(G)$ consists of the pairs of vertices $v = (t_1, \ldots, t_n)$ and $w = (s_1, \ldots, s_n)$ such that $|t_i - s_i| = 1$, for all $1 \leq i \leq n$. Similarly to [11], we have:

Claim The degrees of every vertex $v \in V(G)$ are even, except for $(0, \ldots, 0)$ and $(T_1, \ldots, T_n)$ that are the only vertices that have degree 1.

Proof of the Claim It is simple to verify that $(0, \ldots, 0)$ and $(T_1, \ldots, T_n)$ have degree one.

For $i \in [n]$ and $t \in \{0\} \cup [T_i]$, we say that t is of type:

$$(i, /) \quad \text{if} \quad S_i(t+1) = S_i(t-1) + 2;$$

$$(i, \backslash) \quad \text{if} \quad S_i(t+1) = S_i(t-1) - 2;$$

$$(i, \vee) \quad \text{if } S_i(t+1) = S_i(t-1) \text{ and } \quad S_i(t+1) = S_i(t) + 1;$$

$$(i, \wedge) \quad \text{if } S_i(t+1) = S_i(t-1) \text{ and } \quad S_i(t+1) = S_i(t) - 1.$$

Notice that if $v = (t_1, \ldots, t_n) \in V(G)$ has at least one t_i of type $(i, \vee)$ and at least one t_j of type $(j, \wedge)$ its degree in G has to be equal to zero. This is because $S_i(t_i \pm 1) = S_i(t_i) + 1 = S_j(t_j) + 1$, and $S_j(t_j \pm 1) = S_j(t_j) - 1$, and hence $S_i(t_i \pm 1) - S_j(t_j \pm 1) = 2$ which implies that every possibility for the entries $(t_i \pm 1)$ and $(t_i \pm 1)$ of a neighbor of v would lead to an element that does not belong to $V(G)$. Thus, a necessary condition for the degree of v to be different from 0 is that there is a partition $[n] = A \cup B$ such that for all $j \in B$, t_j is of type $(j, /)$ or $(j, \backslash)$ and, for all $i \in A$ either every t_i is of type $(i, \vee)$ or every t_i is of type $(i, \wedge)$.

In the case that $v = (t_1, \ldots, t_n) \in G$ is such that all t_i are of type $(i, /)$ or $(i, \backslash)$ we have that v has exactly two neighborhoods w and $w' \in V(G)$:

$$w = \left(t_i + \mathbb{1}_{\{t_i \text{ is of type } (i, /)\}} - \mathbb{1}_{\{t_i \text{ is of type } (i, \backslash)\}}\right)_{i=1}^{n},$$
$$w' = \left(t_i - \mathbb{1}_{\{t_i \text{ is of type } (i, /)\}} + \mathbb{1}_{\{t_i \text{ is of type } (i, \backslash)\}}\right)_{i=1}^{n}.$$

In the case that $v = (t_1, .., t_n) \in V(G)$ is such that exactly k entries, say $\{t_{i_j}\}_{j=1}^{k}$ are of type $(i, \vee)$ and the other $n - k$ entries are of type $(i, /)$ or $(i, \backslash)$, we obtain that v has exactly 2^k neighborhoods. This follows by induction on k, by observing that each of the two possibilities $t_{i_j} \pm 1$ imply $S_{i_j}(t_{i_j} \pm 1) = S_{i_j}(t_{i_j}) + 1$. Similarly, this is also true in the case that exactly k distinct $\{t_{i_j}\}_{j=1}^{k}$ are of type $(i, \wedge)$ and the other $n - k$ are of type $(i, /)$ or $(i, \backslash)$. This completes the proof of the claim. □

Let $v^* = (T_1, \ldots, T_n)$ and $o = (0, \ldots, 0)$ be the unique vertices in G that have degree 1. Let $\mathscr{H}$ be the largest connected subgraph of G that contains o. The sum $\sum_{v \in \mathscr{H}} \text{degree}(v)$ is twice the number edges of $\mathscr{H}$, in particular it is an even number. Thus, $\sum_{v \in \mathscr{H} \setminus \{o\}} \text{degree}(v)$ is an odd integer, and this holds if and only if $v^* \in \mathscr{H}$. Hence, o and v^* are in the same connected component of G, which implies the existence of a number $T \in \mathbb{N}$ and path $\gamma : \{0\} \cup [T] \to G$ with $\gamma(0) = o$ and $\gamma(T) = v^*$. The choice $f_i(t) = \pi_{\{i\}}(\gamma(t))$ completes the proof. □

Proof of Proposition 1 Assume that for each $2 \leq j \leq n$, there exist non-negative integers H_j and paths of ω_{I_j}-open sites $\gamma_{I_j} : \{0\} \cup [H_j] \to \pi_{I_j}(B)$ such that: $\pi_{\{1\}} \circ \gamma_{I_j}(0) = 0$, $\pi_{\{1\}} \circ \gamma_{I_j}(H_j) = N$ and for all $0 \leq t < H_j$ we have $\pi_{\{1\}} \circ \gamma_{I_j}(t) < N$.

For each $2 \leq j \leq n$ we define recursively a sequence of times $(\tau_j(0), \tau_j(1), \ldots, \tau_j(T_j))$ as follows:

$$\tau_j(0) = 0;$$
$$\tau_j(t+1) = \inf\{s \in [\tau_j(t) + 1, H_j] \cap \mathbb{N} : |\pi_{\{1\}}(\gamma_{I_j}(s) - \gamma_{I_j}(s-1))| = 1\},$$

and we define T_j as the first time at which $\tau_j(T_j) = H_j$.

Let $S_j := \pi_{\{1\}} \circ \gamma_{I_j} \circ \tau_j$, for each $2 \leq j \leq n$. Then $\{S_j\}_{j=2}^{n}$ satisfies Conditions *i.–iii.* in Lemma 6. Hence, there exist a non negative integer $T > 0$ and $\{f_j\}_{j=2}^{n}$ that satisfy Conditions *a–c* in the same lemma. We claim that the function $\lambda : \{0\} \cup$

$[T] \to B$ denoted by $\lambda(t) = (\lambda_1(t), \dots, \lambda_n(t))$ where

$$\lambda_1(t) = S_2 \circ f_2(t) = \cdots = S_n \circ f_n(t),$$
$$\lambda_j(t) = \pi_{\{j\}} \circ \gamma_{I_j} \circ \tau_j \circ f_j(t), \quad 2 \le j \le n,$$

has the following properties:

i. $\lambda(t)$ is ω_{I_j}-open for all $2 \le j \le n$;
ii. For each $1 \le t \le T$, $\lambda(t)$ and $\lambda(t-1)$ are connected by a path γ_t of sites that are ω_{I_j}-open for all $2 \le j \le n$;
iii. $\lambda_1(0) = 0$ and $\lambda_1(T) = N$.

The proof will be complete once we show that these three conditions are valid. One can check readily that Condition *iii.* holds. So we now prove the validity of the other two.

Validity of Condition i Notice that, for all $t \in [T]$,

$$\pi_{I_j}(\lambda(t)) = \big(\lambda_1(t), \lambda_j(t)\big) = \big(S_j \circ f_j(t), \pi_{\{j\}} \circ \gamma_{I_j} \circ \tau_j \circ f_j(t)\big)$$
$$= \big(\pi_{\{1\}} \circ \gamma_{I_j} \circ \tau_j, \pi_{\{j\}} \circ \gamma_{I_j} \circ \tau_j \circ f_j(t)\big) = \gamma_{I_j} \circ \tau_j \circ f_j(t).$$

Since every site in γ_{I_j} is ω_{I_j}-open, we have $\omega_{I_j}(\pi_{I_j}(\lambda(t))) = \omega_{I_j}(\gamma_{I_j} \circ \tau_j \circ f_j(t)) = 1$.

Validity of Condition ii For each $t \in [T]$ we have either $\lambda_1(t) = \lambda_1(t-1) + 1$ or $\lambda_1(t) = \lambda_1(t-1) - 1$. Let us assume that the former holds (the case when the latter holds can be treated similarly).

There are integers $\{x_j\}_{j=2}^n$ such that

$$\lambda(t) - \lambda(t-1) = (1, x_2, x_3, \dots, x_n),$$

and in particular:

$$\gamma_{I_j} \circ \tau_j \circ f_j(t) - \gamma_{I_j} \circ \tau_j \circ f_j(t-1) = e_1 + x_j e_j.$$

We claim that $\lambda(t) - e_1$ is ω_{I_j}-open for all $2 \le j \le n$. To see this, notice that $\pi_{I_j}(\lambda(t) - e_1) = \gamma_{I_j} \circ \tau_j \circ f_j(t) - e_1$, and that only two possibilities may happen: either $f_j(t) - f_j(t-1) = -1$ or $f_j(t) - f_j(t-1) = +1$. In any case, denoting $a_j = f_j(t) - f_j(t-1)$, we have

$$\gamma_{I_j} \circ \tau_j \circ f_j(t) - e_1 = \gamma_{I_j}(\tau_j \circ f_j(t) - a_j),$$

and hence $\lambda(t) - e_1$ is also ω_{I_j}-open for all $2 \le j \le n$. Let $p_0 = \lambda(t)$, $p_1 = \lambda(t) - e_1$, and $p_k = p_{k-1} - x_k e_k$, for $2 \le k \le n$. In particular, $p_n = \lambda(t-1)$. Let $[p_j, p_{j+1}]$ be the path that goes along the line segment of points $x \in \mathbb{Z}^n$ that has p_j and p_{j+1} as its extremes. We claim that $\gamma_t = \bigcup_{j=1}^n [p_{j-1}, p_j]$ is a path of sites fulfilling Condition *ii.*

Start with y in the line segment $[p_1, p_2]$. Thus, $\pi_1(y) = \pi_1(p_1)$ and it is ω_{I_j}-open for all $2 \leq k \leq n$: The case $k = 2$, follows from the definition of γ_{I_2} and τ_2; for $k > 2$ we have $\omega_{I_k}(\pi_{I_k}(y)) = \omega_{I_k}(\pi_{I_k}(p_1)) = 1$. Inductively, each y in the line segment $[p_j, p_{j+1}]$ satisfies the following properties:

$$\begin{aligned}
\pi_1(y) &= \pi_1(p_1) \\
\omega_{I_k}(\pi_{I_k}(y)) &= \omega_{I_k}(\pi_{I_k}(p_n)) = 1, \text{ for each } 2 \leq k \leq j, \\
\omega_{I_{j+1}}(\pi_{I_{j+1}}(y)) &= 1, \text{ by the definition of } \gamma_{I_{j+1}} \text{ and } \tau_{j+1}, \\
\omega_{I_k}(\pi_{I_k}(y)) &= \omega_{I_k}(\pi_{I_k}(p_1)) = 1, \text{ for each } j+1 < k \leq n.
\end{aligned}$$

This completes the proof. □

4.2 Percolation on a Renormalized Lattice

In this section we will define a percolation process in a renormalized lattice whose sites can be matched to hypercubes from the original lattice, called boxes. We restrict ourselves to the case $k = 2$ so that the projections of theses boxes into the coordinate planes are given by squares. The rough idea is as follows: The boxes are called good depending on whether some crossings occur inside and around some of these projected squares as illustrated in Fig. 1. Taking the side of the boxes to be large enough, we can guarantee that boxes are good with large probability, so that the percolation processes induced by good boxes in the renormalized lattice dominates

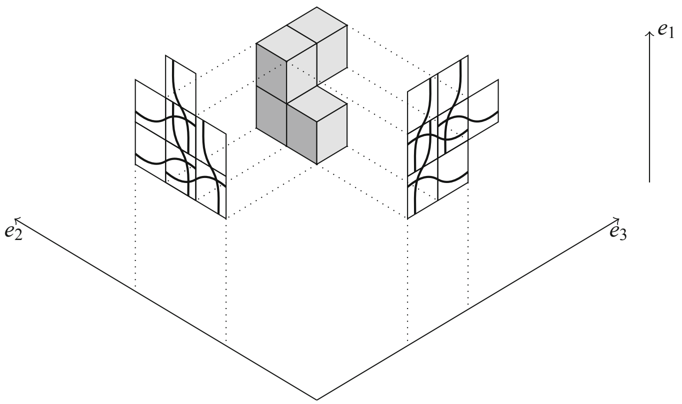


Fig. 1 The projection into the subspace spanned by e_1, e_2 and e_3 of a set of four adjacent boxes that are ω_{I_2}-good and ω_{I_3}-good

a supercritical site percolation process. This implies the existence of arbitrarily long paths of good boxes. The directed nature of the dependencies introduces some complications in showing stochastic domination, and in order to gain some kind of independence we will need to look at oriented portions of the renormalized lattice as the one illustrated in Fig. 3.

Before we proceed and formalize these ideas, we recall a classical fact about crossing events.

Remark 4 Let $p > p_c(\mathbb{Z}^2)$. If $c > 0$ is large enough (depending on p) then

$$\lim_{N\to\infty} \mathbb{P}_p(\mathscr{BT}([0, \lfloor c \log N \rfloor] \times [0, N] \cap \mathbb{Z}^2)) = 1, \tag{15}$$

where and $\lfloor x \rfloor$ denotes the integer part of a real number x.

From now on, given $y \in \mathbb{Z}^n$, we use the notation

$$B(y; N) =: [y_1, y_1 + N - 1] \times \ldots \times [y_n, y_n + N - 1] \cap \mathbb{Z}^n.$$

Furthermore, we define:

$$B_j(y; N) := \pi_{I_j}\big(B(y; N) \cup B(y + Ne_1; N) \cup B(y + Ne_j; N)\big).$$

In the next definition, we still stick to the convention that the first coordinate x_1 measures the height of each $B_j(y; N)$:

Definition 1 Given $\omega_{I_j} \in \{0, 1\}^{\mathbb{Z}^k_{I_j}}$, we say $B_j(y; N)$ is ω_{I_j}-good when

$$\omega_{I_j} \in \mathscr{BT}(\pi_{I_j}(B(y; N) \cup B(y+Ne_1; N))) \cap \mathscr{LR}(\pi_{I_j}(B(y; N) \cup B(y+Ne_j; N))).$$

Furthermore, given $(\omega_{I_j})_{j=2}^n$, we say that $B(y; N)$ is good if for all $j = 2, \ldots, n$, $B_j(y; N)$ is ω_{I_j}-good (see Fig. 1).

For fixed $N \in \mathbb{N}$ we say that $\gamma : \mathbb{N} \to \mathbb{Z}^n$ is a path of good boxes if

$$\begin{aligned} &\text{for each } t \in \mathbb{N},\ B(\gamma(t); N) \text{ is a good box and} \\ &\gamma(t + 1) = \gamma(t) + Ne_{j_t} \text{ for some } j_t \in [n]. \end{aligned} \tag{16}$$

In what follows we will make use of the following lemma whose proof can be done following exactly the same lines as in [7, Lemmas 4.10 and 4.11]. The idea is to iterate the use of Proposition 1 to pass from one good box to the next one following a path contained inside these boxes and whose sites are ω_{I_j}-open for each $2 \le j \le n$ (as done in [7, Lemma 4.10]). This is possible because the definition of good boxes entails the existence of a system of crossings inside the projections of these boxes into the respective $\mathbb{Z}^2_I$ for which Proposition 1 apply as shown in Fig. 1. Being able to pass from one good box to the next adjacent one, all we need to do is to concatenate the paths in order to obtain a path starting in the first good box in the sequence and ending at the last one (as done in [7, Lemma 4.11]).

Lemma 7 *Let $N \in \mathbb{N}$ and $\gamma : \mathbb{N} \to \mathbb{Z}^n$ be a path of good boxes. Then for each $t \in \mathbb{N}$, there exists a path of sites $\{x_0, \dots, x_T\} \subset \cup_{s=0}^{t} B(\gamma(s); N)$ that are ω_{I_j}-open for each $2 \leq j \leq n$, and with $x_0 \in B(\gamma(0); N)$ and $x_T \in B(\gamma(t); N)$.*

Let $p_0 = o$ and $r_0 = 2$. For $N \in \mathbb{N}$ and $t \in \mathbb{Z}$, define recursively

$$p_t = p_{t-1} + Ne_{r_t}, \tag{17}$$

where $r_t \in \{2, 3 \dots, n\}$ is such that $t \equiv r_t - 2 \mod (n-1)$. Roughly speaking, as t increases the values of r_t run through the set $\{2, 3, \dots, n\}$ cyclically. As an example, when $n = 4$ we have $r_0 = 2, r_1 = 3, r_2 = 4, r_3 = 2, r_4 = 3, r_5 = 4$, and so on. As for the points p_t, they form a directed sequence whose increments are segments of length N, each oriented along one of the directions in $\{e_2, e_3, \dots, e_n\}$. The orientation of these segments follow the same cyclic pattern as r_t.

Let $\nu = (\nu(t, x))_{(t,x)\in\mathbb{Z}^2}$ be the random element in $\{0, 1\}^{\mathbb{Z}^2}$ defined as

$$\nu(t, x) := \mathbb{1}[B(p_t + Nxe_1; N) \text{ is good}]. \tag{18}$$

The process ν can be thought as a percolation process in a renormalized (that is, rescaled) lattice where sites are now boxes of the type $B(p_t + Nxe_1; N)$ (for $(t, x) \in \mathbb{Z}^2$). The cyclic nature of r_t and the choice of the sequence p_t as in (17) allows us to derive some properties of this renormalized lattice. On the one hand, the projection into the subspace spanned by $e_2, \dots, e_n$ is given by a spiral sequence of neighboring boxes which is isomorphic to a line of boxes sharing a face, see Fig. 2. On the other hand, the projection into the subspace spanned by e_1, e_i, e_j for $i, j \in \{2, \dots, n\}$ and $i \neq j$ resembles a jagged wall of boxes as illustrated in Fig. 3

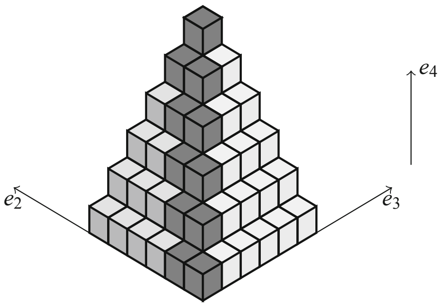


Fig. 2 Let $n = 4$ and $k = 2$. The darker boxes are the projections into the subspace spanned by e_2, e_3, e_4 of boxes $B(p_t + Nxe_1; N)$ for $t = 0, \dots, 15$ and arbitrary x (light gray boxes where added to help visualization). There are only 11 such boxes that are visible from this perspective which correspond to indices t with $r_t = 2$ or $r_t = 4$. The other 5 boxes corresponding to $r_t = 3$ are not visible because they lie behind other boxes colored in light gray

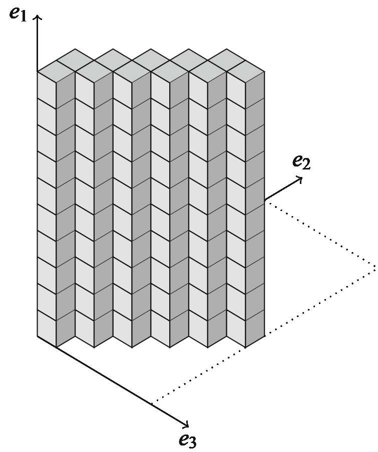


Fig. 3 Let $n = 4$ and $k = 2$. This picture shows the projection into the subspace spanned by e_1, e_2 and e_3 of the boxes $B(p_t + Nxe_1; N)$ for $t = 0, \dots, 15$ and $x = 0, \dots, 9$

which is isomorphic to a plane of adjacent boxes. This specific shape guarantees that, for the process ν, the statuses of distant boxes in this renormalized lattice are independent as they will have disjoint projections into the $\mathbb{Z}^2_{I_j}$ subspaces. This is the content of the next lemma:

Lemma 8 *For the random element ν given as in (18), the variable $\nu(t, x)$ is independent of $\{\nu(s, y): |t - s| \geq 2(n - 1)$ or $|y - x| \geq 2\}$. In particular, the process ν is of class $C(2, \chi, p)$ for some $p > 0$.*

Proof Let $(t, x) \in \mathbb{Z}^2$ be fixed. Since the event that the box $B(y; N)$ is good is measurable with respect to the σ-algebra generated by the family of random variables

$$\bigcup_{j=2}^{n} \{\omega_{I_j}(v) : v \in B_j(y; N)\},$$

all we need to show is that, for each $2 \leq j \leq n$, the sets $B_j(p_t + Nxe_1; N)$ and $B_j(p_s + Nye_1; N)$ are disjoint in both cases $|t - s| \geq 2(n - 1)$ or $|y - x| \geq 2$.

If $|y - x| \geq 2$ we have that for each $2 \leq j \leq n$ the distance between $B_j(p_t + Nxe_1; N)$ and $B_j(p_s + Nye_1; N)$ is at least 1, and this is also true in the case that $|t - s| \geq 2(n - 1)$, since in this case we have that, except for the first coordinate, each coordinate of $p_s - p_t$ has absolute value at least $2N$. □

Once the range of dependency is controlled for the process ν, we use Theorem 5 in order to have it dominated from below by a supercritical Bernoulli process:

Lemma 9 *Assume that for each $2 \leq j \leq n$ we have $p_{I_j} > p_c(\mathbb{Z}^2)$. Then for every $\epsilon > 0$ there exists $N = N(\epsilon) \in \mathbb{N}$ such that the process $\nu = (\nu(t, x))_{(t,x)\in\mathbb{Z}^2}$ (cf. (18)) dominates stochastically a standard Bernoulli site percolation process in $\mathbb{Z}^2$ with parameter $p > 1 - \epsilon$.*

Proof We write $\mathbb{P} = \otimes_{I \in \mathscr{I}(k;n)} \mathbb{P}_{p_I}$. Since for all $(t, x) \in \mathbb{Z}^2$ we have that $\mathbb{P}(\nu(t, x) = 1) = \mathbb{P}(\nu(0, 0) = 1)$, in view of Theorem 5 and the previous Lemma, we only need to show that $\mathbb{P}(\nu(0, 0) = 1)$ can be made arbitrarily close to 1 for suitable large N. Recall that the box $B(o; N)$ is good if for each $2 \le j \le n$, $B_j(o; N)$ is ω_{I_j}-good. Observe that for $j \neq l$, the events $[B_j(o; N)$ is ω_{I_j}-good$]$ and $[B_l(o; N)$ is ω_{I_l}-good$]$ are independent. Hence, all we need to show is that the assumption $p_{I_j} > p_c(\mathbb{Z}^2)$ implies that the probability that each $B_j(o; N)$ is ω_{I_j}-good can be made arbitrarily close to 1 for each j, for a suitable choice of a large integer N, which may depend on p_j. By the FKG inequality, we have:

$$\mathbb{P}(B_j(o; N) \text{ is } \omega_{I_j}\text{-good}) \ge \mathbb{P}_{p_{I_j}}\big(\mathscr{B}\mathscr{T}(\pi_{I_j}(B(y; N) \cup B(y + Ne_1; N))\big)\times$$
$$\mathbb{P}_{p_{I_j}}\big(\mathscr{L}\mathscr{R}(\pi_{I_j}(B(y; N) \cup B(y + Ne_j; N))\big).$$

The fact that each probability in the right-hand side above can be made arbitrarily close to 1 by choosing N sufficiently large follows from the fact that $p_{I_j} > p_c(\mathbb{Z}^2)$ together with classical crossing probability estimates for supercritical Bernoulli site percolation (for instance Eq. (15) in Remark 4 is sufficient). □

4.3 *Proof of Theorems 3 and 2.*

We start this section presenting the proof for Theorem 3. Roughly speaking, it consists of three steps. First find a path spanning a rectangle in the renormalized lattice that is very elongated in the vertical direction. This path can be mapped to a path of good boxes in the original lattice. Lemma 7 allows to obtain a long path of sites in the original lattice that are ω_{I_j} open for every $j = 2, \ldots, n$. Comparison with a supercritical percolation process in the renormalized lattice, shows that this step can be accomplished paying only a constant probability cost. The second step consists of guaranteeing that the sites in this long path are also ω_I open for every index I that does not contain the coordinate 1. The geometry of our construction allows to accomplish this step by paying only a polynomial price in the length of the path. An extra polynomial probability cost needs to be payed in order to require that the long path starts at the origin. The third and last step consists in guaranteeing that the origin does not belong to an infinite connected component. This is accomplished by constructing a closed set surrounding the origin much in the spirit of Lemma 1 . Again only polynomial probability cost is necessary to accomplish this step.

At the end of this section we indicate the modifications that need to be performed in the proof in order to obtain a proof of Theorem 2.

Proof of Theorem 3 Let $\epsilon > 0$ be such that $1 - \epsilon > p_c(\mathbb{Z}^2)$ and $\{\eta(x)\}_{x \in \mathbb{Z}^2}$ be a Bernoulli site percolation process on $\mathbb{Z}^2$ with parameter $1 - \epsilon$, hence supercritical. Fix $N \in \mathbb{N}$ (depending on ϵ) large enough so that the claim in Lemma 9 holds, i.e., the process $\{\nu(x)\}_{x\in\mathbb{Z}^2}$ defined in (18) dominates $\{\eta(x)\}_{x\in\mathbb{Z}^2}$ stochastically. Let

$$\mathscr{O}_1 = \big\{\nu \in \mathscr{B}\mathscr{T}([0, \lfloor c_o \log K\rfloor] \times [0, K] \cap \mathbb{Z}^2)\big\}.$$

In view of Remark 4, we can choose a large constant $c_o > 0$ (depending on ϵ) such that

$$\liminf_{K\to\infty} \mathbb{P}\big(\mathscr{O}_1) \geq \liminf_{K\to\infty} \mathbb{P}\big(\eta \in \mathscr{B}\mathscr{T}([0, \lfloor c_o \log K\rfloor] \times [0, K] \cap \mathbb{Z}^2) > 0. \tag{19}$$

The value of c_o will be kept fixed from now on.

Recall the definition of a path of good boxes as being a directed sequence of sites, each lying at distance N from the preceding one, and such that the corresponding boxes of size N are good (see (16)). In light of Lemma 7, the occurrence of $\mathscr{O}_1$ entails the existence of a path of good boxes $\{z_0, \ldots, z_T\} \subset \mathbb{Z}^n$ with $\pi_{\{1\}}(z_0) = 0$ and $\pi_{\{1\}}(z_T) = NK$, and such that each good box $B(z_i; N)$ is contained in the set $\mathscr{Z}^2(K; N) \subset \mathbb{Z}^n$ defined as

$$\mathscr{Z}^2(K; N) := \bigcup_{\substack{(t,x)\in\mathbb{Z}^2:\\ 0\leq t\leq\lfloor c_o \log K\rfloor, 0\leq x\leq K}} B(p_t + Nxe_1; N). \tag{20}$$

Roughly speaking, the set $\mathscr{Z}^2(K; N)$ is the subset of $\mathbb{Z}^n$ that comprises all the sites inside boxes of size N in a portion of the renormalized lattice resembling a thickening of width N of a jagged rectangular region of side $c_o \log(K)$ and height K. The reader might find it useful to consult Fig. 4 in order to clarify the definition of the set $\mathscr{Z}^2(K; N)$ and the definition of the event $\mathscr{O}_1$. Note however that the picture may be a little bit misleading because for K large, the jagged wall region depicted therein should look very elongated in the e_1 direction.

Let $\mathscr{O}_2$ be the event that every site $z \in \mathscr{Z}^2(K; N)$ is ω_I-open for all the indices $I \in \mathscr{I}(2; n)$ such that $I \cap \{1\} = \varnothing$. Notice that, for every index $I \in \mathscr{I}(2; n)$ that does not include the coordinate 1 each one of the boxes appearing in the l.r.s. of (20) projects into a square in $\mathbb{Z}_I^2$ containing N^2 sites. Moreover, using the cyclic nature of the r_t, we can conclude that as t runs over the interval $0, \ldots, \lfloor c_o \log K\rfloor$, the amount of different projections into each $\mathbb{Z}_I^2$ that one needs to check in order to determine the occurrence of $\mathscr{O}_2$ does not exceed $c' \log K$ for some positive universal constant $c' = c'(n, c_o)$ (for instance $c' = 3(n-1)^{-1}c_o$). See Fig. 5 for an illustration of these projections.

Therefore, we have

$$\mathbb{P}(\mathscr{O}_2) \geq \prod_{\substack{I\in\mathscr{I}(2;n);\\ \{1\}\cap I=\varnothing}} {p_I}^{c'\log(K)N^2} = \exp\Big(c'N^2 \log K \sum_{\substack{I\in\mathscr{I}(2;n);\\ \{1\}\cap I=\varnothing}} \log p_I\Big).$$

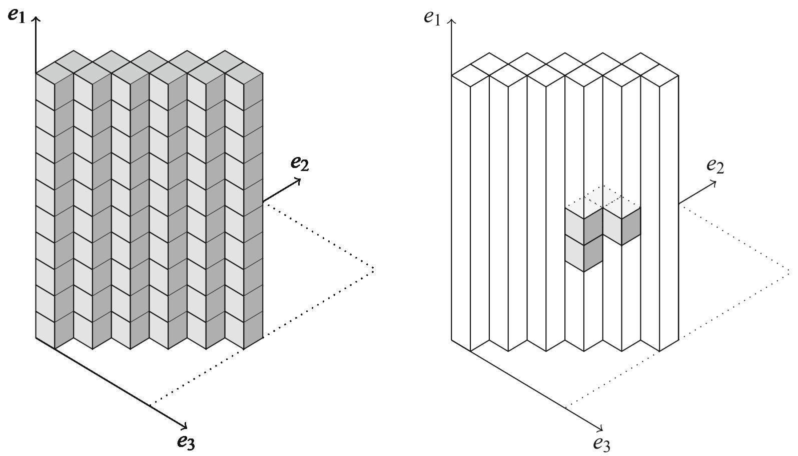


Fig. 4 Let $n = 4$ and $k = 2$. On the left: The projection of the set $\mathscr{X}^2(K; N)$ into the subspace spanned by e_1, e_2 and e_3 in a situation where $K = 9$ and $\lfloor c_o \log K \rfloor = 15$ (only 10 zig-zag steps appear, instead of 15 because indices t for which $r_t = 4$ lead to steps towards a fourth dimension). On the right: the union of the boxes corresponding to a piece of a path of boxes inside $\mathscr{X}(K; N)$. In order for the event $\mathscr{O}_1$ to happen, one needs the existence of such a path crossing the region from bottom to top

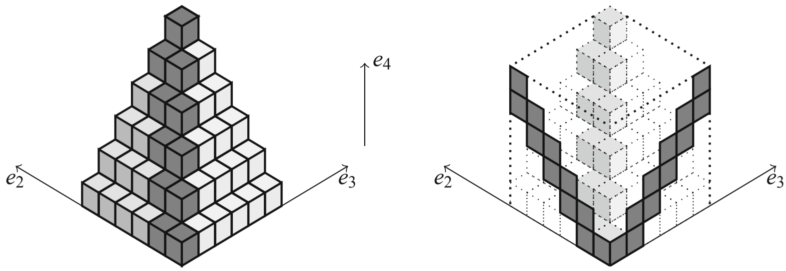


Fig. 5 On the left: the darker region is the projections of the $\mathscr{X}^2(K; N)$ into the subspace spanned by e_2, e_3 and e_4. The spiral contains $c_o\lfloor \log K \rfloor$ boxes of size N. On the right: the projection of the boxes into $\mathbb{Z}^2_{\{2,4\}}$ and $\mathbb{Z}^2_{\{3,4\}}$. These projections contain no more than $c' \log K$ squares with N^2 sites (for some constant $c' = c'(c_o, n) > 0$)

Let us define

$$\alpha_2 := c' N^2 \sum_{\substack{I \in \mathscr{I}(2;n);\\ \{1\} \cap I = \varnothing}} \log\left(\frac{1}{p_I}\right)$$

so that we have

$$\mathbb{P}(\mathscr{O}_2) \geq \exp(-\alpha_2 \log K) = K^{-\alpha_2}$$

which is to say that the probability of $\mathscr{O}_2$ is bounded below by a term that is proportional to a negative power of K with exponent α_2.

By Lemma 7, on the event $\mathscr{O}_1$ there exists a path $\gamma : \{0\} \cup [T] \to \mathscr{X}^2(K; N)$ of sites that are ω_{I_j}−open for all $2 \le j \le n$ such that $\pi_{\{1\}}(\gamma(0)) = 0$ and $\pi_{\{1\}}(\gamma(T)) = NK$. On the event $\mathscr{O}_1 \cap \mathscr{O}_2$, the sites in γ are actually ω-open. Unfortunately, this path may not start at o. In order to fix this issue, let us introduce the event $\mathscr{O}_3$ that all the sites $z \in \mathbb{Z}^n$ with $\pi_{\{1\}}(z) = -1$ and such that $z + e_1 \in \mathscr{X}^2(K; N)$ are ω_{I_j}-open for each $2 \le j \le n$. Since there are no more than $cN^{n-1} \log K$ such sites, $\mathbb{P}(\mathscr{O}_3)$ is also proportional to a negative power of K. Moreover, on the event $\mathscr{O}_2 \cap \mathscr{O}_3$ these sites are ω-open.

Now, on the event $\mathscr{O}_1 \cap \mathscr{O}_2 \cap \mathscr{O}_3$, o and $\gamma(0)$ are connected by a path of ω-open sites $\lambda : \{0\} \cup [J] \to \mathbb{Z}^n$, with $\lambda(0) = o$, $\lambda(J) = \gamma(0)$ and $\pi_{\{1\}}(\lambda(j)) = -1$ for every $j = 1, \dots, J - 1$. In other words, we have:

$$\mathscr{O}_1 \cap \mathscr{O}_2 \cap \mathscr{O}_3 \subset [o \leftrightarrow \partial B(NK)]. \tag{21}$$

Having constructed the events $\mathscr{O}_1$, $\mathscr{O}_2$ and $\mathscr{O}_3$ whose occurrence implies the existence of a long ω-open path starting at the origin, we now construct events on which the cluster containing the origin is finite. This is done by first requiring that the origin in $\mathbb{Z}^2_{\{2,3\}}$ is enclosed by a circuit composed of $\omega_{\{2,3\}}$-closed sites only and then requiring that there exists a set S surrounding the origin in $\mathbb{Z}^{n-2}_{[n]\setminus\{2,3\}}$ whose edges are $\mathscr{P}_{\{2,3\}}(x)$-closed for every for every x inside the circuit.

Indeed, let $\mathscr{O}_4$ be the event in which all sites in the square circuit in $\mathbb{Z}^2_{\{2,3\}}$ given by

$$\partial\Big(([-2, 4Nc_o\lfloor \log K\rfloor + 1] \times [-2, 4Nc_o\lfloor \log K\rfloor + 1]) \cap \mathbb{Z}^2_{\{2,3\}}\Big)$$

are $\omega_{\{2,3\}}$-closed. Since the perimeter of this circuit is less than $cN \log K$, for some $c > 0$, the probability that $\mathscr{O}_4$ occurs is bounded below by a negative power of K. Moreover, on the event $\mathscr{O}_4$, the origin of $\mathbb{Z}^2_{\{2,3\}}$ is surrounded by a $\omega_{\{2,3\}}$-closed circuit, therefore, $\mathscr{V}_{\omega_{\{2,3\}}}(o; \mathbb{Z}^2_{\{2,3\}})$ is finite, which is to say that Condition *i.* in Lemma 1 is satisfied. Note also that the circuit encloses the projection of $\mathscr{X}^2(K; N)$ into $\mathbb{Z}^2_{\{2,3\}}$.

Now, similarly to Condition *ii.* in Lemma 1, let $\mathscr{O}_5$ be the event in which there exists $\mathscr{S} \subset \mathbb{Z}^{n-2}_{[n]\setminus\{2,3\}}$ that surrounds the origin satisfying that $\inf\{\|s\| : s \in \mathscr{S}\} \ge 3NK$ and that each site $s \in \mathscr{S}$ is $\mathscr{P}_{\{2,3\}}(x)$−closed for all $x \in \big([-1, 4Nc_o\lfloor \log K\rfloor] \times [-1, 4Nc_o\lfloor \log K\rfloor]\big) \cap \mathbb{Z}^2_{\{2,3\}}$. Following exactly the same type of Borel–Cantelli argument as in the proof of Lemma 2, we obtain that $\mathbb{P}(\mathscr{O}_5) = 1$.

On the event $\mathscr{O}_4 \cap \mathscr{O}_5$, Conditions *i.-ii.* in the statement of Lemma 1 are satisfied. Therefore

$$\mathscr{O}_4 \cap \mathscr{O}_5 \subset [o \nleftrightarrow \infty]. \tag{22}$$

Combining (21) and (22) we get

$$\mathbb{P}_{\mathbf{p}}(o \leftrightarrow \partial B(NK),\ o \nleftrightarrow \infty) \geq \mathbb{P}(\mathscr{O}_1 \cap \mathscr{O}_2 \cap \mathscr{O}_3 \cap \mathscr{O}_4 \cap \mathscr{O}_5),$$

Since $\mathscr{O}_1$, $\mathscr{O}_2$, $\mathscr{O}_3$, $\mathscr{O}_4$ and $\mathscr{O}_5$ are independent events, the fact that the probabilities of $\mathscr{O}_1$ and $\mathscr{O}_5$ are uniformly bounded below by a positive constant, and that the remaining events have probability proportional to a negative power of K finishes the proof. □

We now indicate a few modifications to the proof of Theorem 3 that lead to Theorem 2.

Sketch of the Proof for Theorem 2 One major ingredient for the proof of power law decay of the truncated connectivity function in the case $k = 2$ to hold in a wider range of parameters $\mathbf{p}$ is Lemma 1 that has a two-dimensional appealing. It is not clear to us how to obtain an analogous counterpart for $k \geq 3$. Moreover, our renormalization arguments that relies on this result provides a suitable N for which the process of good boxes dominates stochastically a supercritical Bernoulli percolation requiring only that the parameters p_{I_j} stay above $p_c(\mathbb{Z}^2)$ (see Lemma 9).

In this line of reasoning, our first modification consists in fixing $N = 1$ and redefining the notion of *good boxes* for boxes of type $B(y; N)$. For $N = 1$, $B(y; 1)$ consists of a single point, i.e., $B(y; 1) = \{y\}$. We say that $B(y; 1)$ is good if y is ω_I-open for all $I \in \mathscr{I}(k; n)$ for which $I \cap \{1\} \neq \varnothing$.

With this notion of *good boxes*, a path of good boxes (as appearing in the statement of Lemma 7) is simply a path of sites that are ω_I-open for all $I \in \mathscr{I}$ that contains 1 as an element, i.e., Lemma 1 holds trivially. Furthermore, Lemma 8 holds with $N = 1$. In particular, if for all $I \in \mathscr{I}(k; n)$ for which $I \cap \{1\} = \{1\}$ we choose all parameters $p_I < 1$ to be close enough to 1, then stochastic domination as in Lemma 9 holds with $1 - \epsilon > p_c(\mathbb{Z}^2)$.

The rest of the proof follows by fixing $N = 1$ and repeating the proof of Theorem 3 with some minor adaptations. □

Appendix

In this appendix we present a proof of Lemma 4 that relies on elementary linear algebra arguments.

Proof of Lemma 4

$$y_j = x_j + x_{n-1} + j x_n, \text{ for each } 1 \leq j \leq n-2.$$

Let $v_1, v_2 \in \mathbb{R}^n$ be the vectors:

$$v_1 = (-1, \ldots, -1, 1, 0),$$

$$v_2 = (-1, -2, -3, \ldots, -(n-2), 0, 1).$$

Let $U : \mathbb{R}^2 \to \mathbb{R}^n$ be the linear application $U(x, y) = xv_1 + yv_2$, and denote

$$\operatorname{Ker} L := \{v \in \mathbb{R}^n : Lv = 0\},$$
$$\operatorname{Ran} U = \{Uu : u \in \mathbb{R}^2\}.$$

We claim that $\operatorname{Ker} L = \operatorname{Ran} U$, and that $\pi_I \circ U$ is injective for all $I \subset [n]$ with $\#I \geq 2$.

In fact, from the definition of L we obtain that $v = (x_1, \dots, x_n) \in \operatorname{Ker} L$ if and only if $Lv = (y_1, \dots, y_{n-2})$ satisfies $y_j = 0$ for all $1 \leq j \leq n-2$, and this equality holds if and only if $x_j = -(x_{n-1} + jx_n)$. In particular, $v \in \operatorname{Ker} L$ if and only if

$$v = x_{n-1}v_1 + x_n v_2.$$

This shows that $\operatorname{Ker} L = \operatorname{Ran} U$.

To prove the second statement of our claim, we begin by observing the fact that: *For each $I \subset [n]$ with $\#I = 2$ the linear application $\pi_I \circ U : \mathbb{R}^2 \to \mathbb{R}^2$ is injective.*

To see that this is true, let $I = \{i, j\}$, with $1 \leq i < j \leq n$. Then the possibilities for the matrix $\pi_I \circ U$ are:

$$(a) \begin{pmatrix} -1 & -i \\ -1 & -j \end{pmatrix}, \ (b) \begin{pmatrix} -1 & -i \\ 1 & 0 \end{pmatrix}, \ (c) \begin{pmatrix} -1 & -i \\ 0 & 1 \end{pmatrix}, \ \text{and } (d) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

where (a) corresponds to $j \leq n-2$; (b) corresponds to $j = n-1$; (c) corresponds to $i \leq n-2$ and $j = n$; (d) corresponds to $i = n-1$. In any case we have $|\det \pi_I \circ U| \geq 1$ which implies injectivity.

It follows now that for all $I \subset [n]$ with $\#I \geq 2$ the linear application $\pi_I \circ U : \mathbb{R}^2 \to \mathbb{R}^{\#I}$ is injective. In fact, if for $u, v \in \mathbb{R}^2$ we have $\pi_I \circ Uu = \pi_I \circ Uv$, then in particular, for each $J \subset I$ with $\#J = 2$ we have $\pi_J \circ Uu = \pi_J \circ Uv$ which implies $u = v$.

Since v_1 and v_2 are linearly independent (over $\mathbb{R}$), by the Gram–Schmidt process there exists $\tilde{w_1}, \tilde{w_2} \in \mathbb{R}^n$ orthogonal and such that their linear span is equal to the linear span of v_1 and v_2:

$$\begin{aligned} \tilde{w_1} &= v_1, \\ \tilde{w_2} &= v_2 - \frac{\langle v_1, v_2 \rangle}{\langle v_1, v_1 \rangle} v_1 \\ &= v_2 + \frac{(2-n)}{2} v_1. \end{aligned}$$

In particular, $w_1 = 2\|\tilde{w_2}\|\tilde{w_1}$ and $w_2 = 2\|\tilde{w_1}\|\tilde{w_2}$ are orthogonal, $\|w_1\| = \|w_2\|$ and both belong to $\mathbb{Z}^n$. Hence the linear application $A : \mathbb{Z}^2 \to \mathbb{R}^n$ given by

$$\begin{aligned} A(x, y) =& x w_1 + y w_2 \\ =& U(2x\|\tilde{w_2}\| + y(2-n)\|\tilde{w_1}\|, 2y\|\tilde{w_1}\|) \end{aligned}$$

is such that $\operatorname{Ran} A \subset \operatorname{Ker} L \cap \mathbb{Z}^n$. Let $H : \mathbb{Z}^2 \to \mathbb{Z}^2$ be the linear application

$$H = \begin{pmatrix} 2\|\tilde{w}_2\| & (2-n)\|\tilde{w}_1\| \\ 0 & 2\|\tilde{w}_1\| \end{pmatrix}.$$

Then H is injective and $A = U \circ H$. Hence for each $I \subset [n]$ with $\#I \geq 2$, the linear application $\pi_I \circ A = \pi_I \circ U \circ H$ is injective, since H and $\pi_I \circ U$ are injective. Defining

$$c = \min_{I \subset [n]} \inf_{\substack{x \in \mathbb{R}^2 \\ \|x\|=1}} \|\pi_I \circ Ax\| > 0$$

completes the proof. □

Acknowledgments The three first authors M.A., M.H. and B.N.B.L. were former PhD students of the fourth author Vladas Sidoravicius. An advanced draft of this paper was produced before Vladas passed away in May 2019. M.A., M.H. and B.N.B.L. would like to thank Vladas for his great intuition, ideas and enthusiasm that played a fundamental role in our academic trajectories and in obtaining the results in this paper. The research of M.H. was partially supported by CNPq grants 'Projeto Universal' (406659/2016-8) and 'Produtividade em Pesquisa' (307880/2017-6) and by FAPEMIG grant 'Projeto Universal' (APQ-02971-17). The research of B.N.B.L. was supported in part by CNPq grant 305811/2018-5 and FAPERJ (Pronex E-26/010.001269/2016). We are also very grateful to Alberto Sarmiento for suggesting a strategy for proving Lemma 4 and to the anonymous referee for pointing out several corrections and for providing valuable suggestions.

References

1. Aizenman, M., Barsky, D.J.: Sharpness of the phase transition in percolation models. Commun. Math. Phys. **108**, 489–526 (1987)
2. Broadbent, R.S., Hammersley, J.M.: Percolation processes: I. Crystals and mazes. Math. Proc. Camb. Philos. Soc. **53**, 629–641 (1957)
3. Duminil-Copin, H., Tassion, V.: A new proof of the sharpness of the phase transition for Bernoulli percolation and the Ising model. Commun. Math. Phys. **343**, 725–745 (2016)
4. Gács, P.: The clairvoyant demon has a hard task. Comb. Probab. Comput. **9**, 421–424 (2000)
5. Grassberger, P.: Universality and asymptotic scaling in drilling percolation. Phys. Rev. E. **95**, 010103 (2017)
6. Grassberger, P., Hilário, M.R., Sidoravicius, V.: Percolation in media with columnar disorder. J. Stat. Phys. **168**, 731–745 (2017)
7. Hilário, M., Sidoravicius, V.: Bernoulli line percolation. Stoch. Proc. Appl. **129**, 5037–5072 (2019)
8. Kantor, Y.: Three-dimensional percolation with removed lines of sites. Phys. Rev. B. **33**, 3522–3525 (1986)
9. Liggett, T.M., Schonmann, R.H., Stacey, A.M.: Domination by product measures. Ann. Probab. **25**, 71–95 (1997)
10. Menshikov, M.V.: Coincidence of critical-points in percolation problems. Soviet Math. Dokl. **25**, 856–859 (1986)

11. Pete, G.: Corner percolation on $\mathbb{Z}^2$ and the square root of 17. Ann. Probab. **36**, 1711–1747 (2008)
12. Schrenk, K.J., Hilário, M.R., Sidoravicius,V., Araújo, N.A.M., Herrmann, H.J., Thielmann, M., Teixeira, A.: Critical fragmentation properties of random drilling: how many holes need to be drilled to collapse a wooden cube? Phys. Rev. Lett. **116**, 055701 (2016)
13. Winkler, P.: Dependent percolation and colliding random walks. Random Struct. Algorithm **16**, 58–84 (2000)

Time Correlation Exponents in Last Passage Percolation

Riddhipratim Basu and Shirshendu Ganguly

In memory of Vladas Sidoravicius

Abstract For directed last passage percolation on $\mathbb{Z}^2$ with exponential passage times on the vertices, let T_n denote the last passage time from $(0, 0)$ to (n, n). We consider asymptotic two point correlation functions of the sequence T_n. In particular we consider $\mathrm{Corr}(T_n, T_r)$ for $r \leq n$ where $r, n \to \infty$ with $r \ll n$ or $n - r \ll n$. Establishing a conjecture from Ferrari and Spohn (SIGMA 12:074, 2016), we show that in the former case $\mathrm{Corr}(T_n, T_r) = \Theta((\frac{r}{n})^{1/3})$ whereas in the latter case $1 - \mathrm{Corr}(T_n, T_r) = \Theta((\frac{n-r}{n})^{2/3})$. The argument revolves around finer understanding of polymer geometry and is expected to go through for a larger class of integrable models of last passage percolation. As a by-product of the proof, we also get quantitative estimates for locally Brownian nature of pre-limits of Airy$_2$ process coming from exponential LPP, a result of independent interest.

Keywords Last passage percolation · Time correlation · KPZ universality class

R. Basu (✉)
International Centre for Theoretical Sciences, Tata Institute of Fundamental Research, Bangalore, India
e-mail: rbasu@icts.res.in

S. Ganguly
Department of Statistics, UC Berkeley, Berkeley, CA, USA
e-mail: sganguly@berkeley.edu

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_5

1 Introduction and Statement of Results

We consider directed last passage percolation on $\mathbb{Z}^2$ with i.i.d. exponential weights on the vertices. We have a random field

$$\omega = \{\omega_v : v \in \mathbb{Z}^2\}$$

where ω_v are i.i.d. standard Exponential variables. For any two points u and v with $u \preceq v$ in the usual partial order, we shall denote by $T_{u,v}$ the last passage time from u to v; i.e., the maximum weight among all weights of all directed paths from u to v (the weight of a path is the sum of the field along the path). By $\Gamma_{u,v}$, we shall denote the almost surely unique path that attains the maximum- this will be called a polymer or a geodesic. This is one of the canonical examples of an integrable model in the so-called KPZ universality class [7, 18], and has been extensively studied also due to its connection to Totally Asymmetric Simple Exclusion process on $\mathbb{Z}$. For notational convenience let us denote (r, r) for any $r \in \mathbb{Z}$ by $\mathbf{r}$ and $T_{\mathbf{0},\mathbf{n}}$ by T_n and similarly $\Gamma_{\mathbf{0},\mathbf{n}}$ by Γ_n. It is well known [18] that $n^{-1/3}(T_n - 4n)$ has a distributional limit (a scalar multiple of the GUE Tracy–Widom distribution), and further it has uniform (in n) exponential tail estimates [1, 24]. Although the scaled and centered field obtained from $\{T_{\mathbf{0},(x,y)}\}_{x+y=2n}$ using the KPZ scaling factors of $n^{2/3}$ in space and $n^{1/3}$ in polymer weight has been intensively studied and the scaling limit as $n \to \infty$ identified to be the Airy$_2$ process (minus a parabola), much less is known about the evolution of the random field in time i.e., across various values of n. However very recently there has been some attempts to understand the latter, see [2, 20–23, 25, 26] for some recent progress.

In this paper, we study two point functions describing the 'aging' properties of the above evolution. More precisely we investigate the correlation structure of the tight sequence of random variables $n^{-1/3}(T_n - 4n)$ across n. In particular, let us define for $r \leq n \in \mathbb{N}$

$$\rho(n, r) =: \mathrm{Corr}(T_n, T_r).$$

We are interested in the dependence of $\rho(n, r)$ on n and r as they become large. Observe that the FKG inequality implies that $\rho(n, r) \geq 0$. Heuristically, one would expect that $\rho(n, r)$ is close to 1 and 0 for $|n - r| \ll n$ and $r \ll n$ respectively.

Our main result in this paper establishes the exponents governing the rate of correlation decay and thus identifies up to constants the asymptotics of ρ in these regimes establishing a prediction from [16]. Namely we show

$$\rho(n, r) = \Theta((\frac{r}{n})^{1/3}) \text{ if } 1 \ll r \ll n \text{ and } \rho(n, r) = 1 - \Theta((\frac{n-r}{n})^{2/3}) \text{ if } 1 \ll n - r \ll n.$$

It turns out that the upper bound in the former case is similar to the lower bound in the latter case, and the lower bound in the former case is similar to to the upper bound in the latter case. We club these statements in the following two theorems.

Theorem 1 *There exists $r_0 \in \mathbb{N}$ and positive absolute constants δ_1, C_1, C_2 such that the following hold.*

(i) For $r_0 < r < \delta_1 n$ and for all n sufficiently large we have

$$\rho(n,r) \le C_1 \left(\frac{r}{n}\right)^{1/3}.$$

(ii) For $r_0 < n - r < \delta_1 n$ and for all n sufficiently large we have

$$1 - \rho(n,r) \le C_2 \left(\frac{n-r}{n}\right)^{2/3}.$$

Theorem 2 *There exists $r_0 \in \mathbb{N}$ and positive absolute constants δ_1, C_3, C_4 such that the following hold.*

(i) For $r_0 < r < \delta_1 n$ and for all n sufficiently large we have

$$\rho(n,r) \ge C_3 \left(\frac{r}{n}\right)^{1/3}.$$

(ii) For $r_0 < n - r < \delta_1 n$ and for all n sufficiently large we have

$$1 - \rho(n,r) \ge C_4 \left(\frac{n-r}{n}\right)^{2/3}.$$

1.1 Note on the History of This Problem and This Paper

These exponents were conjectured in [16] using partly rigorous analysis, and as far as we are aware was first rigorously obtained in an unpublished work of Corwin and Hammond [10] in the context of Airy line ensemble using the Brownian Gibbs property of the same established in [11]. A related work studying time correlation for KPZ equation has since appeared [12]. Days before posting the first version of this paper on arXiv in July 2018, we came across [15] which considers the same problem. Working with rescaled last passage percolation [15] analyzes the limiting quantity $r(\tau) := \lim_{n\to\infty} \mathrm{Corr}(T_n, T_{\tau n})$. They establish the existence of the limit and consider the $\tau \to 0$ and $\tau \to 1$ asymptotics establishing the same exponents as in Theorems 1 and 2. The approach in [15] uses comparison with stationary LPP using exit points [9] together with using weak convergence to Airy process leading to natural variational formulas. In the limiting regime they get a sharper estimate obtaining an explicit expression of the first order term, providing rigorous proofs of some of the conjectures in [16]. In contrast, our approach hinges on using the moderate deviation estimates for point-to-point last passage time to understand local fluctuations in polymer geometry following the approach taken in [3, 6] leading to results for finite n, also allowing us to analyze situations $r \ll n$ or $n-r \ll n$, which can't be read off from weak convergence. Our work is completely independent of [15].

The ideas in this paper has since been further developed in a joint work with Lingfu Zhang [5] to treat the case of flat initial data (i.e., line-to-point last passage percolation) in the $\tau \to 0$ limit, and the remaining conjectured exponent from [16] has been established there. As it turns out, the upper bound in the case of flat initial data requires rather different arguments, but the lower bound in [5] further develops the same line of arguments as in the original version of this paper, and improves upon some of the estimates proved there. As such, in this version, we have decided to omit some of the details of the proof of Theorem 2, and we refer to the relevant steps in [5] instead. We expect this class of ideas and estimates to be crucial in further enhancing our understanding of temporal correlations in the KPZ universality class with more general initial conditions.

1.2 Local Fluctuations of the Weight Profile

In the process of proving Theorems 1 and 2 we prove a certain auxiliary result of independent interest. Namely, we establish a local regularity property of the pre-limiting profile of Airy$_2$ process obtained from the exponential LPP model. We need to introduce some notations before making the formal statement. For $n \in \mathbb{N}$, $s \in \mathbb{Z}$ with $|s| < n$ we define

$$L_{n,s} := T_{\mathbf{0},(n+s,n-s)}.$$

It is known [7] that

$$\mathscr{L}_n(x) := 2^{-4/3} n^{-1/3} (L_{n,x(2n)^{2/3}} - 4n)$$

converges in the sense of finite dimensional distributions to the $\mathscr{A}_2(x) - x^2$ where $\mathscr{A}_2(\cdot)$ denotes the stationary Airy$_2$ process (tightness, and hence weak convergence is also known, see e.g. [14]). It is known that the latter locally looks like Brownian motion [17, 27] and hence one would expect that $\mathscr{L}_n(x) - \mathscr{L}_n(0)$ will have a fluctuation of order $x^{1/2}$ for small x. We prove a quantitative version of the same at all shorter scales.

Theorem 3 *There exist constants $s_0 > 0, z_0 > 0$ and $C, c > 0$ such that the following holds for all $s > s_0$, $z > z_0$ and for all $n > Cs^{3/2}$:*

$$\mathbb{P}\left(\sup_{s':|s'|<s} L_{n,s'} - L_{n,0} \geq zs^{1/2}\right) \leq e^{-cz^{4/9}}.$$

Such an estimate was first obtained in [17] for Brownian last passage percolation using the Brownian Gibbs resampling property of the pre-limiting line ensemble in that model (a more refined version appears in a very recent work [8]). Observe that one would expect the Gaussian exponent z^2 in the upper bound of the probability in the statement of the theorem and that is what is obtained in [17]. We, on the other hand, use a cruder argument to obtain only a stretched exponential decay. However, we have not tried to optimize the exponent 4/9 which is not even sharp for our arguments.

1.3 Key Ideas and Organization of the Paper

Before jumping in to proofs, we present the key reasons driving the exponents and the main ingredients of the proofs of Theorems 1 and 2. Since the reasons governing the behaviour of $\rho(n,r)$ when $r \ll n$ and $\rho(n,r)$ when $n-r \ll n$ are almost symmetric, in this section we will mostly discuss the former case for the sake of brevity. The key realization driving the argument is that Γ_n should overlap significantly with Γ_r up to the region $\{x+y \le 2r\}$. Then at a very high level one can speculate that $\mathrm{Cov}(T_r, T_n)$ should be of the order of the variance of the amount of overlap, which because of the previous sentence should be of the same order as $\mathrm{Var}(T_r) = O(r^{2/3})$ (using the well known sharp estimates of the variance). All of this points to a correlation of the order of $(\frac{r}{n})^{1/3}$.

We now mention a few key ingredients used to make the above heuristic rigorous. The upper bound is relatively straightforward. For convenience, as we shall do throughout the paper, let us denote T_r by X and let $T_n = Z + W$ where Z is the weight of the first part of the polymer Γ_n i.e. the part from $\mathbf{0}$ to the line $x+y=2r$ and W is the weight of the path from $x+y=2r$ to $\mathbf{n}$. See Fig. 1. Let $v=(r+s, r-s)$

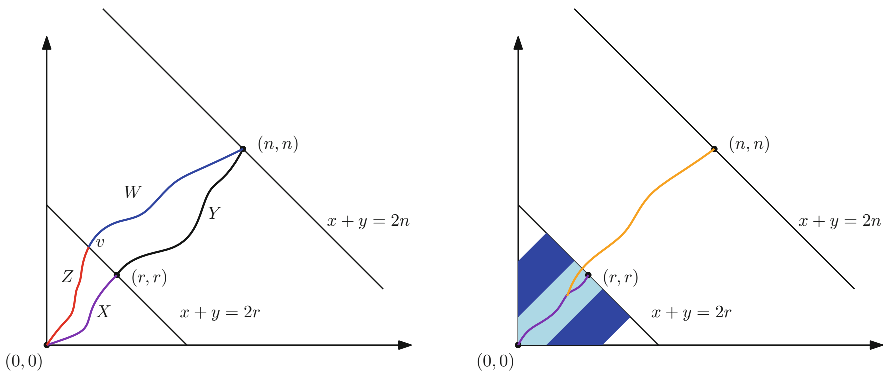


Fig. 1 The figure illustrates the polymers of interest, Γ_r with weight X, $\Gamma_{\mathbf{r},\mathbf{n}}$, with weight Y, Γ_n comprised of $\Gamma_{\mathbf{0},v}$ and $\Gamma_{v,\mathbf{n}}$ with weights Z and W respectively where v is the point of intersection of Γ_n with the line $x+y=2r$. The second figure illustrates our strategy to create barriers (deep blue) around a narrow strip (light blue) to ensure that Γ_r and $\Gamma_{\mathbf{0},v}$ stay localized inside the latter and hence overlaps significantly creating a situation where the covariance between X and $Z+W$ is approximated by the variance of the former

be the vertex at which Γ_n intersects the line $x + y = 2r$. It is well known since the work of Johansson [19] that if r is say $n/2$ then $|s| = O(r^{2/3})$. However the polymer is in some sense self similar and hence one expects that the above result should also hold even when at scales $r \ll n$. Indeed a quantitative version of such a result was established in [6]. This tells us that $|X - Z| = O(r^{1/3})$ by standard results about polymer fluctuations at scale $r^{2/3}$ around the point $\mathbf{r}$. Moreover relying on this we also prove the local Brownian-like square root fluctuations of the distance profile $T_{w,\mathbf{n}}$ as w varies over vertices of the form $(r+s, r-s)$ when $|s| = O(r^{2/3})$ showing that $|W - Y| = O((r^{2/3})^{1/2}) = O(r^{1/3})$ where Y is $T_{\mathbf{r},\mathbf{n}}$ (hence is independent of X). Given the above information, the upper bound, i.e., Theorem 1 is a simple consequence of Cauchy–Schwarz inequality.

However, the lower bound is significantly more delicate since one has to rule out cancellations to show that indeed the heuristic mentioned at the beginning of the section is correct. At a very high level the strategy is to condition on a large part of the noise space in a way which allows us to control cancellations and prove the desired lower bound on $\rho(n, r)$. To do this the first thing to come to our aid is the FKG inequality. If with positive probability β (independent of r, n) the conditioned environment is such that $\rho(n, r) \geq \Theta(\frac{r}{n})^{1/3}$, then since $\rho(n, r) \geq 0$ pointwise on the conditioned environment (using the FKG inequality), averaging over the latter yields the lower bound $\rho(n, r) \geq \beta\Theta(\frac{r}{n})^{1/3}$. Our strategy of choosing the part of the environment to condition on consists of ensuring that, with positive probability, the polymer Γ_r is localized i.e., it is confined to a thin cylinder R_θ of size $r \times \theta r^{2/3}$ for some small θ and ensuring Γ_n essentially agrees with Γ_r up to the line $x + y = 2r$. This is obtained by creating a bad region (barrier) around the thin cylinder making it suboptimal for the polymer to venture out of R_θ. This then implies that under such a conditioning, up to certain correction terms $\mathrm{Cov}(T_r, T_n)$ is equal to $\mathrm{Var}(T_r)$. At this point we prove a sharp estimate on variance of polymer weights constrained to lie in R_θ showing that it scales like $\theta^{-1/2} r^{2/3}$ as θ goes to 0. Thus for θ small enough, the variance term is large enough and dominates all the correction terms yielding the sought lower bound of $\Theta(r^{2/3})$ on the covariance and hence Theorem 2.

We now briefly describe how to use the exact same strategy to bound $\rho(r, n)$ in the regime $n - r \ll n$. We will discuss the more delicate Theorem 2. Note that in this case we are aiming to prove a lower bound on $1 - \rho(n, r)$ and hence an upper bound on $\rho(n, r)$. Thus the natural strategy to adopt would be to show that even after conditioning on T_r, T_n is not completely determined and there is still some fluctuation left. In fact, as expected, our arguments will show that the latter is of the same order as the fluctuation of $T_{\mathbf{r},\mathbf{n}}$ i.e., $\mathrm{Var}(T_n | T_r) = \Theta((n - r)^{2/3})$ on a positive measure part of the space. Thus we get

$$\Theta((n - r)^{2/3}) \leq \inf_{\lambda} \mathrm{Var}(T_n - \lambda T_r) = (1 - \mathrm{Corr}^2(T_r, T_n))\mathrm{Var}(T_n).$$

This, along with the fact that $\mathrm{Var}(T_n) = \Theta(n^{2/3})$, completes the proof.

It is worth emphasizing that while we do crucially make use of the integrability of the exponential LPP model, it is done only in a rather limited nature via the input of weak convergence to Tracy–Widom distribution [18] and the moderate deviation estimates coming from [1, 24]. Therefore we expect our methods to be applicable to a large class of integrable LPP models where such estimates are known. In particular, we do not use any information about the limiting Airy process. As already mentioned, our approach hinges on the fine understanding of the local polymer geometry, following the sequence of recent works [3, 4, 6]. By virtue of being geometric, our proof is also robust, and as already mentioned, similar ideas have already been used in [5] to treat the case of flat initial condition, which does not yet seem accessible by any other method. We extensively draw from some of the estimates derived in those previous works, while introducing some new elements to advance the understanding of polymer geometry. Crucial ingredients include moderate deviations estimates to establish concentration for passage times across parallelograms. This idea originated in [3] and the particular estimates required for the time correlation problems are gathered in [5, Section 4]; we shall be extensively quoting from that source.

1.4 Organization of the Paper

The rest of the paper is organized as follows. We first prove Theorem 3 in Sect. 2. Then we use Theorem 3 to prove Theorem 1 in Sect. 3. Proof of Theorem 2 is done in Sect. 4.

2 Local Fluctuations of Weight Profile: Proof of Theorem 3

As alluded to before, in the case $s = \Theta(n^{2/3})$, one can read off a qualitative version of this result from the limiting Airy process which ceases to provide any relevant information when $s \ll n^{2/3}$. Although it is known that Brownian motion arises as a week limit at some shorter scale [27], we need some finer estimates for finite n. Such a result was indeed achieved in [17] in the special case of Brownian LPP crucially relying on the Brownian Gibbs property of the pre-limiting line ensemble. We shall take a more robust, geometric approach which hinges on establishing that the profile $\{L_{n,s'} - L_{n,0} : |s'| < s\}$ is with high probability determined by the vertex weights in the region

$$\{(x, y) : 2n - C_* s^{3/2} \leq (x + y) \leq 2n\}$$

for some large constant C_*. To this end we have the following proposition.

Proposition 1 *In the set-up of Theorem 3, consider $\Gamma = \Gamma_{\mathbf{0},(n+s',n-s')}$. Let $t \geq 1$ and let $v = (v_1(s', t, s), v_2(s', t, s))$ denote the point at with Γ intersects the anti-diagonal $x + y = 2n - 2t^* s^{3/2}$. There exists $s_0 > 0$, $y_0 > 0$ and $c, C > 0$ such that*

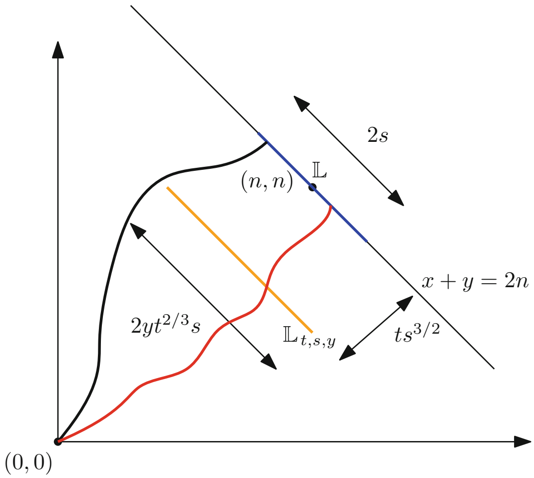


Fig. 2 The figure illustrates the setting of Proposition 1 where the geodesic from $\mathbf{0}$ to any point in an interval $\mathbb{L}$ of length $2s$ centered at $\mathbf{n}$ on the line $x + y = 2n$ is unlikely to intersect the line $x + y = 2(n - ts^{3/2})$ outside the line segment $\mathbb{L}_{t,s,y}$ of length $2yt^{2/3}s$ centered at $(n - ts^{3/2}, n - ts^{3/2})$. Thus it is unlikely that the black path would be a geodesic

the following holds for all $s > s_0$, $t \geq 1$, $y > y_0$ and for all $n > Cr^{3/2}$:

$$\mathbb{P}\left(\sup_{|s'|<s} |v_1(s', t, s) - (n - t^* s^{3/2})| \geq yt^{2/3}s\right) \leq e^{-cy^2}$$

where $t^* = \min\{t, \frac{n}{s^{3/2}}\}$.

See Fig. 2 for an illustration of the setting in Proposition 1 and the associated event.

Proof Clearly, the case $t^* \neq t$ is trivial by the directed-ness of the geodesic. For the other case, observe first that by polymer ordering, it suffices to prove the result for $s' = \pm r$. This case can be read off from the proof of Theorem 3 in [6] (see also Remark 1.3 there about the non-optimality of the exponent).

As in Fig. 2, let $\mathbb{L}$ denote the line segment joining $(n+s, n-s)$ and $(n-s, n+s)$ and $\mathbb{L}_{t,s,y}$ denote the line segment joining $(n - t^*s^{3/2} - yt^{2/3}s, n - t^*s^{3/2} + yt^{2/3}s)$ and $(n - t^*s^{3/2} + yt^{2/3}s, n - t^*s^{3/2} - yt^{2/3}s)$. Clearly on the large probability event (for large y) implied by Proposition 1, the profile:

$$\{L_{n,s'} - L_{n,0} : |s'| < s\}$$

can be upper bounded by using the passage times $T_{u,v}$ where $u \in \mathbb{L}_{t,s,y}$ and $v \in \mathbb{L}$. The next proposition states a concentration result for these passage times around their expectations.

Proposition 2 *Let $\delta \in (0, \frac{1}{3})$ be fixed. Set $\mathbb{L}' := \mathbb{L}_{t,s,t^\delta}$. Then there exists $s_0, y_0 > 0$ and $c > 0$ such that for all $s > s_0$, $y > y_0$ and $t \geq 1$ we have*

$$\mathbb{P}\left(\sup_{u\in\mathbb{L}', v\in\mathbb{L}} |T_{u,v} - \mathbb{E}T_{u,v}| \geq yt^{1/3+\delta/2}s^{1/2}\right) \leq e^{-cy}.$$

The same bound holds for $u = \mathbf{0}$ if $t \neq t^*$.

Proof This follows from [5, Theorem 4.2] by observing that the slope between any two pair of points in $\mathbb{L}$ and $\mathbb{L}'$ remain between $1/2$ and 2 (here we use the fact that $\delta < \frac{1}{3}$, and s is sufficiently large).

We can now complete the proof of Theorem 3. The basic strategy is the following. To bound $L_{n,s'} - L_{n,0}$ we back up a little bit and look at where the geodesic $\Gamma_{\mathbf{0},(n+s',n-s')}$ intersects the line $\mathbb{L}'$ from Proposition 2 for an appropriate choice of the parameters. Calling that u_* the proof proceeds by bounding $|T_{u_*,v} - T_{u_*,\mathbf{n}}|$ and using the simple observation that $L_{n,0} \geq T_{\mathbf{0},u_*} + T_{u_*,\mathbf{n}}$.

Proof of Theorem 3 Let $z > 0$ sufficiently large be fixed. Let $t = z^{4/3}$ and let $\mathscr{A}$ denote the event that

$$\left\{ \sup_{s':|s'|<s} |v_1(s',t,s) - (n - t^* s^{3/2})| \geq z^{10/9} s \right\}$$

Use Proposition 1 with the above value of t and $y = z^{2/9}$ to conclude that $\mathbb{P}(\mathscr{A}) \leq e^{-cz^{4/9}}$ (as z is sufficiently large). We shall now consider two cases separately: (i) $t = t^*$ and (ii) $t \neq t^*$.

In case (i), let $\mathbb{L}'$ be defined as in Proposition 2 with $\delta = 1/6$ (any arbitrary choice for $\delta \in (0, 1/3)$ would work, but would give a different tail exponent) and the choice of t as before. Let $\mathscr{B}$ denote the event that

$$\left\{ \sup_{u \in \mathbb{L}', v, v' \in \mathbb{L}} |T_{u,v} - T_{u,v'}| \geq z s^{1/2} \right\}.$$

Observe now that, for $u = (u_1, u_2)$ and $v = (v_1, v_2)$ as above, we have, by Basu et al. [5, Theorem 4.1], that

$$|\mathbb{E} T_{u,v} - (\sqrt{v_1 - u_1} + \sqrt{v_2 - u_2})^2| \leq C t^{1/3} s^{1/2}$$

for some $C > 0$. By a Taylor expansion, it follows that, for u, v, v' as above we have for some C'

$$\sup_{u \in \mathbb{L}', v, v' \in \mathbb{L}} |\mathbb{E} T_{u,v} - \mathbb{E} T_{u,v'}| \leq C' t^{1/3} s^{1/2} \leq \frac{z}{2} s^{1/2}$$

where the final inequality follows from our choice of t and z sufficiently large. Using Proposition 2 with the choices above, $y = z^{4/9}$, we get from the above that for z sufficiently large, we have $\mathbb{P}(\mathscr{B}) \leq e^{-cz^{4/9}}$. It remains to prove that on $\mathscr{A}^c \cap \mathscr{B}^c$ we have

$$\left\{ \sup_{s':|s'|<s} L_{n,s'} - L_{n,0} \leq z s^{1/2} \right\}.$$

To see this, let s_* with $|s_*| \le s$ be such that

$$L_{n,s_*} = \sup_{s':|s'|<s} L_{n,s'}.$$

Let $v := (n+s_*, n-s_*)$. On $\mathscr{A}^c$, the geodesic $\Gamma_{\mathbf{0},(n+s_*,n-s_*)}$ intersects the line segment $\mathbb{L}'$, let u_* be the intersection point. On $\mathscr{B}^c$ we have $|T_{u_*,v} - T_{u_*,\mathbf{n}}| \le zs^{1/2}$. The claim is established by observing that $L_{n,0} \ge T_{\mathbf{0},u_*} + T_{u_*,\mathbf{n}}$.[1]

In case (ii), we proceed as before but now notice that $u_* = \mathbf{0}$. The same argument now can be repeated with $\mathscr{B}'$ defined as

$$\left\{ \sup_{v,v'\in\mathbb{L}} |T_{\mathbf{0},v} - T_{\mathbf{0},v'}| \ge zs^{1/2} \right\}.$$

Observe now that since $t \ne t^*$, we must have $n \le ts^{3/2}$. Using [5, Theorem 4.1] as before it follows that for some $C > 0$

$$\sup_{v,v'\in\mathbb{L}} |\mathbb{E}T_{\mathbf{0},v} - \mathbb{E}T_{\mathbf{0},v'}| \le Cn^{1/3} \le Ct^{1/3}s^{1/2} \le \frac{z}{2}s^{1/2}$$

where the last inequality follows as before by taking z sufficiently large and our choice of t. Using the above and Proposition 2 as before we show that $\mathbb{P}(\mathscr{B}') \le e^{-cz^{4/9}}$. The rest of the proof is identical with the previous case.

3 Proof of Upper Bounds

In this section we shall prove Theorem 1 using Theorem 3 and Proposition 1. As we shall see, the proofs of parts (i) and (ii) rely on much of the same ingredients. Before proceeding further let us introduce some notation that will be used throughout this section.

Before diving in to the proofs we adopt the convention of ignoring the values of the vertices $\{\omega_{(x,y)} : x + y = 2r\}$. This would enable us to write cleaner equations of the form $T_{\mathbf{n}} = T_{\mathbf{0},v} + T_{v,\mathbf{n}}$ where v is the unique vertex $\Gamma_n \cap \{x + y = 2r\}$. However since by definition, the random v can be one of $2r$ possible vertices, whose maximum value is no more than $\log r$ with an exponential tail, it does not create any change in the computations throughout the paper since all the objects that we deal with, have fluctuations of the order of $r^{1/3}$. We shall adopt this convention throughout the remainder of this paper, and not comment further on this topic. It will be easy to verify the minor details in each case, and we leave that to the reader.

For any path γ, we shall denote by $\ell(\gamma)$, the weight of the path. Let $\Gamma := \Gamma_n$ denote the polymer from $\mathbf{0}$ to $\mathbf{n}$. Let $v = (v_1, v_2)$ denote the point at which Γ

[1] We shall ignore the contribution of the vertex u_*, one can check that this does not change any of the asymptotics.

intersects the line $\{x + y = 2r\}$. Recall from Sect. 1 that $T_r := T_{\mathbf{0},\mathbf{r}}$. Let us define (see Fig. 1)

$$X := T_r, \quad Y := T_{\mathbf{r},\mathbf{n}}, \tag{1}$$
$$Z := \ell(\Gamma_{\mathbf{0},v}) \text{ and } W := \ell(\Gamma_{v,\mathbf{n}}).$$

Thus by definition $T_n = Z + W$.[2] Finally we shall denote by X^* the weight of the polymer, denoted by Γ^*, from $\mathbf{0}$ to the line $\{x + y = 2r\}$.

We shall need some preparatory results. First we want to show that $(Z - X)_+$ is tight at scale $r^{1/3}$. Observing that $X^* \geq Z$, this is a consequence of [5, Theorem 4.1] that $r^{-1/3}(X^* - X)$ has stretched exponential tails: for all y large enough and for all r large enough

$$\mathbb{P}(X^* - X \geq yr^{1/3}) \leq e^{-cy^{1/3}}. \tag{2}$$

The next lemma shall show that $W - Y$ is also typically of order $r^{1/3}$. Notice that if $r \ll n$, now we can no-longer replace W by the weight of the line-to-point polymer from the line $\{x + y = 2r\}$ to $\mathbf{n}$. This is where we shall need the full power of Proposition 1 and Theorem 3.

Lemma 1 *There exists positive constants r_0, y_0 and $C, c > 0$ such that for all $r > r_0$ and $y > y_0$ and $n > Cr$ we have*

$$\mathbb{P}(W - Y > yr^{1/3}) \leq e^{-cy^{1/3}}.$$

Proof For $z > 0$, let $\mathscr{A}_z$ denote the event $|v_1 - r| \geq zr^{2/3}$ and $\mathscr{B}_z$ denote the event that

$$\sup_{|s| \leq zr^{2/3}} T_{(r+s,r-s),\mathbf{n}} - T_{\mathbf{r},\mathbf{n}} \geq yr^{1/3}.$$

Clearly for every $z > 0$,

$$\mathbb{P}(W - Y > yr^{1/3}) \leq \mathbb{P}(\mathscr{A}_z) + \mathbb{P}(\mathscr{B}_z).$$

The lemma follows by taking $z = y^{1/6}$ and using Proposition 1 and Theorem 3 to bound $\mathbb{P}(\mathscr{A}_z)$ and $\mathbb{P}(\mathscr{B}_z)$ respectively. Note that in the last application, Theorem 3 is applied for the inverted ensemble i.e., replace $\mathbf{n}$ by $\mathbf{0}$, $\mathbf{0}$ by $\mathbf{n}$ and $\mathbf{r}$ by $\mathbf{n} - \mathbf{r}$.

We can now prove the following proposition which immediately implies Theorem 1, (i) as $\text{Var}\, T_n = \Theta(n^{2/3})$, and $\text{Var}\, T_r = \Theta(r^{2/3})$.

---

[2]This is first of the many situations we ignore the weights on the line $x + y = 2r$, as mentioned above we shall not comment on this issue henceforth.

Proposition 3 *There exists absolute constants r_0, δ_1 and C such that we have for all $r_0 < r < \delta_1 n$ and n sufficiently large*

$$\mathrm{Cov}(T_n, T_r) \le Cr^{2/3}.$$

Proof We need to upper bound

$$\mathrm{Cov}(X, Z + W) = \mathrm{Cov}(X, Z) + \mathrm{Cov}(X, W).$$

we bound the two terms separately. Clearly, by Cauchy–Schwarz inequality and the observation $\mathrm{Var}\, X = \Theta(r^{2/3})$, to prove $\mathrm{Cov}(X, Z) \le Cr^{2/3}$, it suffices to show that $\mathrm{Var}\, Z = O(r^{2/3})$. Now notice that, $\mathrm{Var}(Z) \le 2(\mathrm{Var} X + \mathbb{E}(X - Z)^2)$. Observing $Y - W \le Z - X \le X^* - X$, and using (2) and Lemma 1 it follows that $\mathbb{E}(X - Z)^2 = O(r^{2/3})$ which in turn implies $\mathrm{Cov}(X, Z) \le Cr^{2/3}$ for some absolute constant C.

For the second term in the above decomposition observe that

$$\mathrm{Cov}(X, W) = \mathrm{Cov}(X, W - Y)$$

because X and Y are independent. Using Cauchy–Schwarz inequality again, it suffices to show that $\mathbb{E}(Y - W)^2 = O(r^{2/3})$. Observing as before that $W - Y \ge X - Z \ge X - X^*$, this follows from (2) and Lemma 1. This completes the proof of the proposition.

To prove Theorem 1, (ii) we shall need the following easy observation.

Observation 1 *For any two random variables U and V we have*

$$\mathrm{Var}(U - V) \ge (1 - \mathrm{Corr}^2(U, V))\mathrm{Var}(U).$$

The observation follows from noticing that

$$(1 - \mathrm{Corr}^2(U, V))\mathrm{Var}(U) = \inf_{\lambda \in \mathbb{R}} \mathrm{Var}(U - \lambda V) \le \mathrm{Var}(U - V).$$

Using Observation 1, the following Proposition immediately implies Theorem 1, (ii).

Proposition 4 *There exists $r_0 \in \mathbb{N}$ and positive absolute constants δ_1, C, for all r such that $\delta_1 n > (n - r) > r_0$, and all n sufficiently large we have*

$$\mathrm{Var}\ (T_n - T_r) \le C(n - r)^{2/3}.$$

Proof Recalling X, Y, Z, W as defined at the beginning of this section, we need to upper bound $\mathrm{Var}(Z + W - X)$. Expanding we get that,

$$\mathrm{Var}(Z + W - X) = \mathrm{Var}(Z - X) + \mathrm{Var}(W) + 2\mathrm{Cov}(Z - X, W).$$

We shall show each of the terms above is $O((n-r)^{2/3})$ separately. In fact, by Cauchy–Schwarz inequality it suffices to only show that bound for the first two terms. Notice that the picture is same as before except the roles of r and $(n-r)$ has been reversed. Using the proof of Lemma 1 we can now show that

$$\mathbb{P}(Z - X \geq y(n-r)^{1/3}) \leq e^{-cy^{1/4}}$$

and using (2) again it follows that

$$\mathbb{P}(W - Y \geq y(n-r)^{1/3}) \leq e^{-cy^{1/4}}$$

for all y sufficiently large. As in the proof of Proposition 3, this is then used to argue that $\mathrm{Var}(Z - X) = O((n-r)^{2/3})$, and $\mathbb{E}[W-Y]^2 = O((n-r)^{2/3})$, which together with the observation that $\mathrm{Var}\, Y = O((n-r)^{2/3})$ completes the proof of the proposition.

4 Proof of Lower Bounds

We now move towards proving Theorem 2. As in the proof of Theorem 1, parts (i) and (ii) of Theorem 2 have rather similar proofs as well (after exchanging the roles of r and $n-r$). In this section we describe in detail the line of argument leading to the proof of Theorem 2, (i). We shall first complete the proof modulo the key result Proposition 6. We shall give a sketch of how the same strategy is used to prove Theorem 2, (ii). The final subsection will be dedicated to the proof of Proposition 6.

For the readers' benefit, we recall briefly the strategy outlined in Sect. 1.3. By the FKG inequality, it should suffice to obtain a lower bound on the conditional correlation on an event with probability bounded uniformly below. By the trivial observation $\mathrm{Cov}(X, X+Y) = \Theta(r^{2/3})$, a very natural way to construct such an event is to ask that v is very close to $\mathbf{r}$ which will imply $X \approx Z$ and $Y \approx W$ (using Theorem 3). However one needs to be careful so that there will be enough fluctuation left in the conditional environment. To this end, it turns out one can construct such an event measurable with respect to the configuration outside a thin strip of width $\Theta(r^{2/3})$ around the straightline joining $\mathbf{0}$ to $\mathbf{r}$.

For $\theta > 0$, let $R_\theta \subseteq \mathbb{Z}^2$ be defined as follows:

$$R_\theta := \{(x, y) \in \mathbb{Z}^2 : 0 \leq x + y \leq 2r \text{ and } |x - y| \leq \theta r^{2/3}\}.$$

Let $\omega_\theta = \{\omega_v : v \in [\![0, n]\!]^2 \setminus R_\theta\}$ denote a weight configuration outside R_θ. Let $\mathscr{F}_\theta$ denote the σ-algebra generated by the set of all such configurations Ω_θ. Observe that events measurable with respect to $\mathscr{F}_\theta$ can be written as subsets of Ω_θ, and we shall often adopt this interpretation without explicitly mentioning it. The major step in the proof is the following proposition.

Proposition 5 *There exist absolute positive constants* $\beta, \delta_1, \theta, C > 0$ *sufficiently small such that for* $\delta_1 n > r > r_0$ *and* n *sufficiently large there exists an event* $\mathscr{E}$ *measurable with respect to* $\mathscr{F}_\theta$ *with* $\mathbb{P}(\mathscr{E}) \geq \beta$ *and the following property: for all weight configuration* $\omega \in \mathscr{E}$ *we have*

$$\mathrm{Cov}(T_n, T_r \mid \omega) > Cr^{2/3}.$$

The proof of Theorem 2, (i) using Proposition 5 is straightforward.

Proof of Theorem 2, (i) Observe that for each fixed weight configuration $\omega \subset \Omega_\theta$ on the vertices outside R_θ, both T_n and T_r are increasing in the weight configuration on R_θ. Observe also that $\mathbb{E}[T_n \mid \mathscr{F}_\theta]$ and $\mathbb{E}[T_r \mid \mathscr{F}_\theta]$ are both again increasing in the configuration ω. Applying the FKG inequality twice (in the third and fifth lines of the following computation) together with Proposition 5 (in the third line of the following computation while dealing with the integral over $\mathscr{E}$) then implies

$$\begin{aligned}
\mathbb{E}T_nT_r &= \mathbb{E}\left(\mathbb{E}[T_nT_r \mid \mathscr{F}_\theta]\right)\\
&= \int_{\mathscr{E}} \mathbb{E}[T_nT_r \mid \mathscr{F}_\theta]d\omega + \int_{\mathscr{E}^c} \mathbb{E}[T_nT_r \mid \mathscr{F}_\theta]d\omega\\
&\geq \int_{\mathscr{E}} \mathbb{E}[T_n \mid \mathscr{F}_\theta]\mathbb{E}[T_r \mid \mathscr{F}_\theta]d\omega + C\beta r^{2/3} + \int_{\mathscr{E}^c} \mathbb{E}[T_n \mid \mathscr{F}_\theta]\mathbb{E}[T_r \mid \mathscr{F}_\theta]d\omega\\
&= \mathbb{E}\left(\mathbb{E}[T_n \mid \mathscr{F}_\theta]\mathbb{E}[T_r \mid \mathscr{F}_\theta]\right) + C\beta r^{2/3}\\
&\geq \mathbb{E}[T_n]\mathbb{E}[T_r] + C\beta r^{2/3};
\end{aligned}$$

which is what we set out to prove.

4.1 Constructing a Suitable Environment

The key step in the proof of Proposition 5, is the construction of $\mathscr{E}$, towards which we now move. For easy reference we recall the notations already introduced in Sect. 3, that we will use again.

$$\begin{aligned}
X &:= T_r, \quad Y := T_{\mathbf{r},\mathbf{n}}, Z := \ell(\Gamma_{\mathbf{0},v}) \text{ and } W := \ell(\Gamma_{v,\mathbf{n}}).\\
X^* &:= \max\{\ell(\Gamma_{\mathbf{0},w}) : w \in \mathbb{L}_r\},
\end{aligned}$$

where $\mathbb{L}_r$ denote the line $\{x + y = 2r\}$. We shall also denote by X_* (resp. X_θ) the weight of the best path from $\mathbf{0}$ to $\mathbf{r}$ that does not exit $R_{2\theta}$ (resp. R_θ). Finally for $\phi > \theta$, $\mathbb{L}_{r,\phi}$ shall denote the line segment joining $(r - \phi r^{2/3}, r + \phi r^{2/3})$ and $(r + \phi r^{2/3}, r - \phi r^{2/3})$. We shall denote by X_ϕ the weight of the best path from $\mathbf{0}$ to $\mathbb{L}_{r,\phi}$, and by Y_ϕ the weight of the best path from $\mathbb{L}_{r,\phi}$ to $\mathbf{n}$.

The event $\mathscr{E}$ will depend on a number of parameters ϕ_0, ϕ, L, c_0 (and naturally θ), the choices of which shall be specified later.

The event will consist of two major parts.

1. **Regular fluctuation of the profile** $\{T_{w,\mathbf{n}} : w \in \mathbb{L}_{r,\phi}\}$**:** Let $\mathscr{E}_1$ denote the event that

$$\left\{ \sup_{w \in \mathbb{L}_{r,\phi_0}} T_{w,\mathbf{n}} - T_{\mathbf{r},\mathbf{n}} \leq \phi_0^{1/2} \log^9(\theta^{-1}) r^{1/3} \right\} \cap$$

$$\left\{ \sup_{w \in \mathbb{L}_{r,\phi} \setminus \mathbb{L}_{r,\phi_0}} T_{w,\mathbf{n}} - \sqrt{|w_1 - w_2|} \log^9(\theta^{-1}) \leq T_{\mathbf{r},\mathbf{n}} \right\},$$

where $w = (w_1, w_2)$. Our choice of parameters (see below) would ensure $\phi_0 \ll \phi$. Observe that $\mathscr{E}_1$ only depends on the weight configuration above the line $\mathbb{L}_r$.

2. **Barrier around** R_θ: Let U_1 (resp. U_2) denote a $r \times (\phi - \theta) r^{2/3}$ rectangle whose one set of parallel sides are aligned with the lines $x + y = 0$ and $x + y = 2r$ respectively and whose left (resp. left right) side coincides with the right (resp. left) side of R_θ.[3] For any point $u = (u_1, u_2) \in \mathbb{Z}^2$, let $d(u) := u_1 + u_2$. Also, for any region U, and points $u, v \in U$, let us denote, by $T_{u,v}^U$ to be the weight of the best path from u to v that does not exit U. Let $\mathscr{E}_2$ denote the following event measurable with respect to the configuration in U_1:

$$T_{u,u'}^{U_1} - \mathbb{E} T_{u,u'} \leq -L r^{1/3} \; \forall u, u' \in U_1 \text{ with } |d(u) - d(u')| \geq \frac{r}{L}.$$

Let $\mathscr{E}_3$ denote the same event with U_1 replaced by U_2. We set $\mathscr{E}_4 := \mathscr{E}_2 \cap \mathscr{E}_3$.

4.1.1 Choice of Parameters

We need to fix our choice of parameters appearing in the definitions of the above events before proceeding to proving probability bounds for the same. Throughout the sequel c_0 is a small enough universal constant, we shall choose θ to be an arbitrarily small constant; and $L \gg \phi \gg \phi_0$. We need to choose ϕ_0 poly-logarithmic in θ^{-1}, ϕ a large inverse power of θ, and L a much larger inverse power of θ depending on ϕ. For concreteness we shall fix $\phi_0 = \log^{10}(\frac{1}{\theta}), \phi = (\frac{1}{\theta})^{30}$ and $L = \phi^{30}$. Given all of these we shall take r sufficiently large, and r/n sufficiently small. Throughout the remainder of this paper we shall work with this fixed choice of parameters.

[3] In keeping with the often used practice, left and right are defined after rotating the picture counter-clockwise by 45 degrees, so that the line $x = y$ becomes vertical.

4.2 Construction of $\mathscr{E}$

We are now ready to define the event $\mathscr{E}$. First we define certain nice events conditioned on $\mathscr{E}_4$ towards the proof of Proposition 5.

1. Let $\mathscr{E}_5$ denote the set of all $\omega = \omega_\theta \in \mathscr{E}_1 \cap \mathscr{E}_4$ such that

$$\mathbb{E}[(X^* - X_*)^2 \mid \omega] \leq 10r^{2/3},$$

2. Let $\mathscr{E}_6$ denote the set of all $\omega = \omega_\theta \in \mathscr{E}_4 \cap \mathscr{E}_1$ such that

$$\mathbb{E}[(Z + W - Y - X_*)^2 \mid \omega] \leq 40\phi_0^2 r^{2/3}.$$

3. Let $\mathscr{E}_7$ denote the set of all $\omega \in \mathscr{E}_4 \cap \mathscr{E}_1$ such that

$$\text{Var}\ (X_* \mid \omega) \geq c_0\theta^{-1/2}r^{2/3}$$

   where c_0 is a sufficiently small constant to be chosen appropriately later (independent of θ) and θ will be chosen sufficiently small.

   We shall set

$$\mathscr{E} := \mathscr{E}_5 \cap \mathscr{E}_6 \cap \mathscr{E}_7. \tag{3}$$

4.3 Proof of Proposition 5

It remains to prove Proposition 5 using the $\mathscr{E}$ defined above. First we need to state the desired lower bound for $\mathbb{P}(\mathscr{E})$.

Proposition 6 *There exists $\beta > 0$ depending on all parameters such that $\mathbb{P}(\mathscr{E}) > \beta$.*

Deferring the proof of this proposition to Sect. 4.4, we first finish the proof of Proposition 5.

Proof of Proposition 5 Let $\mathscr{E}$ be as defined above. By Proposition 6 we know that $\mathbb{P}(\mathscr{E})$ is bounded below as required. Fix $\omega = \omega_\theta \in \mathscr{E}$. Observe that Y is a deterministic function of ω. Using linearity of covariance and Cauchy–Schwarz inequality, we have for each $\omega \in \mathscr{E}$,

$$\begin{aligned}\text{Cov}(X, Z + W \mid \omega) &= \text{Cov}(X, Z + W - Y \mid \omega)\\ &= \text{Cov}(X_*, Z + W - Y \mid \omega)\\ &\quad + \text{Cov}(X - X_*, Z + W - Y \mid \omega)\\ &= \text{Var}(X_* \mid \omega)\end{aligned}$$

$$
\begin{aligned}
&+\mathrm{Cov}(X_*, Z+W-Y-X_* \mid \omega)\\
&+\mathrm{Cov}(X-X_*, X_* \mid \omega)\\
&+\mathrm{Cov}(X-X_*, Z+W-Y-X_* \mid \omega)\\
&\geq \mathrm{Var}(X_* \mid \omega) - \sqrt{\mathrm{Var}(X^*)}\\
&\times \left(\sqrt{\mathrm{Var}(X-X_* \mid \omega)} + \sqrt{\mathrm{Var}(Z+W-Y-X_* \mid \omega)}\right)\\
&-\sqrt{\mathrm{Var}(X-X_* \mid \omega)}\sqrt{\mathrm{Var}(Z+W-Y-X_* \mid \omega)}.
\end{aligned}
$$

By definition of $\mathscr{E}_5$, and the observation that $X^* \geq X \geq X_*$ we get that for each $\omega \in \mathscr{E}$, $\mathrm{Var}(X - X_* \mid \omega) \leq 10r^{2/3}$. By definition of $\mathscr{E}_6$, $\mathrm{Var}(Z + W - Y - X_* \mid \omega) = O(\phi_0^2 r^{2/3})$. The proof is completed by the definition of $\mathscr{E}_7$, observing that by our choices of parameters $\theta^{-1/4} \gg \phi_0$ (whch ensures that the first term dominates in the above expression).

We now illustrate how the proof of Theorem 2, (ii) can be completed along the same lines. We shall only provide a sketch.

Proof of Theorem 2, (ii) First observe that in the notation of the above proof, using Cauchy–Schwarz inequality, and the fact that Y is a deterministic function of ω, we have, as above, that for all $\omega \in \mathscr{E}$,

$$
\mathrm{Var}(Z+W \mid \omega) \geq \mathrm{Var}(X_* \mid \omega) - 2\sqrt{\mathrm{Var}(Z+W-Y-X_* \mid \omega)}\sqrt{\mathrm{Var}(X_* \mid \omega)}.
$$

By definition of $\mathscr{E}_6$ and $\mathscr{E}_7$, we get that for θ sufficiently small and for all $\omega \in \mathscr{E}$, we have $\mathrm{Var}(Z + W \mid \omega) \geq c(\theta)r^{2/3}$ for some $c(\theta) > 0$. Now we make the same definitions as before, but interchange the roles of r and $(n - r)$. Let the event corresponding to $\mathscr{E}$ be now denoted $\widetilde{\mathscr{E}}$. The analogue of Proposition 6 and the above observation now implies that for $1 \ll n - r \ll n$ there exists a positive probability set $\widetilde{\mathscr{E}}$ such that for each $\omega \in \widetilde{\mathscr{E}}$, $\mathrm{Var}(T_n \mid \omega) \geq c(n-r)^{2/3}$ (and T_r is a deterministic function of $\omega \in \widetilde{\mathscr{E}}$). This implies for some constant $c' > 0$ we have

$$
c'(n-r)^{2/3} \leq \inf_{\lambda} \mathrm{Var}(T_n - \lambda T_r) = (1 - \mathrm{Corr}^2(T_r, T_n))\mathrm{Var}(T_n);
$$

which completes the proof.

The remainder of the paper is devoted to the proof of Proposition 6.

4.4 Proof of Proposition 6

The proof has two parts. First we consider the event $\mathscr{E}_1 \cap \mathscr{E}_4$ and show that it has probability bounded below. Then we show that conditional on $\mathscr{E}_1 \cap \mathscr{E}_4$ each of the events $\mathscr{E}_5$, $\mathscr{E}_6$ and $\mathscr{E}_7$ has probability close to one. Since $\mathscr{E}_1$ and $\mathscr{E}_4$ are independent it suffices to lower bound their probabilities separately.

Lemma 2 *There exists positive constants r_0, δ_1 such that for all $\delta_1 n > r > r_0$, we have*

$$\mathbb{P}(\mathscr{E}_1) \geq 1 - e^{-c \log^4(\theta^{-1})}.$$

Proof Recall the two events whose intersection $\mathscr{E}_1$ consists of. That the first of those has probability at least $1 - e^{-c \log^4(\theta^{-1})}$ is an immediate consequence of Theorem 3. The probability lower bound for the second event also follows by writing the line segment $\mathbb{L}_{r,\phi}$ as an increasing union over $\mathbb{L}_{r,i}$ for $i = 1, 2 \ldots, \phi$, applying Theorem 3 for each and taking a union bound.

The next lemma, quoted from [5] without proof, shows that $\mathscr{E}_4$ occurs with positive probability.

Lemma 3 ([5, Lemma 6.5]) *There exists $\epsilon = \epsilon(\phi, L) > 0$ such that $\mathbb{P}(\mathscr{E}_4) > \epsilon$.*

Let us now move towards bounding the conditional probabilities of $\mathscr{E}_5$, $\mathscr{E}_6$ and $\mathscr{E}_7$ given $\mathscr{E}_1 \cap \mathscr{E}_4$. Notice that $\mathscr{E}_5$ is independent of $\mathscr{E}_1$ and hence for those it suffices to consider conditional probability given $\mathscr{E}_4$ only. We need the following result from [5].

Lemma 4 ([5, Lemma 6.6]) *There exists positive constants r_0, δ_1 such that for all $\delta_1 n > r > r_0$, and θ sufficiently small, we have*

$$\mathbb{E}[(X^* - X_*)^2 \mid \mathscr{E}_4] \leq r^{2/3}.$$

Lemma 5 *There exists positive constants r_0, δ_1 such that for all $\delta_1 n > r > r_0$, we have*

$$\mathbb{E}[(Z + W - Y - X_*)^2 \mid \mathscr{E}_1 \cap \mathscr{E}_4] \leq 4\phi_0^2 r^{2/3}.$$

We shall come back to the proof of Lemma 5 at the end of this subsection.

Lemma 6 *There exists positive constants r_0, δ_1 and c_0 such that for all $\delta_1 n > r > r_0$, we have*

$$\mathbb{P}[\text{Var}[X_* \mid \omega_\theta] \geq c_0 \theta^{-1/2} r^{2/3} \mid \mathscr{E}_1 \cap \mathscr{E}_4] \geq 0.9.$$

Essentially the same statement is proved in [5, (43)] and we shall omit the proof. See the proofs of [5, Lemma 6.9, Lemma 6.10].

We can now complete the proof of Proposition 6.

Proof of Proposition 6 Observe that by Markov inequality (and the fact that $\mathscr{E}_1$ is independent of $\mathscr{E}_4$ and $\mathscr{E}_5$) we have $\mathbb{P}(\mathscr{E}_5 \mid \mathscr{E}_1 \cap \mathscr{E}_4) \geq 0.9$ and $\mathbb{P}(\mathscr{E}_6 \mid \mathscr{E}_1 \cap \mathscr{E}_4) \geq 0.9$ and $\mathbb{P}(\mathscr{E}_7 \mid \mathscr{E}_1 \cap \mathscr{E}_4) \geq 0.9$ using Lemmas 4, 5, and 6 respectively. Combined, these give $\mathbb{P}(\mathscr{E} \mid \mathscr{E}_1 \cap \mathscr{E}_4) \geq 0.7$. Observe further that by Lemmas 2 and 3 and the fact that $\mathscr{E}_1$ and $\mathscr{E}_4$ are independent implies that for θ sufficiently small we have $\mathbb{P}(\mathscr{E}_1 \cap \mathscr{E}_4) \geq \epsilon/2$. The proof of the proposition is completed by choosing $\beta = \epsilon/4$.

It remains to prove Lemma 5. It is in spirit similar (in fact somewhat easier) to the proof of [5, Lemma 6.7] but several ingredients are different.

Proof of Lemma 5 Let A denote the event that the point v where the geodesic Γ_n from $\mathbf{0}$ to $\mathbf{n}$ intersects $\mathbb{L}_r$ lies in $\mathbb{L}_{r,\phi_0}$. We write

$$\begin{aligned}
&\mathbb{E}[(Z+W-Y-X_*)^2 \mid \mathscr{E}_1 \cap \mathscr{E}_4] \\
&\quad = \mathbb{E}[(Z+W-Y-X_*)^2 1_A \mid \mathscr{E}_1 \cap \mathscr{E}_4] + \mathbb{E}[(Z+W-Y-X_*)^2 1_{A^c} \mid \mathscr{E}_1 \cap \mathscr{E}_4] \\
&\quad \leq \mathbb{E}[(Z+W-Y-X_*)^2 1_A \mid \mathscr{E}_1 \cap \mathscr{E}_4] + \mathbb{E}[(Z+W-Y-X_\theta)^2 1_{A^c} \mid \mathscr{E}_1 \cap \mathscr{E}_4]
\end{aligned}$$

where the inequality is a consequence of $0 \leq Z+W-Y-X_* \leq Z+W-Y-X_\theta$.

To bound the first term, we notice that, on A, $|Z-Y| \leq \sup_{w \in \mathbb{L}_{r,\phi_0}} |T_{w,\mathbf{n}} - Y|$, and consequently

$$\begin{aligned}
&\mathbb{E}[(Z+W-Y-X_*)^2 1_A \mid \mathscr{E}_1 \cap \mathscr{E}_4] \\
&\quad \leq O\left(\mathbb{E}\left[\sup_{w \in \mathbb{L}_{r,\phi_0}} |T_{w,\mathbf{n}} - Y|^2 \mid \mathscr{E}_1\right] + \mathbb{E}[(X^* - X_*)^2 \mid \mathscr{E}_4]\right) \\
&\quad = O(\phi_0 r^{2/3})
\end{aligned}$$

where in the first inequality we use the fact that $\sup_{w \in \mathbb{L}_{r,\phi_0}} |T_{w,\mathbf{n}} - Y|^2$ is independent of $\mathscr{E}_4$ and the last inequality is a consequence of Theorem 3, Lemmas 2 and 4.

Now for the second term, using Cauchy–Schwarz inequality we get

$$\mathbb{E}[(Z+W-Y-X_\theta)^2 1_{A^c} \mid \mathscr{E}_1 \cap \mathscr{E}_4] \leq \mathbb{P}[A^c \mid \mathscr{E}_1 \cap \mathscr{E}_4]^{1/2} \mathbb{E}[(Z+W-Y-X_\theta)^4 \mid \mathscr{E}_1 \cap \mathscr{E}_4]^{1/2}.$$

Now we claim that $\mathbb{P}(A \mid \mathscr{E}_1 \cap \mathscr{E}_4) \geq 1 - e^{-\log^2(\theta)}$. This is proved in Lemma 7 below. Also notice that since $\mathscr{E}_1$ and $\mathscr{E}_4$ are independent, Lemma 2 implies that for θ sufficiently small we have $\mathbb{E}[(Z+W-Y-X_\theta)^4 \mid \mathscr{E}_1 \cap \mathscr{E}_4] \leq 2\mathbb{E}[(Z+W-Y-X_\theta)^4 \mid \mathscr{E}_4]$. Further observe that the event $\mathscr{E}_4$ is decreasing in the configuration R_θ, and the FKG inequality implies that conditioning on $\mathscr{E}_4$ makes the configuration outside R_θ stochastically smaller. Since $(Z+W-Y-X_\theta)$ is positive and increasing in the configuration outside R_θ, we thus have

$$\mathbb{E}[(Z+W-Y-X_\theta)^4 \mid \mathscr{E}_4] \leq \mathbb{E}[(Z+W-Y-X_\theta)^4].$$

Finally, as $Z \le X^*$ and $Z + W - Y - X_\theta \ge 0$

$$\begin{aligned} E[(Z + W - Y - X_\theta)^4] &\le \mathbb{E}[(X^* + W - Y - X_\theta)^4] \\ &\le O(\mathbb{E}[(X^* - X_\theta)^4] + \mathbb{E}[(W - Y)^4]) \\ &= O(\theta^{-4} r^{4/3}). \end{aligned}$$

For the last equality we use [5, Proposition 4.5] to show $\mathbb{E}[(X_\theta - 4r)^4] = O(\theta^{-4} r^{4/3})$, use [5, Theorem 4.1] to get $\mathbb{E}[(X^* - 4r)^4] = O(r^{4/3})$ and deduce $\mathbb{E}[(W - Y)^4] = O(r^{4/3})$ from Lemma 1 as in the proof of Proposition 3. By taking θ sufficiently small, this concludes the proof of the proposition modulo Lemma 7 below.

Lemma 7 *In the set up of the proof of Proposition 6, we have* $\mathbb{P}(A \mid \mathscr{E}_1 \cap \mathscr{E}_4) \ge 1 - e^{-\log^2(\theta)}$ *for all* θ *sufficiently small.*

For this proof we make numerous uses of the estimate in [5, Theorem 4.2] which states that for an $r \times r^{2/3}$ rectangle (or parallelogram) R and for pairs of $u, w \in R$ such that the slope joining u, w is bounded away from 0 and infinity we have $\inf_{u,w} r^{-1/3}(T_{u,w} - \mathbb{E}T_{u,w})$ and $\sup_{u,w} r^{-1/3}(T_{u,w} - \mathbb{E}T_{u,w})$ both have stretched exponential tails.

Proof We shall construct a number of large probability events which together will imply A. Let $\mathscr{A}_{\mathrm{loc},\phi}$ denote the event that for some $w \in \mathbb{L}_r \setminus \mathbb{L}_{r,\phi^{1/2}}$ we have

$$T_{\mathbf{0},w} + T_{w,\mathbf{n}} \ge X_\theta + T_{\mathbf{r},\mathbf{n}}.$$

Let $\mathscr{A}_{\theta,1}$ denote the event that for all $v' = (v'_1, v'_2) \in R_\theta$ with $2r - 2\theta^{3/2} r \le d(v') \le 2r - \theta^{3/2} r$, we have

$$T^{R_\theta}_{v',\mathbf{r}} \ge 2(2r - d(v')) - \theta^{1/2} r^{1/3} \log^5(\frac{1}{\theta}).$$

Let $\mathscr{A}_{\theta,2}$ denote the event that for all v' as above and for all $w \in \mathbb{L}_{r,\phi^{1/2}} \setminus \mathbb{L}_{r,\phi_0}$ we have

$$T_{v',w} - \mathbb{E}T_{v',w} \le \theta^{1/2} r^{1/3} \log^{10}(\frac{1}{\theta}).$$

Finally, let $\widetilde{\mathscr{E}}$ denote the event that for any $w' \in \mathbb{L}_{r,\phi^{1/2}} \setminus \mathbb{L}_{r,\phi_0}$ and the geodesic Γ' from $\mathbf{0}$ to w' there exists $v' \in R_\theta \cap \Gamma'$ with $2r - 2\theta^{3/2} r \le d(v') \le 2r - \theta^{3/2} r$ such that from 0 to v', Γ' is entirely contained in $R_{2\theta}$.

We first claim that $\mathscr{E}_1 \cap \widetilde{\mathscr{E}} \cap \mathscr{A}_{\theta,1} \cap \mathscr{A}_{\theta,2} \cap (\mathscr{A}_{\mathrm{loc},\phi})^c \subseteq A$. Indeed, observe first that $(\mathscr{A}_{\mathrm{loc},\phi})^c$ implies that $v := \Gamma \cap \mathbb{L}_r \in \mathbb{L}_{r,\phi^{1/2}}$. Then notice that for any $w' = (w'_1, w'_2) \in \mathbb{L}_{r,\phi^{1/2}} \setminus \mathbb{L}_{r,\phi_0}$, $\widetilde{\mathscr{E}}$ implies that

$$T_{\mathbf{0},w'} - X_* \leq \sup_{v'}[T_{v',w'} - T^{R_\theta}_{v',\mathbf{r}}]$$

where the supremum is taken over all $v' \in R_\theta$ such that $2r - 2\theta^{3/2}r \leq d(v') \leq 2r - \theta^{3/2}r$.

Recall, for w' as above, the lower bound on $Y - T_{w',\mathbf{n}}$ given by the definition of $\mathscr{E}_1$. Using this together with the fact that $\mathbb{E}T_{v',w'} \leq 2(2r - d(v')) - \frac{|w'_1 - w'_2|^2}{50\theta^{3/2}r}$ (this is a consequence of the moderate deviation estimate [5, Theorem 4.1]) and the definitions of $\mathscr{A}_{\theta,1}$ and $\mathscr{A}_{\theta,2}$, it follows that on $\mathscr{E}_1 \cap \mathscr{A}_{\theta,1} \cap \mathscr{A}_{\theta,2}$ we have

$$\sup_{v'}[T_{v,w'} - T^{R_\theta}_{v',\mathbf{r}}] \leq Y - T_{w',\mathbf{n}}$$

where the supremum over v' is as before. It therefore follows that on $\mathscr{E}_1 \cap \mathscr{E}' \cap \mathscr{A}_{\theta,1} \cap \mathscr{A}_{\theta,2}$ we have

$$T_{\mathbf{0},w'} - X_* \leq Y - T_{w',\mathbf{n}}$$

for each $w' \in \mathbb{L}_{r,\phi^{1/2}} \setminus \mathbb{L}_{r,\phi_0}$. This completes the proof of the claim.

Now the barrier event $\mathscr{E}_4$ is designed in such a way that a path from $\mathbf{0}$ to w' as above is penalised more heavily than a path constrained to stay within R_θ (as $L \gg \phi \gg \frac{1}{\theta}$). Formalising this, [5, Lemma 6.12] implies (the event $\widetilde{\mathscr{E}}$ defined there is slightly different, where the starting point of the geodesic is also allowed to vary around $\mathbf{0}$ but the same proof works) $\mathbb{P}((\widetilde{\mathscr{E}})^c \mid \mathscr{E}_4) \leq e^{-\log^3(1/\theta)}$. It follows from [5, Theorem 4.2] that $\mathbb{P}(\mathscr{A}_{\theta,1})^c \leq e^{-\log^{5/2}(1/\theta)}$ for all θ small. Next, notice that by dividing $\mathbb{L}_{r,\phi^{1/2}} \setminus \mathbb{L}_{r,\phi_0}$ into intervals of length $\theta r^{2/3}$, applying [5, Theorem 4.2] and taking a union bound and using the FKG inequality it follows that $\mathbb{P}((\mathscr{A}_{\theta,2})^c \mid \mathscr{E}_4) \leq e^{-\log^3(1/\theta)}$. It remains to upper bound $\mathbb{P}(\mathscr{A}_{\mathrm{loc},\phi} \mid \mathscr{E}_4)$.

To this end, set $S_j := \mathbb{L}_{r,j+1} \setminus \mathbb{L}_{r,j}$. Our objective is to show that with high probability, $Y + X_\theta \geq \sup_{w \in S_j} T_{\mathbf{0},w} + \sup_{w \in S_j} T_{w,\mathbf{n}}$. Let $\mathscr{C}_j$ denote the event that $\sup_{w \in S_j} T_{\mathbf{0},w} - X_\theta \geq \inf_{w \in S_j} Y - T_{w,\mathbf{n}}$. Clearly, for $j > \phi^{1/2}$, we can upper bound $\mathbb{P}(\mathscr{C}_j)$ by

$$\mathbb{P}(\sup_{w \in S_j} T_{\mathbf{0},w} - 4r \geq -0.001 j^2 r^{1/3})$$
$$+ \mathbb{P}(X_\theta \leq 4r - 0.001 j^2 r^{1/3}) + \mathbb{P}(\inf_{w \in S_j} Y - T_{w,\mathbf{n}} \leq -0.002 j^2 r^{1/3}).$$

Using [5, Theorem 4.2] for $j < 0.9r^{1/3}$ and [5, (13)] together with a union bound for $j \geq 0.9r^{1/3}$, we can show that the first probability is upper bounded by $e^{-cj^{3/2}}$ (and the same is true conditionally on $\mathscr{E}_4$ by the FKG inequality). By Basu et al. [5, Theorem 4.2] and a simple concentration inequality for sums of $\theta^{-3/2}$ many independent subexponential variables at scale $\theta^{1/2}r^{1/3}$ (see the proof of [5, Proposition 4.5]) we get that the second probability is upper bounded by $e^{-cj^2\theta}$, whereas the third probability, by Theorem 3, is upper bounded by $e^{-cj^{2/3}}$. Notice also that the second and third events above are independent of $\mathscr{E}_4$. Summing over all $j > \phi^{1/2}$ and using that ϕ is a large power of θ^{-1} gives the result gives that $\mathbb{P}(\mathscr{A}_{\text{loc}}^{\phi} \mid \mathscr{E}_4) \leq e^{-\log^3(1/\theta)}$.

Combining all these together and using $\mathscr{E}_1$ is independent of $\mathscr{E}_4$ together with Lemma 3 gives us $\mathbb{P}(A \mid \mathscr{E}_1 \cap \mathscr{E}_4) \geq 1 - e^{-\log^2(\theta)}$ for all θ sufficiently small, as desired.

4.5 A Note on the Variance of Constrained Last Passage Time

Before concluding we also comment that the proof of Lemma 6 (see the proof of [5, Lemma 6.9]) can be used to obtain the sharp order of variance of the weight of the best path constrained to stay within a thin cylinder. In particular one can show that for $\theta \leq 1$ and r sufficiently large, Var $X_\theta = \Theta(\theta^{-1/2}r^{2/3})$ answering a question raised in [13]. The lower bound in the above statement can be proved using a simpler version of the argument used in the proof of [5, Lemma 6.9]. Upper bound follows from a Poincaré inequality argument after revealing $\theta^{3/2}r \times \theta r^{2/3}$ rectangles one by one and using [5, Theorem 4.2].

Acknowledgments We thank Alan Hammond and Ivan Corwin for discussing the results in [10], and Alan Hammond for extensive discussions around the results of [17] and Theorem 3. We also thank an anonymous referee for several useful comments and suggestions. RB is partially supported by an ICTS-Simons Junior Faculty Fellowship, a Ramanujan Fellowship (SB/S2/RJN-097/2017) from the Science and Engineering Research Board, and by ICTS via project no. 12-R&D-TFR-5.10-1100 from DAE, Govt. of India. SG is partially supported by a Sloan Research Fellowship in Mathematics and NSF Award DMS-1855688.

References

1. Baik, J., Ferrari, P.L., Péché, S.: Convergence of the two-point function of the stationary TASEP. In: Griebel, M. (ed.), Singular Phenomena and Scaling in Mathematical Models, pp. 91–110. Springer, Berlin (2014)
2. Baik, J., Liu, Z.: Multi-point distribution of periodic TASEP (2017, preprint). arXiv:1710.03284
3. Basu, R., Sidoravicius, V., Sly, A.: Last passage percolation with a defect line and the solution of the Slow Bond Problem (2014). arXiv 1408.3464

4. Basu, R., Ganguly, S., Hammond, A.: The competition of roughness and curvature in area-constrained polymer models. Commun. Math. Phys. **364**(3), 1121–1161 (2018)
5. Basu, R., Ganguly, S., Zhang, L.: Temporal correlation in last passage percolation with flat initial condition via brownian comparison (preprint, 2019). arXiv:1912.04891
6. Basu, R., Sarkar, S., Sly, A.: Coalescence of geodesics in exactly solvable models of last passage percolation. J. Math. Phys. **60**, 093301 (2019)
7. Borodin, A., Ferrari, P.: Large time asymptotics of growth models on space-like paths I: PushASEP. Electron. J. Probab. **13**, 1380–1418 (2008)
8. Calvert, J., Hammond, A., Hegde, M.: Brownian structure in the KPZ fixed point (preprint, 2019). arXiv:1912.00992
9. Cator, E., L.P.R. Pimentel, On the local fluctuations of last-passage percolation models. Stoch. Process. Their Appl. **125**(2), 538–551 (2015)
10. Corwin, I., Hammond, A.: Correlation of the Airy_2 process in time (Unpublished)
11. Corwin, I., Hammond, A.: Brownian Gibbs property for Airy line ensembles. Invent. Math. **195**(2), 441–508 (2014)
12. Corwin, I., Ghosal, P., Hammond, A.: KPZ equation correlations in time (preprint, 2019). arXiv:1907.09317
13. Dey, P.S., Peled, R., Joseph, M.: Longest increasing path within the critical strip (preprint). https://arxiv.org/abs/1808.08407
14. Ferrari, P.L., Occelli, A.: Universality of the goe Tracy-Widom distribution for TASEP with arbitrary particle density. Electron. J. Probab. **23**, 24pp. (2018)
15. Ferrari, P.L., Occelli, A.: Time-time covariance for last passage percolation with generic initial profile. Math. Phys. Anal. Geom. **22**(1), 1 (2019)
16. Ferrari, P.L., Spohn, H.: On time correlations for KPZ growth in one dimension. SIGMA **12**, 074 (2016)
17. Hammond, A.: Brownian regularity for the Airy line ensemble, and multi-polymer watermelons in Brownian last passage percolation. Mem. Am. Math. Soc. (to appear, 2019). https://www.ams.org/cgi-bin/mstrack/accepted_papers/memo
18. Johansson, K.: Shape fluctuations and random matrices. Commun. Math. Phys. **209**(2), 437–476 (2000)
19. Johansson, K.: Transversal fluctuations for increasing subsequences on the plane. Probab. Theory Relat. Fields **116**(4), 445–456 (2000)
20. Johansson, K.: Two time distribution in brownian directed percolation. Commun. Math. Phys. **351**(2), 441–492 (2017)
21. Johansson, K.: The long and short time asymptotics of the two-time distribution in local random growth (preprint, 2019). arXiv:1904.08195
22. Johansson, K.: The two-time distribution in geometric last-passage percolation. Probab. Theor. Relat. Fields. **175**, 849–895 (2019). https://link.springer.com/article/10.1007/s00440-019-00901-9
23. Johansson, K., Rahman, M.: Multi-time distribution in discrete polynuclear growth (preprint, 2019). arXiv:1906.01053
24. Ledoux, M., Rider, B., et al.: Small deviations for beta ensembles. Electron. J. Probab. **15**, 1319–1343 (2010)
25. Liu, Z.: Multi-time distribution of TASEP (preprint, 2019). arXiv:1907.09876
26. Matetski, K., Quastel, J., Remenik, D.: The KPZ fixed point (preprint, 2017). arXiv:1701.00018
27. Pimentel, L.P.R.: Local behaviour of airy processes. J. Stat. Phys. **173**(6), 1614–1638 (2018)

On the Four-Arm Exponent for 2D Percolation at Criticality

Jacob van den Berg and Pierre Nolin

This paper is dedicated to the memory of Vladas Sidoravicius, whose enthusiasm and dynamism have been very stimulating to us.

Abstract For two-dimensional percolation at criticality, we discuss the inequality $\alpha_4 > 1$ for the polychromatic four-arm exponent (and stronger versions, the strongest so far being $\alpha_4 \geq 1 + \frac{\alpha_2}{2}$, where α_2 denotes the two-arm exponent). We first briefly discuss five proofs (some of them implicit and not self-contained) from the literature. Then we observe that, by combining two of them, one gets a completely self-contained (and yet quite short) proof.

Keywords Critical percolation · Arm exponents

MSC2010 Subject Classifications 60K35, 82B43

Pierre Nolin was partially supported by a GRF grant from the Research Grants Council of the Hong Kong SAR (project CityU11306719).

J. van den Berg (✉)
Centrum Wiskunde en Informatica (CWI), Amsterdam, The Netherlands

VU University Amsterdam, Amsterdam, The Netherlands
e-mail: J.van.den.Berg@cwi.nl

P. Nolin
Department of Mathematics, City University of Hong Kong, Kowloon Tong, Hong Kong
e-mail: bpmnolin@cityu.edu.hk

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_6

1 Introduction

In this paper we focus on critical site percolation on the square lattice $(\mathbb{Z}^2, \mathbb{E}^2)$. The vertices of this lattice are the points in $\mathbb{R}^2$ with integer coordinates, and the edges in $\mathbb{E}^2$ connect all pairs of vertices $v, v' \in \mathbb{Z}^2$ with $\|v - v'\|_2 = 1$ ($\|.\|_2$ denoting the usual Euclidean norm). However, note that the results would also hold on any two-dimensional lattice with enough symmetries, such as the honeycomb lattice, and also for bond percolation.

We are interested in upper bounds for the probability that two disjoint clusters connect neighbors of the origin to distance n, i.e. in lower bounds on the corresponding exponent. This exponent is called two-arm exponent in [3] (a paper concerning dimensions ≥ 2), but in two dimensions it is the same as what is usually called four-arm exponent: two open arms, one for each of the two open clusters, separated by two closed arms (ensuring that these two clusters are indeed not connected by an open path). We denote the corresponding exponent by α_4. In the particular case of site percolation on the triangular lattice, this exponent is known to be equal to $\frac{5}{4}$ [17], and this is widely believed to hold for all "nice" two-dimensional lattices (for site percolation, as well as for bond percolation).

For the square lattice it has been known for quite some time that $\alpha_4 > 1$. This strict inequality is related to the so-called noise sensitivity of certain percolation phenomena (see Sects. 4.1 and 4.2). This inequality (and stronger versions) has an interesting history, due to the diversity of the problems where four-arm probabilities (and their analogs in higher dimensions) played, play, or might play, a role (for instance, the uniqueness of the infinite cluster and the famous conjecture that $\theta(p_c) = 0$ for every dimension).

The first paper from which a proof of $\alpha_4 > 1$ can be (implicitly) obtained is (as several authors have mentioned, but without giving details) Kesten's celebrated scaling relations paper [10]. We discuss in some detail in Sect. 3 how to do this. This method is quite technical and assumes much percolation background. Readers without such background are advised to skip that section.

In Sect. 4 we discuss parts of four other papers in the literature which, sometimes implicitly, provide a proof (some of them of the stronger result $\alpha_4 \geq 1 + \frac{\alpha_2}{2}$). Those proofs avoid the heavy near-critical machinery from [10]. However, in most of these papers the four-arm inequality came up as a by-product or a necessary ingredient, and the authors have not always strived for optimizing simplicity or length of the proof. Several of the proofs use a concentration result (which for this inequality is not needed) and/or a so-called arm-separation result: a result by Kesten which, although intuitively appealing, has a rather long and cumbersome proof.

A natural question is whether there is a short and self-contained proof that can be given in the first part of an introductory course on percolation theory, right after presenting the classical Russo-Seymour-Welsh result on crossing probabilities. We observed that one gets such a proof by following a special case of a proof by Garban in Appendix B of [15] (which is inspired by a general inequality of [14], see also [6]), with modifications and ingredients from Cerf's arguments in [3]. This proof is presented in Sect. 5. It gives the stronger version of the inequality mentioned above, as stated more precisely in Theorem 1 below, but it is probably also, essentially, the shortest self-contained proof of the weaker version $\alpha_4 > 1$.

Theorem 1 *For site percolation on the square lattice* $(\mathbb{Z}^2, \mathbb{E}^2)$ *at criticality (*$p = p_c^{site}(\mathbb{Z}^2)$*), the following inequality between the two- and four-arm exponents, denoted by (resp.)* α_2 *and* α_4*, holds:*

$$\alpha_4 \geq 1 + \frac{\alpha_2}{2}. \tag{1}$$

We want to stress again that Theorem 1 is not new, but that the proof presented in Sect. 5 (a modification and combination of other proofs) is arguably the most self-contained. It does not use Kesten's arm-separation results [10]: in fact, it only uses pre-1980 percolation, namely the Russo-Seymour-Welsh result that at criticality, "box-crossing probabilities are bounded away from 0 and 1".

1.1 Organization of the Paper

In Sect. 2, we set notation, and we recall the properties of critical percolation in 2D that we are going to use. We then comment on earlier (explicit or implicit) proofs of the inequality $\alpha_4 > 1$ (or even of (1)) in Sects. 3 and 4, before turning to the self-contained proof of Theorem 1 in Sect. 5.

2 Two-Dimensional Percolation at Criticality

2.1 Setting and Notations

Recall that we work with the square lattice $G = (V, E)$, with set of vertices $V = \mathbb{Z}^2$, and set of edges $E = \mathbb{E}^2$ connecting any two vertices which are at a Euclidean distance 1 apart (i.e. differing along exactly one coordinate, by ± 1). Two vertices $v, v' \in \mathbb{Z}^2$ are adjacent (or neighbors) if they are connected by an edge, i.e. $\{v, v'\} \in E$, and we write it $v \sim v'$. For a subset of vertices $A \subseteq V$, its inner and outer vertex boundaries are defined as, respectively,

$$\partial^{\text{in}} A := \left\{ v \in A \ : \ v \sim v' \text{ for some } v' \in V \setminus A \right\}$$

and $\partial^{\text{out}} A := \partial^{\text{in}}(V \setminus A)$. The matching lattice $G^* = (V^*, E^*)$, or simply *-lattice, is obtained from G by adding the two diagonal edges to each face, as shown on Fig. 1 (Left), and we use the notation $\sim^*$ for adjacency on G^*. A path (resp *-path) of length $k \geq 1$ on G (resp. G^*) is a finite sequence of vertices $v_0, v_1, \ldots, v_k$

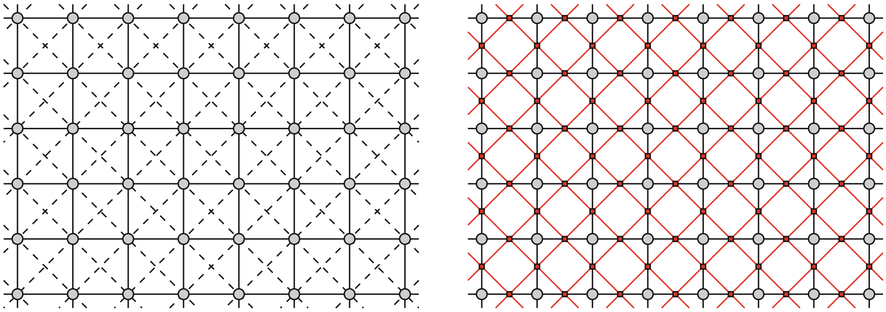


Fig. 1 *Left:* This figure shows the square lattice G, as well as the *-lattice obtained by adding the two diagonal edges (in dashed line) to every face of G. *Right:* This figure depicts, in red, the medial lattice of G

such that $v_i \sim v_{i+1}$ (resp. $v_i \sim^* v_{i+1}$) for all $i = 0, \dots, k-1$. We denote by $B_n := [-n, n]^2$ the ball of radius $n \geq 0$ around 0 for the L^∞ norm $\|.\| = \|.\|_\infty$, and by $A_{n_1,n_2} := B_{n_2} \setminus B_{n_1-1}$ the annulus with radii $0 \leq n_1 < n_2$ centered at 0.

We also introduce the medial lattice $G^\diamond = (V^\diamond, E^\diamond)$ of G, for which a vertex $e^\diamond \in V^\diamond$ is located at the middle of every edge $e \in E$, and two such vertices $e^\diamond$, $e'^\diamond$ in $V^\diamond$ are connected by an edge *if and only if* the corresponding edges e, e' are incident to a common vertex in V: see Fig. 1 (Right).

Bernoulli site percolation on G with parameter $p \in [0, 1]$ is obtained by declaring each vertex $v \in V$ either open or closed, with respective probabilities p and $1 - p$, independently of the other vertices. We denote by $\Omega := \{0, 1\}^V$ the set of configurations $(\omega_v)_{v \in V}$, where $\omega_v = 1$ if v is open, and $\omega_v = 0$ if v is closed. We write $\mathbb{P}_p$ for the product measure with parameter p on Ω.

Two vertices $v, v' \in V$ are connected (resp. *-connected) if there exists a path (resp. *-path) of length k, for some $k \geq 1$, along which all vertices are open (resp. closed), and we use the notation $v \leftrightarrow v'$ (resp. $v \leftrightarrow^* v'$). More generally for $A, A' \subseteq V$, $A \leftrightarrow A'$ (resp. $A \leftrightarrow^* A'$) means that there exist $v \in A$ and $v' \in A'$ such that $v \leftrightarrow v'$ (resp. $v \leftrightarrow^* v'$). Open vertices can be grouped into maximal connected components, that we call open clusters, and we denote by $\mathscr{C}(v)$ the open cluster containing a given $v \in V$ (with $\mathscr{C}(v) = \emptyset$ if v is closed). Closed *-clusters are defined in a similar way.

Exploration processes turn out to be an important ingredient in the proofs below. Such processes determine the outer boundary of an open cluster by revealing it in a step-by-step manner: all the open vertices along it, together with all the adjacent closed vertices (and discovering no other vertices). As shown on Fig. 2, they can be seen as edge-self-avoiding paths on the medial lattice $G^\diamond$.

Site percolation of G displays a phase transition at a percolation threshold $p_c = p_c^{\text{site}}(G)$: for all $p < p_c$ there exists almost surely (a.s.) no infinite open cluster and a unique infinite closed *-cluster, while for all $p > p_c$ there is a.s. a unique infinite open cluster but no infinite closed *-cluster. In the present paper, we are concerned with the critical regime $p = p_c$, where neither infinite open clusters nor infinite

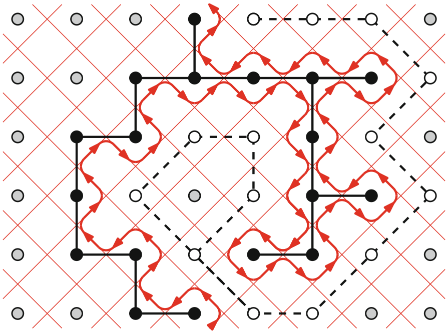


Fig. 2 An exploration process on G following an "interface" between open and closed sites. It can be seen as an edge-self-avoiding path on the medial graph $G^\diamond$ of G (each edge of $G^\diamond$ is followed at most once, although some vertices may be visited several times). The black and white vertices are revealed during the exploration, and respectively open and closed, while the grey vertices are left unexplored

closed *-clusters do exist. We refer the reader to the classical references [7, 8] for more background on percolation theory.

Finally, the cardinality of a set S is denoted by $|S|$, and for an event E, its indicator function $\mathbb{1}_E$ is defined by: $\mathbb{1}_E(\omega) = 1$ if $\omega \in E$, and $\mathbb{1}_E(\omega) = 0$ otherwise.

2.2 Critical Regime

We now recall classical definitions and properties concerning Bernoulli percolation at the critical point p_c.

If $R = [x_1, x_2] \times [y_1, y_2]$ (for some integers $x_1 < x_2$, $y_1 < y_2$) is a rectangle on the lattice, we denote by $\mathscr{C}_H(R)$ (resp. $\mathscr{C}^*_H(R)$) the existence of an open path (resp. closed *-path) in R connecting the left side $\{x_1\} \times [y_1, y_2]$ and the right side $\{x_2\} \times [y_1, y_2]$. The classical Russo-Seymour-Welsh (RSW) theory states that

$$\mathbb{P}_{p_c}\big(\mathscr{C}_H([0,4n] \times [0,n])\big) \geq \delta_4 \quad \text{and} \quad \mathbb{P}_{p_c}\big(\mathscr{C}^*_H([0,4n] \times [0,n])\big) \geq \delta_4 \tag{2}$$

for some universal $\delta_4 > 0$. Using standard arguments, (2) implies that for $\delta' = (\delta_4)^4 > 0$,

$$\mathbb{P}_{p_c}(B_n \leftrightarrow \partial^{\text{in}} B_{2n}) \leq 1 - \delta'. \tag{3}$$

For $1 \leq n_1 < n_2$, let $\mathscr{C}_{n_1,n_2}$ denote the collection of open clusters in B_{n_2} connecting B_{n_1} and $\partial^{\text{in}} B_{n_2}$. For future reference, observe that for some universal $c_1 < \infty$:

$$\text{for all } n \geq 1,\ \ell \geq n, \quad \mathbb{E}_{p_c}\big[|\mathscr{C}_{n,n+\ell}|^2\big] \leq c_1. \tag{4}$$

Indeed, we know from (3) that $\mathbb{P}_{p_c}(|\mathscr{C}_{n,n+\ell}| \geq 1)$ is bounded away from 0 and 1, uniformly in n and $\ell \geq n$. Hence, by the BK inequality, $|\mathscr{C}_{n,n+\ell}|$ is (uniformly in n and $\ell \geq n$) stochastically dominated by a geometrically distributed random variable, which gives (4).

Let $k \geq 1$, we consider the alternating sequence $\sigma_k = (oco\ldots) \in \{o, c\}^k$, where o and c stand for "open" and "closed" respectively. In an annulus $A = A_{n_1,n_2}$ $(0 \leq n_1 < n_2)$, let $\mathscr{A}_k(A)$ be the event that there exist k disjoint paths $(\gamma_i)_{1\leq i\leq k}$ in A, in counter-clockwise order, each connecting two vertices v and v' with $\|v\| = n_1$ and $\|v'\| = n_2$, and with respective types prescribed by σ_k (i.e. γ_i is an open path if i is odd, and a closed *-path if i is even). We write

$$\pi_k(n_1, n_2) := \mathbb{P}_{p_c}\big(\mathscr{A}_k(A_{n_1,n_2})\big), \tag{5}$$

and in particular $\pi_k(n) := \pi_k(\tilde{k}, n)$, where $\tilde{k}$ is the smallest integer for which $|\partial^{\text{in}} B_{\tilde{k}}| \geq k$. Note that in this paper we consider only the cases $k = 1, 2, 4$, for which $\tilde{k} = 0, 1, 1$ respectively. Finally, we introduce the k-arm (polychromatic, unless $k = 1$) exponent

$$\alpha_k := -\limsup_{n\to\infty} \frac{\log \pi_k(n)}{\log n}. \tag{6}$$

It follows from standard constructions again (based on (2)) that

$$\text{for all } k \geq 1, \quad \alpha_k \in (0, \infty).$$

Remark 1

1. These arm exponents are known rigorously in the particular case of site percolation on the triangular lattice: $\alpha_1 = \frac{5}{48}$ [11], and for all $k \geq 2$, $\alpha_k = \frac{k^2-1}{12}$ [17]. It is widely believed that these exponents should have the same values on other two-dimensional lattices such as the square lattice, considered in this paper.
2. Adding certain "macroscopic" restrictions concerning the endpoints of the arms (for instance, in the case of four arms, that one endpoint is on the "north" side of B_n, and one on the west, one on the south, and one on the east side) does not increase the corresponding exponent. This "arm-separation result" was an important technical intermediate result by Kesten in his paper on scaling relations [10]. Its proof is quite long and far from easy.

3 Proof from Kesten's Scaling Relations (1987)

In this section, we point out how the inequality $\alpha_4 > 1$ can be extracted from the results of [10]. To the best of our knowledge, this paper is where the inequality $\alpha_4 > 1$ was first (implicitly) proved. Note that in this part, we assume much more percolation knowledge than in the rest of our paper, and the explanation below is mainly meant for specialists.

Other authors have already observed that the inequality $\alpha_4 > 1$ (or even better bounds on α_4) can be obtained from [10]. For instance, the paper [2] (that we discuss in more detail below, see Sect. 4.1) says in Remark 4.2:

> Although this is better than the general bound . . . , a somewhat better bound can be extracted from Kesten's. . .

But as far as we know, the authors did not write details about *how* to obtain it from [10].

At first sight, doing so requires the assumption that some exponents exist. More explicitly, we assume first the existence of α_1 (i.e. that the limit superior in (6) can be replaced by an actual limit), which implies that there is $\delta > 0$ such that

$$\mathbb{P}_{p_c}(|\mathscr{C}(0)| \geq n) = n^{-\frac{1}{\delta}+o(1)} \quad \text{as } n \to \infty.$$

In addition, we need to assume the existence of α_4, or equivalently of $\nu > 0$ such that $L(p) = |p - p_c|^{-\nu+o(1)}$ as $p \to p_c$, where the characteristic length L is defined by $L(p) := \min\{n \geq 1 : \mathbb{P}_p(\mathscr{C}_H([0,n] \times [0,n])) \leq 0.001\}$ (resp. ≥ 0.999) for $p < p_c$ (resp. $p > p_c$).

Corollary 2 in [10] then states the inequality $\nu \geq \frac{\delta+1}{\delta}$. This inequality follows from previous results in [10], combined with either of the following two inequalities, as $p \nearrow p_c$:

$$\frac{\mathbb{E}_p\big[|\mathscr{C}(0)|^2\big]}{\mathbb{E}_p\big[|\mathscr{C}(0)|\big]} \geq (p_c - p)^{-2+o(1)} \tag{7}$$

(see (3) in [4], Section 5), or

$$\mathbb{E}_p\big[|\mathscr{C}(0)|\big] \geq (p_c - p)^{-2(\delta-1)/\delta+o(1)} \tag{8}$$

(see [12], Theorem 1.3). Note that in [10], these inequalities (7) and (8) are stated in terms of the critical exponents corresponding to the quantities in their l.h.s., usually denoted by Δ_2 and γ (respectively).

Hence, we have in particular $\nu > 1$. From the scaling relation $(2 - \alpha_4)\nu = 1$ (which follows from (4.28) and (4.33) in [10]), we can thus obtain $2 - \alpha_4 < 1$, so the desired inequality $\alpha_4 > 1$. Moreover, we can actually get $\alpha_4 \geq 1 + \frac{\alpha_1}{2}$, by following more closely the previous sequence of inequalities and using the relation $\frac{2}{\delta+1} = \alpha_1$, proved in [9] (see the two sentences below (1.20) in [10], and note that in the notations of this paper, $\frac{1}{\delta_r}$ refers to the exponent α_1).

Even if we do not assume the existence of some exponents, a large part of the results in [10] can still be stated and established. In particular, one has the scaling relation

$$|p - p_c| L(p)^2 \pi_4(L(p)) \asymp 1 \tag{9}$$

as $p \to p_c$ (see (4.28) and (4.33) in [10], or Proposition 34 in [13]). However, after closer inspection it is not immediately clear how to obtain the inequality $\alpha_4 \geq 1 + \frac{\alpha_1}{2}$ (or even $\alpha_4 > 1$).

We now explain how to obtain this inequality from the proof of (7) in [4]. Note that if we try to follow the proof of (8) in [12] instead, a difficulty arises. Indeed, the hypothesis (1.17) of Theorem 1.3 in [12] amounts to a lower bound on $\mathbb{P}_p(|\mathscr{C}(0)| \geq n)$, while our definition of α_1 involves an upper bound. As a consequence, we could not see how to use the reasonings in this paper (although it may be possible, we have not tried very hard).

Even though the paper [4] (see Section 5) assumes the existence of exponents, we were able to fix this issue, and we now sketch briefly how to do it. For that, we use the (now-classical) scaling relations

$$\chi(p) = \mathbb{E}_p\big[|\mathscr{C}(0)|\big] \asymp L(p)^2 \pi_1(L(p))^2 \quad \text{and} \quad \mathbb{E}_p\big[|\mathscr{C}(0)|^2\big] \asymp L(p)^4 \pi_1(L(p))^3 \tag{10}$$

as $p \nearrow p_c$ (this is (1.25) in [10], for $t = 1$ and $t = 2$ respectively). In addition, one also has

$$\frac{d\chi(p)}{dp} \asymp L(p)^2 \pi_4(L(p)) \cdot \chi(p). \tag{11}$$

Indeed, this can be proved by estimating $\frac{d}{dp}\mathbb{P}_p(0 \leftrightarrow v)$ for each $v \in \mathbb{Z}^2$, and then using similar reasonings as in [10]. For $p < p_c$, these relations can be combined with the following inequality from [4] (see p. 266):

$$\mathbb{E}_p\big[|\mathscr{C}(0)|^2\big] \geq \frac{K}{\chi(p)} \left(\frac{d\chi(p)}{dp} \right)^2, \tag{12}$$

for some universal constant $K \in (0, \infty)$. Hence, we get

$$\pi_4(L(p)) \leq K^{-1/2} L(p)^{-1} \pi_1(L(p))^{1/2}. \tag{13}$$

Since $L(p) \to \infty$ as $p \nearrow p_c$, this gives the desired inequality between α_1 and α_4.

As a conclusion, we want to stress that one drawback of this approach is that it requires the arm-separation result mentioned in Remark 1(2). Also, we used quite heavy results on the behavior of percolation near criticality to deduce an inequality which is purely about the behavior at criticality. Proofs "staying at criticality" are arguably more satisfying.

4 Other Proofs in the Literature

We now discuss four papers in the literature which show lower bounds on α_4 without using the quite heavy near-critical results in Kesten's paper [10].

The first three papers do this for bond percolation on the square lattice, and they are related to questions of noise sensitivity for a configuration at criticality. Presumably, after small modifications they also work for site percolation. We keep using the same notation $\pi_4(n)$ etcetera as we did for site percolation. These papers are: a paper by Benjamini et al. [2] (Sect. 4.1), a paper by Schramm and Steif [16] (Sect. 4.2), and an appendix by Garban in a paper by Schramm and Smirnov [15] (Sect. 4.3). For some of the results in these papers, we also refer the reader to Sections 6.2.2 and 8.5 in the book [6] by Garban and Steif.

Finally, we discuss a paper by Cerf [3] (Sect. 4.4), which is written for site percolation on the square lattice (and, more generally, on the hypercubic lattice $\mathbb{Z}^d$ in any $d \geq 2$). Contrary to the above-mentioned papers, this paper is mostly concerned with dimensions $d \geq 3$, but, as we explain, it still yields interesting properties in dimension $d = 2$.

Each of these papers uses some kind of exploration procedure in its proof of $\alpha_4 > 1$. And each of the first three papers uses Kesten's arm-separation result (see Remark 1(2)). The proofs from [2] and [3] use a concentration inequality, but the proofs in [16] and [15] do not. The main contribution by Garban in [15] is a multi-scale version of Theorem 1 (see Lemma 5 below).

The proofs in [16] and [15] seem to be, partly or indirectly, influenced by [2], but none of these three papers appears to be influenced by Aizenman et al. [1] or Gandolfi et al. [5]. On the other hand, [3] is influenced by these last two papers, but it seems to be completely independent of [2, 15, 16].

Throughout this section the percolation parameter is equal to the bond or site (depending on the context) percolation threshold on the square lattice, and we omit it from our notation.

4.1 The Benjamini-Kalai-Schramm Paper (1999)

The paper [2] is the first to give (for bond percolation on the square lattice) a proof of $\alpha_4 > 1$ without using the near-critical percolation results of [10].

Consider the event $A = A_m = \mathscr{C}_H([0, m+1] \times [0, m])$, and recall that an edge e is said to be pivotal for A if changing the state of e changes the occurrence, or not, of A. The following is shown in [2], where the only percolation knowledge used in the proof is the classical consequence from RSW that there exist $\rho, C > 0$ such that:

$$\text{for all } n \geq 1, \quad \mathbb{P}_{p_c}(0 \leftrightarrow \partial B_n) \leq C n^{-1/\rho} \tag{14}$$

(which follows immediately from (3)).

Proposition 1 ([2], equation (4.2) and Remark 4.2) *There is a constant $C > 0$ such that: for all $m \geq 1$,*

$$I(A) \leq C m^{1-1/3\rho} (\log m)^{3/2}, \tag{15}$$

where $I(A)$ is the expected number of pivotal edges for the event A.

It follows from Kesten's arm-separation result that each edge in, say, the $\frac{m}{2} \times \frac{m}{2}$ square centered in the middle of the large box has a probability of order $\pi_4(m)$ to be pivotal. Since the expected number of pivotal edges in that square is smaller than or equal to the l.h.s. of (15), we get $m^2\pi_4(m) \leq C'm^{1-1/3\rho}(\log m)^{3/2}$ (for some constant C') and hence,

$$\pi_4(m) \leq C'm^{-1-1/3\rho}(\log m)^{3/2}. \tag{16}$$

Recalling the meaning of ρ, this gives, in our earlier notation,

$$\alpha_4 \geq 1 + \frac{\alpha_1}{3}. \tag{17}$$

Proposition 1 is used in [2] to show that these box-crossing events are noise sensitive. An event $E \subseteq \Omega := \{0,1\}^n$ is said to be noise-sensitive if, roughly speaking, the following holds. For a large fraction of the configurations $\omega \in \Omega$, knowing ω does not significantly help to predict whether a perturbed configuration ω' (obtained from ω by randomly and independently flipping with small probability the "bits" ω_i, $i = 1, \ldots, n$) belongs to the event E.

The proof of Proposition 1 is somewhat spread over different locations in the paper. As indicated above, the main concern of the paper is noise sensitivity. The paper contains some theorems of an "algebraic" flavour (involving discrete Fourier analysis), which give, for a quite general setting (i.e. not specifically for percolation) sufficient conditions for noise sensitivity. This type of results, combined with Proposition 1, is essential to conclude noise sensitivity of the box-crossing events, but it is not needed for the proof of Proposition 1 itself. This makes it a bit hard to locate precisely those ingredients in the paper needed for the proof of Proposition 1 itself.

Another type of results in the paper is of a more probabilistic nature and gives, again in a quite general setting, upper bounds for the total influence, which can then be used to check if the earlier mentioned conditions for noise sensitivity hold. One of the latter results, used for the proof of Proposition 1, is the following Lemma 1. Let us first explain the notation in that lemma.

As before, $\Omega = \{0,1\}^n$, and the probability distribution considered is the product distribution with parameter $\frac{1}{2}$ (i.e. the uniform distribution on Ω). For a function $f : \Omega \to [0,1]$, and a subset K of $\{1, \ldots, n\}$, the notation $I_K(f)$ is used for $\sum_{k \in K} I_k(f)$, where

$$I_k(f) = \frac{1}{2^n} \sum_{\omega \in \Omega} \left| f(\omega) - f(\omega^{(k)}) \right|,$$

with $\omega^{(k)}$ the configuration obtained from ω by flipping ω_k (note that if f is the indicator function of an event, then $I_k(f)$ is the probability that k is pivotal for that event).

Finally, M_K is the majority function for K, which takes the value 1 if the family $(\omega_i)_{i\in K}$ has more 1's than 0's, the value -1 if it has more 0's than 1's, and the value 0 otherwise.

Lemma 1 ([2], Corollary 3.2 and Theorem 3.1) *Let* $K \subseteq \{1,\dots,n\}$*, and* $f : \Omega \to [0,1]$ *be monotone. Then, for some universal constant* C*,*

$$I_K(f) \leq C\sqrt{|K|}\,\mathbb{E}\big[f M_K\big]\left(1+\sqrt{-\log \mathbb{E}\big[f M_K\big]}\right). \tag{18}$$

The proof of Lemma 1 is self-contained and not very long (about one page), but certainly not obvious: it is a clever and surprising combination of nice elementary observations and standard concentration-like inequalities.

The other important ingredient in [2] for the proof of Proposition 1 is the following. This ingredient is very specific to the percolation setting mentioned before. Consider the $(m+1)\times m$ box in Proposition 1 and the crossing event A there.

Lemma 2 ([2], Two Lines Before equation (4.2)) *For each subset* K *of the set of edges in the right half of the* $(m+1)\times m$ *box,*

$$\mathbb{E}\big[\mathbb{1}_A M_K\big] \leq C\, m^{-1/3\rho}\log m, \tag{19}$$

where C *is some universal constant.*

Before we say a few words about the proof of Lemma 2, let us first see how Proposition 1 follows. Combining Lemmas 2 and 1 gives immediately

$$I_K(A) \leq C\sqrt{|K|}\, m^{-1/3\rho}(\log m)^{3/2}$$

for each subset K of the set of edges in the right half of the $(m+1)\times m$ box. By symmetry, it then also holds for every K in the left half of the box, and hence (with C replaced by $C\sqrt{2}$) for every K. Taking for K the set of all edges of the box gives Proposition 1.

As to the proof of Lemma 2, it is practically self-contained; the only percolation knowledge that it uses is (14). The main ingredients of the proof of Lemma 2 are an exploration argument (for the existence of a horizontal crossing in the box), and some necessary quantitative work, again (as in the proof of Lemma 1) including some concentration-like inequalities. The main idea in the proof is that, to detect whether or not there is a horizontal crossing, typically a very small portion of K is inspected. Indeed, in a simple exploration procedure, starting on the left side of the box, only edges of which at least one endpoint is connected to the left side of the box are inspected. Since each edge e of K is at a distance $\geq m/2$ from the left side of the box, the probability that it is inspected is at most of order $m^{-1/\rho}$. Using this it is shown that, typically, the "surplus" of 0's or 1's on the part of K inspected by the algorithm is much smaller than that on the rest of K, and therefore is unlikely to be decisive for the value of M_K. The mentioned concentration-like inequalities are used to make this precise.

4.2 Four-Arm Results in the Schramm-Steif Paper (2010)

The paper [16] studies the set of times at which an infinite cluster appears in a critical dynamical 2D percolation model. Noise sensitivity plays an important role in that study.

Some intermediate key results in this paper are stated in terms of discrete Fourier analysis (w.r.t. the Fourier-Walsh expansion). One such result is Theorem 1.8 in the paper. Let $\Omega = \{0, 1\}^n$ and let $f : \Omega \to \mathbb{R}$ be a function. Theorem 1.8 gives, for each $k \leq n$, an upper bound for the sum of the squares $\hat{f}(S)^2$ of the Fourier coefficients, over $S \subseteq \{1, \ldots, n\}$ with $|S| = k$. In the case where $k = 1$ and f is the indicator function of an increasing event A, one can use (as mentioned in the remark below Theorem 4.1 in [16]) that $\hat{f}(\{i\})$ is equal to the probability that i is pivotal for A. For that special case, Theorem 1.8 in [16] is as follows.

Lemma 3 (Special Case of [16], Theorem 1.8) *Let $\Omega = \{0, 1\}^n$ and let $E \subseteq \Omega$ be an increasing event. Further, let A be a randomized algorithm which determines, by a step-by-step procedure, whether a configuration ω belongs to E or not, and where at each step of the procedure, the value of exactly one ω_i is "revealed" (the choice of i may depend on the values of the ω_j's that have already been inspected at that stage). The algorithm stops as soon as it is known whether E occurs or not. Let δ_A be the maximum over all $i \in \{1, \ldots, n\}$ of the probability that i is inspected. Then*

$$\sum_{i=1}^{n} \mathbb{P}\big(i \text{ is pivotal for } E\big)^2 \leq \delta_A \, \mathbb{P}(E). \tag{20}$$

The proof of Theorem 1.8 in [16] is not long, and it is reasonably self-contained but quite subtle.

Another result in [16] which is relevant for obtaining bounds on four-arm probabilities is Theorem 4.1 in that paper. It gives a suitable "decision algorithm" A for the event that there is a horizontal open crossing of an $R \times R$ square. This algorithm needed special care because δ_A is the maximum revealment probability over *all* edges in the square (not only the edges in the concentric $\frac{R}{3} \times \frac{R}{3}$ square). More precisely, Theorem 4.1 says (in our notation) the following.

Lemma 4 ([16], Theorem 4.1) *For the above mentioned crossing event for site percolation on the triangular lattice, there is an algorithm A with $\delta_A \leq R^{-\frac{1}{4}+o(1)}$. For the similar event for bond percolation on the square lattice, there exists a constant $a > 0$ and an algorithm A with $\delta_A \leq R^{-a+o(1)}$.*

The paper [16] gives a proof for the statement on the triangular lattice, and says that the proof of the statement for the square lattice is similar. Note that the value $\frac{1}{4}$ in Lemma 4 is the two-arm exponent α_2 on the triangular lattice. From the proof of the lemma, it is not clear whether, in the case of the square lattice, we may take $a = \alpha_2$ in the above theorem. However, this *is* clear for the weaker lemma where δ_A is replaced by the maximum revealment probability over the edges in the earlier mentioned $\frac{R}{3} \times \frac{R}{3}$ square. Combining that weaker lemma with a suitable modification of Lemma 3 (where for E we take the event that there is an open crossing of an $R \times R$ square, we replace the sum in the l.h.s. of (20) by the smaller sum restricted to the vertices in the concentric $\frac{R}{3} \times \frac{R}{3}$ box, and δ_A is replaced as mentioned above), and then using Kesten's arm-separation result, gives $R^2 \pi_4(R)^2 \leq R^{-\alpha_2 + o(1)}$, and hence Theorem 1. See Corollary A.4 of [18] for such modifications.

4.3 The Result of Garban (2011)

In Appendix B of the paper [15] by Schramm and Smirnov, Garban gives a "multi-scale bound" on the four-arm probability for bond percolation on $\mathbb{Z}^2$. More precisely, let ε be such that there is a constant $c' > 0$ for which: for all $1 \leq m \leq n$, $\pi_2(m, n) \leq c' \left(\frac{m}{n}\right)^{2\varepsilon}$. The following is proved in [15].

Lemma 5 ([15], Appendix B) *There is a constant $c > 0$ such that:*

$$\text{for all } 1 \leq m \leq n, \quad \pi_4(m, n) \leq c \left(\frac{m}{n}\right)^{1+\varepsilon}. \tag{21}$$

For the special case $m = 1$, this gives $\alpha_4 \geq 1 + \frac{\alpha_2}{2}$. A nice aspect of Garban's proof is that it is completely focused on the problem in question, while the mentioned four-arm results in [2] and [16] were in some sense (versions of) intermediate results needed in the proof of some other results.

Interestingly, Garban says that:

> [The case $m = 1$] can be extracted from [10] as well as [2] or [16].

In fact, following his proof, but (roughly speaking) taking everywhere $m = 1$, is considerably simpler than extracting a full proof for that case from the mentioned papers. Apart from the fact that it uses Kesten's arm-separation results, it is probably the shortest and most elegant proof that $\alpha_4 \geq 1 + \frac{\alpha_2}{2}$. It avoids concentration results (which were used in Cerf's computation, see the next section). As Garban indicates, a key part in his proof, in that special case $m = 1$, is essentially an application of (or almost "equivalent" to the proof of) a quite general inequality of [14] (see also the remark following the proof of Proposition 6.6 in Section 8.5 of [6]).

4.4 A Result by Cerf (2015)

Lemma 5.2 in [3], that we now state in any dimension $d \geq 2$, gives the following result (recall that $\mathscr{C}_{n,n+\ell}$ is the collection of open clusters in $B_{n+\ell}$ connecting B_n and $\partial^{\text{in}} B_{n+\ell}$).

Lemma 6 ([3], Lemma 5.2) *Let $d \geq 2$, and consider site percolation on the hypercubic lattice $\mathbb{Z}^d$. For all $p \in (0,1)$, $n \geq 1$ and $\ell \geq 0$,*

$$\mathbb{P}_p\big(\mathscr{A}_4(A_{1,2n+\ell})\big) \leq \frac{2d(\log n)}{\sqrt{|B_n|}} \mathbb{E}_p\left[\sqrt{|\mathscr{C}_{n,n+\ell}|}\right] + \frac{4d}{p(1-p)} |B_n|^2 e^{-2(\log n)^2 p^2 (1-p)^2}. \tag{22}$$

Note that this result holds for any $p \in (0,1)$. For our purpose, we will restrict, but only later, to $d = 2$ and $p = p_c^{\text{site}}(\mathbb{Z}^2)$.

The proof of this lemma in [3] is completely self-contained, it assumes no percolation knowledge at all. It is a nice mixture of arguments with a combinatorial flavor, and application of a concentration inequality (see our comments later in this section). As Cerf remarks, a version of this result, with only the parameter n, not ℓ (or, more precisely, with $\ell = 0$), is somewhat hidden in the arguments of Gandolfi et al. [5] and Aizenman et al. [1], to prove the uniqueness of the infinite open cluster.

Following [3], taking $\ell = 0$ in (22) and using the trivial upper bound $|\partial^{\text{in}} B_n| \asymp n^{d-1}$ for $|\mathscr{C}_{n,n+\ell}|$ gives

$$\mathbb{P}_p\big(\mathscr{A}_4(A_{1,2n})\big) \leq c\, \frac{\log n}{\sqrt{n}}, \tag{23}$$

where c depends on the dimension d only.

The main contribution in [3] is to "bootstrap" (22) in a clever way: the inequality (23) is used to improve the above-mentioned trivial upper bound for $\mathbb{E}_p\big[\sqrt{|\mathscr{C}_{n,n+\ell}|}\big]$, which is then plugged into (22) to get an improvement of (23), then leading to an even better bound for $\mathbb{E}_p\big[\sqrt{|\mathscr{C}_{n,n+\ell}|}\big]$, and so on. The introduction by Cerf of the extra parameter ℓ seems to provide the flexibility needed to do this bootstrapping.

As pointed out in [3], for $d = 2$ the final result obtained in this way is $\alpha_4 \geq \frac{11}{21}$, which looks disappointing. However, the main focus in the paper is on dimensions $d \geq 3$, where the "bootstrapping" that we just explained does give interesting new results.

Nevertheless, it may be worth mentioning that, as we observed, (22) (and a modified version obtained from small changes in its proof) is also useful for the case $d = 2$ (even without using the bootstrapping), as we point out now.

First, note that for $d = 2$ and $p = p_c^{\text{site}}(\mathbb{Z}^2)$, $\mathbb{E}_p\big[\sqrt{|\mathscr{C}_{n,2n}|}\big]$ is uniformly bounded in n (so bootstrapping makes no sense for $d = 2$). So, for $d = 2$, (22), now with

$\ell = n$, actually gives

$$\pi_4(3n) \le \tilde{c}\,\frac{\log n}{n}$$

for some constant $\tilde{c}$, and hence $\alpha_4 \ge 1$.

As we point out next, one can, with a very small modification in the proof of (22), obtain $\alpha_4 \ge 1 + \frac{\alpha_1}{2}$. Lines 8–9 in Section 5 of [3] give an upper bound for the quantity

$$\sum_{C \in \mathscr{C}} \sqrt{|\bar{C} \cap B_n|}, \tag{24}$$

where $\mathscr{C} = \mathscr{C}_{n,2n}$ and we denote $\bar{C} := C \cup \partial^{\text{out}} C$. Namely (by Jensen's inequality), this quantity is at most

$$\sqrt{|\mathscr{C}|}\sqrt{\sum_{C \in \mathscr{C}} |\bar{C} \cap B_n|}, \tag{25}$$

which, since every vertex v belongs to at most $2d$ subsets $\bar{C}$ with $C \in \mathscr{C}$, is at most $\sqrt{|\mathscr{C}|}\sqrt{2d}\sqrt{|B_n|}$. So for the expectation of the sum in (24):

$$\mathbb{E}_{p_c}\Big[\sum_{C \in \mathscr{C}} \sqrt{|\bar{C} \cap B_n|}\Big] \le \mathbb{E}_{p_c}\Big[\sqrt{|\mathscr{C}|}\Big]\sqrt{2d}\sqrt{|B_n|}, \tag{26}$$

which is used later in [3] to obtain (22).

The "very small modification" that we meant is the following: by the Cauchy-Schwarz inequality, the expectation of (25) is at most

$$\mathbb{E}_{p_c}\Big[\sqrt{|\mathscr{C}|}\sqrt{\sum_{C \in \mathscr{C}} |\bar{C} \cap B_n|}\Big] \le \sqrt{\mathbb{E}_{p_c}[|\mathscr{C}|]}\sqrt{\mathbb{E}_{p_c}\Big[\sum_{C \in \mathscr{C}} |\bar{C} \cap B_n|\Big]}. \tag{27}$$

Since every $v \in \bigcup_{C \in \mathscr{C}} (\bar{C} \cap B_n)$ has an open path to $\partial^{\text{in}} B_{2n}$, the expectation in the second factor in (27) above is at most $2d|B_n|\pi_1(n)$. So we get that the expectation of (24) is at most

$$\mathbb{E}_{p_c}\Big[\sum_{C \in \mathscr{C}} \sqrt{|\bar{C} \cap B_n|}\Big] \le \sqrt{\mathbb{E}_{p_c}\big[|\mathscr{C}|\big]}\sqrt{2d}\sqrt{|B_n|}\sqrt{\pi_1(n)}. \tag{28}$$

Comparing this with the r.h.s. of (26) (and recalling that, for $d = 2$, $\mathbb{E}_{p_c}[|\mathscr{C}|]$ is uniformly bounded), we see that we made appear an extra factor $\sqrt{\pi_1(n)}$. This then also causes the same additional factor in the first term in the r.h.s. of (22), and yields

$$\alpha_4 \geq 1 + \frac{\alpha_1}{2}.$$

Finally, one gets (still for $d = 2$) a further improvement by considering, in the proof in [3], instead of $\bar{C}$, the set of vertices $\tilde{C} := C^* \cup C'$, where

$$C^* := \left\{ v^* \in \partial^{\text{out}} C \cap B_{2n} \ : \ v^* \leftrightarrow^* \partial^{\text{in}} B_{2n} \right\},$$

and

$$C' := \left\{ v \in C \ : \ v \sim v^* \text{ for some } v^* \in C^* \right\},$$

and then using that from every $v \in \bigcup_{C \in \mathscr{C}} (\tilde{C} \cap B_n)$, one can find an open path and a closed *-path (starting from neighbors of v) to $\partial^{\text{in}} B_{2n}$. This now produces, instead of the above-mentioned $\sqrt{\pi_1(n)}$, an extra factor $\sqrt{\pi_2(n)}$ in the first term in the r.h.s. of (22), so that we get

$$\alpha_4 \geq 1 + \frac{\alpha_2}{2}.$$

Comparing the case $m = 1$ of Garban's proof (mentioned in Sect. 4.3) of this inequality with the proof in [3] of (22), we observe that the latter avoids Kesten's arm-separation result, and is thus more self-contained. It uses a large-deviation argument which makes it longer, and which is, presumably, only useful for the case $d \geq 3$.

In the next section, we give a short and self-contained proof of $\alpha_4 \geq 1 + \frac{\alpha_2}{2}$, which can be considered as a combination of the proof of (21) (in the special case $m = 1$) in [15] and the proof of (22) in [3].

5 A Self-contained Proof of Theorem 1, Based on Garban's and Cerf's Arguments

5.1 Introductory Remarks

We follow Garban's proof for the result in Sect. 4.3, but restrict to the case $m = 1$, and replace the *event that there is a horizontal crossing of a box*, by the *number of connected components crossing an annulus*. The proof of Theorem 1 obtained in this way is, in some sense, a mixture of Garban's argument and that by Cerf: it still exploits, as in Garban's proof (which, as said, was inspired by O'Donnell and Servedio [14]), the full power of symmetry provided by involving the notion of pivotality, while it also uses the advantage of considering the number of crossings of an annulus (as Cerf did) instead of the event (considered by Garban) that there is a horizontal crossing of a box. This enables one to avoid Kesten's arm-separation result (we do not see how to avoid that result in the proof of Lemma 5 for a general $m \geq 1$). To underline the flexibility of the method, we deal with *site* percolation on the square lattice (which has less symmetry than bond percolation on that lattice), with parameter $p_c = p_c^{\text{site}}(\mathbb{Z}^2)$.

5.2 Proof

Let n be a positive integer, and let $\Omega = \{0, 1\}^{B_{2n}}$ be the set of all configurations of open and closed vertices in the box B_{2n}. Let $Z = |\mathscr{C}_{n,2n}|$ be the number of open clusters in B_{2n} that have at least one vertex in each of B_n and $\partial^{\text{in}} B_{2n}$. From (4), we know that for some universal $\bar{c} > 0$ (independent of n),

$$\mathbb{E}_{p_c}\left[Z^2\right] \leq \bar{c}^2. \tag{29}$$

Note that if we close an open vertex in B_{n-1}, the value of Z does not decrease. Let $v_1, v_2, \dots$ be a list of the vertices in B_{n-1}. For each $1 \leq j \leq |B_{n-1}|$, define the random variable C_j as follows:

$$C_j = \begin{cases} -(1 - p_c) & \text{if } v_j \text{ is open,} \\ p_c & \text{if } v_j \text{ is closed.} \end{cases}$$

In the remainder of this proof, $\{v_j \text{ is pivotal}\}$ denotes the event that if the state of v_j is changed, then the value of Z changes as well. More precisely,

$$\left\{v_j \text{ is pivotal}\right\} := \left\{\omega \in \Omega \ : \ Z(\omega^{(j)}) \neq Z(\omega)\right\},$$

where $\omega^{(j)}$ denotes the configuration obtained from ω by "flipping" ω_{v_j}.

We now consider an exploration procedure Γ which counts the number Z of open clusters in $\mathscr{C}_{n,2n}$. Roughly speaking, Γ is constructed so as to follow successively the boundaries (as depicted on Fig. 2) of all open clusters in B_{2n} that intersect $\partial^{\text{in}} B_{2n}$, starting from $\partial^{\text{in}} B_{2n}$. It has the property that each time it reaches a "fresh" vertex, the state of this vertex is revealed, open with probability p_c and closed with probability $1 - p_c$, independently of all information obtained so far in the procedure. We refer to Fig. 3, which shows an intermediate stage of this procedure, and where the vertices pivotal for Z are marked.

We let

$$Y_j := \mathbb{1}_{v_j \text{ is visited by } \Gamma}.$$

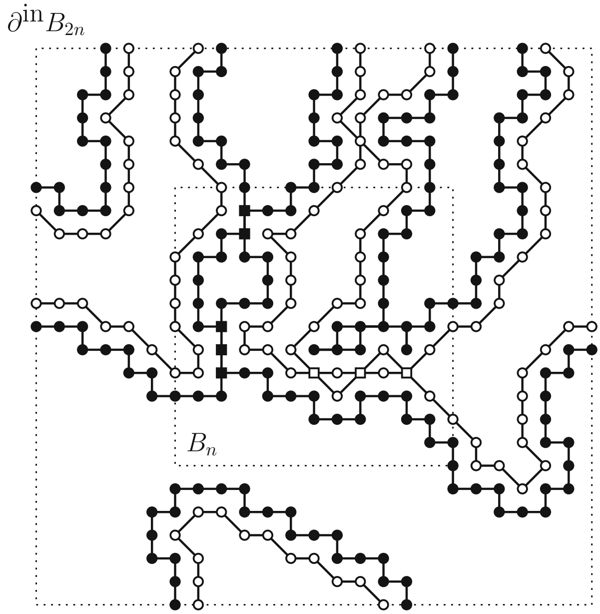


Fig. 3 This figures shows part of the exploration procedure Γ, which explores iteratively the "interfaces" between open clusters and closed *-clusters connected to $\partial^{\text{in}} B_{2n}$. The black vertices are open, and the white ones are closed. The vertices indicated with a square are pivotal for Z: changing the state of such a vertex v would increase or decrease the value of Z, depending on whether v is open or closed, respectively

Note that for each vertex v visited by Γ (and away from $\partial^{\text{in}} B_{2n}$), it is possible to find an open path and a closed *-path from neighbors (or *-neighbors) of v to $\partial^{\text{in}} B_{2n}$. Since each vertex in B_n is at a distance at least n from $\partial^{\text{in}} B_{2n}$, we obtain

$$\mathbb{E}_{p_c}\big[Y_j\big] \leq c\,\pi_2(n) \tag{30}$$

for some constant $c > 0$.

By the nature of the exploration path (the next step of the path depends only on the states of the vertices hit by the path so far),

$$Y_j(\omega^{(j)}) = Y_j(\omega). \tag{31}$$

In particular, C_j and Y_j are independent, and $\mathbb{E}_{p_c}[C_j Y_j] = 0$. For essentially the same reason, if v_i and v_j are two distinct vertices, then, at the first step in the procedure that one of these two vertices is hit, the Y- and C-values of the other vertex are conditionally independent, given all information obtained during the exploration so far. Because of this (and a similar argument for the case where neither v_i nor v_j is hit), we get:

$$\text{for all } i \neq j, \quad \mathbb{E}_{p_c}\big[(C_i Y_i)(C_j Y_j)\big] = 0. \tag{32}$$

We now study $\mathbb{E}_{p_c}[ZC_jY_j]$ (this is analogous to Garban's proof, but with Z instead of the indicator function of a crossing event). Clearly,

$$\mathbb{E}_{p_c}\big[ZC_jY_j\big] = \mathbb{E}_{p_c}\big[ZC_jY_j\ \mathbb{1}_{v_j \text{ is pivotal}}\big] + \mathbb{E}_{p_c}\big[ZC_jY_j\ \mathbb{1}_{v_j \text{ is not pivotal}}\big]. \tag{33}$$

Let $\omega \in \Omega$. As is easy to check (using (31)), we have

$$\mathbb{P}_{p_c}(\omega)C_j(\omega)Y_j(\omega) = -\mathbb{P}_{p_c}(\omega^{(j)})C_j(\omega^{(j)})Y_j(\omega^{(j)}).$$

On the one hand, if $\omega \in \{v_j \text{ is not pivotal}\}$, then $\omega^{(j)} \in \{v_j \text{ is not pivotal}\}$ as well, and $Z(\omega) = Z(\omega^{(j)})$. Hence, the contribution of the pair $(\omega, \omega^{(j)})$ to the second term in the r.h.s. of (33) is 0, from which it follows that this term is equal to 0. On the other hand, if $\omega \in \{v_j \text{ is pivotal}\}$, the state of v_j must be explored by Γ. Hence, the first term of (33) is equal to $\mathbb{E}_{p_c}\big[ZC_j\ \mathbb{1}_{v_j \text{ is pivotal}}\big]$.

Now let $\omega \in \{v_j \text{ is pivotal}\}$, and suppose that $\omega_{v_j} = 1$, so that $C_j(\omega) = -(1-p_c)$. Then also $\omega^{(j)} \in \{v_j \text{ is pivotal}\}$, but $C_j(\omega^{(j)}) = p_c$. It follows that the contribution of the pair $(\omega, \omega^{(j)})$ to the first term in (33) is $p_c(1-p_c)q\,(Z(\omega^{(j)}) - Z(\omega))$, where $q = q(\omega)$ denotes the probability of the configuration $(\omega_v)_{v \in B_{2n} \setminus \{v_j\}}$ (note that $q(\omega) = q(\omega^{(j)})$). Using that $Z(\omega^{(j)}) - Z(\omega) \geq 1$, and summing over all configurations in the event $\{v_j \text{ is pivotal}\}$, we obtain that the first term in the r.h.s. of (33) is larger than or equal to

$$p_c(1-p_c)\mathbb{P}_{p_c}(v_j \text{ is pivotal}).$$

By the above, and also observing that $\mathbb{P}_{p_c}(v_j \text{ is pivotal}) \geq \pi_4(3n)$ (indeed, if a vertex $v \in B_n$ has four arms to distance $3n$, then it has four arms to $\partial^{\text{in}} B_{2n}$, and so it is pivotal for Z), we conclude that

$$\mathbb{E}_{p_c}\big[ZC_jY_j\big] = \mathbb{E}_{p_c}\big[ZC_j\ \mathbb{1}_{v_j \text{ is pivotal}}\big] \geq p_c(1-p_c)\pi_4(3n). \tag{34}$$

The sum over j of the l.h.s. of (34) satisfies (for some constant $\hat{c} > 0$):

$$\begin{aligned}
\sum_j \mathbb{E}_{p_c}\big[ZC_jY_j\big] &\leq \sqrt{\mathbb{E}_{p_c}[Z^2]\mathbb{E}_{p_c}\Big[\Big(\sum_j C_jY_j\Big)^2\Big]} \\
&\leq \bar{c}\sqrt{\mathbb{E}_{p_c}\Big[\Big(\sum_j C_jY_j\Big)^2\Big]} = \bar{c}\sqrt{\mathbb{E}_{p_c}\Big[\sum_j C_j^2Y_j^2\Big]} \\
&\leq \bar{c}\sqrt{\mathbb{E}_{p_c}\Big[\sum_j Y_j^2\Big]} = \bar{c}\sqrt{\sum_j \mathbb{E}_{p_c}[Y_j]} \\
&\leq \hat{c}\sqrt{n^2\pi_2(n)} = \hat{c}\,n\sqrt{\pi_2(n)},
\end{aligned}$$

where the four inequalities follow, respectively, from the Cauchy-Schwarz inequality, (29), the fact that $|C_j| \leq 1$, and (30), and where the first equality follows from (32), and the second one from the fact that $Y_j^2 = Y_j$.

Since the sum over j of the r.h.s. of (34) is of order $n^2\pi_4(3n)$, we get that, for some universal constant $\tilde{c}$,

$$\pi_4(3n) \leq \frac{\tilde{c}}{n}\sqrt{\pi_2(n)}.$$

This (using also that π_4 is decreasing) completes the proof of Theorem 1.

Acknowledgments We thank Christophe Garban for valuable comments related to Sects. 4.2 and 4.3. Many of our discussions on the subject of this paper took place at the University of Cambridge in the Fall of 2019, and we thank the Department of Pure Mathematics and Mathematical Statistics for its hospitality.

References

1. Aizenman, M., Kesten, H., Newman, C.M.: Uniqueness of the infinite cluster and continuity of connectivity functions for short and long range percolation. Commun. Math. Phys. **111**(4), 505–531 (1987)
2. Benjamini, I., Kalai, G., Schramm, O.: Noise sensitivity of Boolean functions and applications to percolation. Inst. Hautes Études Sci. Publ. Math. **90**, 5–43 (1999)
3. Cerf, R.: A lower bound on the two-arms exponent for critical percolation on the lattice. Ann. Probab. **43**(5), 2458–2480 (2015)
4. Durrett, R., Nguyen, B.: Thermodynamic inequalities for percolation. Commun. Math. Phys. **99**(2), 253–269 (1985)
5. Gandolfi, A., Grimmett, G., Russo, L.: On the uniqueness of the infinite cluster in the percolation model. Commun. Math. Phys. **114**(4), 549–552 (1988)
6. Garban, C., Steif, J.E.: Noise Sensitivity of Boolean Functions and Percolation, Volume 5 of Institute of Mathematical Statistics Textbooks. Cambridge University Press, New York (2015)
7. Grimmett, G.: Percolation, Volume 321 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 2nd edn. Springer, Berlin (1999)
8. Kesten, H.: Percolation Theory for Mathematicians, Volume 2 of Progress in Probability and Statistics. Birkhäuser, Boston (1982)
9. Kesten, H.: A scaling relation at criticality for 2D-percolation. In: Percolation Theory and Ergodic Theory of Infinite Particle Systems (Minneapolis, Minn., 1984–1985), Volume 8 of IMA Vol. Math. Appl., pp. 203–212. Springer, New York (1987)
10. Kesten, H.: Scaling relations for 2D-percolation. Commun. Math. Phys. **109**(1), 109–156 (1987)
11. Lawler, G.F., Schramm, O., Werner, W.: One-arm exponent for critical 2D percolation. Electron. J. Probab. **7**(2), 13 pp. (2002)
12. Newman, C.M.: Inequalities for γ and related critical exponents in short and long range percolation. In: Percolation Theory and Ergodic Theory of Infinite Particle Systems (Minneapolis, Minn., 1984–1985), Volume 8 of IMA Vol. Math. Appl., pp. 229–244. Springer, New York (1987)
13. Nolin, P.: Near-critical percolation in two dimensions. Electron. J. Probab. **13**(55), 1562–1623 (2008)

14. O'Donnell, R., Servedio, R.A.: Learning monotone decision trees in polynomial time. SIAM J. Comput. **37**(3), 827–844 (2007)
15. Schramm, O., Smirnov, S.: On the scaling limits of planar percolation. Ann. Probab. **39**(5), 1768–1814 (2011). With an appendix by Christophe Garban
16. Schramm, O., Steif, J.E.: Quantitative noise sensitivity and exceptional times for percolation. Ann. Math. (2) **171**(2), 619–672 (2010)
17. Smirnov, S., Werner, W.: Critical exponents for two-dimensional percolation. Math. Res. Lett. **8**(5–6), 729–744 (2001)
18. Vanneuville, H.: Annealed scaling relations for Voronoi percolation. Electron. J. Probab. **24**(39), 71 pp. (2019)

Universality of Noise Reinforced Brownian Motions

Jean Bertoin

Abstract A noise reinforced Brownian motion is a centered Gaussian process $\hat{B} = (\hat{B}(t))_{t\geq 0}$ with covariance

$$\mathbb{E}(\hat{B}(t)\hat{B}(s)) = (1-2p)^{-1} t^p s^{1-p} \quad \text{for} \quad 0 \leq s \leq t,$$

where $p \in (0, 1/2)$ is a reinforcement parameter. Our main purpose is to establish a version of Donsker's invariance principle for a large family of step-reinforced random walks in the diffusive regime, and more specifically, to show that $\hat{B}$ arises as the universal scaling limit of the former. This extends known results on the asymptotic behavior of the so-called elephant random walk.

Keywords Reinforcement · Brownian motion · Invariance principle · Elephant random walk

Mathematics Subject Classification 60G50, 60G51, 60K35

1 Introduction

This work concerns a rather simple real-valued and centered Gaussian process $\hat{B} = (\hat{B}(t))_{t\geq 0}$ with covariance function

$$\mathbb{E}(\hat{B}(t)\hat{B}(s)) = \frac{t^p s^{1-p}}{1-2p} \qquad \text{for } 0 \leq s \leq t, \tag{1}$$

where $p \in (0, 1/2)$ is a fixed parameter. Recently, this process has notably appeared as the scaling limit for diffusive regimes of the so-called elephant random

J. Bertoin (✉)
Institute of Mathematics, University of Zurich, Zürich, Switzerland
e-mail: jean.bertoin@math.uzh.ch

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_7

walk, a simple random walk with memory that has been introduced by Schütz and Trimper [21]. Just as the standard Brownian motion B corresponds to the integral of a white noise, $\hat{B}$ can be thought of as the integral of a reinforced version of the white noise, hence the name *noise reinforced Brownian motion*. Here, reinforcement means that the noise tends to repeat itself infinitesimally as time passes; we refer to [19] for a survey of various models of stochastic processes with reinforcement and their applications, and to [16] and works cited therein for more recent contributions in this area. The parameter p should be interpreted as the strength of the reinforcement; specifically, it represents the probability that an infinitesimal portion of the noise is a repetition. In the limiting case $p = 0$ without reinforcement, one just recovers the standard Brownian motion.

Our purpose here is twofold. We will first present several basic properties of $\hat{B}$ that mirror well-known facts for the standard Brownian motion, even though the laws of B and $\hat{B}$ are mutually singular. We will then establish a version of Donsker's invariance principle. That is, we will show that any so-called step-reinforced random walk with reinforcement parameter p, whose typical step has a finite second moment, converges after the usual centering and rescaling to a noise reinforced Brownian motion.

This invariance principle for step-reinforced random walks has been established previously for elephant random walks, that is in the special case when the typical step has the Rademacher law; see [2, 9]. Technically, [9] uses Skorokhod's embedding and relies crucially on the assumption that the increments take values in $\{-1, +1\}$, whereas [2] uses limit theorems for generalized Pólia urns due to Janson [14]. The latter approach is not available for arbitrary step distributions either, notably as one would need to work with urn models having types in an infinite space (the real line $\mathbb{R}$ to be more specific), and how to deal with this kind of urns is still an open problem; see Remark 4.1 in [14].

We shall therefore follow here a different method and rather embed a step-reinforced random walk in a branching process with types in $\mathbb{R}$, using a time-substitution via an independent Yule process. This yields a remarkable martingale, whose quadratic variation can then be estimated from well-known properties of Yule processes. In turn, this enables the application of the martingale functional central limit theorem (later on referred to as martingale FCLT).

2 Some Basic Properties

We start by observing from (1) that the noise reinforced Brownian motion admits a simple representation as a Wiener integral, namely

$$\hat{B}(t) = t^p \int_0^t s^{-p} \mathrm{d}B(s), \qquad t \geq 0, \tag{2}$$

where $B = (B(s))_{s\geq 0}$ is a standard Brownian motion. Note that equivalently,

$$\hat{B} \text{ has the same law as } \left(\frac{t^p}{\sqrt{1-2p}} B(t^{1-2p})\right)_{t\geq 0}. \tag{3}$$

It follows immediately from (3) and the classical law of the iterated logarithm for B (see, e.g. Theorem II.1.19 in [20]) that

$$\limsup_{t\to\infty} \frac{\hat{B}(t)}{\sqrt{2t \ln \ln t}} = \limsup_{t\to 0+} \frac{\hat{B}(t)}{\sqrt{2t \ln \ln(1/t)}} = \frac{1}{\sqrt{1-2p}} \qquad \text{a.s.} \tag{4}$$

In particular, we see that for different reinforcement parameters p, the distributions of noise reinforced Brownian motions, say on the time interval [0, 1], yield laws on $\mathscr{C}([0, 1])$ which are mutually singular, and are also singular with respect to the Wiener measure.

We deduce from (2) by stochastic calculus that $\hat{B}$ is a semi-martingale which solves the stochastic differential equation

$$\mathrm{d}\hat{B}(t) = \mathrm{d}B(t) + \frac{p}{t}\hat{B}(t)\mathrm{d}t, \qquad \hat{B}(0) = 0. \tag{5}$$

This shows that $\hat{B}$ is actually a time-inhomogeneous diffusion process with quadratic variation $\langle \hat{B}\rangle(t) = t$. In this direction, we also infer from (1) that the process

$$\hat{b}(t) = \hat{B}(t) - t^{1-p}\hat{B}(1) \qquad \text{for } 0 \leq t \leq 1,$$

is independent of $\hat{B}(1)$. Hence, for any $x \in \mathbb{R}$,

$$\hat{b}(t) + t^{1-p}x = \hat{B}(t) + t^{1-p}(x - \hat{B}(1)) \qquad \text{for } 0 \leq t \leq 1,$$

is a version of the bridge of $\hat{B}$ from 0 to x with unit duration. These properties should be viewed as the reinforced versions of the classical construction of Brownian bridges; see for instance [20] on page 37.

We further see from (1) that the scaling property,

$$\text{for every } c > 0, \ \left(c^{-1}\hat{B}(c^2 t)\right)_{t\geq 0} \text{ has the same law as } \hat{B}, \tag{6}$$

as well as the time-inversion property,

$$\left(t\hat{B}(1/t)\right)_{t>0} \text{ has the same law as } \left(\hat{B}(t)\right)_{t>0},$$

both hold. Needless to say, these two properties are also fulfilled by the standard Brownian motion; see, e.g. Proposition I.1.10 in [20]. In this vein, we also point at a remarkable connection with Ornstein-Uhlenbeck processes: the process

$$\hat{U}(t) = \mathrm{e}^{-t/2}\hat{B}(\mathrm{e}^t)\,, \qquad t \in \mathbb{R}$$

is a stationary Ornstein-Uhlenbeck process with infinitesimal generator

$$\mathscr{G}f(x) = \frac{1}{2}f''(x) + (p - 1/2)xf'(x),$$

where $f \in \mathscr{C}^2(\mathbb{R})$. This can be checked by stochastic calculus from (5), or directly by observing from (1) and (6) that $\hat{U}$ is a stationary Gaussian process with covariance

$$\mathbb{E}(\hat{U}(t)\hat{U}(0)) = \frac{\mathrm{e}^{(p-1/2)t}}{1-2p} \qquad \text{for } t \geq 0.$$

On the other hand, several classical results for the standard Brownian motion plainly fail for its reinforced version. For instance, the increments of $\hat{B}$ are clearly not independent, and the time-reversal property (e.g. Exercise 1.11 in Chapter I of [20]) also fails.

3 An Invariance Principle with Reinforcement

The main purpose of this section is to point out that noise reinforced Brownian motions arise as universal scaling limits of a large class of random walks with step reinforcement. We first recall some features on the so-called elephant random walk, where the story begins.

The *elephant random walk* has been introduced by Schütz and Trimper [21] as a discrete-time nearest neighbor process with memory on $\mathbb{Z}$; it can be depicted as follows. Fix some $q \in (0, 1)$ and call q the memory parameter. Imagine that a walker (an elephant) makes a first step in $\{-1, +1\}$ at time 1; then at each time $n \geq 2$, it selects randomly a step from its past. With probability q, the elephant repeats this step, and with complementary probability $1 - q$, it makes the opposite step. Note that for $q = 1/2$, the elephant merely follows the path of a simple symmetric random walk. The elephant random walk has generated much interest in the recent years, we refer notably to [2, 3, 7–9, 17, 18], see also [1, 4–6, 12] for variations, and references therein for further related works. A remarkable feature is that the large time asymptotic behavior of an elephant random walk is diffusive when the memory parameter q is less than $3/4$ and super-diffusive when $q > 3/4$.

Remark 1 The laws of the iterated logarithm (4) for a noise reinforced Brownian motion bear the same relation to that for the elephant random walk in the diffusive regime (Corollary 1 in [9] and Theorem 3.2 in [3]; beware that the memory parameter denoted by p there corresponds to $q = (1 + p)/2$ in the present notation as it will be stressed below), as the law of the iterated logarithm for the standard Brownian motion due to P. Lévy does to that for the simple random walk due to A.Y. Khintchine.

We are interested in a generalization where the distribution of a typical step of the walk is arbitrary, that we call *step reinforced random walk*. Fix a parameter $p \in (0,1)$, called the *reinforcement parameter*;[1] at each discrete time, with probability p, a step reinforced random walk repeats one of its preceding steps chosen uniformly at random, and otherwise, i.e. with probability $1-p$, it has an independent increment with a fixed distribution. More precisely, consider a sequence $X_1, X_2, \dots$ of i.i.d. copies of a random variable X in $\mathbb{R}$ and define recursively $\hat{X}_1, \hat{X}_2, \dots$ as follows. Let $(\varepsilon_i : i \geq 2)$ be an independent sequence of Bernoulli variables with parameter p. We set first $\hat{X}_1 = X_1$, and next for $i \geq 2$, we let $\hat{X}_i = X_i$ if $\varepsilon_i = 0$, whereas we define $\hat{X}_i$ as a uniform random sample from $\hat{X}_1, \dots, \hat{X}_{i-1}$ if $\varepsilon_i = 1$. Finally, the sequence of the partial sums

$$\hat{S}(n) = \hat{X}_1 + \cdots + \hat{X}_n, \qquad n \in \mathbb{N},$$

is referred to as a step reinforced random walk. We stress that in general, the Markov property fails for $\hat{S}$ (even though it may hold for certain specific step distributions).

When the typical step X has the Rademacher law, i.e. $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = 1/2$, Kürsten [18] (see also [10]) pointed out that $\hat{S}$ is a version of the elephant random walk with memory parameter $q = (p+1)/2$ in the present notation. When X has a symmetric stable distribution, $\hat{S}$ is the so-called shark random swim which has been studied in depth by Businger [6]. More general versions when the distribution of X is infinitely divisible have been considered in [5].

The main result of Businger [6] is that the large time asymptotic behavior of a shark random swim exhibits a phase transition similar to that for the elephant random walk, but for a different critical parameter. We shall now extend this to a large class of step reinforced random walks. In the sequel we implicitly rule out the degenerate case when the typical step variable X is a constant. We start with the super-diffusive regime for which all the ingredients are already in Section 3.1 in [6].

Theorem 1 *Let $p \in (1/2, 1)$, and suppose that $X \in L^2(\mathbb{P})$. Then*

$$\lim_{n\to\infty} \frac{\hat{S}(n) - n\mathbb{E}(X)}{n^p} = L \qquad \text{in } L^2(\mathbb{P}),$$

where L is some non-degenerate random variable.

[1] Beware that we assumed $p < 1/2$ in the preceding sections. The first part of the present section (super-diffusive regime) does not involve any noise reinforced Brownian motion, and the case $p \geq 1/2$ is allowed. In the second part of this section (diffusive regime), we shall again focus on the case $p < 1/2$ and noise reinforced Brownian motions will then re-appear.

Proof By centering and normalizing, we may assume without loss of generality that $\mathbb{E}(X) = 0$ and $\mathrm{Var}(X) = 1$. One first introduces for every $i, n \in \mathbb{N}$, the number $r_{i,n}$ of repetitions of the variable X_i in the reinforced sequence $\hat{X}_1, \ldots, \hat{X}_n$ (in particular, $r_{i,n} = 0$ when either $\varepsilon_i = 1$ or $i > n$). We stress that repetition numbers only depend on the Bernoulli variables ε_j and on the uniform random samples among the previous steps, and are hence independent of the variables X_j. We know from Lemmas 3 and 5 in [6] that for each $i \in \mathbb{N}$

$$\lim_{n\to\infty} n^{-p} r_{i,n} = R_i, \qquad \text{a.s.}$$

where R_i is some non-degenerated random variable with $\sum_i \mathbb{E}\left(R_i^2\right) < \infty$.

In particular, $\sum_i R_i^2 < \infty$ a.s., and since the variables X_i are i.i.d. centered and with unit variance, the sum $\sum_i R_i X_i = L$ is well-defined a.s. (as a martingale limit, conditionally on the R_i). More precisely, $\mathbb{E}(L) = 0$ and $\mathbb{E}(L^2) = \sum_i \mathbb{E}\left(R_i^2\right)$.

To conclude, we recall from Equation (6) in [6] that

$$\lim_{n\to\infty} \sum_i \mathbb{E}\left(\left|n^{-p} r_{i,n} - R_i\right|^2\right) = 0. \tag{7}$$

By the construction of the step reinforced random walk, we have

$$\hat{S}(n) = \sum_i r_{i,n} X_i.$$

Since the variables $r_{i,n}$ and R_i are independent of the X_i, and further $\mathbb{E}(X) = 0$ and $\mathbb{E}(X^2) = 1$, there is the identity

$$\mathbb{E}\left(|n^{-p}\hat{S}(n) - L|^2\right) = \mathbb{E}\left(\sum_i \left|n^{-p} r_{i,n} - R_i\right|^2\right).$$

An appeal to (7) completes the proof.

We then turn our attention to the diffusive regime, and obtain a version of Donsker's invariance principle for step reinforced random walks. In this direction, we refer to Chapter VI in [13] for background on the Skorokhod topology and weak convergence of stochastic processes.

Theorem 2 *Let $p \in (0, 1/2)$, and suppose that $X \in L^2(\mathbb{P})$. Then as $n \to \infty$, the sequence of processes*

$$\frac{\hat{S}(\lfloor tn \rfloor) - tn\mathbb{E}(X)}{\sqrt{n\mathrm{Var}(X)}}, \qquad t \geq 0$$

converges in distribution in the sense of Skorohod towards a noise reinforced Brownian motion $\hat{B}$ with reinforcement parameter p.

Theorem 2 will be established in the next section, first under the additional assumption that the typical step is bounded, and then in the general situation. Our approach relies essentially on the martingale FCLT, which also plays a key role for urn models (see [11, 14]), as well in the works of Coletti et al. [8, 9] and of Bercu [3] on the elephant random walk. For the sake of simplicity, we consider here the one-dimensional setting only, however the argument could be readily adapted to $\mathbb{R}^d$ using the Cramér-Wold device; see also [4].

We first embed the step reinforced random walk in a branching process as follows. Let $Y = (Y_t)_{t\geq 0}$ be a standard Yule process, that is Y is a pure birth process started from $Y_0 = 1$, with birth rate n from any state $n \in \mathbb{N}$. We further assume that Y and $\hat{S}$ are independent, and will mainly work with the time-changed process $\hat{S}(Y_\cdot)$. It may be worth dwelling a bit on our motivation for introducing this time-substitution; even though this discussion will not be used in the proof, it nonetheless provides a useful guiding line for our approach. Roughly speaking, Y describes a population model started from a single ancestor, where individuals are eternal, and each begets a child at unit rate, independently of the other individuals. Those individuals are naturally enumerated in the increasing order of their birth time, in particular the ancestor is the first individual. We decide to assign the type $\hat{X}_n$ to the n-th individual, and write

$$\mathbf{Z}_t(\mathrm{d}x) = \sum_{n=1}^{Y_t} \delta_{\hat{X}(n)}(\mathrm{d}x), \qquad x \in \mathbb{R},$$

for the point process of the types of individuals alive at time t. We see from basic properties of independent exponential clocks that $\mathbf{Z} = (\mathbf{Z}_t)_{t\geq 0}$ is a (multitype) branching process, in which each individual begets a child at unit rate, independently of the other individuals. A child is either a clone of its parent, an event which occurs with probability p, or a mutant, an event which occurs with probability $1-p$. If a child is a clone, then its type is the same as that of its parent, whereas if it is a mutant, its type is given by an independent copy of X. In this setting, there are the identities

$$Y_t = \mathbf{Z}_t(\mathbf{1}) \quad \text{and} \quad \hat{S}(Y_t) = \mathbf{Z}_t(\mathrm{Id}), \tag{8}$$

with the notation $\mathbf{1}(x) = 1$, $\mathrm{Id}(x) = x$ and $\mathbf{Z}_t(f) = \int_{\mathbb{R}} f(x)\mathbf{Z}_t(\mathrm{d}x)$. The function $\mathbf{z} \mapsto \mathbf{z}(\mathbf{1})$ on the space of point measures $\mathbf{z}$ is always an eigenfunction for the infinitesimal generator of the branching process $\mathbf{Z}$ for the eigenvalue 1. The key point is that when the typical step is centered, $\mathbb{E}(X) = 0$, the function $\mathbf{z} \mapsto \mathbf{z}(\mathrm{Id})$ is also an eigenfunction, now for the eigenvalue p.

4 Proof of the Invariance Principle

This section is devoted to the Proof of Theorem 2. Without loss of generality, we henceforth assume that $p \in (0, 1/2)$ and $X \in L^2(\mathbb{P})$ with $\mathbb{E}(X) = 0$, and set $\sigma^2 = \mathbb{E}(X^2)$. We shall first give some preliminary estimates related to a remarkable martingale, then we shall establish Theorem 2 under the additional assumption that the variable X is bounded, and finally we shall show how this constraint can be removed.

4.1 On a Remarquable Martingale

Recall that we implicitly assume that X is centered.

Lemma 1 *The process*

$$M(t) = \mathrm{e}^{-pt}\hat{S}(Y_t), \quad \textit{for } t \geq 0,$$

is a square integrable martingale with finite variation.

Proof Obviously, the set of jump times of M is discrete. Since M decays continuously between two consecutive jump times, *a fortiori* its paths have finite variation. Plainly, $\mathbb{E}(\hat{S}(n)^2) \leq n^2\mathbb{E}(X^2) = n^2\sigma^2$, and since $\hat{S}$ and Y are independent with $\mathbb{E}(Y_t^2) < \infty$, $M(t)$ is indeed square integrable.

We next point out that for any $n \in \mathbb{N}$,

$$\begin{aligned}\mathbb{E}(\hat{X}_{n+1} \mid \hat{X}_1, \ldots, \hat{X}_n) &= (1-p)\mathbb{E}(X) + p\frac{\hat{X}_1 + \cdots + \hat{X}_n}{n} \\ &= p\frac{\hat{S}(n)}{n}\end{aligned}$$

(this observation is also the starting point of the analysis of the elephant random walk in [3, 8, 9]). Since the Yule process has precisely jump rate n from the state n, this entails that

$$\hat{S}(Y_t) - p\int_0^t \hat{S}(Y_s)\mathrm{d}s, \quad \text{for } t \geq 0$$

is a martingale, and our statement then follows from elementary stochastic calculus.

We next derive numerical bounds for the second moment of the supremum process of the martingale M.

Lemma 2 *For every $t \geq 0$, one has*

$$\mathbb{E}\left(\sup_{0\leq s\leq t} M^2(s)\right) \leq \frac{4\sigma^2}{1-2p}\mathrm{e}^{(1-2p)t}.$$

Proof Since M has finite variation, its square-bracket process can be expressed in the form

$$[M](t) = \int_{(0,t]} \mathrm{e}^{-2ps}|\hat{X}_{Y_s}|^2 \mathrm{d}Y_s; \tag{9}$$

see for instance Theorem 26.6(viii) in [15]. Recall that the instantaneous jump rate of the Yule process at time s equals Y_s and that the sequence $\hat{X}_1, \hat{X}_2, \dots$ is independent of the Yule process with $\mathbb{E}(\hat{X}_j^2) = \sigma^2$ for every $j \geq 1$. It follows that

$$\mathbb{E}([M](t)) = \sigma^2 \int_0^t \mathrm{e}^{-2ps}\mathbb{E}(Y_s)\mathrm{d}s = \frac{\sigma^2}{1-2p}\left(\mathrm{e}^{(1-2p)t} - 1\right),$$

where for the second equality, we used $\mathbb{E}(Y_s) = \mathrm{e}^s$. This yields our claim by an appeal to the Burkholder-Davis-Gundy inequality; see, e.g. Theorem 26.12 in [15].

We then obtain bounds for the second moments of the supremum process of the step reinforced random walk itself, which will be useful later one.

Corollary 1 *For every $n \geq 2$, we have*

$$\mathbb{E}\left(\max_{k\leq n} |\hat{S}(k)|^2\right) \leq \frac{4\mathrm{e}^a\sigma^2}{1-2p}n,$$

with $a = -\inf_{0<x\leq 1/2} x^{-1}\ln(1-x)$.

Proof Since $\hat{S}$ and Y are independent and Y is a counting process, we have for every $n \geq 1$ and $t \geq 0$ that

$$\mathbb{E}(\max_{k\leq n} |\hat{S}(k)|^2)\mathbb{P}(Y_t > n) \leq \mathbb{E}(\sup_{s\leq t} |\hat{S}(Y_s)|^2) \leq \mathrm{e}^{2pt}\mathbb{E}\left(\sup_{0\leq s\leq t} M^2(s)\right).$$

Take $t = \ln n$ and recall that Y_t hat the geometric distribution with parameter $\mathrm{e}^{-t} = 1/n$. So

$$\mathbb{P}(Y_t > n) = (1-1/n)^n \geq \mathrm{e}^{-a} \qquad \text{for all } n \geq 2,$$

and we conclude the proof using Lemma 2.

4.2 Proof of Theorem 2 When X Is Bounded

In this section, we shall prove Theorem 2 under the additional assumption that the typical step X is a bounded variable. We first estimate the angle-bracket $\langle M \rangle$ of M, and in this direction, we introduce

$$\hat{V}(n) = \hat{X}(1)^2 + \cdots + \hat{X}(n)^2, \qquad n \in \mathbb{N}.$$

Lemma 3 *Assume that $\|X\|_\infty < \infty$. We have*

$$\mathbb{E}\left(\left|e^{-pt}\hat{V}(Y_t) - (1-p)\sigma^2 \int_0^t e^{-ps} Y_s \mathrm{d}s\right|^2\right) = o(e^{2(1-p)t}) \qquad \textit{as } t \to \infty.$$

Proof Just as in the proof of Lemma 1, we note the identity

$$\mathbb{E}(\hat{X}(n+1)^2 \mid \hat{X}(1), \dots, \hat{X}(n)) = p\frac{\hat{V}(n)}{n} + (1-p)\sigma^2,$$

and get by stochastic calculus that the process

$$M'(t) = e^{-pt}\hat{V}(Y_t) - (1-p)\sigma^2 \int_0^t e^{-ps} Y_s \mathrm{d}s, \qquad t \geq 0$$

is a square integrable martingale with finite variation and square bracket

$$[M'](t) = \int_{(0,t]} e^{-2ps} |\hat{X}(Y_s)|^4 \mathrm{d}Y_s, \qquad t \geq 0.$$

We compute its expected value and get

$$\begin{aligned}\mathbb{E}(X^4)\mathbb{E}\left(\int_{(0,t]} e^{-2ps} \mathrm{d}Y_s\right) &= \mathbb{E}(X^4) \int_0^t e^{(1-2p)s} \mathrm{d}s \\ &= o(e^{2(1-p)t}),\end{aligned}$$

where for the first equality, we used that the instantaneous jump rate of the Yule process at time s equals Y_s and $\mathbb{E}(Y_s) = e^s$.

We next recall that the asymptotic behavior of the Yule process is described by

$$\lim_{t\to\infty} e^{-t} Y_t = \tau \qquad \text{a.s.} \tag{10}$$

where τ is a standard exponential variable. Needless to say, τ is independent of $\hat{S}$.

Corollary 2 *Assume that $\|X\|_\infty < \infty$. The angle-bracket $\langle M\rangle$ of the martingale M in Lemma 1 fulfills*

$$\lim_{t\to\infty} \mathrm{e}^{-(1-2p)t}\langle M\rangle(t) = \frac{\sigma^2\tau}{1-2p} \qquad \textit{in probability.}$$

Proof Recall that the square-bracket process of M is given by (9); if follows readily that its predictable compensator is

$$\langle M\rangle(t) = \int_{(0,t]} \mathrm{e}^{-2ps}\left(p\hat{V}(Y_s) + (1-p)\sigma^2 Y_s\right)\mathrm{d}s, \qquad t \geq 0. \tag{11}$$

Since $2p < 1$, our statement now derives from (10), (11) and Lemma 3.

The angle-bracket $\langle M\rangle$ is a continuous strictly increasing bijection from $\mathbb{R}_+$ to $\mathbb{R}_+$ a.s., see (11), and we write T for the inverse bijection. We introduce for each $n \in \mathbb{N}$ the process

$$N_n(t) = n^{-1/2}M_{T(nt)} = n^{-1/2}\mathrm{e}^{-pT(nt)}\hat{S}(Y_{T(nt)}), \qquad t \geq 0.$$

We are in position of applying the martingale FCLT.

Proposition 1 *Assume that $\|X\|_\infty < \infty$. As $n \to \infty$, the sequence of processes N_n converges in distribution in the sense of Skorohod to a standard Brownian motion.*

Proof Plainly, each N_n is a square-integrable martingale with angle-bracket $\langle N_n\rangle(t) = t$, and obviously its maximum jump is asymptotically negligible in $L^2(\mathbb{P})$ since X is bounded. This enables us to apply the martingale FCLT; see e.g. Theorem 2.1 in [22], and also Section VIII.3 in [13] for more general versions.

We shall actually need a slightly stronger version of Proposition 1 in which the convergence holds conditionally on the variable τ in (10).

Corollary 3 *Assume that $\|X\|_\infty < \infty$. As $n \to \infty$, the sequence of pairs (τ, N_n) converges in distribution in the sense of Skorohod to (τ, B) where B is a standard Brownian motion independent of τ.*

Proof First fix $t > 0$ and consider a random variable $A(t)$ which is measurable with respect to the sigma-algebra $\sigma(M_{T(s)} : 0 \leq s \leq t)$. The statement with $A(t)$ replacing τ follows from the martingale FCTL applied to the sequence of processes $(N_n(s+t/n))_{s\geq 0}$, just as in Proposition 1. Corollary 3 can then be deduced, provided that we can choose $A(t)$ such that

$$\lim_{t\to\infty} A(t) = \tau \quad \text{a.s.} \tag{12}$$

Specifically, consider any continuous functional F on the Skorohod space with $|F| \leq 1$, and any continuous and bounded function f on $\mathbb{R}$. Taking (12) for granted, for any $\eta > 0$ arbitrarily small, we can first choose $t > 0$ sufficiently large so that

$$\mathbb{E}(|f(A(t)) - f(\tau)|) \leq \eta,$$

and then $n(t) \in \mathbb{N}$ such that

$$|\mathbb{E}\,(f(A(t))F(N_n)) - \mathbb{E}(f(A(t)))\mathbb{E}(F(B)))| \leq \eta \qquad \text{for all } n \geq n(t).$$

We can conclude from the triangle inequality that

$$|\mathbb{E}\,(f(\tau)F(N_n)) - \mathbb{E}(f(\tau))\mathbb{E}(F(B)))| \leq 3\eta \qquad \text{for all } n \geq n(t).$$

We now have to construct variables $A(t)$ so that (12) holds, and in this direction, we first assume that $\mathbb{P}(X = 0) = 0$. All the steps of $\hat{S}$ are non-zero a.s., and $Y_{T(t)}$ is then the total number of jumps of the process $M_{T(\cdot)}$ on the time interval $[0, t]$ (recall that this takes the initial jump at time 0 into account), and hence is measurable with respect to $\sigma(M_{T(s)} : 0 \leq s \leq t)$. On the other hand, we deduce from Corollary 2 that

$$T(t) = \frac{1}{1-2p} \ln\left(\frac{(1-2p)t}{\sigma^2\tau}\right) + o(1) \qquad \text{as } t \to \infty, \tag{13}$$

and then from (10) that

$$Y_{T(t)} \sim ((1-2p)t\sigma^{-2})^{1/(1-2p)}\tau^{-2p/(1-2p)} \qquad \text{as } t \to \infty. \tag{14}$$

The construction of the sought variables $A(t)$ is plain from (14).

The case when $\mathbb{P}(X = 0) = a \in (0, 1)$ only requires a minor modification. We can no longer identify $Y_{T(t)}$ with the total number of jumps of the process $M_{T(\cdot)}$ on the time interval $[0, t]$. Nonetheless the latter is now close to $(1-a)Y_{T(t)}$ for $t \gg 1$, and we can conclude just as above.

We now have all the ingredients needed for the proof of the invariance principle.

Proof (Proof of Theorem 2 When X Is Bounded) For the sake of simplicity, we further assume in this proof that $\sigma^2 = 1$, which induces no loss of generality. Set

$$s = s(t) = (\tau^{-2p}t)^{1/(1-2p)} \quad \text{and} \quad k = k(n) = ((1-2p)n)^{1/(1-2p)},$$

so (14) and (13) yield respectively

$$Y_{T(nt)} \sim ks \quad \text{and} \quad \mathrm{e}^{-pT(nt)} \sim (ks/\tau)^{-p}.$$

Since $n = k^{1-2p}/(1-2p)$ and $t = \tau^{2p}s^{1-2p}$, we deduce from Corollary 3 and the very definition of the Skorohod topology involving changes of time (see, e.g. Section VI.1 in [13]) that as $k \to \infty$, the sequence of processes $k^{-1/2}(\hat{S}(\lfloor ks \rfloor))_{s \geq 0}$

converges in distribution in the sense of Skorohod towards

$$\left(\frac{\tau^{-p}}{\sqrt{1-2p}} s^p B(\tau^{2p} s^{1-2p})\right)_{s\geq 0},$$

where B is a standard Brownian motion independent of τ. By the scaling property of Brownian motion, the process displayed above has the same distribution as

$$\left(\frac{s^p}{\sqrt{1-2p}} B(s^{1-2p})\right)_{s\geq 0},$$

and we complete the proof with (3).

4.3 Reduction to the Case When X Is Bounded

In this section, we only assume that $X \in L^2(\mathbb{P})$ with $\mathbb{E}(X) = 0$. We shall complete the proof of Theorem 2 by a truncation argument, for which some notation is needed.

For every $b > 0$, we set

$$X^{(b)} = \mathbf{1}_{|X|\leq b} X - \mathbb{E}(X\mathbf{1}_{|X|\leq b}),$$

so $X^{(b)}$ is a centered and bounded variable; we write $\sigma^{(b)}$ for its standard deviation. Similarly, we set

$$\hat{X}_n^{(b)} = \mathbf{1}_{|\hat{X}_n|\leq b} \hat{X}_n - \mathbb{E}(X\mathbf{1}_{|X|\leq b}) \quad \text{and} \quad \hat{S}^{(b)}(n) = \hat{X}_1^{(b)} + \cdots + \hat{X}_n^{(b)}.$$

Clearly, $\hat{S}^{(b)}$ is a version of the step-reinforced random walk with typical step distributed as $X^{(b)}$. The latter being bounded, an application of Theorem 2 as proven in the preceding section shows that there is the convergence in distribution in the sense of Skorohod

$$n^{-1/2}\hat{S}^{(b)}(\lfloor \cdot n \rfloor) \Longrightarrow \sigma^{(b)}\hat{B}(\cdot) \qquad \text{as } n \to \infty, \tag{15}$$

where $\hat{B}$ denotes a noise reinforced Brownian motion with reinforcement parameter p.

Recall that the topology of weak convergence on the set of probability measures on the Skorokhod space of càdlàg paths is metrizable. Since plainly $\lim_{b\to\infty} \sigma^{(b)} = \sigma$, it follows readily from (15) that we have also

$$n^{-1/2}\hat{S}^{(b(n))}(\lfloor \cdot n \rfloor) \Longrightarrow \sigma\hat{B}(\cdot) \qquad \text{as } n \to \infty \tag{16}$$

for some sequence $(b(n))$ of positive real number that tends to ∞ slowly enough.

Then consider

$$\check{S}^{(b(n))}(n) = \hat{S}(n) - \hat{S}^{(b(n))}(n),$$

and observe that

$$\check{S}^{(b(n))}(n) = \check{X}_1^{(b(n))} + \cdots + \check{X}_n^{(b(n))}$$

with

$$\check{X}_n^{(b(n))} = \mathbf{1}_{|\hat{X}_n| > b(n)} \hat{X}_n - \mathbb{E}(X \mathbf{1}_{|X| > b(n)}).$$

In turn, $\check{S}^{(b(n))}$ is also a step-reinforced random walk, now with typical step distributed as $X - X^{(b(n))}$. The latter is a centered variable in $L^2(\mathbb{P})$, and if we write $\varsigma^{(b(n))}$ for its standard deviation, then clearly $\lim_{n\to\infty} \varsigma^{(b(n))} = 0$ since $b(n)$ tends to ∞. We deduce from Corollary 1 that for any $t > 0$,

$$\lim_{n\to\infty} n^{-1} \mathbb{E}\left(\max_{k \le nt} |\check{S}^{(b(n))}(k)|^2 \right) = 0. \tag{17}$$

We now see from (17) and the Markov inequality that the requirement 3.30 on page 316 in [13] holds for $Z^n_\cdot = n^{-1/2} \check{S}^{(b(n))}(\lfloor \cdot n \rfloor)$. This enables us to apply Lemma 3.31 there with $Y^n_\cdot = n^{-1/2} \hat{S}^{(b(n))}(\lfloor \cdot n \rfloor)$, and we conclude that

$$n^{-1/2} \hat{S}(\lfloor \cdot n \rfloor) = Y^n_\cdot + Z^n_\cdot \Longrightarrow \sigma \hat{B}(\cdot) \qquad \text{as } n \to \infty.$$

The Proof of Theorem 2 is now complete.

References

1. Baur, E.: Baur, E.: On a class of random walks with reinforced memory. J. Stat. Phys. **181**, 772–802 (2020). https://doi.org/10.1007/s10955-020-02602-3
2. Baur, E., Bertoin, J.: Elephant random walks and their connection to Pólya-type urns. Phys. Rev. E **94**, 052134 (2016)
3. Bercu, B.: A martingale approach for the elephant random walk. J. Phys. A **51**(1), 015201, 16 (2018)
4. Bercu, B., Laulin, L.: On the multi-dimensional elephant random walk. J. Stat. Phys. **175**(6), 1146–1163 (2019)
5. Bertoin, J.: Noise reinforcement for Lévy processes. Ann. Inst. Henri Poincaré B **56**, 2236–2252 (2020). https://doi.org/10.1214/19-AIHP1037
6. Businger, S.: The shark random swim (Lévy flight with memory). J. Stat. Phys. **1723**, 701–717 (2018)
7. Coletti, C.F., Papageorgiou, I.: Asymptotic analysis of the elephant random walk (2020). arXiv:1910.03142

8. Coletti, C.F., Gava, R., Schütz, G.M.: Central limit theorem and related results for the elephant random walk. J. Math. Phys. **58**(5), 053303, 8 (2017)
9. Coletti, C.F., Gava, R., Schütz, G.M.: A strong invariance principle for the elephant random walk. J. Stat. Mech. Theory Exp. **12**, 123207, 8 (2017)
10. González-Navarrete, M., Lambert, R.: Non-Markovian random walks with memory lapses. J. Math. Phys. **5911**, 113301, 11 (2018)
11. Gouet, R.: Martingale functional central limit theorems for a generalized Pólya urn. Ann. Probabil. **213**, 1624–1639 (1993)
12. Gut, A., Stadtmueller, U.: Elephant random walks with delays (2019). arXiv:1906.04930
13. Jacod, J., Shiryaev, A.N.: Limit Theorems for Stochastic Processes. Grundlehren der mathematischen Wissenschaften. Springer, Berlin (2002)
14. Janson, S.: Functional limit theorems for multitype branching processes and generalized pólya urns. Stoch. Process. Their Appl. **110**(2), 177–245 (2004)
15. Kallenberg, O.: Foundations of Modern Probability. Probability and Its Applications (New York), 2nd edn. Springer, New York (2002)
16. Kious, D., Sidoravicius, V.: Phase transition for the once-reinforced random walk on $\mathbb{Z}^d$-like trees. Ann. Probab. **464**, 2121–2133 (2018)
17. Kubota, N., Takei, M.: Gaussian fluctuation for superdiffusive elephant random walks. J. Stat. Phys. **177**(6), 1157–1171 (2019)
18. Kürsten, R.: Random recursive trees and the elephant random walk. Phys. Rev. E **93**(3), 032111, 11 (2016)
19. Pemantle, R.: A survey of random processes with reinforcement. Probab. Surveys **4**, 1–79 (2007)
20. Revuz, D., Yor, M.: Continuous Martingales and Brownian Motion. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 293, 3rd ed. Springer, Berlin (1999)
21. Schütz, G.M., Trimper, S.: Elephants can always remember: exact long-range memory effects in a non-markovian random walk. Phys. Rev. E **70**, 045101 (2004)
22. Whitt, W.: Proofs of the martingale FCLT. Probab. Surv. **4**, 268–302 (2007)

Geodesic Rays and Exponents in Ergodic Planar First Passage Percolation

Gerandy Brito and Christopher Hoffman

Abstract We study first passage percolation on the plane for a family of invariant, ergodic measures on $\mathbb{Z}^2$. We prove that for all of these models the asymptotic shape is the ℓ_1 ball and that there are exactly four infinite geodesics starting at the origin a.s. In addition we determine the exponents for the variance and wandering of finite geodesics. We show that the variance and wandering exponents do not satisfy the relationship of $\chi = 2\xi - 1$ which is expected for independent first passage percolation.

Keywords First passage percolation · Fluctuation exponent · Variance exponent

1 Introduction

First passage percolation is a widely studied model in statistical physics. One of the main reasons for interest in first passage percolation is that it is believed that, for independence passage times (and under mild assumptions on the common distribution) the model belongs to the KPZ universality class. The study of first passage percolation has centered on the three main sets of questions below. (Precise definitions are given in the next two sections.)

1. **Asymptotic shape**. Cox and Durrett proved that every model of first passage percolation has an asymptotic shape $\mathbf{B} \subset \mathbb{R}^2$ which is convex and has the symmetries of $\mathbb{Z}^2$ [5]. We would like to determine $\mathbf{B}$ or at least describe some

G. Brito
Georgia Institute of Technology, College of Computing, Atlanta, GA, USA
e-mail: gbrito3@gatech.edu

C. Hoffman (✉)
Department of Mathematics, University of Washington, Seattle, WA, USA
e-mail: hoffman@math.washington.edu

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_8

of its properties. In particular is the asymptotic shape is strictly convex and is its boundary differentiable?
2. **Infinite geodesics from the origin**. Are there infinitely many one-sided infinite geodesics that start at (0, 0)? Do these geodesics all have asymptotic directions?
3. **Variance and wandering exponents**. For any $\lambda \geq 0$ does there exist a variance exponent $\chi = \chi(\lambda)$ such that

$$\text{Var}(T(0, (n, \lambda n)) = n^{2\chi + o(1)}?$$

Does there exist a wandering exponent $\xi = \xi(\lambda)$ such that with high probability every edge in $\gamma(\mathbf{0}, (n, \lambda n))$ is within distance $n^{\xi + o(1)}$ of the line segment connecting $\mathbf{0}$ and $(n, \lambda n)$? Do χ and ξ satisfy the universal scaling relation

$$\chi = 2\xi - 1?$$

It is widely believed that (under mild assumptions) in independent first passage percolation the answer to all of these questions is yes. However in our models we show that the answer all of these questions is at least somewhat different than the answers that are expected for the independent case. Thus our model shows that universality cannot be expected to hold for all models of ergodic first passage percolation. Our results are as follows.

1. For all of our models the asymptotic shape $\mathbf{B}$ is the unit ball in the ℓ_1-norm.
2. Our models have exactly four one-sided infinite geodesics starting from the origin a.s., each of which meander through a quadrant.
3. For each value of λ we calculate exact variance and wandering exponents of the geodesic from $\mathbf{0}$ to $(n, \lambda n)$. For all $\lambda > 0$ the variance exponent χ is zero while the wandering exponent is 1. For $\lambda = 0$ we get variance and wandering exponents that satisfy $0 < \chi = \xi < 1$. In neither of these cases do the exponents satisfy the universal scaling relation $\chi = 2\xi - 1$.

It is already known that there exist models of ergodic first passage percolation whose behavior is different from what is expected for independent first passage percolation. Häggström and Meester showed that for any set $\mathbf{B} \subset \mathbb{R}^2$ which is bounded, convex and has non-empty interior and all the symmetries of $\mathbb{Z}^2$ there is a model of ergodic first passage percolation that has $\mathbf{B}$ as its limiting shape [7]. The examples we construct show that there are models of ergodic first passage percolation that have anomalous geodesic structures. More interestingly our models have anomalous variance and wandering exponents and these exponents depend on the direction. We are not aware of any other non-trivial models of ergodic first passage percolation where the variance and wandering exponents have been explicitly calculated.

2 Background on First Passage Percolation

In first passage percolation, a nonnegative variable is associated to each edge of a given graph. These variables give rise to a random metric space. Among the fundamental objects of study of this metric space are the scaling properties of balls and the structure of geodesics. By planar first passage percolation, we refer to the model on the lattice, denoted by $\mathbb{Z}^2$, which has vertex set $V = \{(x, y)\, x, y \in \mathbb{Z}\}$ and edge set $\mathscr{E} = \{(v, w) : |v - w| = 1\} \subset V \times V$ where $|\cdot|$ denotes the ℓ_1 distance. A configuration of $\mathbb{Z}^2$ is simply a function from the edge set to the nonnegative real numbers:

$$\mathbf{t} : \mathscr{E} \to [0, \infty). \tag{1}$$

We will use the more common notation $\mathbf{t}_e$ for $\mathbf{t}(e)$. If ν is a probability measure on $[0, \infty)^{\mathscr{E}}$, we denote the probability space with state space $[0, \infty)^{\mathscr{E}}$ and measure ν by $FPP(\nu)$. The number $\mathbf{t}_e$ can be seen as the passage time or length of the edge e. Given a configuration $\mathbf{t}$ on $\mathbb{Z}^2$ and a path $\pi = \{e_i\}_{i=1}^k$ the length of π is

$$\Gamma(\pi) = \sum_{i=1}^{k} \mathbf{t}_{e_i}.$$

The distance between two vertices u and v is denoted by $T(u, v)$ and it is defined as

$$T(u, v) = \inf \Gamma(\pi) \tag{2}$$

where the inf is taken over the set of all paths connecting u and v. It is not hard to check that $(\mathbb{Z}^2, T(\cdot, \cdot))$ is a pseudometric space for any configuration. Furthermore, if the values $\mathbf{t}_e$ are all bigger than zero then $T(\cdot, \cdot)$ is a metric. As we will see in Sect. 4, our measures will be bounded away from zero, so for the rest of the paper $(\mathbb{Z}^2, T(\cdot, \cdot))$ is a metric space. The ball of radius R centered at u is

$$B(u, R) = \{v \in V : \ T(u, v) < R\}. \tag{3}$$

Cox and Durrett [5] studied the behavior of large balls after scaling. They proved that, if $\mathbf{t}_e \sim \nu$ satisfying

$$\mathbb{E}(\min\{\mathbf{t}_1^2, \mathbf{t}_2^2, \mathbf{t}_3^2, \mathbf{t}_4^2\}) < \infty \tag{4}$$

for independent copies of $\mathbf{t}_e$, and the mass at zero is less than the threshold for bond percolation then there is a non-empty set $\mathbf{B}$, compact, convex and symmetric with respect to the origin such that, for any $\epsilon > 0$

$$\mathbb{P}\left((1 - \epsilon)\mathbf{B} \subset \frac{B(\mathbf{0}, R)}{R} \subset (1 + \epsilon)\mathbf{B} \text{ for all large } R\right) = 1. \tag{5}$$

Boivin extended this to a wide class of ergodic models of first passage percolation [4].

The question of which compact sets can be obtained as limit in FPP is almost entirely open for the i.i.d. case. Interestingly, when we consider the bigger set of stationary and ergodic measures on $(\mathbb{R}_+)^{\mathscr{E}}$ it was proved by Haggstrom and Meester [7] that any compact, convex, symmetric (with respect to the origin) set is the limiting shape for a stationary and ergodic measure, not necessarily i.i.d. It is worth mentioning that the limiting shape $\mathbf{B}$ is the unit ball of a norm $\|\cdot\|_\nu$. This norm can be computed as follows. First, we extend $T(\cdot,\cdot)$ to $\mathbb{R}^2$ by setting $T(u,v) = T(\lfloor u \rfloor, \lfloor v \rfloor)$ for all $u, v \in \mathbb{R}^2$. Here we slightly abused notation using the floor function on points. The reader should understand this by applying it to each coordinate. It can be shown, under mild assumptions on the distribution of $\mathbf{t}_e$, that the norm $\|x\|_\nu$ satisfies

$$\|x\|_\nu = \lim_n \frac{T(\mathbf{0}, nx)}{n}$$

where the limit exists a.s. and in L_1 for every fixed $x \in \mathbb{R}^2$. The set $\mathbf{B}$

$$\mathbf{B} = \{x \in \mathbb{R}^2 : \ \|x\|_\nu \leq 1\}. \tag{6}$$

A *geodesic* between u and v is a path that realizes the infimum in (2). We denote geodesics by $\gamma(u,v)$. Geodesics aren't always unique. A simple condition to guarantee such property for independent edge weights is to consider continuous distribution for $\mathbf{t}_e$. A *geodesic ray* is an infinite path $\{v_0, v_1, v_2, \ldots\}$ such that every finite sub-path is a geodesic between its endpoints. We consider two geodesic rays to be distinct if they intersect in only finitely many edges. We denote by $\mathscr{T}_0$ the set of all geodesic rays starting at the origin. Ahlberg and Hoffman [1] recently showed that for a wide class of measures the cardinality of $\mathscr{T}_0$ is constant almost surely, possibly ∞.

3 Statement of Results

The limiting shape $\mathbf{B}$ is closely related to the number and geometry of geodesic rays for ergodic FPP. Let $sides(\mathbf{B})$ denote the number of sizes of $\mathbf{B}$ if it is a polygon, and infinity otherwise, Hoffman [8, 9] proved that, for any $k \leq sides(\mathbf{B})$ there exist k geodesic rays almost surely, for *good measures*, see Sect. 4 for details. In particular, his results imply that there exist at least four geodesics a.s. When $\mathbf{B}$ is a polygon, little is known about existence of geodesics rays in the direction of the corners of $\mathbf{B}$. Recently, Alexander and Berger [2] exhibit a model for which the limiting shape is an octagon and all (possibly infinitely many) geodesic rays are directed along the coordinate axis. Our first result shows that our model has exactly four geodesic rays a.s.. To the best of our knowledge, this is the first known FPP model for which $|\mathscr{T}_0|$ is finite.

Theorem 1 *There exists a family of measures* $\{\nu_\alpha\}_{0<\alpha<0.2}$ *such that* $|\mathscr{T}_0| = 4$ ν_α-*almost surely.*

Our next result is about the direction of geodesic rays. We start with a definition. The *direction*, $Dir(\Pi)$, of a sequence of (not necessarily distinct) points $\Pi = \{v_k,\ k \geq 0\}$ is the set of limits of $\{v_k/|v_k|,\ v_k \in \Pi\}$. If $|v_k| \to \infty$ then $Dir(\Pi)$ is a connected subset of S^1. Damron and Hanson [6] were the first to prove directional results for geodesic rays for *good measures* that also have the upward finite energy property. Their results are also dependent on the geometry of **B** in the following way. We say that a linear functional $\rho : \mathbb{R}^2 \to \mathbb{R}$ is tangent to **B** if the line $\{x \in \mathbb{R}^2 :\ \rho(x) = 1\}$ is tangent to **B** at a point of differentiability of the boundary of **B**. In view of Eq. (6), we can write the intersection of this tangent line and the boundary of **B** as a set in S^1:

$$D_\rho = \{x \in S^1 :\ \rho(x) = \|x\|_\nu\}. \tag{7}$$

Damron and Hanson [6, Theorem 1.1] states that for any functional ρ, tangent to **B**, there is an element $\gamma \in \mathscr{T}_0$ satisfying $Dir(\gamma) \subset D_\rho$. Because of the differentiability condition, their result gives no information about the behavior around corners of **B**. For our family of measures $\{\nu_\alpha\}$ we are able to completely characterize the directions of geodesic rays.

Theorem 2 *Fix* α *such that* $0 < \alpha < 0.2$ *and consider* $FPP(\nu_\alpha)$*. Let* ρ *be a linear functional tangent to the* ℓ_1*-ball. There is exactly one geodesic with generalized direction equal to* D_ρ*.*

Lastly, we turn our attention to the geometry of finite geodesics. We follow the classical approach and study it both from the random and the geometric point of view. For the first one, the most basic analysis comes from understanding the variance of $T(\mathbf{0}, x)$. As stated informally in the introduction, it is believed that there exist a universal exponent that governs this quantity. For our model, we show the existence of a constant $\chi = \chi(\lambda)$ and a universal constant K such that

$$\frac{1}{K} \leq \frac{Var(T(\mathbf{0}, (n, \lambda n)))}{n^{2\chi}} \leq K,$$

see Lemmas 8 and 10 for the formal statements. We point out that our results are strong enough that they satisfy any reasonable definition of the variance exponent, in particular those suggested in [3].

From the geometric perspective, we look at how far is $\gamma(\mathbf{0}, (n, \lambda n))$ from the straight line connecting the origin and the point $(n, \lambda n)$. As in the case of the variance exponent, it is widely believed that there is a universal constant, denoted by ξ, such that the maximal distance between $\gamma(\mathbf{0}, (n, \lambda n))$ and the line through the origin and $(n, \lambda n)$ is of order n^ξ. So, if one sets $\mathrm{Cyl}(x_n^\lambda, N)$ to be the set of points within (Euclidean) distance N from the line connecting $\mathbf{0}$ and $x_n^\lambda = (n, \lambda n)$, it is expected that $N = n^\xi$ is the right scale of cylinder to contain $\gamma(\mathbf{0}, x_n^\lambda)$. To formally capture this property we adopt the definition in [10] which we reproduce below.

Definition 1 For $0 \leq \lambda \leq 1$ and $x_n^\lambda = (n, \lambda n)$ we set

$$\xi_\lambda = \inf\{\alpha > 0 : \liminf_n \mathbb{P}(\gamma(\mathbf{0}, x_n^\lambda) \subset \mathrm{Cyl}(x_n^\lambda, n^\alpha)) > 0\}.$$

We compute this exponent in any direction, and confirm non-universality of FPP for invariant, but not necessarily identically distributed, edge weight.

Theorem 3 *Fix α such that $0 < \alpha < 0.2$ and consider $FPP(\nu_\alpha)$. In every direction not parallel to the coordinate axes we have the variance exponent $\chi = 0$ and the wandering exponent $\xi = 1$. Parallel to the coordinate axes the two exponents are equal with $\chi = \xi = \frac{\log 5}{\log 5 - \log \alpha}$. In no direction do the exponents satisfy the universal scaling relation.*

Remark 1 This is the content of Lemmas 10 and 11 for the coordinate directions and Lemma 8 and Propositions 2 and 3 for the non-coordinate directions. The reader will notice that these results lead to stronger statements. In particular, we can deduce that, for $0 < \lambda \leq 1$

$$\lim_n \mathbb{P}(\gamma(\mathbf{0}, x_n^\lambda) \subset \mathrm{Cyl}(x_n^\lambda, n^{1+\epsilon})) = 1$$

and

$$\lim_n \mathbb{P}(\gamma(\mathbf{0}, x_n^\lambda) \subset \mathrm{Cyl}(x_n^\lambda, n^{1-\epsilon})) = 0$$

and similarly for the coordinate directions and $\xi = 0$. In particular, we believe that for our model the results of Theorem 3 apply to any reasonable definition of the variance and wandering exponents.

3.1 Organization of the Paper

The rest of the paper is organized as follows. In Sect. 4 we define the measures ν_α and show its main properties. The proof of Theorem 1 is given in Sect. 5 where the limiting shape is also determined. Section 6 is devoted to the study of the directional properties of geodesic rays and the proof of Theorem 2. In Sects. 7 and 8 we prove our final theorem which determines the exponents in all directions. This is the content of Lemmas 10 and 11 for the coordinate directions and Lemma 8 and Propositions 2 and 3 for the non-coordinate directions.

4 Construction of the Measures $\{\nu_\alpha\}$

In this section we construct a family of measures $\{\nu_\alpha\} \subset \mathscr{M}((\mathbb{R}_+)^{\mathscr{E}})$, indexed by a parameter $0 < \alpha < 0.2$. We state their main properties and study the behavior of geodesics for $FPP(\nu_\alpha)$.

Let $\Omega = \{0, 1, 2, 3, 4\}^{\mathbb{N}}$. Let $\sigma : \Omega \to \Omega$ be the 5-adic adding machine: σ adds one (mod 5) to the first coordinate. If the result is not zero then we leave all the other coordinates unchanged. If the result is zero then we add one to the second coordinate. We repeat until we get the first non-zero coordinate. All subsequent coordinates are left unchanged. Thus

$$\sigma(0, 1, 2, \dots) = (1, 1, 2, \dots) \quad \text{and} \quad \sigma(4, 4, 2, 1, \dots) = (0, 0, 3, 1, \dots).$$

We also adopt the convention $\sigma(444\dots) = 000\dots$

It follows that $\sigma : \Omega \to \Omega$ is uniquely ergodic with respect to the uniform measure on Ω. We use this map to form a $\mathbb{Z}^2$ action of $\Omega \times \Omega$ as follows. Let $L : \Omega \to \mathbb{N}$ given by

$$L(\omega) = \min\{i : \omega(i) > 0\}. \tag{8}$$

For $(\omega_1, \omega_2) \in \Omega \times \Omega$ fixed and $v = (x(v), y(v)) \in \mathbb{Z}^2$ we define:

$$k(v, v + \mathsf{e}_1) = L(\sigma^{y(v)}(\omega_1)) \tag{9}$$

and

$$k(v, v + \mathsf{e}_2) = L(\sigma^{x(v)}(\omega_2))$$

where $\mathsf{e}_1, \mathsf{e}_2$ are the vectors in the canonical base of $\mathbb{R}^2$. The following set will be referred to often in the paper, so we highlight its definition now.

Definition 2 We denote the set of edges $e \in \mathscr{E}$ such that $k(e) \geq j$ as the j-grid.

It is helpful to visualize the j-grid. Note that, by definition, these subgraphs are nested:

$$1\text{-grid} \supseteq 2\text{-grid} \supseteq \dots$$

Also, since edges e and e' in the same horizontal or vertical line satisfy $k(e) = k(e')$, it is not hard to see that for each j, the subgraph induced by the j-grid is isomorphic to $5^j\,\mathbb{Z}^2$.

We are ready to define the measure ν_α. Let $\{X_{j,e}\}_{j\in\mathbb{N}, e\in\mathscr{E}}$ be a set of independent random variables where $X_{j,*}$ has the uniform distribution over the interval

$[0, \frac{\alpha^j(1-\alpha)}{1000}]$. We take

$$\mathbf{t}_e = 1 + \alpha^{k(e)} + X_{k(e),e}$$

where ω_1 and ω_2 are chosen uniformly i.i.d. and independent of the $\{X_{j,e}\}$. For technical reasons, we let $0 < \alpha < 0.2$. Note that for every edge e

$$1 \leq \mathbf{t}_e \leq 1.3. \tag{10}$$

We set ν_α to be the resulting measure on $(\mathbb{R}_+)^{\mathscr{E}}$.

Remark 2 The measures ν_α fall into the class of *good* measures introduced in [8] and [9].

We recall the definition of good measures. A measure $\mathbb{P}$ is good if:

(a) $\mathbb{P}$ is ergodic with respect to the translations of $\mathbb{Z}^2$.
(b) $\mathbb{P}$ has all the symmetries of $\mathbb{Z}^2$.
(c) $\mathbb{P}$ has unique passage times.
(d) The distribution of $\mathbb{P}$ on an edge has finite $2 + \epsilon$ moment.
(e) The limiting shape is bounded.

The construction of ν_α is done so properties (a)–(e) are easy to check.

Informally, we think of a realization of $FFP(\nu_\alpha)$ as building a series of horizontal and vertical highways on the nearest neighbor graph of $\mathbb{Z}^2$. The value of ω_i, $i = 1, 2$, determines where the origin lies with respect to these highways. By construction, edges in the j-grid are faster (i.e.: have smaller passage time) than edges in any j'-grid for $j > j'$. Hence, a geodesics ray is expected to follow one grid until it encounters a faster one. Then the geodesic continues along edges of the faster grid. Globally, we expect to see rays with longer segments parallel to the axes as they move away from the origin. We also suspect that the length of these horizontal or vertical segments is roughly determined by the value of the j-grid they are part of. We formalize this intuition in the next sections.

5 Structure of Finite Geodesics

In this section we present several properties of geodesics in $FPP(\nu_\alpha)$. The first lemmas describe the geometric properties of finite geodesics along vertices in the k-grid, recall Definition 2.

Lemma 1 *Let $C = C(x, y, k)$ be a square of side 5^k with lower left vertex (x, y) such that all the edges in its boundary are in the k-grid. Consider two vertices v and w in the boundary of C. Then*

(i) $\gamma(v, w)$ is completely contained in C.

(ii) If v and w lie in the same or adjacent sides of C, $\gamma(v, w)$ lies in the boundary of C.

Proof We argue by contradiction. Assume there are vertices v and w in the boundary of C such that $\gamma(v, w)$ intersects the complement of C. Because a subpath of a geodesic is also a geodesic, we can assume that the edges of $\gamma(v, w)$ lie entirely in the complement of C, by considering a segment of $\gamma(v, w)$ completely in the complement of C and taking v and w to be its end points. Let d denote the length (the ℓ_1-distance) of the shortest path along the boundary of C connecting v and w.

If the maximal distance from a vertex in $\gamma(v, w)$ to C is less than 5^k then, by construction, all edges in $\gamma(v, w)$ will lie on the $(k-1)$-grid at most, hence, have passage time at least $1+\alpha^{k-1}$. Then the passage time of $\gamma(v, w)$ is at least

$$d(1+\alpha^{k-1}) > d(1+\alpha^k) + d\alpha^k$$

The right hand side is an upper bound for the passage time of the path from v to w along the boundary of C. We conclude that going along the boundary of C will be a shortest path from v to w. Hence, there should be a vertex in $\gamma(v, w)$ at distance at least 5^k of C. Then the passage time of $\gamma(v, w)$ is at least

$$2(5^k) + d$$

where the factor of two appears since we move away from C at least 5^k edges and come back to C, crossing another 5^k edges. Observe that $d \le 2(5^k)$. Hence,

$$d(1+\alpha^k) + d\alpha^k = d + 2d\alpha^k < d + 2(5^k),$$

using that $2\alpha^k < 1$ as long as $\alpha < 1/5$. The left hand side above is an upper bound on the passage time of a path connecting v and w along the boundary of C. This concludes the proof of part (i).

To prove (ii), assume that v lies in the left side of C and consider two cases for w.

- **Case 1: w lies also on the left hand side or the horizontal sides of C, but it is not a corner on the right hand side**. By (i) we know $\gamma(v, w)$ is contained in C. If $\gamma(v, w)$ uses edges in the interior of C, we can assume, changing v and w if necessary, that the entire geodesic lies in the interior. This implies that all edges in $\gamma(v, w)$ have passage times at least $(1+\alpha^{k-1}) > 1+\alpha^k + \frac{\alpha^k(1-\alpha)}{1000} \ge t_e$ for all e in the boundary of C. Hence, a path along the boundary will have smaller passage times, which shows that $\gamma(v, w)$ lies on the boundary.
- **Case 2: w is a corner on the right hand side of C**. We compare the path along the boundary of C to any path π which traverses edges in the interior of C. By case 1 we can assume that π exits the boundary of C on the left hand side and rejoins the boundary of C on the right hand side. Observe that π will have to traverse at least 5^k many edges horizontally, because v and w are on opposite

sizes of C, and at least $|y(w) - y(v)|$ many edges vertically. This leads to the lower bound

$$\Gamma(\pi) \geq 5^k(1 + \alpha^{k-1}) + |y(w) - y(v)|.$$

A path on the boundary of C connecting v and w has length at most:

$$5^k(1 + \alpha^k) + |y(w) - y(v)|(1 + \alpha^k) + 2(5)^k \frac{\alpha^k(1 - \alpha)}{1000}.$$

The last summand is an upper bound on the sum of the random portion of the path's distance. To conclude it suffices to show that

$$5^k(1+\alpha^{k-1})+|y(w)-y(v)| \geq 5^k(1+\alpha^k)+|y(w)-y(v)|(1+\alpha^k)+2(5)^k \frac{\alpha^k(1 - \alpha)}{1000}.$$

This inequality is equivalent to

$$2(5)^k \left(1 - \alpha - \frac{\alpha(1 - \alpha)}{1000}\right) \geq |y(w) - y(v)|\alpha$$

which follows directly since $5^k \geq |y(w) - y(v)|$ and $1 - \alpha - \frac{\alpha^k(1-\alpha)}{1000} \geq \alpha$ for $0 < \alpha < \frac{1}{5}$.

Corollary 1 *In the setting of Lemma 1, let v and w be any vertices in the boundary. Assume that $\gamma(v, w)$ visits a corner of C. Then $\gamma(v, w)$ is completely contained in the boundary of C.*

Proof Let v' be a vertex in $\gamma(v, w)$ which is in the corner of C. Then v' is in two sides and both the other two sides are adjacent to one of these two sides. Then both the pairs v and v' and v' and w lie in (the same or) adjacent sides of C. Thus the corollary follows from Lemma 1 applied to $\gamma(v, v')$ and $\gamma(v', w)$.

We extend the result above to a large rectangle in the next lemma.

Lemma 2 *Let $M = M(\alpha, k)$ be an integer such that $\alpha^k M > 1$, and let $R = R(x, y, k)$ be a rectangle with vertices (x, y); $(x, y + 5^k)$; $(x + 5^k M, y)$; $(x + 5^k M, y + 5^k)$ such that all the edges in its sides are in the k-grid. Let v and w be vertices in the boundary of R such that at least one is on one of the shorter sides of R. If $\gamma(v, w)$ is contained in R, then it is contained in the k-grid.*

Remark 3 The lemma above is still true if the largest side of the rectangle is parallel to the y-axis.

Remark 4 The lemma above confirms that, once a geodesics enters a fast grid, it will not visit slower edges anymore: the only edges parallel to the x-axis in $\gamma(v, w)$ are in the boundary of R. It may traverse edges in the interior of R but only parallel to the y-axis and on the k-grid.

Proof To fix ideas, assume v lies on the left side of R. Notice that R can be divided into M squares of side 5^k, each satisfying the condition of Lemma 1, namely, each has boundary edges in the k-grid. We denote these squares by $C_1,\ C_2\ldots, C_M$ from left to right. Also, for $1 \le j \le M$ denote v_j and w_j the first and last vertex that $\gamma(v, w)$ visits in C_j, respectively, when this intersection is not empty. We will again split the proof into cases.

- **Case 1: w lies on the common boundary of C_1 and R.** This is the content of Lemma 1.
- **Case 2: w lies in one of the larger (horizontal) sides of R.** This case follows by induction on M, with case 1, or Lemma 1, being the initial step. Note that in this case we may traverse edges in the interior of R, but only in the boundary of C_j for some values of j, thus we are still on the k-grid.
- **Case 3: w is in the right side of R.** Assume $\gamma(v, w)$ visits a vertex on the horizontal sides of R, say v'. Then by **case 2** applied to $\gamma(v, v')$ and $\gamma(w, v')$ we conclude that $\gamma(v, w)$ is on the union of the boundaries of the C_j, which is a subset of the k-grid. If such vertex v' does not exist, we deduce that all horizontal edges (i.e.: parallel to the y-axis) on $\gamma(v, w)$ are in the interior of R. Then its length will be at least:

$$5^k M(1 + \alpha^{k-1}),$$

since all edges in the interior are in the $(k-1)$-grid at most. The shortest path on the boundary of R connecting v and w has length bounded above by

$$5^k(M + 1)(1 + \alpha^k + \frac{\alpha^k(1-\alpha)}{1000}).$$

It can be checked that, for our choice of M it holds $5^k M(1 + \alpha^{k-1}) \ge 5^k(M + 1)(1+\alpha^k)+5^k(M+1)(\alpha^k \frac{1-\alpha}{1000})$, which yields the desired contradiction. We have proved that $\gamma(v, w)$ lies on the union of the boundaries of C_j, which proves the lemma.

Proposition 1 *Let v, w vertices that are end points of edges in the k-grid, satisfying $|v - w| \ge 5^{2k} M$, for an integer M such that $\alpha^k M > 1$. Then the geodesic $\gamma(v, w)$ is contained in the k-grid.*

Proof Suppose that $\gamma(v, w) = \{e_1, e_2, \ldots, e_t\}$ contains at least one edge outside the k-grid. Let

$$m = \min\{s \ge 1 : \ e_s \ \text{ is not on the } k\text{-grid}\}.$$

By definition, one endpoint of e_m lies in the k-grid and one endpoint lies outside of it. Let R^{e_m} be the unique rectangle defined by the following constraints:

(a) The boundary of R^{e_m} is a subset of the k-grid. Furthermore, the length of the sizes of R^{e_m} are 5^k and $5^k M$ for a natural number M such that $\alpha^k M > 1$.

Fig. 1 The rectangle R^{e_m} in **case 2** of the proof of proposition 1: the geodesic will connect vertices v and v' traversing edges in the k-grid only, which forces it to visit a corner of R^{e_m}, denoted by c. Then the geodesics follows from c to w traversing e_m, which is not the fastest path, by Lemma 2

(b) The boundary of R^{e_m} intersects e_m.
(c) R^{e_m} contains e_m.
(d) The larger sides of R^{e_m} are parallel to e_m.

Note that the conditions in (a) are as those in Lemma 2. Since e_m is the first edge on $\gamma(v, w)$ not on the k-grid, exactly one endpoint of it is in this grid, which is the only possible intersection in condition (b). We will denote this vertex by v'. Finally, all of the conditions ensure that we have a unique choice of R^{e_m} (Fig. 1).

We consider two cases:

- **Case 1:** w **is in the complement of** R^{e_m}. Then in order to reach w, $\gamma(v, w)$ has to exit R^{e_m} for the last time at some vertex w' in its boundary. Since one endpoint of e_m is inside R^{e_m} we have that $w' \neq v'$. By Lemma 2 or Remark 3 the geodesic from v' to w' is contained in the k-grid and hence it cannot traverse e_m. This contradicts our assumption.
- **Case 2:** w **is in** R^{e_m}. We start making two simple observations. First, because e_m is parallel to the longer sides of R^{e_m}, we deduce that v' is on one of its shorter sides. Second, from the assumption that v and w are far away from each other, we conclude that v is in the complement of R^{e_m}. By definition of m, all edges e_s, $s < m$ are in the k-grid. Traverse $\gamma(v, w)$ from v until we get to v'. We claim that we must visit one of the corners of the side of R^{e_m} that contains v'. To see this, simply observe that removing the two corners of such side disconnects v' from v in the graph induced by the k-grid (because v is in the complement of R^{e_m}). Call the visited corner c and let C be the square of size length 5^k completely inside R^{e_m} with one corner equal to c. Let γ' be the intersection of C and $\gamma(c, w)$. Note that $e_m \in \gamma'$. Now γ' is connected, and its endvertices are c, a corner of C, and one vertex on its boundary. In the square C, because c is a corner, it lies on the same or an adjacent side of such vertex, and thus by Lemma 1 (ii) we get that γ' is completely on the boundary of C, which contradicts the definition of e_m and finishes the proof.

To prepare the ground for our next lemma, we draw a few conclusions from Proposition 1. First, notice that any geodesic ray γ will have infinitely many vertices in the k-grid, for all k. If v_k is the first such vertex, it follows that *all* edges in γ after v_k are in the corresponding grid. Applying the same reasoning we conclude that the intersection of γ and the k-grid is an infinite connected set. The vertices v_k break γ into slower edges, those with passage time of order $1 + \alpha^k$, and faster edges, with passage time of order $1 + \alpha^{k+1}$. We turn our attention to a set of special vertices and introduce the following definition.

Definition 3 Let v be a vertex of $\mathbb{Z}^2$. We denote by $\mathscr{V}_k(v)$ the square in the plane containing v with the following properties:

(a) The boundary of $\mathscr{V}_k(v)$ is in the k-grid.
(b) The area of $\mathscr{V}_k(v)$ is 5^{2k}.
(c) If v lies in the intersection of two or more such squares, $\mathscr{V}_k(v)$ is the only one to the right and/or above v.

Denote by $v_i^k(v)$, $1 \leq i \leq 4$ the corners of $\mathscr{V}_k(v)$, starting at the upper right and going counterclockwise.

Conditions (a) and (b) imply that from all bounded regions in the plane with v in its interior and boundary a subset of the k-grid, $\mathscr{V}_k(v)$ is the one with smaller area. Condition (c) handles the case when v is in the boundary of such region. When there is no confusion, we will drop the dependence on v in $\mathscr{V}_k$ and v_i^k. The importance of these vertices is explained in the next lemma.

Lemma 3 *Let γ be a geodesic ray starting at v. For each k, there is at least one value $1 \leq i \leq 4$ such that $v_i^k \in \gamma$.*

Proof Consider $w \in \gamma$ be a vertex in the k-grid such that $d(v, w) \geq 5^{3k}M$, for M which was defined in the proof of Lemma 2. The existence of w can be deduced from the fact that the k-grid is isomorphic to $5^k\mathbb{Z}^2$ and then we can find infinite closed paths on it disconnecting v from infinity. Since γ is an infinite path it will intersect the k-grid infinitely many times. Let $\hat{v} \in \mathscr{V}_k(v)$ denotes the first vertex in the k-grid that we encounter while going along γ, starting at v. We have $d(v, \hat{v}) \leq 5^k$ and thus, by Proposition 1, the subpath from $\hat{v}$ to w is contained in the k-grid. Thus, the last vertex that γ visits in $\mathscr{V}_k$ is one of its corners.

We are ready to prove Theorem 1.

5.1 Proof of Theorem 1

Assume there exits five different $\gamma_i \in \mathscr{T}_0$, $i \in \{1, 2, 3, 4, 5\}$. Then there is a (random) ball B centered at v sufficiently large such that any two of these five geodesic rays only intersect in the interior of B. Take k large such that $\mathscr{V}_k = \mathscr{V}_k(\mathbf{0})$ has its four corners in the complement of B. By Lemma 1 each γ_i will visit at least one corner of $\mathscr{V}_k$. This contradicts the intersection property since $\mathscr{V}_k$ has four corners. This proves

$$|\mathscr{T}_0| \leq 4$$

Since ν_α is good, it follows from [9, Theorem 1.2] that $|\mathscr{T}_0| \geq 4$ and the result follows.

This result allows us to determine the shape. A direct proof of the shape is also very short.

Corollary 2 *The limiting shape of $FPP(\nu_\alpha)$ is the ℓ_1-ball.*

Proof From Theorem 1 and [9, Theorem 1.2] we have that the limiting shape of $FPP(\nu_\alpha)$ is either proportional to the ℓ_1-ball or the ℓ_∞-ball. As every edge has passage time at least one the limiting shape must be contained in the ℓ_1-ball. But the speed in the coordinate directions is one so the limiting shape must be the ℓ_1-ball.

6 Direction of the Geodesic Rays and Proof of Theorem 2

Our goal in this section is to completely characterize $Dir(\gamma)$ for each geodesic ray γ in $FPP(\nu_\alpha)$. We start by combining Theorem 1 and recent results of Ahlberg and Hoffman [1] to get further information about the geodesic rays.

Throughout this section, we will write $\mathscr{V}_k$ to refer to $\mathscr{V}_k(\mathbf{0})$. Similarly, the corners of $\mathscr{V}_k$ will be denoted by $\{v_i^k\}_{i=1}^k$, see Definition 3. For $1 \leq i \leq 4$, denote by $C_i = \{v_i^k,\ k \in \mathbb{N}\}$ the set of corners lying in the ith quadrant of the coordinate plane.

For any geodesic γ recall that $\text{Dir}(\gamma) \subset S^1$. In the remainder of the section we will slightly abuse notation by considering $\text{Dir}(\gamma) \subset [0, 2\pi)$.

Lemma 4 *With ν_α probability one the following holds: for each $1 \leq i \leq 4$ there is a unique geodesic ray γ_i, starting at the origin, such that the angle $(i - \frac{1}{2})\frac{\pi}{2} \in Dir(\gamma_i)$ and $Dir(\gamma_i)$ is in the ith quadrant.*

Proof For each quadrant there is a linear function ρ_i whose level set $D_{\rho_i}(z) = 1$ (see Eq. (7) for the definition of D_ρ) is the intersection of the boundary of the ℓ_1-ball with the ith quadrant. By Theorems 1.11 and 4.6 of [6] for each i there is a geodesic whose Busemann function is asymptotically linear with growth rate D_{ρ_i} and whose $Dir(\gamma_i)$ is contained in the ith quadrant. As there are only four geodesics a.s., these geodesics are unique. We denote by γ_i the only geodesic ray directed on the ith quadrant.

For any $v, w \in \mathbb{Z}^d$ we have for all k sufficiently large that $\mathscr{V}_k(v) = \mathscr{V}_k(w)$. Thus by Lemma 3 we have that the geodesics are coalescing. Thus $Dir(\gamma_i)$ is an almost sure invariant subset of S_i. Either

$$\mathbb{P}(Dir(\gamma_1) \cap [0, \pi/4] \neq \emptyset) \geq 0$$

or

$$\mathbb{P}(Dir(\gamma_1) \cap [\pi/4, \pi/2] \neq \emptyset) \geq 0.$$

By symmetry they must both be greater than zero. By shift invariance they both must have probability one. As $\mathrm{Dir}(\gamma_1)$ is connected subset of $[0, \pi/2]$ then $\pi/4 \in \mathrm{Dir}(\gamma_1)$. The same argument works for the other three quadrants.

Lemma 5 *With ν_α probability one it holds that $v_i^k \in \gamma_i$ for $1 \leq i \leq 4$ and for all but finitely many k.*

Proof For each k there exists an i such that both coordinates of v_i^k are at least $5^k/2$ in absolute value. For such values of k and i, we have $Dir(v_i^k) \in (i-1)\pi/2 + (.1, \pi/2 - 0.1)$. Let K be large enough such that for each $k > K$ we have that v_i^k for each i is in a distinct geodesic (the existence of such K follows from Theorem 1 and Lemma 3). Also, for any i and any vertex $v \in \gamma_i$ such that $|v| \geq \min_j |v_j^k|$ we have $Dir(v) \in (i-1)\pi/2 + (-0.01, \pi/2 + 0.01)$. Then for this particular i we have that $v_i^k \in \gamma_i$. From this we can conclude that for all other $j \neq i$ we have that $v_j^k \in \gamma_j$ as well.

Lemma 6 *Let $\omega_1, \omega_2 \in \Omega$ be sampled uniformly i.i.d.. The position of the origin in the interior of $\mathscr{V}_k$ is completely determined by the first k entries of ω_i, $i = 1, 2$.*

Remark 5 Lemma 6 can be interpreted as follows: a realization of ω_1, ω_2 determines which edges are in the k-grid via Eqs. (8) and (9), and thus determines $\mathscr{V}_k = \mathscr{V}_k(\mathbf{0})$. This region is a square of size 5^k and by definition the origin is one of the 5^{2k} vertices in it that do not lie on the top or right sides (see Definition 3). The lemma above tells us that it is enough to know the first k coordinates of ω_i, $i = 1, 2$ to determine which of those 5^2k vertices is the origin. Note that, to know the region $\mathscr{V}_k$ itself we need to know all coordinates of ω_i $i = 1, 2$.

Proof We will prove the lemma by induction on k. Let $\{e_j\}$ be the canonical base of Ω. The entries of e_j satisfies: $(e_j)_k = \delta_{\{k=j\}}$.

For $k = 1$, there are 25 possible positions of the origin within $\mathscr{V}_1$. Assign to each of those vertices a pair (a, b) given by the distance from it to the bottom side and left side of $\mathscr{V}_1$, respectively. This is a surjective map from the set of vertices in $\mathscr{V}_1$ and $\{0, 1, 2, 3, 4\}^2$. We can check now that the origin the vertex with label (a, b) if and only if: $(\omega_1)_1 = a$ and $(\omega_2)_1 = b$. This proves the initial case. To prove the general case, consider $\mathscr{V}_k$ divided into 25 squares of side 5^{k-1}. We will prove next that the pair $((\omega_1)_k, (\omega_2)_k)$ is enough to determine in which of these squares the origin is. To see this, we argue similarly to the case $k = 1$. Notice that each of the 25 squares can be encode by a pair (a, b) given by the distance to the bottom and left side of $\mathscr{V}_k$, respectively. We can check that the origin lies in the square labeled (a, b) if and only if $(\omega_1)_k = a$ and $(\omega_2)_k = b$. Using the induction hypothesis the proof will follow.

Lemma 7 *Denote by* θ_1^k *the argument of* v_1^k. *Fix* $\theta \in (0, \pi/2)$. *For any* $\epsilon > 0$ *there are infinitely many values of* k *such that*

$$|\theta - \theta_1^k| < \epsilon.$$

Proof We will do the case $\theta = 0$. We want to show that infinitely many v_1^k are inside the cone bounded by the lines $\theta = 0$ and $\theta = \epsilon$. Let t be a natural number such that $\frac{5^{-t}}{1-5^{-t}} < \tan(\epsilon)$. For large values of k, denote by E_k the event:

$$(\omega_1)_{k-t+1} = (\omega_1)_{k-t+2} = \cdots = (\omega_1)_k = 4$$

$$(\omega_2)_{k-t+1} = (\omega_2)_{k-t+2} = \cdots = (\omega_2)_k = 0.$$

In words, this corresponds to t coordinates been simultaneously equal to 0 and 4 in ω_1 and ω_2, respectively. If follows by Borel-Cantelli that $\{E_k\}$ happens infinitely often. By Lemma 6, this event corresponds to the origin being in the top left 5^{k-t} square in $\mathscr{V}_k$. Then

$$0 < \theta_1^k \leq \arctan\left(\frac{5^{k-t}}{5^k - 5^{k-t}}\right) < \epsilon.$$

which completes the proof.

6.1 Proof of Theorem 2

Let ρ be a functional tangent to the ℓ_1-ball. Associate to ρ a set C_i of corners, in the natural way. By Lemma 4 there is a unique geodesic ray, γ_i, with the property that $Dir(\gamma_i) \subset (i-1)(\pi/2) + (0, \pi/2)$. By Lemma 7 we can find points $u \in Dir(\gamma_i)$ as close as we want to the endpoints of $(i-1)(\pi/2) + (0, \pi/2)$. Since $Dir(\gamma_i)$ is connected we conclude that $(i-1)(\pi/2) + (0, \pi/2) \subset Dir(\gamma_i)$. It follows now that $(i-1)\pi/2 + [0, \pi/2] = Dir(\gamma_i)$.

7 Exponents in Non-coordinate Directions

The next two sections are devoted to the proof of Theorem 3. We start by showing that $T(x, y)$ is well concentrated. Denote the origin by $\mathbf{0}$.

Lemma 8 *Let* $0 < \lambda \leq 1$ *be a fixed constant and* $x_n^\lambda = (n, \lambda n)$. *There exists a constant* $C = C(\lambda)$ *such that*

$$|x_n^\lambda|_1 \leq T(\mathbf{0}, x_n^\lambda) \leq |x_n^\lambda|_1 + C.$$

Proof The lower bound follows from the fact that $t_e \geq 1$. For the upper bound, we construct a path from $\mathbf{0}$ to x_n^λ satisfying the desired inequality. Consider the squares $\{\mathcal{V}_k(\mathbf{0})\}$ and $\{\mathcal{V}_k(x_n^\lambda)\}$ for $1 \leq k \leq N-1$ where N is the minimum t such that the projections of $\mathcal{V}_t(\mathbf{0})$ and $\mathcal{V}_t(x_n^\lambda)$ onto either the x or y axes have nonempty intersection. Note that this definition implies that $n \leq (2/\lambda)5^N$.

Next, we choose a corner in $\{\mathcal{V}_k(\mathbf{0})\}$ for each k that is closest to x_n^λ, and similarly we choose a corner in $\{\mathcal{V}_k(x_n^\lambda)\}$ closest to $\mathbf{0}$. We have a sequence of vertices:

$$\mathbf{0}, v_1, v_2, \ldots, v_{N-1}, w_{N-1}, \ldots, w_2, w_1, x_n^\lambda$$

where v_i, w_i are the corners chosen above in $\mathcal{V}_i(\mathbf{0})$ and $\mathcal{V}_i(x_n^\lambda)$, respectively. Our path is the concatenation of paths joining consecutive vertices on this sequence such that the path connecting v_i and v_{i+1} (and w_{i+1} and w_i) is on the i-grid, and the path joining v_{N-1} and w_{N-1} is on the $(N-1)$-grid. All of these subpaths are taken to have the minimal possible number of edges.

Note that for each i, all edges in the subpaths from v_i to v_{i+1} have passage times bounded by $1 + 2\alpha^i$. Also, we cross at most $2 \cdot 5^{i+1}$ many edges between v_i and v_{i+1}. An analogous analysis extends to the vertices w_j. Between v_{N-1} and w_{N-1} we have at most $2n$ edges with weights at most $1 + 2\alpha^{N-1}$. The total length of our path is at most $|x_n^\lambda|_1$. We put this together to conclude that

$$T(\mathbf{0}, x_n^\lambda) \leq |x_n^\lambda|_1 + 20\sum_{j=1}^{\infty} 5^j\alpha^j + n2\alpha^N \leq |x_n^\lambda|_1 + C' + (2/\lambda)5^N 2\alpha^N = |x_n^\lambda|_1 + C$$

as $\alpha < 1/5$.

7.1 The Wandering Exponent

Proposition 2 *Let* $0 < \lambda \leq 1$. *Let* $Cyl(x_n^\lambda, cn)$ *be the set of all points within distance* cn *of the line segment connecting* $\mathbf{0}$ *and* $x_n^\lambda = (n, \lambda n)$. *There exists* $c = c(\lambda) > 0$ *such that for all* n *sufficiently large* $\gamma(\mathbf{0}, (n, \lambda n))$ *is not contained in* $Cyl(x_n^\lambda, cn)$.

Proof Let P be a path from $\mathbf{0}$ to $(n, \lambda n)$ with all of its vertices in $\mathrm{Cyl}(x_n^\lambda, cn)$. Let $j = \lfloor \log_5(\lambda n) \rfloor - 2$. Let j' be the smallest integer such that no horizontal (or vertical) line segment of length $5^{j'-3}$ lies entirely in $\mathrm{Cyl}(x_n^\lambda, cn)$. Note that if c is small and n is sufficiently large then $j > j' + 5$. Let z_1 be the first (closest to $\mathbf{0}$) vertex of P in the j-grid and let z_2 be the last (closest to $(n, \lambda n)$) vertex of P in the j-grid. Note that by the choice of j we have that both the x and y coordinates of z_2 are at least 5^j greater than the respective x and y coordinates of z_1. Thus there exists a northeast directed path P'' from z_1 to z_2 that is contained entirely in the j-grid. We will show that there exists a path P' which is not contained in $\mathrm{Cyl}(x_n^\lambda, cn)$ which is a faster path from $\mathbf{0}$ to $(n, \lambda n)$. P' will agree with P from $\mathbf{0}$ to z_1 and from z_2 to $(n, \lambda n)$. Between z_1 and z_2 the path P' is P''. The choice of j insures that this is possible and $|z_1 - z_2| > \frac{1+\lambda}{2}n$.

As every edge of P' between z_1 and z_2 is in the j-grid, the sum of the passage times of all of these edges is at most

$$|z_1 - z_2|(1 + 2 \cdot \alpha^j).$$

We now show that this is faster than P so this path is not a geodesic.

Define a sequence $\{z_1^i\}_{i=0}^k$ with $z_1^0 = z_1$ and $z_1^k = z_2$ with each z_1^i (with $0 < i < k$) the first time that P hits a new vertical line on the j'-grid. Note that k is at least $5^5 > 1000$. If $0 < i < k - 1$ and the path P between z_1^i and z_1^{i+1} hits another vertical line (besides the start and end lines) in the j'-grid then it has at least $3 \cdot 5^{j-1}$ horizontal edges. Similarly we can see that between z_1^i and z_1^{i+1} the path P hits at two horizontal lines in the $(j-1)$-grid. By the choice of j' and $\lambda \leq 1$ we have $|z_1^i - z_1^{i+1}| \leq 3 \cdot 5^{j'}$ and the difference in the x coordinate is $5^{j'}$. As all edges have passage times between 1 and 2 then P is not a geodesic.

Otherwise as $\lambda \leq 1$ this segment from z_1 to z_2 of P contains edges on the j'-grid on at most one vertical and two horizontal lines. By the choice of j' each of these lines contains at most $5^{j'-3}$ edges in $\text{Cyl}(x_n^\lambda, cn)$. Thus this segment of P contains at most $3 \cdot 5^{j'-3}$ edges in the j'-grid and at least 5^j edges in total. Thus at least 85% of the edges in this segment of P are not in the j'-grid and have passage times at least $1 + \alpha^{j'}$. As this applies to all but the first and last segments, at least 80% of the edges on P from z_1 to z_2 are not in the j'-grid. As above the first and last segments have at most $3 \cdot 5^{j-3}$ edges and thus make up less than three percent of the length of P from z_1 to z_2 (see Fig. 2).

Thus the total passage time for P between z_1 and z_2 is at least

$$|z_1 - v_2|(1 + 0.8 \cdot \alpha^{j'}) > |v_1 - v_2|(1 + 2\alpha^j).$$

Thus the passage time along P is more than the passage time along P' and P is not a geodesic. This proves that the geodesic does not lie in $\text{Cyl}(x_n^\lambda, cn)$.

Proposition 3 *Let* $0 < \lambda \leq 1$. *Remember that* $Cyl(x_n^\lambda, 10n)$ *is the set of all points within distance* $10n$ *of the line segment connecting* $\mathbf{0}$ *and* $(n, \lambda n)$. *For all n sufficiently large* $\gamma(\mathbf{0}, (n, \lambda n))$ *is contained in* $Cyl(x_n^\lambda, 10n)$.

Proof If a path P from $\mathbf{0}$ to $(n, \lambda n)$ is not in $\text{Cyl}(x_n^\lambda, 10n)$ then the length of P is at least $10n$. But as $\lambda \leq 1$ there is a path P' which is in $\text{Cyl}(x_n^\lambda, cn)$ from $\mathbf{0}$ to $(n, \lambda n)$ of length at most $2n$. As every edge has weight at most 2 the length of P' is at most $4n$ and P is not the geodesic from $\mathbf{0}$ to $(n, \lambda n)$.

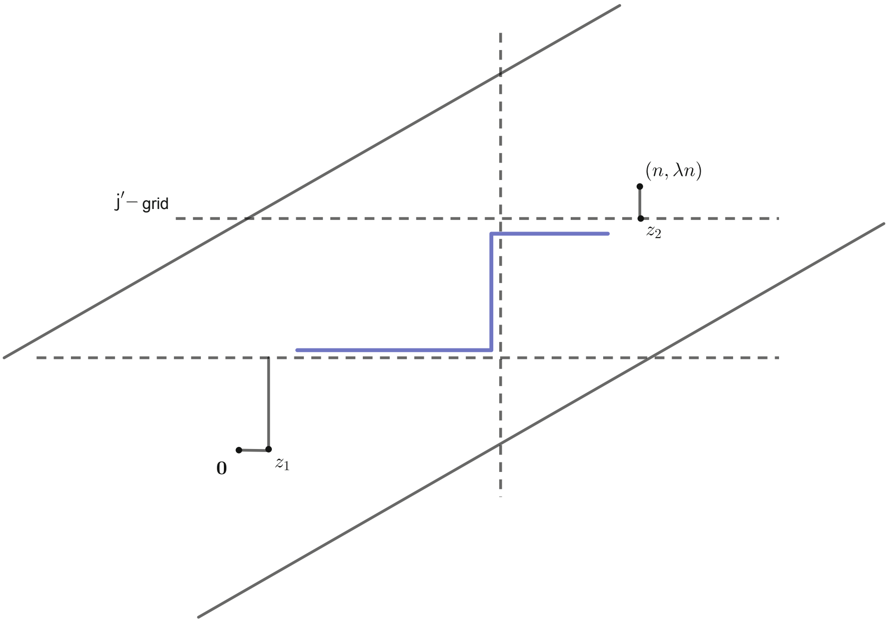


Fig. 2 Proof of Proposition 2: the path P' hits one vertical and two horizontal lines in the j'-grid. Now, by definition of j', at most $3 \cdot 5^{j'-3}$ edges in the geodesic from z_1 to z_2 are on the j'-grid. Furthermore, by the choice of j, z_1 and z_2 most of the edges of the path P are in the segment connecting these two vertices. We then put all these ingredients together to lower bound the passage time of the entire path

8 Exponents in the Coordinate Direction

In this section we consider $\gamma(\mathbf{0}, (n, 0))$. Define

$$\beta = \frac{\log 5}{\log 5 - \log \alpha} < 1. \tag{11}$$

Note that, for any j we can write

$$\alpha^{\beta j} = (5^j)^{\beta - 1}. \tag{12}$$

Lemma 9 *There exists universal constants C and N such that for all $c > C$ and $n > N$ we have*

1. $T(\mathbf{0}, (n, 0)) \geq n$
2. $\mathbb{P}(T(\mathbf{0}, (n, 0)) \leq n + 0.01n^{\beta}) > 10^{-9}$
3. $\mathbb{P}(T(\mathbf{0}, (n, 0)) \geq n + 0.02n^{\beta}) > 10^{-9}$ *and*
4. $\mathbb{P}(T(\mathbf{0}, (n, 0)) \geq n + 10n^{\beta}) = 0.$

Proof The first inequality is true because all passage times are at least 1.

For the second inequality we define Γ_l to be the following path from $(0, 0)$ to $(n, 0)$. The start of Γ_l goes northeast from $(0, 0)$ to the line $y = y_l$, where y_l is the lowest non-negative number such that the line $y = y_l$ is in the l-grid. Suppose we have defined the path to the point (x', y') where both the lines $x = x'$ and $y = y'$ are in the l'-grid. Then we extend the path so that it goes east to the $(l' + 1)$-grid and then north to the $(l' + 1)$-grid. We continue until we have hit the line $y = y_l$. The final portion of Γ_l is defined in a symmetric manner. It goes northwest from $(n, 0)$ to the line $y = y_\ell$. Then Γ_l connects these two pieces by moving horizontally along the line $y = y_\ell$.

Given n, choose j such that

$$5^j \leq n < 5^{j+1}. \tag{13}$$

Let Q be the event that there exists $y^* \in [0, 0.001 \cdot 5^{\beta j}]$ with the line $y = y^*$ in the $(\lfloor \beta j \rfloor + 5)$-grid. If Q occurs then $\Gamma_{\lfloor \beta j \rfloor + 5}$ contains:

1. at most $n + 0.002 \cdot 5^{\beta j}$ edges,
2. at most $20 \cdot 5^k$ edges in the k-grid but not the $(k + 1)$-grid for all $k < \lfloor \beta j \rfloor + 5$, and,
3. at most n edges in the $(\lfloor \beta j \rfloor + 5)$-grid.

If Q occurs, from $1 - 3$ above and the definition of the $X_{k(e),e}$, we have

$$\begin{aligned}
\sum_{e \in \Gamma_{\lfloor \beta j \rfloor + 5}} \alpha^{k(e)} + X_{k(e),e} &\leq \sum_{e \in \Gamma_{\lfloor \beta j \rfloor + 5}} 1.001\alpha^{k(e)} \\
&\leq 1.001 \sum_k \left(20 \cdot 5^k \alpha^k\right) + 1.001 n\alpha^{\lfloor \beta j \rfloor + 5} \\
&\leq C + 0.001 \cdot 5^{\beta j - 1} \\
&\leq C + 0.001 n^\beta.
\end{aligned}$$

Then, if Q occurs

$$\begin{aligned}
T((0, 0), (n, 0)) &\leq T(\Gamma_{\lfloor \beta j \rfloor + 5}) \\
&\leq |\Gamma_{\lfloor \beta j \rfloor + 5}| + \sum_{e \in \Gamma_{\lfloor \beta j \rfloor + 5}} \alpha^{k(e)} + X_{k(e),e} \\
&\leq n + 0.002 \cdot 5^{\beta j} + C + 0.001 n^\beta \\
&\leq C + n + 0.004 n^\beta \\
&\leq n + 0.01 n^\beta.
\end{aligned}$$

Then

$$\mathbb{P}(T(\mathbf{0}, (n, 0)) \le n + 0.01n^{\beta}) \ge \mathbb{P}(Q) \ge 0.001 \cdot 5^{-5} \ge 2 \cdot 10^{-9}$$

and the result follows.

The fourth inequality follows in much the same way as the second except we do not assume that the event Q occurs. In this case we have that $\Gamma_{\lfloor \beta j \rfloor}$ contains

1. at most $n + 2 \cdot 5^{\beta j}$ edges
2. at most $20 \cdot 5^{k}$ edges in the k-grid but not the $(k+1)$-grid for all $k < \lfloor \beta j \rfloor$ and
3. at most n edges in the $(\lfloor \beta j \rfloor)$-grid.

Then a similar calculation as above proves the claim.

For the third inequality we note that if there does not exist y_0 such that $|y_0| \le 0.1 \cdot 5^{\beta j}$ such that the line $y = y_0$ is in the $\lfloor \beta j \rfloor$-grid and if Γ' be any path from $(0, 0)$ to $(n, 0)$ in the cylinder $\{(x, y) : |y| \le 0.1 \cdot 5^{\beta j}\}$, then

$$\begin{aligned} T(\Gamma') &\ge n(1 + \alpha^{\beta j - 1}) \\ &\ge n + \frac{1}{\alpha} n \alpha^{\beta j} \\ &\ge n + \frac{1}{\alpha} n (5^{j})^{\beta - 1} \\ &\ge n + \frac{1}{\alpha} n (5^{j})^{\beta} (5^{j})^{-1} \\ &\ge n + \frac{1}{\alpha} (5^{j})^{\beta} \\ &\ge n + \frac{1}{\alpha} (5^{j+1})^{\beta} 5^{-\beta} \\ &\ge n + \frac{1}{\alpha 5^{\beta}} n^{\beta} \\ &\ge n + n^{\beta}. \end{aligned}$$

Now let Γ'' be any path from $(0, 0)$ to $(n, 0)$ not contained in the cylinder $\{(x, y) : |y| \le 0.1 \cdot 5^{\beta j}\}$. Then by (11) and (13)

$$\begin{aligned} T(\Gamma'') &\ge n + 0.2 \cdot 5^{\beta j} \\ &\ge n + 0.2 \cdot (5^{j+1})^{\beta} 5^{-\beta} \\ &\ge n + 0.04 n^{\beta}. \end{aligned}$$

As any path from $(0, 0)$ to $(n, 0)$ falls into one of these two categories we have that

$$T((0, 0), (n, 0)) \ge \min(n + n^{\beta}, n + 0.04n^{\beta}) = n + 0.04n^{\beta}.$$

This happens with probability at least

$$1 - \frac{0.3 \cdot 5^{\beta j}}{5^{\lfloor \beta j \rfloor}} \geq 10^{-9}.$$

We use Lemma 9 to show that the variance exponent is β along the axes.

Lemma 10 *There exists $K > 0$ such that for all n sufficiently large*

$$\frac{1}{K} n^{2\beta} < Var(T(\mathbf{0}, (n, 0))) < K n^{2\beta}.$$

Proof The lower bound follows directly from parts 2 and 3 from Lemma 9. The upper bound follows from parts 1 and 4.

For any K define $\mathrm{Cyl}((n, 0), K)$ be the subgraph with vertices $\{(x, y) : |y| \leq K\}$ and all edges between two vertices in the set. Now we show that the fluctuation exponent is also β.

Lemma 11 *For any $\epsilon > 0$*

$$\mathbb{P}(\gamma(\mathbf{0}, (n, 0)) \text{ is contained in } Cyl((n, 0), n^{\beta-\epsilon}) = o(1).$$

Also

$$\mathbb{P}(\gamma(\mathbf{0}, (n, 0)) \text{ is not contained in } Cyl((n, 0), 10n^{\beta}) = 0.$$

Proof Define

$$j = j(n) = \max\{k(e) : \ e \text{ is an horizontal edge in } \gamma(\mathbf{0}, (n, 0))\}.$$

We first notice that all horizontal edges in $\gamma(\mathbf{0}, (n, 0))$ with $k(e) = j$ are contained in the horizontal line that is furthest away from the x-axis. Consider a path P that goes up to the $j + 1$ grid and connects $\mathbf{0}$ and $(n, 0)$. We have

$$T(\mathbf{0}, (n, 0)) \leq T(P).$$

by definition.

For any $\epsilon > 0$ and for all n sufficiently large we will show that

$$\mathbb{P}(\exists \text{ a path } P \text{ from } \mathbf{0} \text{ to } (n, 0) \text{ contained in } \mathrm{Cyl}((n, 0), n^{\beta-\epsilon})$$
$$\text{with } T(P) \leq n + 10n^{\beta}) = o(1).$$

There are at least n horizontal edges in any path from $\mathbf{0}$ to $(n, 0)$. If all the horizontal edges in $\mathrm{Cyl}((n, 0), n^{\beta-\epsilon})$ have passage time at least $1 + 10n^{\beta-1}$ then the passage time across any path from $\mathbf{0}$ to $(n, 0)$ entirely contained in $\mathrm{Cyl}((n, 0), n^{\beta-\epsilon})$ has passage time at least $n + 10n^{\beta}$.

By part 4 of Lemma 9 the event that

$$\gamma(\mathbf{0}, (n, 0)) \text{ is contained in } \mathrm{Cyl}((n, 0), n^{\beta-\epsilon})$$

is contained in the event that

$$\text{there exists a path } P \text{ from } \mathbf{0} \text{ to } (n, 0) \text{ contained in } \mathrm{Cyl}((n, 0), n^{\beta-\epsilon})$$
$$\text{with } T(P) \leq n + 10n^{\beta}.$$

This last event is in turn contained in the event that

$$\text{there exists a horizontal edge in } \mathrm{Cyl}((n, 0), n^{\beta-\epsilon})$$
$$\text{with passage time at most } 1 + 10n^{\beta-1}.$$

This requires that there is a line of the form $y = l$ which is in the $(\lfloor \beta j \rfloor - 3)$-grid with $l \in [-n^{\beta-\epsilon}, n^{\beta-\epsilon}]$. By the choice of β and j the probability of this is at most

$$\frac{2n^{\beta-\epsilon} + 1}{5^{\beta j-3}} \leq Cn^{-\epsilon}.$$

The upper bound follows from part 4 of Lemma 9 and the fact that all passage times are at least 1.

8.1 *Proof of Theorem 3*

For the non-coordinate directions, $\chi = 0$ follows directly from Lemma 8 and $\xi = 1$ follows combining Propositions 2 and 3. For the coordinate directions, $\chi = \xi = \beta$ is a consequence of Lemmas 10 and 11.

Acknowledgments The authors are grateful to the anonymous referee whose suggestions improved the presentation of our work. G.B. thanks Ken Alexander and Michael Damron for fruitful conversations at the early stages of this project. C.H. was supported by grant DMS-1712701.

References

1. Ahlberg, D., Hoffman, C.: Random coalescing geodesics in first-passage percolation (2016). arXiv:1609.02447v1

2. Alexander, K.S., Berger, Q.: Geodesics toward corners in first passage percolation. J. Stat. Phys. **172**(4), 1029–1056 (2018). MR 3830297
3. Auffinger, A., Damron, M., Hanson, J.: 50 Years of First-Passage Percolation, vol. 68. American Mathematical Soc., Providence (2017)
4. Boivin, D.: First passage percolation: the stationary case. Probab. Theory Relat. Fields **86**(4), 491–499 (1990). MR 1074741
5. Cox, J.T., Durrett, R.: Some limit theorems for percolation processes with necessary and sufficient conditions. Ann. Probab. **9**, 583–603 (1981)
6. Damron, M., Hanson, J.: Busemann functions and infinite geodesics in two-dimensional first-passage percolation. Commun. Math. Phys. **325**(3), 917–963 (2014)
7. Häggström, O., Meester, R.: Asymptotic shapes for stationary first passage percolation. Ann. Probab. **23**(4), 1511–1522 (1995). MR 1379157 (96m:60237)
8. Hoffman, C.: Coexistence for Richardson type competing spatial growth models. Ann. Appl. Probab. **15**(1B), 739–747 (2005). MR 2114988 (2005m:60235)
9. Hoffman, C.: Geodesics in first passage percolation. Ann. Appl. Probab. **18**(5), 1944–1969 (2008)
10. Newman, C.M., Piza, M.S.T.: Divergence of shape fluctuations in two dimensions. Ann. Probab. **23**(3), 977–1005 (1995)

Avalanches in Critical Activated Random Walks

Manuel Cabezas and Leonardo T. Rolla

Abstract We consider Activated Random Walks on $\mathbb{Z}$ with totally asymmetric jumps and critical particle density, with different time scales for the progressive release of particles and the dissipation dynamics. We show that the cumulative flow of particles through the origin rescales to a pure-jump self-similar process which we describe explicitly.

Keywords Self-organized criticality · Absorbing-state phase transitions · Avalanches · Scaling limits · Duality · Brownian web · Critical flow

MSC 82C27, 60K35, 82C23, 60K40

1 Introduction

The totally asymmetric Activated Random Walk (ARW) dynamics on $\mathbb{Z}$ is a continuous-time conservative system made of active and passive particles, where each active particle jumps from x to $x + 1$ at rate 1, and spontaneously decays to a passive state at rate $0 < \lambda < \infty$. Active particles reactivate passive particles instantly when they occupy the same site, in particular active particles at the same site prevent each other from decaying. This model has received increasing attention [1, 3–5, 7, 9, 12–18], see [11] for a self-contained introduction.

This model displays a phase transition in terms of the density of particles: Let $\eta(x)$ denote the initial number of particles at $x \in \mathbb{Z}$ and assume that the initial configuration is i.i.d. with mean $\zeta = \mathbb{E}[\eta(0)]$. If $\zeta > \frac{\lambda}{1+\lambda}$ then the system can

M. Cabezas
Universidad Católica de Chile, Santiago, Chile
e-mail: mncabeza@mat.puc.cl

L. T. Rolla (✉)
IMAS-Conicet, Buenos Aires, Argentina and NYU-Shanghai, Shanghai, China
e-mail: leorolla@dm.uba.ar

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_9

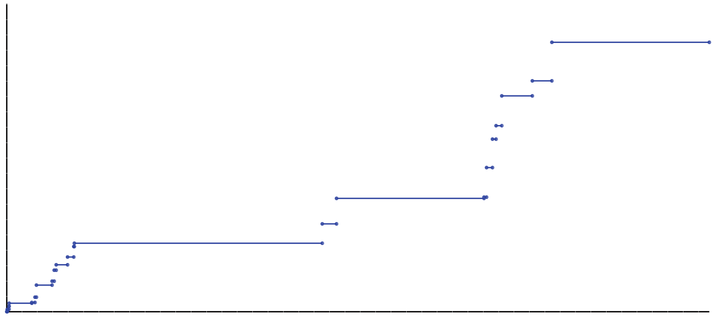


Fig. 1 Simulation of the avalanche process with $n = 10^5$, $\zeta = 0.808$ and $\rho = 0.1$

sustain a non-vanishing density of active particles, whereas, if $\zeta \leqslant \frac{\lambda}{1+\lambda}$, the density of active particles decays to 0, see [11].

In this paper we study the flow process defined as follows. Let $\eta_0 = \eta(0)\delta_0$, that is, the configuration where site 0 has $\eta(0)$ particles and other sites are vacant. We then run the above dynamics starting from η_0 until we get a configuration η'_0 without active particles, which we call *stable*. Finally, we let C_0 denote the number of particles which jump from 0 to $+1$ during this evolution. In the next step, we take $\eta_1 = \eta'_0 + \eta(-1)\delta_{-1}$, that is, we add $\eta(-1)$ particles to the site -1. Again we run the above dynamics starting from η_1 until we get a stable configuration η'_1, and let C_1 denote the total number of particles which jump from 0 to $+1$. In the same fashion, we define $\eta_n = \eta'_{n-1} + \eta(-n)\delta_{-n}$, stabilize η_n obtaining η'_n, and define C_n as the number of particles of η'_n found to the right of site 0. Finally, the *flow process* is defined as $(C_n)_{n=0,1,2,\ldots}$.

Note that the sequence $(\eta'_n)_n$ is a non-homogeneous Markov process with respect to its natural filtration. The transition probabilities are determined by the common distribution of $\eta(x)$ and by the dynamics run between each pair of steps, which is parametrized by λ. Also, C_n can be read from η'_n, but $(C_n)_n$ is not Markovian.

If the system is subcritical (i.e., $\zeta < \frac{\lambda}{1+\lambda}$) then C_n is eventually constant. If it is supercritical, $\frac{C_n}{n}$ tends to a positive number. In the critical case, none of the above happens: C_n diverges but $\frac{C_n}{n}$ vanishes, and one expects the system to have a non-trivial scaling limit (Fig. 1).

Let $\sigma_s^2 = \zeta - \zeta^2$ denote the variance of a Bernoulli variable with parameter ζ, and let $\sigma_p^2 = \mathbb{E}[\eta(0)^2] - \zeta^2 \geqslant \sigma_s^2$. Consider the critical case $\zeta = \frac{\lambda}{1+\lambda}$, assume that $\sigma_p^2 < \infty$ and define $\rho := \frac{\sigma_s}{\sigma_p} \in (0, 1]$.

Theorem 1 *There exists a stochastic process* $(\mathscr{C}_x^\rho)_{x \geqslant 0}$ *such that*

$$\frac{1}{\sigma_p}\left(\varepsilon\, C_{\lfloor \varepsilon^{-2} x \rfloor}\right)_{x \geqslant 0} \xrightarrow{\mathrm{d}} \left(\mathscr{C}_x^\rho\right)_{x \geqslant 0}$$

in the Skorohod J_1 *metric as* $\varepsilon \to 0$.

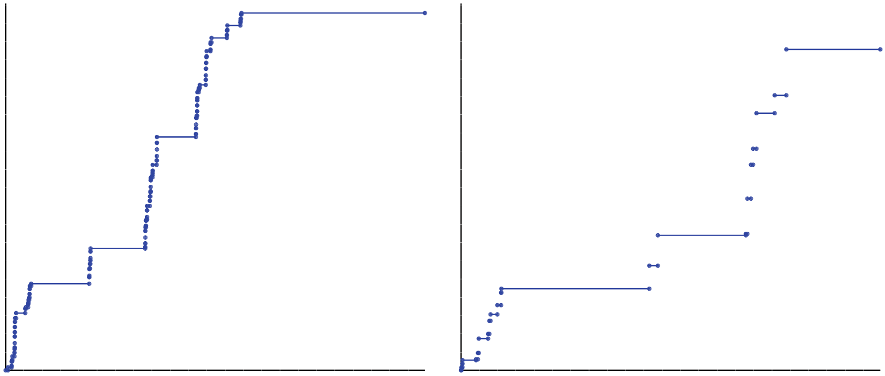


Fig. 2 comparison of an avalanche process with $\rho = 0$ and $\rho = 0.1$

Remark 1 The study of critical flows for the ARW started in [7], where the extreme case $\lambda = \infty$ was studied. In this case, $\rho = 0$ and the limiting process $\big(\mathscr{C}_x^0\big)_{x\geqslant 0}$ is the running maximum of a Brownian motion.[1] See Fig. 2 for a comparison.

Below we will give a description of the stochastic process $\big(\mathscr{C}_x^\rho\big)_{x\geqslant 0}$ in terms of a family of coalescing reflected correlated Brownian motions. We point out that the superscript ρ in the limiting process cannot be reduced to a multiplicative factor, and these processes are indeed very different as ρ varies, see Fig. 3.

Before that, we state a qualitative property of the scaling limit which explains why we call it an avalanche process. Being a scaling limit, it is scale invariant.

Theorem 2 *The process $\big(\mathscr{C}_x^\rho\big)_{x\geqslant 0}$ consists of pure jumps. Its jump times accumulate at 0 (as $\mathscr{C}^\rho$ is scale-invariant) but are otherwise discrete.*

The limiting process $\mathscr{C}^\rho$ can be informally described as follows. Suppose $0 < x_0 < \cdots < x_k$, we want to sample $\mathscr{C}_x^\rho$ for $x = x_0, \dots, x_k$. Let $(\mathscr{P}_x)_{x\leqslant 0}$ be a backward Brownian motion, with diffusivity constant equal to 1, started at the origin. Then, for each $i = 0, \dots, k$, we run a Brownian motion $(\mathscr{R}_t^i)_{t\geqslant -x_i}$ started at $\mathscr{R}_{-x_i}^i = \mathscr{P}_{-x_i}$, with diffusivity ρ, reflected from below by the graph of $\mathscr{P}$. We let the $k+1$ reflected Brownian motions diffuse independently (except for the reflection) until they coalesce. Finally, $\mathscr{C}_{x_i}$ is given by $\mathscr{R}_0^i$ for $i = 0, \dots, k$.

Finally, we would like to say a word on the metric of Theorem 1. The convergence being in the Skorohod J_1-metric implies that each jump in the limiting process must be matched by a corresponding jump in the discrete process. Hence, the discontinuities in $\mathscr{C}^\rho$ (guaranteed by Theorem 2) correspond to abrupt increments, avalanches in the discrete process $(C_i)_{i\in\mathbb{N}}$.

[1]In this case, the same convergence as stated in Theorem 1 follows from the more complicated, continuous-time analysis done in [7], or alternatively from the arguments presented here.

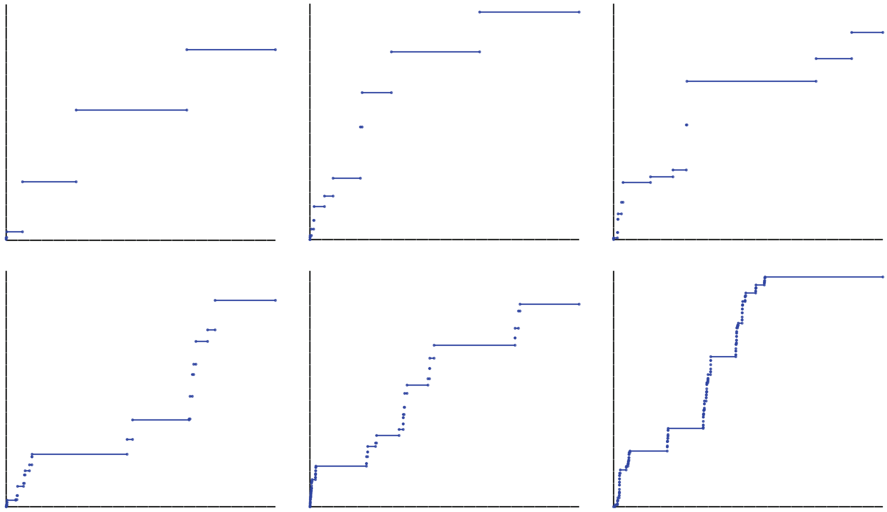


Fig. 3 Simulations of $(C_n)_n$ with $\rho = 1.000, 0.500, 0.300, 0.100, 0.051, 0.000$. The steps are 100,000 for the first four graphs, then 50,000 and 25,000. For large ρ, the process seems more jumpy, and as ρ gets smaller the size of jumps become smaller and jumps tend do cluster together, until finally at the extreme case $\rho = 0$ studied in [7] the process becomes continuous

The entire literature on ARW would probably not exist were it not for our dear friend Vladas Sidoravicius. He was always in contact with prominent physicists, including Ronald Dickman, bringing interesting cutting-edge problems to the Probability community and enthusiastically promoting them. In particular, he has been advertising the Activated Random Walk model since the early 2000s. In 2007, he proposed this problem to the second author as part of his PhD studies, which finally resulted in [12]. During our joyful meetings in the early 2010s, we worked on predecessors of the current paper [7, 8].

This paper is organized as follows. In Sect. 2 we give the formal definition of the limiting process $\mathscr{C}^\rho$. In Sect. 3 we define the ARW and state the Abelian property of its site-wise construction. In Sect. 4 we introduce the sequential stabilization that will be used throughout the article. In Sect. 5 we prove convergence of finite dimensional distributions to later get full convergence (in the J_1 Skorohod metric) in Sect. 6. Finally, in Sect. 7 we show that $\mathscr{C}^\rho$ is a pure-jump process when $\rho \neq 0$.

2 Formal Definition of the Limiting Process

Next, we will describe formally the finite-dimensional distributions of the limiting process. For convenience, and building the connection with the upcoming proofs, we describe the differences $\mathscr{R}^i - \mathscr{P}$, instead, given by the $B^{+,i}$ below.

Let $k \in \mathbb{N}_0$ and $-x_k \leqslant -x_{k-1} \leqslant \cdots < -x_0 \leqslant 0$. Let $(\mathscr{S}^i_x)_{x \geqslant -x_i}, i = 0, \ldots, k$. be independent Brownian motions started at 0, $(\mathscr{S}^i_{-x_i} = 0)$ with diffusion coefficient σ_s. Let also $(\mathscr{P}_x)_{x \leqslant 0}$ be a backwards Brownian motion started at 0 $(\mathscr{P}_0 = 0)$ with diffusion coefficient σ_p and independent of $\mathscr{S}^i, i = 0, \ldots, k$. For $i = 0, \ldots, k$ and $x \geqslant -x_i$ define

$$\tilde{B}^i_x := \mathscr{P}_x - \mathscr{P}_{-x_i} - \mathscr{S}^i_x,$$

so that $(\tilde{B}^i_x)_{x \geqslant -x_i}$ are Brownian motions started at 0, that is $(\tilde{B}^i_{-x_i} = 0)$, with diffusion coefficient $r = \sqrt{\sigma_s^2 + \sigma_p^2}$. For $i = 0, \ldots, k$ and $x \geqslant -x_i$ we define

$$\tilde{B}^{+,i}_x := \tilde{B}^i_x - \inf\{\tilde{B}^i_s : s \in [-x_i, x]\} \tag{1}$$

so that $(\tilde{B}^{+,i}_x)_{x \geqslant -x_i}, i = 0, \ldots, k$ are reflected Brownian motions started at 0 with diffusion coefficient r. Let $B^{+,k} := \tilde{B}^{+,k}$ and, for $i = k-1, k-2, \ldots, 0$, let

$$\tau_i := \inf\{x \geqslant -x_i : \tilde{B}^{+,i}_x = B^{+,i+1}_x\} \quad \text{and} \quad B^{+,i}_x := \begin{cases} \tilde{B}^{+,i}_x & : x \in [-x_i, \tau_i), \\ B^{+,i+1}_x & : x \geqslant \tau_i. \end{cases} \tag{2}$$

Then, for each $k \in \mathbb{N}$ and each sequence $0 \leqslant x_1 \leqslant \cdots \leqslant x_k$ we have that

$$(\mathscr{C}^\rho_{x_0}, \ldots, \mathscr{C}^\rho_{x_k}) \overset{d}{=} (\tfrac{1}{\sigma_p} B^{+,0}_0, \ldots, \tfrac{1}{\sigma_p} B^{+,k}_0),$$

where $\overset{d}{=}$ denotes equality in distribution.

The above description is self-consistent, in the sense that removing points from the $\{x_0, \ldots, x_k\}$ does not affect the distribution of the remaining points. Since the resulting process is non-decreasing, we can already deduce the existence of a càdlàg process $\left(\mathscr{C}^\rho_x\right)_{x \geqslant 0}$ with these finite-dimensional distributions.

In Fig. 4, we see the graph of $\mathscr{C}^\rho$ on the right-hand side ($x \geqslant 0$) and the elements of this construction on the left side. There is a black curve starting at (0, 0) moving backwards (from right to left). This curve is the Brownian motion $\mathscr{P}$ with diffusivity σ_p. From each point on the black we can start a red path moving forward (from left to right). This red path is a Brownian with diffusivity σ_s, reflected against the black curve. Different red paths diffuse independently until they meet, and coalesce after that. To find the value of $\mathscr{C}^\rho$ at a certain $x_0 > 0$, we follow the red path which starts at $(-x_0, \mathscr{P}_{-x_0})$ until it hits the vertical axis $\{x = 0\}$. The value $\mathscr{C}^\rho_{x_0}$ is the height coordinate of the red path at this terminal point.

An alternative description is the following. By considering the dual of the red paths one obtains a family of blue paths. These paths start from each point on the positive vertical axis and move from right to left. Blue paths coalesce when they meet, and they terminate upon hitting the black curve. Each terminal point

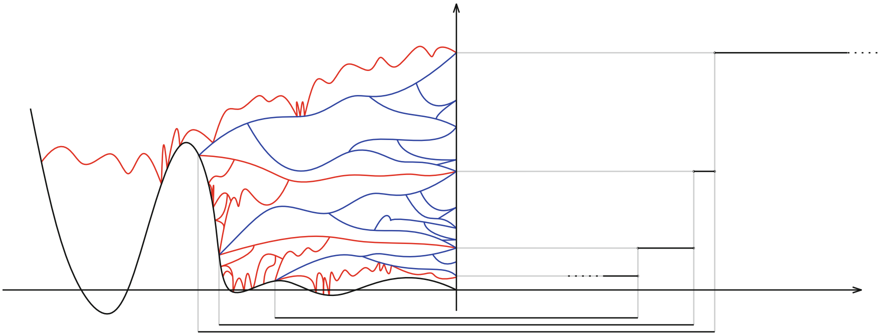


Fig. 4 Black path, blue paths and dual red paths on the left, $\mathscr{C}^\rho$ on the right

determines an interval given by the initial height of the blue paths which terminate there. This interval corresponds to the jumps of $\mathscr{C}^\rho$, see Fig. 4. On the left side we see the intervals determined by the blue paths. The extremes of the interval (down and up) correspond to the different (left and right) limits at a discontinuity of $\mathscr{C}^\rho$. The x coordinate of the jumps is (the reflection of) the x coordinate of the terminal point of the blue paths. This description is technically simpler, and will be used to prove convergence with respect to the J_1 metric in Sect. 6.

3 Explicit Construction and Abelian Property

In this section we give a more formal definition of the ARW dynamics, and briefly recall the site-wise construction. For details, see [11].

3.1 Notation and ARW Dynamics

Let $\mathbb{N} = \{0, 1, 2, \dots\}$ and $\mathbb{N}_\mathfrak{s} = \mathbb{N}_0 \cup \{\mathfrak{s}\}$, where $\mathfrak{s}$ represents a sleeping particle. For convenience we define $|\mathfrak{s}| = 1$, and $|n| = n$ for $n \in \mathbb{N}$, and write $0 < \mathfrak{s} < 1 < 2 < \cdots$. Also define $\mathfrak{s} + 1 = 2$ and $n \cdot \mathfrak{s} = n$ for $n \geqslant 2$ and $\mathfrak{s}$ if $n = 1$.

The ARW dynamics $(\eta_t)_t$ is defined as follows. A site x is *unstable* if it has active particles, i.e., if $\eta_t(x) \geqslant 1$. At each unstable site x, a clock rings at rate $(1+\lambda)\,|\eta_t(x)|$. When this clock rings, site x is *toppled*, which means that the system goes through the transition $\eta \to \mathfrak{t}_{x\mathfrak{s}}\eta$ with probability $\frac{\lambda}{1+\lambda}$, otherwise $\eta \to \mathfrak{t}_{x\mathfrak{m}}\eta$

with probability $\frac{1}{1+\lambda}$. These transitions are given by

$$\mathfrak{t}_{x\mathfrak{m}}\eta(z) = \begin{cases} \eta(x) - 1, & z = x, \\ \eta(y) + 1, & z = x + 1, \\ \eta(z), & \text{otherwise}, \end{cases} \qquad \mathfrak{t}_{x\mathfrak{s}}\eta(z) = \begin{cases} \eta(x) \cdot \mathfrak{s}, & z = x, \\ \eta(z), & \text{otherwise}, \end{cases}$$

and only occur when $\eta(x) \geqslant 1$. The operator $\mathfrak{t}_{x\mathfrak{s}}$ represents a particle at x trying to fall asleep, which effectively happen if there are no other particles present at x. Otherwise, by definition of $n \cdot \mathfrak{s}$, the system state does not change. The operator $\mathfrak{t}_{x\mathfrak{m}}$ represents a particle jumping from x to $x+1$, where possible activation of a sleeping particle previously found at $x + 1$ is represented by the convention that $\mathfrak{s} + 1 = 2$.

3.2 Site-Wise Representation and Abelian Property

We now define a field of *instructions* to be read by the active particles. The instructions $\mathscr{I} = (\mathfrak{t}^{x,j})_{x\in\mathbb{Z},j\in\mathbb{N}}$ are i.i.d. with $\mathbb{P}[\mathfrak{t}^{x,j} = \mathfrak{t}_{x\mathfrak{m}}] = \frac{1}{1+\lambda}$ and $\mathbb{P}[\mathfrak{t}^{x,j} = \mathfrak{t}_{x\mathfrak{s}}] = \frac{\lambda}{1+\lambda}$. Using a field of instructions, the operation of *toppling* a site x consists in applying the first instruction available at x, and discarding it so that the next unused instruction at x becomes available. Toppling a site is *legal* if it is unstable.

The ARW dynamics can be recovered from the initial configuration and instructions as follows. Suppose that every active particle carries a clock which rings according to a Poisson process. Different particles carry independent clocks. When a clock rings for some particle, we topple the site where it is located. For a system with finite initial configuration, the process obtained this way has the same distribution as the one described above.

The *Abelian property* reads as follows. For each finite set $V \subseteq \mathbb{Z}$ and initial configuration, if two legal sequences of topplings are contained in V and make each site in V stable, then the resulting configuration is the same. Using the Abelian property, we can answer many questions of the ARW model by choosing the order in which the sites topple instead of using the Poisson clocks.

4 Sequential Stabilization

Let $L \in \mathbb{N}$ and, for $x = -L, \ldots, 0$, let $N_L(x)$ denote the number of particles which jump from x (to $x + 1$) when $[-L, 0]$ is stabilized. Moreover, due to the Abelian property, one has that

$$C_L = N_L(0),$$

in the sense that $(C_n)_{n\in\mathbb{N}}$ will have the same distribution as the process defined in Sect. 1. Indeed, we can first stabilize V_0, then V_1, and so on until V_L. Then writing k for the number of times that 0 is toppled, $N_L(0)$ will be given by the number of $\mathfrak{t}_{0\mathfrak{m}}$ found among $\mathfrak{t}^{0,1}, \dots, \mathfrak{t}^{0,k}$.

In what follows, we will decompose $N_L(x)$ as the sum of two processes, one accounting for the randomness of the initial configuration and another for the randomness of the sleeping instructions. Recall that $\tilde{B}^{+,i}$ is a reflected Brownian motion with diffusion coefficient $r = \sqrt{\sigma_s^2 + \sigma_p^2}$.

Proposition 1 *For all $i = 0, \dots, k$ we have that*

$$\left(\varepsilon\, N_{\lfloor \varepsilon^{-2} x_i \rfloor}(\lfloor \varepsilon^{-2} x \rfloor)\right)_{x \geqslant -x_i} \xrightarrow{\mathrm{d}} (\tilde{B}_x^{+,i})_{x \geqslant -x_i} \tag{3}$$

as $\varepsilon \to 0$ in the metric of uniform convergence on compact intervals of time.

For the proof of the proposition above, we will need to describe the sequential stabilization as a reflected random walk. We will show that for each $L \in \mathbb{N}$ the process $(N_L(z))_{z\geqslant -L}$ is distributed as a random walk started at zero and reflected at zero.

We stabilize η on $[-L, 0]$ as follows. Topple site $z = -L$ until it is stable, and denote by $Y_L(-L)$ the indicator of the event that the last particle remains passive on $z = -L$. In case $\eta(-L) = 0$, sample $Y_L(-L)$ independently of anything else. By the Abelian property, we have that

$$N_L(-L) := [\eta(-L) - Y_L(-L)]^+.$$

Note that, after stabilizing $z = -L$, there are $N_L(-L) + \eta(-L+1)$ particles at $z = -L+1$. Now topple site $z = -L+1$ until it is stable, and denote by $Y_L(-L+1)$ the indicator of the event that the last particle remains passive on $z = -L+1$. If there are no particles in z, sample $Y_L(z)$ independent of everything else. Continue this procedure for $z = -L+2, \dots, 0$. Let

$$T_L(x) = \sum_{y=-L}^{x} (\eta(y) - Y_L(y))$$

and observe that

$$N_L(x) = T_L(x) - \inf_{y=-L,\dots,x} T_L(y).$$

Write $L = L(\varepsilon) = \lfloor \varepsilon^{-2} x_i \rfloor$. Note that $Y_L(-L), \dots, Y_L(0), \eta(-L), \dots, \eta(0)$ are independent. Hence, the increments $(\eta(x) - Y_L(x))_{x\geqslant -L}$ of T_L are i.i.d. Since we are assuming that $\zeta = \frac{\lambda}{1+\lambda}$ and $\sigma_p < \infty$, each term in the sum has mean zero and finite variance, so it follows by Donsker's invariance principle that

$$\Big(\varepsilon\, T_L(\varepsilon^{-2}x)\Big)_{x\in[-x_i,0]} \xrightarrow{\mathrm{d}} (\tilde{B}^i_x)_{x\in[-x_i,0]} \tag{4}$$

as $\varepsilon \to 0$ in the uniform metric. Since the reflection map $\tilde{B}^i \mapsto \tilde{B}^{+,i}$ in (1) is continuous, the above convergence implies (3).

5 Convergence of Finite-Dimensional Projections

Let $k \in \mathbb{N}_0$ and $-x_k \leqslant -x_{k-1} \leqslant \cdots < -x_0 \leqslant 0$. Recall the definition of $(B_0^{+,0}, B_0^{+,1}, \ldots, B_0^{+,k})$ in Sect. 2. In this section we prove the following.

Theorem 3 (Finite-Dimensional Convergence) *We have*

$$\big(\varepsilon\, C_{\lfloor\varepsilon^{-2}x_0\rfloor}, \varepsilon\, C_{\lfloor\varepsilon^{-2}x_1\rfloor}, \ldots, \varepsilon\, C_{\lfloor\varepsilon^{-2}x_k\rfloor}\big) \xrightarrow{\mathrm{d}} (B_0^{+,0}, B_0^{+,1}, \ldots, B_0^{+,k}),$$

as $\varepsilon \to 0$.

Recalling the construction in the previous section, since $C_{\lfloor\varepsilon^{-2}x_i\rfloor} = N_{\lfloor\varepsilon^{-2}x_i\rfloor}(0)$, it is certainly enough to show joint convergence of the counting processes

$$\bigg(\Big(\varepsilon N_{\lfloor\varepsilon^{-2}x_i\rfloor}(\lfloor\varepsilon^{-2}x\rfloor)\Big)_{x\in[-x_i,0]}\bigg)_{i=0,\ldots,k} \xrightarrow{\mathrm{d}} \bigg(\Big(B_x^{+,i}\Big)_{x\in[-x_i,0]}\bigg)_{i=0,\ldots,k}.$$

To keep exposition simpler, we consider the case $k = 1$. There are no differences when considering larger k except for more cluttered notation. For $x \leqslant 0$, we define

$$P(x) := \sum_{y=x}^{0} (\zeta - \eta(y))$$

Given $L \in \mathbb{N}$ and $x \geqslant -L$, let

$$S_L(x) := \sum_{y=-L}^{x} (Y_L(y) - \zeta)\,,$$

where $Y_L(y)$ is as in Sect. 4. Actually, we can construct the processes P, S_L and Y_L jointly for different L's. That is, for $-L_1 < -L_0 < 0$, define

$$T_{L_i}(x) = \sum_{y=-L_i}^{x} \big(\eta(y) - Y_{L_i}(y)\big) = P(x) - P(-L_i) - S_{L_i}(x) \quad i = 0, 1 \tag{5}$$

and

$$N_{L_i}(x) = T_{L_i}(x) - \min_{y\in[-L_i,x]} T_{L_i}(y) \quad i = 0, 1.$$

The key observation is that the η terms in (5) are common for $i = 0, 1$, whereas the Y terms are independent until $\theta := \inf\{y : N_{L_0}(y) = N_{L_1}(y)\}$. After θ, the Y terms are also common for $i = 0, 1$.

We will need a modified version of the process T_{L_0}, whose S-component remains independent of T_{L_1} even after θ. Let

$$\tilde{Y}_{L_0}(y) := \begin{cases} Y_{L_0}(y) & \text{if } y \leqslant \theta, \\ \tilde{s}(y) & \text{if } y > \theta, \end{cases}$$

where $\tilde{s}$ are i.i.d. Bernoulli random variables with parameter ζ, independent of everything else. Let

$$\tilde{S}_{L_0}(x) := \sum_{i=-L_0}^{x} \left(\tilde{Y}_{L_0}(x) - \zeta\right)$$

and

$$\tilde{T}_{L_0}(x) := \sum_{i=-L_0}^{x} \left(\eta(x) - \tilde{Y}_{L_0}(x)\right) = P(x) - P(-L_0) - \tilde{S}_{L_0}(x).$$

By Donsker's invariance principle (taking $L_i = \lfloor \varepsilon^{-2} x_i \rfloor$), the triple

$$\left((\varepsilon P(\lfloor\varepsilon^{-2}x\rfloor))_{x\leqslant 0}, (\varepsilon S_{\lfloor\varepsilon^{-2}x_1\rfloor}(\lfloor\varepsilon^{-2}x\rfloor))_{x\geqslant -x_1}, (\varepsilon \tilde{S}_{\lfloor\varepsilon^{-2}x_0\rfloor}(\lfloor\varepsilon^{-2}x\rfloor))_{x\geqslant -x_0}\right)$$

converges in distribution to three independent Brownian motions

$$\left((\mathscr{P}_x)_{x\leqslant 0}, (\mathscr{S}^1_t)_{x\geqslant -x_1}, (\tilde{\mathscr{S}}^0_t)_{x\geqslant -x_0}\right)$$

with diffusion coefficients $\sigma_p, \sigma_s, \sigma_s$ respectively and started at 0 (i.e., $\mathscr{P}_0 = \mathscr{S}^1_{-x_1} = \tilde{\mathscr{S}}^0_{-x_0} = 0$).

Let

$$\mathscr{T}^1_x := \mathscr{P}_x - \mathscr{P}_{-x_1} - \mathscr{S}^1_x, \quad 0 \geqslant x \geqslant -x_1$$

and

$$\tilde{\mathscr{T}}^0_x := \mathscr{P}_x - \mathscr{P}_{-x_0} - \tilde{\mathscr{S}}^0_x, \quad x \geqslant -x_0.$$

Let also

$$\mathscr{N}_x^1 = \mathscr{T}_x^1 - \inf_{s \in [-x_1, x]} \mathscr{T}_s^1$$

and

$$\tilde{\mathscr{N}}_x^0 = \tilde{\mathscr{T}}_x^0 - \inf_{s \in [-x_0, x]} \tilde{\mathscr{T}}_s^0.$$

Define

$$N_x^{\varepsilon,0} := \varepsilon \, N_{\lfloor \varepsilon^{-2} x_0 \rfloor}(\varepsilon^{-2} x),$$

also define $\tilde{N}_x^{\varepsilon,0}$, $N_x^{\varepsilon,1}$, $\tilde{T}_x^{\varepsilon,0}$ and $T_x^{\varepsilon,1}$ analogously. As in the proof of Proposition 1, by invariance principle and continuity of the map $(f(t))_{t \geqslant 0} \mapsto (\inf_{s \leqslant t} f(s))_{t \geqslant 0}$ under the uniform metric, we have that

$$\begin{aligned} &((\tilde{T}_x^{\varepsilon,0})_{x \geqslant -x_0}, (T_x^{\varepsilon,1})_{x \geqslant -x_1}, (\tilde{N}_x^{\varepsilon,0})_{x \geqslant -x_0}, (N_x^{\varepsilon,1})_{x \geqslant -x_1}) \\ &\qquad \text{converges in distribution, as } \varepsilon \to 0 \text{ to} \\ &((\tilde{\mathscr{T}}_x^0)_{x \geqslant -x_0}, (\mathscr{T}_x^1)_{x \geqslant -x_1}, (\tilde{\mathscr{N}}_x^0)_{x \geqslant -x_0}, (\mathscr{N}_x^1)_{x \geqslant -x_1}) \end{aligned} \tag{6}$$

uniformly over compacts. By the Skorohod representation theorem we can (and will) assume that the convergence above holds almost surely.

We still have to show that $N^{\varepsilon,0}$ converges to $\mathscr{N}^0$ defined now. Let

$$\tau := \inf\{x \geqslant -x_0 : \tilde{\mathscr{N}}_x^0 = \mathscr{N}_x^1\} \text{ and } \mathscr{N}_x^0 := \begin{cases} \tilde{\mathscr{N}}_x^0, & x \in [-x_0, \tau), \\ \mathscr{N}_x^1, & x \geqslant \tau. \end{cases}$$

To prove $N^{\varepsilon,1} \to \mathscr{N}^1$ we consider the coalescing time of discrete processes. Writing

$$\tau^\varepsilon := \inf\{x \geqslant -x_0 : \tilde{N}_x^{\varepsilon,0} = N_x^{\varepsilon,1}\},$$

we have

$$N^{\varepsilon,0} = \begin{cases} \tilde{N}_x^{\varepsilon,0}, & x \in [-x_0, \tau), \\ N_x^{\varepsilon,1}, & x \geqslant \tau^\varepsilon, \end{cases}$$

so to conclude the proof it suffices to show that $\tau^\varepsilon \to \tau$ a.s.

Since the first time that two paths meet is a lower semi-continuous function of the paths, we have $\liminf_{\varepsilon \to 0} \tau^\varepsilon \geqslant \tau$ a.s. It remains to show that

$$\limsup_{\varepsilon \to 0} \tau^\varepsilon \leqslant \tau \quad \text{a.s.} \tag{7}$$

Since $N_{\lfloor \varepsilon^{-2} x_1 \rfloor}(x) - \tilde{N}_{\lfloor \varepsilon^{-2} x_0 \rfloor}(x) \geqslant 0$ at $x = -x_0$ and this difference only jumps by 0 or ± 1, it suffices to show the following claim:

$$\text{Given any } \delta > 0, \text{ a.s., for all } \varepsilon \text{ small enough, there is } x \leqslant \tau + \delta \text{ with } N_x^{\varepsilon,1} \leqslant \tilde{N}_x^{\varepsilon,0}. \tag{8}$$

Since τ is a stopping time for the pair $(\mathscr{T}^1, \mathscr{T}^0)$, by the strong Markov property the processes $(\Delta_x)_{x \geqslant 0} := (\mathscr{T}^1_{\tau+x} - \mathscr{T}^1_\tau)_{x \geqslant 0}$ and $(\tilde{\Delta}_x)_{x \geqslant 0} := (\tilde{\mathscr{T}}^0_{\tau+x} - \mathscr{T}^1_{\tau+x})_{x \geqslant 0}$ are distributed as Brownian motions started at value 0. Hence, a.s. there is a point $z^* \in [0, \delta]$ such that $\tilde{\Delta}_{z^*} > 0$. Moreover, since $\mathrm{d}\mathscr{N} \geqslant \mathrm{d}\mathscr{T}$ due to reflection, we either have $\mathscr{N}^1_{\tau+z^*} = \mathscr{N}^1_\tau + \Delta_{z^*}$ or $\mathscr{N}^1_{\tau+z^*} > \mathscr{N}^1_\tau + \Delta_{z^*}$. We will distinguish between those two cases. In the first case

$$\mathscr{N}^1_{\tau+z^*} = \mathscr{N}^1_\tau + \Delta_{z^*} = \tilde{\mathscr{N}}^0_\tau + \Delta_{z^*} < \tilde{\mathscr{N}}^0_\tau + \Delta_{z^*} + \tilde{\Delta}_{z^*} \leqslant \tilde{\mathscr{N}}^0_{\tau+z^*}. \tag{9}$$

In particular, we have obtained the strict inequality

$$\mathscr{N}^1_{\tau+z^*} < \tilde{\mathscr{N}}^0_{\tau+z^*}.$$

By this inequality and (6), it follows that, for n large enough,

$$N^{\varepsilon,1}_{\tau+z^*} < \tilde{N}^{\varepsilon,0}_{\tau+z^*}.$$

In this first case, (8) follows directly from this inequality.

The case

$$\mathscr{N}^1_{\tau+z^*} > \mathscr{N}^1_\tau + \Delta_{z^*}$$

is subtler. We work it by observing that the above inequality is equivalent to

$$\mathscr{T}^1_{\tau+z^*} - \min_{x \in [-x_1, \tau+z^*]} \mathscr{T}^1_x > \mathscr{T}^1_\tau - \min_{x \in [-x_1, \tau]} \mathscr{T}^1_x + \mathscr{T}^1_{\tau+z^*} - \mathscr{T}^1_\tau,$$

which, in turn, is equivalent to

$$\min_{x \in [-x_1, \tau+z^*]} \mathscr{T}^1_x < \min_{x \in [-x_1, \tau^*]} \mathscr{T}^1_x.$$

From (6), this implies that, for ε small enough,

$$\min_{x \in [-x_1, \tau+z^*]} T^{\varepsilon,1}_x < \min_{x \in [-x_1, \tau]} T^{\varepsilon,1}_x,$$

in which case there is $x^* \in [\tau, \tau + z^*]$ such that

$$T^{\varepsilon,1}_{x^*} = \min_{s \in [-x_1, x^*]} T^{\varepsilon,1}_s,$$

whence

$$N_{x^*}^{\varepsilon,1} = 0 \leqslant \tilde{N}_{x^*}^{\varepsilon,0}.$$

Hence (8) holds, and this completes the proof of Theorem 3.

6 Convergence in the J_1 Metric

In this section we prove the following.

Theorem 4 *The family of processes* $(\varepsilon C_{\lfloor \varepsilon^{-2} x \rfloor})_{x \geqslant 0}$ *indexed by* $\varepsilon \in (0, 1]$ *is tight in the Skorohod* J_1 *metric.*

Since the processes $(\varepsilon\, C_{\lfloor \varepsilon^{-2} x \rfloor})_{x \geqslant 0}$ are non-decreasing for each ε, by Theorem 3 they converge in the M_1-metric to a process $(\mathscr{C}_x)_{x \geqslant 0}$. Moreover, in order to get tightness in the J_1 metric, it suffices (see [6, Theorem 13.2]) to show that, for each $K < \infty$ and $\gamma > 0$ fixed,

$$\lim_{R \to \infty} \limsup_{\varepsilon \to 0} \mathbb{P}\left[\sup_{x \in [0,K]} |C_x^{\varepsilon}| \geqslant R \right] = 0 \tag{10}$$

and

$$\lim_{\delta \to 0} \limsup_{\varepsilon \to 0} \mathbb{P}\left[\sup |C_{x_1}^{\varepsilon} - C_{x_0}^{\varepsilon}| \wedge |C_{x_1}^{\varepsilon} - C_{x_2}^{\varepsilon}| \geqslant \gamma \right] = 0. \tag{11}$$

where the last supremum is taken over triples $x_0 \leqslant x_1 \leqslant x_2$ which satisfy $x_2 \leqslant K$ and $x_2 - x_0 \leqslant \delta$. Note that (10) follows directly from Theorem 3. We only have to establish (11). The event whose probability we want to control is that of having two macroscopic jumps in a short interval.

For this proof we will use a graphical construction for the flow process, which we now proceed to describe. First, consider the *black path* $B : (-\infty, 0] \to \mathbb{Z}$, given by $B(x) = \sum_{y=x}^{0} \eta(y)$, where we recall that $\eta(y)$ is the initial number of particles at y. As in Fig. 5, for every lattice point $(x, y) \in (-\infty, 0] \times \mathbb{N}$ above the graph of B, place a red arrow which either points to $(x+1, y-1)$ or to $(x+1, y)$, independently with probability $\frac{\lambda}{1+\lambda} = \zeta$ and $\frac{1}{1+\lambda} = 1 - \zeta$, respectively. Let $x \in (-\infty, 0]$. We denote by $R_x : [x, 0] \to \mathbb{Z}$ the path which starts at $(x, B(x))$, follows the red arrows and is reflected on the black path B. The resulting path will be referred as the *red path*. In Fig. 5, the red path is depicted in bold red. Note that, by associating the red arrows which point down with sleeping instructions and the horizontal red arrows with no-sleep instructions, we see that $R_x(0)$ is distributed as $N_{-x}(0) = C_x$. Moreover, one can use the same black path and red arrows to get any joint distribution $(C_{x_0}, \ldots, C_{x_k})$, $-x_k \leqslant -x_{k-1} \leqslant \cdots \leqslant -x_0 \leqslant 0$: simply consider the collection of red paths $R_{-x_k}, \ldots, R_{-x_0}$ constructed using the same black path

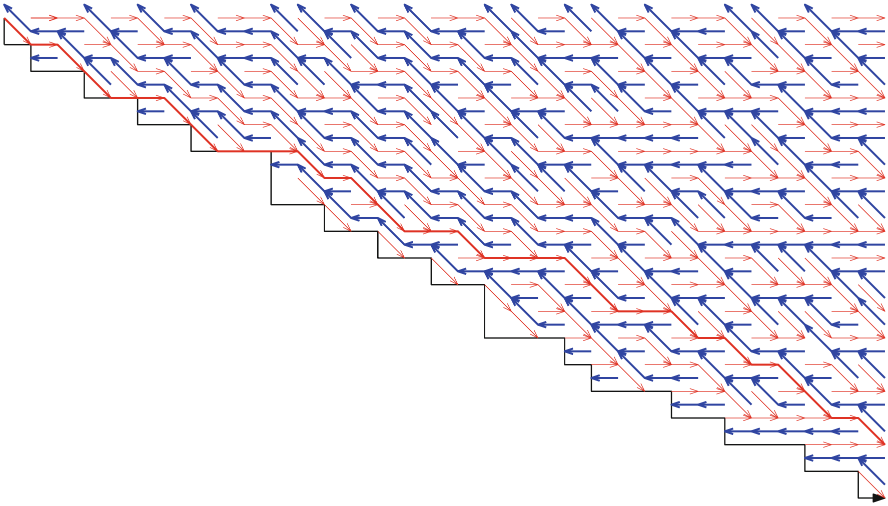


Fig. 5 Black path, red paths and dual blue paths

and the same red arrows. The random vector $(R_{-x_0}(0), \ldots, R_{-x_k}(0))$ will have the distribution of $(C_{x_0}, \ldots, C_{x_k})$. Observe that the red paths are independent until they meet, after which they coalesce.

Now, consider a collection of *blue arrows* which is dual to the red arrows. That is, for every $(x, y) \in \mathbb{Z}^2$ above the graph of the black path, there is a (backwards) blue arrow emanating from $(x, y - \frac{1}{2})$. That blue arrow points to $(x - 1, y + \frac{1}{2})$ if the red arrow staring at $(x - 1, y)$ points to $(x, y - 1)$. Otherwise, the blue arrows points to $(x - 1, y - \frac{1}{2})$ (See Fig. 5). By following the blue arrows we can construct a collection of *blue paths* which are dual to the red ones. The blue paths should be read from right to left. The blue paths jump at each time with probability ζ. Also, the blue paths are killed when they meet the black path. For any $y \in \mathbb{Z}$, let A_y be the blue path starting at $(0, y + \frac{1}{2})$. Observe that, by construction, the blue paths cannot intersect the red paths.

In the following reasoning, we will consider jointly two different blue paths A_{y_1}, A_{y_2}. For convenience, we will consider a modified version of the processes $\tilde{A}_{y_1}, \tilde{A}_{y_2}$ such that they evolve independently after collision (do not coalesce). Also, they don't get killed at intersecting the black path, they evolve independently. We do this because now $\tilde{A}_{y_1}, \tilde{A}_{y_2}$ and B are three independent random walks and we can apply Donsker's invariance principle.

We consider the (modified) blue path $\tilde{A}_y$ as a stochastic process $(\tilde{A}_y(x))_{x \leqslant 0}$, where, for any $x \leqslant 0$, $\tilde{A}_y(x)$ is the position at x of the (modified) blue path started at $(0, y + \frac{1}{2})$.

Let

$$(\tilde{A}^{\varepsilon}_{y_1}(x))_{x\leqslant 0} := (\varepsilon \tilde{A}_{\lfloor \varepsilon^{-1} y_1 \rfloor}(\lfloor \varepsilon^{-2} x \rfloor) - \zeta x)_{x\leqslant 0}$$

$$(\tilde{A}^{\varepsilon}_{y_2}(x))_{x\leqslant 0} := (\varepsilon \tilde{A}_{\lfloor \varepsilon^{-1} y_2 \rfloor}(\lfloor \varepsilon^{-2} x \rfloor) - \zeta x)_{x\leqslant 0}$$

$$(B^{\varepsilon}(x))_{x\leqslant 0} := (\varepsilon B(\lfloor \varepsilon^{-2} x \rfloor) - \zeta x)_{x\leqslant 0}.$$

By Donsker's invariance principle, we get that

$$((\tilde{A}^{\varepsilon}_{y_1}(x))_{x\leqslant 0}, (\tilde{A}^{\varepsilon}_{y_2}(x))_{x\leqslant 0}, (\tilde{B}^{\varepsilon}(x))_{x\leqslant 0}) \to ((\tilde{\mathscr{A}}_{y_1}(x))_{x\leqslant 0}, (\tilde{\mathscr{A}}_{y_2}(x))_{x\leqslant 0}, (\mathscr{B}(x))_{x\leqslant 0}), \tag{12}$$

where $\tilde{\mathscr{A}}_{y_1}$, $\tilde{\mathscr{A}}_{y_2}$, $\mathscr{B}$ are three (time-reversed) independent Brownian motions started at y_1, y_2 and 0 respectively. By the Skorohod Representation theorem, we can (and will) assume that the convergence holds almost surely.

It follows that, under the event in (11), both $\tilde{A}^{\varepsilon}_{y_1}$ and $\tilde{A}^{\varepsilon}_{y_2}$ (when read from 0 to $-\infty$) intersect B^{ε} for the first time in a time window of length smaller than δ (see Fig. 6). By (12), the probability of the event above converges to the probability that two independent Brownian motions $\tilde{\mathscr{A}}_{y_1}$, $\tilde{\mathscr{A}}_{y_2}$ intersect a third B.m., $\mathscr{B}$, also independent, in a time window of length smaller than δ. Using the continuity of the Brownian motion, we get that, as $\delta \to 0$, this converges to the probability that three independent Brownian motions eventually meet at the same point at the same time. The latter probability is 0 and this will finish the proof. To complete the proof, we have to explain how to choose the initial points of the blue paths y_1, y_2.

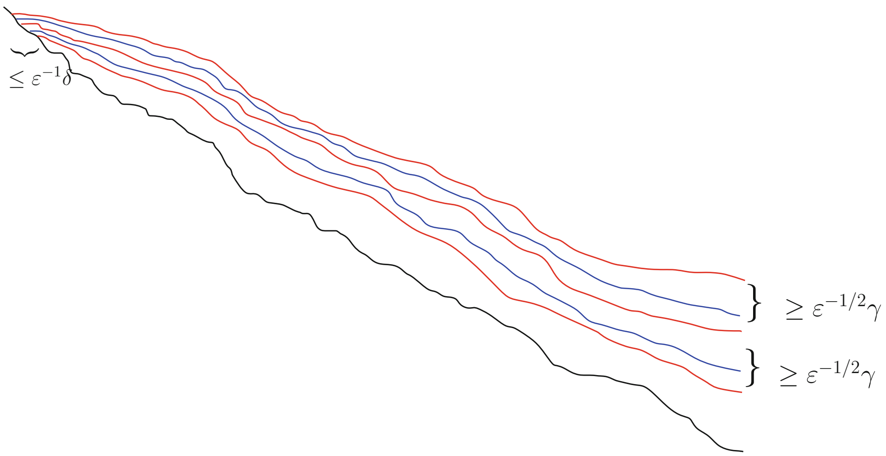


Fig. 6 Event $\mathfrak{B}(\varepsilon, \delta, \gamma, M, K)$ in terms of black, red and blue paths

We will choose M large but fixed. It suffices to deal with the case $C_K^\varepsilon \geqslant M$, since Theorem 3 readily implies that

$$\lim_{M\to\infty} \limsup_{\varepsilon\to 0} \mathbb{P}\left[C_K^\varepsilon > M\right] = 0.$$

Therefore, it is enough to show that, for every $M > 0$ and $\gamma > 0$,

$$\lim_{\delta\to 0} \limsup_{\varepsilon\to 0} \mathbb{P}\left[\sup_{\substack{x_0\leqslant x_1\leqslant x_2\in[0,K]\\ x_2-x_0\leqslant\delta}} (C_{x_1}^\varepsilon - C_{x_0}^\varepsilon) \wedge (C_{x_2}^\varepsilon - C_{x_1}^\varepsilon) \geqslant \gamma;\, C_{x_2}^\varepsilon \leqslant M\right] = 0. \tag{13}$$

Let $\mathfrak{B}(\varepsilon, \delta, \gamma, M, K)$ be the event inside the above probability (see Fig. 6). If $\mathfrak{B}(\varepsilon, \delta, \gamma, M, K)$ holds, then, there exists $x_0^* < x_1^* < x_2^*$ with $x_2^* - x_0^* \leqslant \delta$ such that both $C_{x_1^*}^\varepsilon - C_{x_0^*}^\varepsilon$ and $C_{x_2^*}^\varepsilon - C_{x_1^*}^\varepsilon$ are greater than γ. Moreover, $C_{x_2^*}^\varepsilon \leqslant M$.

Now, we partition $[0, M]$ into intervals of size $\frac{\gamma}{2}$. We consider the intervals $I_i = [\frac{\gamma i}{2}, \frac{\gamma(i+1)}{2}]$, $i = 0, \dots, \lceil\frac{2M}{\gamma}\rceil$. By the discussion on the paragraph above, there must be indices $i_1 < i_2 \in \{0, \dots, \lceil\frac{2M}{\gamma}\rceil\}$ such that, I_{i_1} is contained in $[C_{x_0^*}^\varepsilon, C_{x_1^*}^\varepsilon]$ and I_{i_2} is contained in $[C_{x_1^*}^\varepsilon, C_{x_2^*}^\varepsilon]$. Let $\mathfrak{B}_{i_1,i_2}^{\varepsilon,\delta}$ be the event described just above.

Then,

$$\begin{aligned}\mathbb{P}[\mathfrak{B}(\varepsilon, \delta, \gamma, M, K)] &\leqslant \sum_{i_1<i_2\in\{0,\dots,\lceil\frac{2M}{\gamma}\rceil\}} \mathbb{P}\left[\mathfrak{B}_{i_1,i_2}^{\varepsilon,\delta}\right] \\ &\leqslant \left(\left\lceil\frac{2M}{\gamma}\right\rceil + 1\right)^2 \sup_{i_1<i_2\in\left\{0,\dots,\left\lceil\frac{2M}{\gamma}\right\rceil\right\}} \mathbb{P}\left[\mathfrak{B}_{i_1,i_2}^{\varepsilon,\delta}\right].\end{aligned} \tag{14}$$

We now bound the probability of $\mathfrak{B}_{i_1,i_2}^{\varepsilon,\delta}$ for $i_1 < i_2$. Since blue paths do not intersect red paths under $\mathfrak{B}_{i,j}^{\varepsilon,\delta}$, any blue path started between $C_{x_0}^\varepsilon$ and $C_{x_1}^\varepsilon$ will remain between $C_{x_0}^\varepsilon$ and $C_{x_1}^\varepsilon$ and therefore will intersect B^ε at some point $z_1^* \in [-x_1, -x_0]$. Hence, since I_{i_1} is contained in $[C_{x_0}^\varepsilon, C_{x_1}^\varepsilon]$, we have that $\tilde{A}_{\frac{\gamma}{2}i_1}^\varepsilon$ intersects B^ε in $[-x_1, -x_0]$. Analogously, $\tilde{A}_{\frac{\gamma}{2}i_2}^\varepsilon$ intersects B^ε in $[-x_2, -x_1]$. Hence, since $x_2 - x_1 \leqslant \delta$, both $\tilde{A}_{\frac{\gamma}{2}i_1}^\varepsilon$ and $\tilde{A}_{\frac{\gamma}{2}i_2}^\varepsilon$ intersect B^ε in a time interval of size δ.

That is, let $\tau_i^\varepsilon := \inf\{t \geqslant 0 : \tilde{A}_{\frac{\gamma}{2}i}^\varepsilon(t) = B^\varepsilon(t)\}$. Then

$$\mathfrak{B}_{i_1,i_2}^{\varepsilon,\delta} \subseteq \{|\tau_{i_1}^\varepsilon - \tau_{i_2}^\varepsilon| \leqslant \delta\}.$$

Therefore, by (12),

$$\limsup_{\varepsilon\to 0} \mathbb{P}\left[\mathfrak{B}^{\varepsilon,\delta}_{i_1,i_2}\right] \leqslant \mathbb{P}\left[|\tau_{i_1} - \tau_{i_2}| \leqslant \delta\right]$$

where $\tau_i := \inf_{t\geqslant 0}\{\mathscr{A}_{\frac{\gamma}{2}i}(t) = \mathscr{B}(t)\}$.

Hence,

$$\lim_{\delta\to 0}\limsup_{\varepsilon\to 0} \mathbb{P}[\mathfrak{B}^{\varepsilon,\delta}_{i_1,i_2}] \leqslant \lim_{\delta\to 0} \mathbb{P}[|\tau_{i_1} - \tau_{i_2}| \leqslant \delta] = \mathbb{P}[\tau_{i_1} = \tau_{i_2}].$$

The proof is finished by noticing that $\tau_{i_1} = \tau_{i_2}$ implies that the three-dimensional Brownian motion $(\mathscr{A}_{\frac{\gamma}{2}i_1}, \mathscr{A}_{\frac{\gamma}{2}i_2}, \mathscr{B})$ intersects the line $\{(x, y, z) \in \mathbb{R}^3 : x = y = z\}$, and that event has zero probability.

7 The Scaling Limit Is a Pure-Jump Process

In this section we prove Theorem 2. By Theorem 1, the non-decreasing càdlag process $\left(\mathscr{C}^{\rho}_x\right)_{x\geqslant 0}$ is well-defined. Following the description of the previous section, it can be constructed directly from a black Brownian motion $\mathscr{B} = (\mathscr{B}_x)_{x\leqslant 0}$ started at $\mathscr{B}_0 = 0$ and blue coalescing Brownian motions $(\mathscr{A}_y(x))_{x\leqslant 0}$ starting from $\mathscr{A}_y(0) = y$ and independent of $\mathscr{B}$, having diffusion coefficients 1 and ρ respectively.

The construction is as follows. First observe that the red paths can be recovered from the black and blue paths, so they are not needed in the construction. Moreover, blue paths are independent of each other until they coalesce, and independent of the black path until they are killed by it. Furthermore, since blue paths coalesce, killing them upon meeting the black path is irrelevant and we can disconsider it. In the scaling limit, this collection of blue paths converges to the a family of paths [2, 10] consisting of independent coalescing paths $(\mathscr{A}_y(x))_{x\leqslant 0}$ indexed by $y > 0$, each one started from $\mathscr{A}_y(0) = y$.

This family satisfies the following. Let $T_y = \inf\{x \geqslant 0 : \mathscr{A}_y(-x) = \mathscr{B}_{-x}\}$. Then a.s. $0 < T_y < \infty$ for every $y > 0$, $y \mapsto T_y$, is non-decreasing, the set of values $\{T_y : y \in [\delta, K]\}$ is finite for every $0 < \delta < K < \infty$, $\lim_{y\to 0^+} T_y = 0$, and $\lim_{y\to\infty} T_y = \infty$.

To be self-contained, let us justify the statements in the previous paragraph more carefully. Write $\mathbb{Q}^*_+ = \{y_n\}_{n\in\mathbb{N}}$. Take $\mathscr{A}_{y_1} = (\mathscr{A}_{y_1}(x))_{x\leqslant 0}$ starting from $\mathscr{A}_{y_1}(0) = y_1$. For each n, take $\mathscr{A}_{y_n} = (\mathscr{A}_{y_n}(x))_{x\leqslant 0}$ starting from $\mathscr{A}_{y_n}(0) = y_n$, independent of $\mathscr{A}_{y_1}, \dots, \mathscr{A}_{y_{n-1}}$ until the first point (that is, highest x) where it meets one of them, and equal to that one after such time (that is, for lower values of x). Now for $y > 0$ rational, let $T_y = \inf\{x \geqslant 0 : \mathscr{A}_y(-x) = \mathscr{B}_{-x}\}$. Then a.s. $0 < T_y < \infty$ for every y. By coalescence, $\mathscr{A}_y \leqslant \mathscr{A}_{y'}$ for $y < y'$, hence $y \mapsto T_y$, is non-decreasing. Furthermore, using Borel-Cantelli one can show that $\lim_{y\to 0^+} T_y = 0$

and $\lim_{y\to\infty} T_y = \infty$. Finally, by well-known properties of coalescing Brownian motions [2, 10], for each $0 < a < b < \infty$ and $\varepsilon > 0$, the set $\{\mathscr{A}_y(-\varepsilon) : a < y < b\}$ is a.s. finite. As a consequence of the two last properties, the set $\{T_y : y \in [\delta, K]\}$ is finite for every $0 < \delta < K < \infty$.

To conclude, following the description of the previous section, we can define

$$\mathscr{C}_x^\rho := \inf\left\{y > 0 : T_y > x\right\}, \quad x \geqslant 0.$$

By the remarks of the previous paragraph, a.s. for every $0 < a < b < \infty$ the process $\left(\mathscr{C}_x^\rho\right)_{x\in[a,b]}$ takes only finitely many values, and therefore it is a pure-jump process.

Acknowledgments We thank V. Sidoravicius and J. Yu for helpful discussions. Manuel Cabezas was supported by Iniciativa Científica Milenio NC120062. This project was supported by MathAmSud grant RSPSM.

References

1. Amir, G., Gurel-Gurevich, O.: On fixation of activated random walks. Electron. Commun. Probab. **15**, 119–123 (2010). http://dx.doi.org/10.1214/ECP.v15-1536
2. Arratia, R.: Coalescing Brownian Motions on the Line. Ph.D. Thesis, University of Wisconsin, Madison (1979)
3. Asselah, A., Schapira, B., Rolla, L.T.: Diffusive bounds for the critical density of activated random walks (2019). Preprint. arXiv:1907.12694
4. Basu, R., Ganguly, S., Hoffman, C.: Non-fixation for conservative stochastic dynamics on the line. Comm. Math. Phys. **358**, 1151–1185 (2018). http://dx.doi.org/10.1007/s00220-017-3059-7
5. Basu, R., Ganguly, S., Hoffman, C., Richey, J.: Activated random walk on a cycle. Ann. Inst. Henri. Poincaré Probab. Stat. **55**, 1258–1277 (2019). http://dx.doi.org/10.1214/18-aihp918
6. Billingsley, P.: Convergence of Probability Measures. Wiley Series in Probability and Statistics: Probability and Statistics, 2 edn. Wiley, New York (1999). http://dx.doi.org/10.1002/9780470316962
7. Cabezas, M., Rolla, L.T., Sidoravicius, V.: Non-equilibrium phase transitions: activated random walks at criticality. J. Stat. Phys. **155**, 1112–1125 (2014). http://dx.doi.org/10.1007/s10955-013-0909-3
8. Cabezas, M., Rolla, L.T., Sidoravicius, V.: Recurrence and density decay for diffusion-limited annihilating systems. Probab. Theory Relat. Fields **170**, 587–615 (2018). http://dx.doi.org/10.1007/s00440-017-0763-3
9. Dickman, R., Rolla, L.T., Sidoravicius, V.: Activated random walkers: Facts, conjectures and challenges. J. Stat. Phys. **138**, 126–142 (2010). http://dx.doi.org/10.1007/s10955-009-9918-7
10. Fontes, L.R.G., Isopi, M., Newman, C.M., Ravishankar, K.: The Brownian web: characterization and convergence. Ann. Probab. **32**, 2857–2883 (2004). http://dx.doi.org/10.1214/009117904000000568
11. Rolla, L.T.: Activated random walks on Z^d. Probab. Surveys. **17**, 478–544 (2020). http://dx.doi:10.1214/19-PS339
12. Rolla, L.T., Sidoravicius, V.: Absorbing-state phase transition for driven-dissipative stochastic dynamics on Z. Invent. Math. **188**, 127–150 (2012). http://dx.doi.org/10.1007/s00222-011-0344-5

13. Rolla, L.T., Tournier, L.: Non-fixation for biased activated random walks. Ann. Inst. H. Poincaré Probab. Statist. **54**, 938–951 (2018). http://dx.doi.org/10.1214/17-AIHP827
14. Shellef, E.: Nonfixation for activated random walks. ALEA Lat. Am. J. Probab. Math. Stat. **7**, 137–149 (2010). http://alea.impa.br/articles/v7/07-07.pdf
15. Sidoravicius, V., Teixeira, A.: Absorbing-state transition for stochastic sandpiles and activated random walks. Electron. J. Probab. **22**, 33 (2017). http://dx.doi.org/10.1214/17-EJP50
16. Stauffer, A., Taggi, L.: Critical density of activated random walks on transitive graphs. Ann. Probab. **46**, 2190–2220 (2018). http://dx.doi.org/10.1214/17-AOP1224
17. Taggi, L.: Absorbing-state phase transition in biased activated random walk. Electron. J. Probab. **21**, 13 (2016). http://dx.doi.org/10.1214/16-EJP4275
18. Taggi, L.: Active phase for activated random walks on $\mathbb{Z}^d$, $d \geqslant 3$, with density less than one and arbitrary sleeping rate. Ann. Inst. Henri. Poincaré Probab. Stat. **55**, 1751–1764 (2019). http://dx.doi.org/10.1214/18-aihp933

An Overview of the Balanced Excited Random Walk

Daniel Camarena, Gonzalo Panizo, and Alejandro F. Ramírez

Abstract The balanced excited random walk, introduced by Benjamini, Kozma and Schapira in 2011, is defined as a discrete time stochastic process in $\mathbb{Z}^d$, depending on two integer parameters $1 \le d_1, d_2 \le d$, which whenever it is at a site $x \in \mathbb{Z}^d$ at time n, it jumps to $x \pm e_i$ with uniform probability, where $e_1, \ldots, e_d$ are the canonical vectors, for $1 \le i \le d_1$, if the site x was visited for the first time at time n, while it jumps to $x \pm e_i$ with uniform probability, for $1 + d - d_2 \le i \le d$, if the site x was already visited before time n. Here we give an overview of this model when $d_1 + d_2 = d$ and introduce and study the cases when $d_1 + d_2 > d$. In particular, we prove that for all the cases $d \ge 5$ and most cases $d = 4$, the balanced excited random walk is transient.

Keywords Excited random walk · Transience

AMS 2010 subject classification Primary 60G50, 82C41; secondary 60G42

1 Introduction

We consider an extended version of the balanced excited random walk introduced by Benjamini, Kozma and Schapira in [1]. The balanced excited random walk is defined in any dimension $d \ge 2$, and depends on two integers $d_1, d_2 \in \{1, \ldots, d\}$. For each $1 \le i \le d$, let $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ be the canonical vector whose i-th coordinate is 1, while all other coordinates are 0. We define the process $(S_n : n \ge 0)$,

D. Camarena · G. Panizo
Instituto de Matemática y Ciencias Afines, Universidad Nacional de Ingeniería, Lima, Perú
e-mail: vcamarenap@uni.pe; gpanizog@uni.edu.pe

A. F. Ramírez (✉)
Facultad de Matemáticas, Pontificia Universidad Católica de Chile, Santiago, Chile

NYU-ECNU Institute of Mathematical Sciences at NYU Shanghai, Shanghai, China
e-mail: aramirez@mat.uc.cl

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_10

called the *balanced excited random walk* on $\mathbb{Z}^d$ as a mixture of two simple random walks, with the initial condition $S_0 = 0$: if at time n, S_n visits a site for the first time, with probability $1/(2d_1)$, at time $n + 1$ it performs a simple random walk step using one of the first d_1 coordinates, so that for all $1 \leq i \leq d_1$,

$$\mathbb{P}[S_{n+1} - S_n = \pm e_i | \mathscr{F}_n, S_n \neq S_j \text{ for all } 0 \leq j < n] = \frac{1}{2d_1},$$

where $\mathscr{F}_n$ is the σ-algebra generated by $S_0, \dots, S_n$; on the other hand, if at time n, S_n visits a site it has previously visited, at time $n + 1$ it performs a simple random walk using one the last d_2 coordinates, so that for all $d - d_2 + 1 \leq i \leq d$,

$$\mathbb{P}[S_{n+1} - S_n = \pm e_i | \mathscr{F}_n, S_n = S_j \text{ for some } 0 \leq j < n] = \frac{1}{2d_2}.$$

We call this process S the $M_d(d_1, d_2)$-random walk. In [1], this random walk was considered in the case when $d_1 + d_2 = d$, which we call the *non-overlapping* case. Here we will focus on the *overlapping* case corresponding to $d_1 + d_2 > d$.

We say that the $M_d(d_1, d_2)$-random walk is transient if any site is visited only finitely many times, while we say that it is recurrent if it visits every site infinitely often. Since a random walk $M_d(d_1, d_2)$ is not Markovian, in principle could be neither transient nor recurrent.

For the non-overlapping case, in 2011 in [1] it was shown that the $M_4(2, 2)$-random walk is transient, while in 2016, Peres, Schapira and Sousi in [7], showed that the $M_3(1, 2)$-random walk is transient, but the transience of $M_3(2, 1)$-random walk is still an open question.

The main result of this article is the following theorem concerned with the overlapping case.

Theorem 1 *For every* (d, d_1, d_2) *with* $d \geq 4$, $1 \leq d_1, d_2 \leq d$, $d_1 + d_2 > d$ *and* $(d, d_1, d_2) \neq (4, 3, 2)$, *the* $M_d(d_1, d_2)$*-random walk is transient.*

Theorem 1 has a simple proof for $d \geq 7$, for all admissible values of d_1 and d_2. Let $r := d_1 + d_2 - d$. Note that if $r \geq 3$ then the walk is transient, since its restriction to the r overlapping coordinates is at least a 3-dimensional simple symmetric random walk with geometrically bounded holding times. We will argue in the next two paragraphs that the walk is also transient if $d_1 - r \geq 3$ or if $d_2 - r \geq 3$. Assuming for the moment that each of the three inequalities $r \geq 3$, $d_1 - r \geq 3$ or $d_2 - r \geq 3$ implies transience, note that if none of them holds we have that $d = d_1 + d_2 - r \leq 6$. We conclude that for $d \geq 7$ the walk is transient for all admissible values of d_1 and d_2.

Case $d_2 - r \geq 3$ We claim that with probability 1 the fraction of times when the random walk uses the last $d_2 - r$ coordinates is asymptotically bounded from below by a positive constant and therefore, the random walk is transient. To see this note that whenever the walk makes 3 consecutive steps, the probability that in at least one of these steps it visits a previously visited (old) site is bounded away from 0. Indeed, if in two consecutive steps the walk visits two previously unvisited (new) sites then with probability $1/(2d_1)$ it backtracks in the next step and, thus, visits an old site.

Case $d_1 - r \geq 3$ We claim that the number of times the random walk uses the first $d_1 - r$ coordinates goes to infinity as $n \to \infty$, which is enough to prove transience. Denote by r_n the number of points in the range of the walk at time n. We will show that $r_n \to \infty$ as $n \to \infty$ a.s. For $k \geq 1$, let $n_k = \inf\{n \geq 0 : r_n = k\}$. We will argue that if $n_k < \infty$ then with probability one $n_{k+1} < \infty$. Note that $r_{n_k} = k$, S_{n_k} is a new site, and there are $k-1$ other sites in the range. Let $n_k < \infty$ and A_1 be the event that in the next k steps the walk jumps only in positive coordinate directions. On A_1, at times $n_k+1, n_k+2, \ldots, n_k+k$ the walk visits k distinct sites of $\mathbb{Z}^d - \{S_{n_k}\}$. Among these sites there are at most $k-1$ old sites. Therefore, on the event $A_1 \cap \{n_k < \infty\}$ the walk will necessarily visit a new site and $n_{k+1} \leq n_k + k < \infty$. Note that the probability of A_1 (given $n_k < \infty$) is 2^{-k}. If A_1 does not occur, then we consider the next k steps and define A_2 to be the event that in these next k steps the walk jumps only in the positive coordinate directions, and so on. Since, conditional on $n_k < \infty$, the events $A_1, A_2, \ldots$ are independent and each has probability 2^{-k}, we conclude that $n_{k+1} < \infty$ with probability one.

Therefore, to complete the proof of Theorem 1 we have only to consider the cases $d = 4, 5, 6$. It will be shown below that the cases $d = 5, 6$ and several cases in $d = 4$, can be derived in an elementary way sometimes using the trace condition of [6]. In a less straightforward way the cases $M_4(2, 4)$ and $M_4(4, 2)$ can be treated through the methods of [1]. The case $M_4(2, 3)$ which is more involved, can be treated through a modification of methods developed by Peres, Schapira and Sousi [7] for the $M_3(1, 2)$-random walk through good controls on martingale increments by sequences of geometric i.i.d. random variables. It is not clear how the above mentioned methods could be applied to the $M_4(3, 2)$-random walk to settle down the transience-recurrence question for it, so this case remains open.

In Sect. 2 we will give a quick review of the main results that have been previously obtained for the non-overlapping case of the balanced excited random walk. In Sect. 3, we will prove Theorem 1. In Sect. 3.1, we will introduce the trace condition of [6], which will be used to prove the cases $d = 5, 6$ and several cases in dimension $d = 4$. In Sect. 3.2, we will prove the transience of the random walks $M_4(2, 4)$ and $M_4(4, 2)$. While in Sect. 3.3, we will consider the proof of the transience of the $M_4(2, 3)$-random walk.

2 Overview of the Balanced Excited Random Walk

The balanced excited random walk was introduced in its non-overlapping version by Benjamini, Kozma and Schapira in [1]. A precursor of the balanced excited random walk, is the *excited random walk*, introduced by Benjamini and Wilson in 2003 [2], which is defined in terms of a parameter $0 < p < 1$ as follows: the random walk $(X_n : n \geq 0)$ has the state space $\mathbb{Z}^d$ starting at $X_0 = 0$; whenever the random walk visits a site for the first time, it jumps with probability $(1+p)/2d$ in direction e_1, probability $(1-p)/2d$ in direction $-e_1$ and with probability $1/2d$ in the other directions; whenever the random walk visits a site which it already visited previously it jumps with uniform probability in directions $\pm e_i$, $1 \le i \le d$. Benjamini and Wilson proved in [2] that the model is transient for $d > 1$. A central limit theorem and a law of large numbers for $d > 1$ was proven in [3] and [4]. A general review of the model can be found in [5]. Often the methods used to prove transience, the law of large numbers and the central limit theorem for the excited random walk, are based on the ballisticity of the model (the fact that the velocity is non-zero), through the use of regeneration times. This means that most of these methods are not well suited to study the balanced excited random walk, which is not ballistic. For the moment, a few results have been obtained for the balanced excited random walk, where basically for each case a different technique has been developed. The first result, obtained by Benjamini, Kozma and Schapira in [1] for the $M_4(2, 2)$ case, is the following theorem.

Theorem 2 (Benjamini et al. [1]) *The* $M_4(2, 2)$*-random walk is transient.*

The proof of Theorem 2 is based on obtaining good enough estimates for the probability that a 2-dimensional random walk returns to its starting point in a time interval $[n/c(\log n)^2, cn]$, for some constant $c > 0$, and on the range of the random walk. This then allows to decouple using independence the first 2 coordinates from the last 2 ones. In this article, we will apply this method to derive the transience in the $M_4(4, 2)$ and $M_4(2, 4)$ cases of Theorem 1.

In 2016, Peres, Sousi and Schapira in [7], considered the case $M_3(1, 2)$ proving the following result.

Theorem 3 (Peres et al. [7]) *The* $M_3(1, 2)$*-random walk is transient.*

The approach developed in [7] to prove Theorem 3, starts with conditioning on all the jumps of the last two coordinates, and then looking at the first coordinate at the times when the last two move, which gives a martingale. It is then enough to obtain good estimates on the probability that this martingale is at 0 at time n. The proof of the $M_4(2, 3)$-random walk case of Theorem 1, is based on a modification of the method used to prove Theorem 3, where a key point is to obtain appropriate bounds for martingale increments (which will correspond to the first coordinate of the movement of the $M_4(2, 3)$-random walk) in terms of i.i.d. sequences of geometric random variables.

3 Proof of Theorem 1

We will divide the proof of Theorem 1 in three steps. With the exception of the cases $M_4(1, 4)$, $M_4(4, 1)$, $M_4(2, 4)$, $M_4(4, 2)$ and $M_4(2, 3)$, we will use an important result of Peres et al. [6]. The cases $M_4(1, 4)$ and $M_4(4, 1)$ will be derived as those in dimension $d \ge 7$. For the cases $M_4(2, 4)$ and $M_4(4, 2)$ we will show how the argument of [1] can be adapted. And the case $M_4(2, 3)$ is handled as in [7].

3.1 The Trace Condition

Here we will recall the so called trace condition of [6] which is a general condition under which a generalized version of the balanced random walk is transient, and see how it can be used to prove Theorem 1 for the remaining $d \geq 5$ cases and case $M_4(3,4)$, $M_4(4,3)$, $M_4(3,3)$ and $M_4(4,4)$.

Given $d \geq 1$ and $m \geq 1$, consider probability measures $\mu_1, \ldots, \mu_m$ on $\mathbb{R}^d$ and for each $1 \leq i \leq m$, let $(\xi_n^i : n \geq 1)$ be an i.i.d. sequence of random variables distributed according to μ_i. We say that a stochastic process $(\ell_k : k \geq 0)$ is an *adapted rule* with respect to a filtration $(\mathscr{F}_n : n \geq 0)$ of the process, if for each $k \geq 0$, ℓ_k is $\mathscr{F}_k$-measurable. We now define the random walk $(X_n : n \geq 0)$ generated by the probability measures $\mu_1, \ldots, \mu_m$ and the adapted rule ℓ by

$$X_{n+1} = X_n + \xi_{n+1}^{\ell_n}, \qquad \text{for} \quad n \geq 0.$$

Let μ be a measure on $\mathbb{R}^d$. μ is called of *mean* 0 if $\int x d\mu = 0$. The measure μ is said to have β *moments* if for any random variable Z distributed according to μ, $||Z||$ has moment of order β. The covariance matrix of μ, $Var(\mu)$, is defined as the covariance of Z.

Given a matrix A, we call $\lambda_{max}(A)$ its maximal eigenvalue and A^t its transpose. In [6], the following result was proven.

Theorem 4 (Peres et al. [6]) *Let $\mu_1, \ldots, \mu_m$ be measures in $\mathbb{R}^d$, $d \geq 3$, with zero mean and $2+\beta$ moments, for some $\beta > 0$. Assume that there is a matrix A such that the trace condition is satisfied:*

$$tr(A\ Var(\mu_i)\ A^t) > 2\lambda_{max}(A\ Var(\mu_i)\ A^t)$$

for all $1 \leq i \leq m$. Then any random walk X generated by these measures and any adapted rule is transient.

It follows from Theorem 4, that whenever $d_1 \geq 3$ and $d_2 \geq 3$, the trace condition is satisfied, with $A = I$, for the two corresponding covariance matrices associated to the motions in first d_1 and last d_2 dimensions, and hence the $M_d(d_1, d_2)$-random walk is transient. Hence, by the discussion right after the statement of Theorem 1 in Sect. 1, we see that the only cases which are not covered by Theorem 4, correspond to

$$d_1 - r \leq 2,\ r \leq 2 \quad \text{and} \quad d_2 - r \leq 2, \tag{1}$$

and

$$\min\{d_1, d_2\} \leq 2.$$

But (1) implies that $\max\{d_1, d_2\} \leq 2 + r$. Thus,

$$d_1 + d_2 = \max\{d_1, d_2\} + \min\{d_1, d_2\} \leq 4 + r,$$

so that $d = d_1 + d_2 - r \leq 4$. This proves the transience for all the cases when $d \geq 5$. Now note that in dimesion $d = 4$ the random walks $M_4(3, 3)$, $M_4(3, 4)$, $M_4(4, 3)$ and $M_4(4, 4)$ satisfy $d_1 \geq 3$ and $d_2 \geq 3$, so that the trace condition of [6] is satisfied.

Finally, the random walks $M_4(1, 4)$ and $M_4(4, 1)$ satisfy $d_1 - r \geq 3$ or $d_2 - r \geq 3$, so that they are also transient.

3.2 The Random Walks $M_4(2, 4)$ *and* $M_4(4, 2)$

Consider the $M_4(4, 2)$-random walk and call r_n the cardinality of its range at time n. Let us use the notation $S = (X, Y)$ for the $M_4(4, 2)$-random walk, where X are the first two components and Y the last two ones. We will also call $r_n^{(1)}$ the number of times up to time n that the random walk jumped using the X coordinates while it was at a site that it visited for the first time and $r_n^{(2)} := r_n - r_n^{(1)}$. In analogy with Lemma 1 of [1], we have the following result.

Lemma 1 *For any* $M > 0$ *and each* $i = 1, 2$*, there exists a constant* $C > 0$ *such that*

$$\mathbb{P}[n/(C \log n)^2 \leq r_n^{(i)} \leq 99n/100] = 1 - o\left(n^{-M}\right). \tag{2}$$

Proof First note that in analogy to the proof Lemma 1 of [1], we have that

$$\mathbb{P}[n/(C \log n)^2 \leq r_n \leq 99n/100] = 1 - o\left(n^{-M}\right).$$

Since each time the random walk is at a newly visited site with probability $1/2$ it jumps using the X random walk and with probability $1/2$ the Y random walk, by standard large deviation estimates, we deduce (2).

Now note that

$$\{(X_k, Y_k) : k \geq 1\} = \{(U_1(r_{k-1}^{(1)}),\ U_2(r_{k-1}^{(2)}) + V(k - r_{k-1})) : k \geq 1\}, \tag{3}$$

where U_1, U_2 and V are three independent simple random walks in $\mathbb{Z}^2$. It follows from the identity (3) and Lemma 1 used to bound the components $r_n^{(1)}$ and $r_n^{(2)}$ of the range of the walk, that

$$\begin{aligned}\mathbb{P}[0 \in \{S_n, \ldots, S_{2n}\}] &\leq \mathbb{P}[0 \in \{U(n/(C \log n)^2), \ldots, U(2n)\}] \\ &\times \mathbb{P}[0 \in \{W(n/(C \log n)^2), \ldots, W(2n)\}] + o(n^{-M}),\end{aligned} \tag{4}$$

where U and W are simple symmetric random walks on $\mathbb{Z}^2$. At this point, we recall Lemma 2 of [1].

Lemma 2 (Benjamini et al. [1]) *Let U be a simple random walk on $\mathbb{Z}^2$ and let $t \in [n/(\log n)^3, 2n]$. Then*

$$\mathbb{P}[0 \in \{U(t), \dots, U(2n)\}] = O\left(\frac{\log\log n}{\log n}\right).$$

Combining inequality (4) with Lemma 2, we conclude that there is a constant $C > 0$ such that for any $n > 1$ (see Proposition 1 of [1])

$$\mathbb{P}[0 \in \{S_n, \dots, S_{2n}\}] \le C\left(\frac{\log\log n}{\log n}\right)^2.$$

Hence,

$$\sum_{k=0}^{\infty} \mathbb{P}[0 \in \{S_{2^k}, \dots, S_{2^{k+1}}\}] < \infty,$$

and the transience of the $M_4(4, 2)$-random walk follows form Borel–Cantelli. A similar argument can be used to prove the transience of the $M_4(2, 4)$-random walk.

3.3 The $M_4(2, 3)$-Random Walk

Here we will follow the method developed by Peres, Schapira and Sousi in [7]. We first state Proposition 2.1 of [7].

Proposition 1 (Peres et al. [7]) *Let $\rho > 0$ and $C_1, C_2 > 0$. Let M be a martingale with quadratic variation V and assume that $(G_k : k \ge 0)$ is a sequence of i.i.d. geometric random variables with mean C_1 such that for all $k \ge 0$,*

$$|M_{k+1} - M_k| \le C_2 G_k. \tag{5}$$

For all $n \ge 1$ and $1 \le k \le \log_2(n)$ let $t_k := n - \frac{n}{2^k}$ and

$$A_k := \left\{ V_{t_{k+1}} - V_{t_k} \ge \rho \frac{t_{k+1} - t_k}{(\log n)^{2a}} \right\}.$$

Suppose that for some $N \ge 1$ and $1 \le k_1 < \dots < k_N < \log_2(n)/2$ one has that

$$\mathbb{P}\left(\cap_{i=1}^{N} A_{k_i}\right) = 1. \tag{6}$$

Then, there exists constant $c > 0$ and a positive integer n_0 such that for all $a \in (0, 1)$ and $n \geq n_0$ one has that

$$\mathbb{P}(M_n = 0) \leq \exp\left(-cN/(\log n)^a\right).$$

Remark 1 Proposition 1 is slightly modified with respect to Proposition 2.1 of [7] since we have allowed the mean C_1 of the geometric random variables to be arbitrary and the bound (5) to have an arbitrary constant C_2.

Let us now note that the $M_4(2, 3)$-random walk $(S_n : n \geq 0)$ can be defined as follows. Suppose $(\zeta_n : n \geq 1)$ is a sequence of i.i.d. random variables taking each of the values $(0, \pm 1, 0, 0)$, $(0, 0, \pm 1, 0)$ and $(0, 0, 0, \pm 1)$ with probability $1/6$, while $(\xi_n : n \geq 1)$ is a sequence of i.i.d. random variables (independent from the previous sequence) taking each of the values $(0, \pm 1, 0, 0)$ and $(\pm 1, 0, 0, 0)$ with probability $1/4$. Define now recursively, $S_0 = 0$, and

$$S_{n+1} = S_n + \Delta_{n+1}$$

where the step is

$$\Delta_{n+1} = \begin{cases} \xi_{r_n}, & \text{if } r_n = r_{n-1} + 1 \\ \zeta_{n+1-r_n}, & \text{if } r_n = r_{n-1} \end{cases}$$

and $r_n = \#\{S_0, \ldots, S_n\}$ as before is the cardinality of the range of the random walk at time n (note that formally $r_{-1} = 0$).

Let us now write the position at time n of the $M_4(2, 3)$ random walk as

$$S_n = (X_n, Y_n, Z_n, W_n).$$

Define recursively the sequence of stopping times $(\tau_k : k \geq 0)$ by $\tau_0 = 0$ and for $k \geq 1$,

$$\tau_k := \inf\{n > \tau_{k-1} : (Z_n, W_n) \neq (Z_{n-1}, W_{n-1})\}.$$

Note $\tau_k < \infty$ a.s. for all $k \geq 0$. Furthermore, the process $(U_k : k \geq 0)$ defined by

$$U_k = (Z_{\tau_k}, W_{\tau_k}),$$

is a simple random walk in dimension $d = 2$, and is equal to the simple random walk with steps defined by the last two coordinates of ζ. Let us now call P_U the law of S conditionally on the whole U process. Note that the first coordinate $\{X_n : n \geq 0\}$ is an $\mathscr{F}_n := \sigma\{\Delta_k : k \leq n\}$-martingale with respect to P_U, since

$$E_U(X_{n+1} - X_n \mid \mathscr{F}_n) = 1_{\{r_n = r_{n-1}+1\}} E(\xi_{r_n} \cdot e_1 \mid \mathscr{F}_n, U),$$

U is $\sigma(\zeta_k : k \geq 1)$-measurable as it is defined only in terms of the sequence $\left(\zeta_k 1_{\{\pi_{34}(\zeta_k)\neq 0\}}\right)_{k\geq 1}$, ($\pi_{34}$ being the projection in the 3rd and 4th coordinates), and

$$E[\,\xi_{r_n} \cdot e_1 \,|\, \mathscr{F}_n, (\zeta_k : k \geq 1)] = 0,$$

by independence. Hence, $\{M_m : m \geq 0\}$ with $M_m := X_{\tau_m}$, is a $\mathscr{G}_m$-martingale with respect to P_U, where $\mathscr{G}_m := \mathscr{F}_{\tau_m}$. To prove the theorem, it is enough to show that $\{(M_n, U_n) : n \geq 0\}$ is transient (under P). Let us call $r_U(n)$ the cardinality of the range of the random walk U at time n. For each $n \geq 0$ and $k \geq 0$, let

$$t_k := n - n/2^k \tag{7}$$

and

$$\mathscr{K} := \left\{k \in \{1, \ldots, (\log n)^{3/4}\} : r_U(t_{k+1}) - r_U(t_k) \geq \rho(t_{k+1} - t_k)/\log n\right\}. \tag{8}$$

We will show that

$$\begin{aligned} P(M_n = U_n = 0) = E[P_U(M_n = 0)1\{|\mathscr{K}| \geq \rho(\log n)^{3/4}, U_n = 0\}] \\ +E[P_U(M_n = 0)1\{|\mathscr{K}| < \rho(\log n)^{3/4}, U_n = 0\}], \end{aligned} \tag{9}$$

is summable in n, for $\rho = \rho_0$ chosen appropriately. At this point, let us recall Proposition 3.4 of [7], which is a statement about simple symmetric random walks.

Proposition 2 (Peres et al. [7]) *For $k \geq 1$, consider t_k as defined in (7). Then, for $\mathscr{K}$ as defined in (8), we have that there exist positive constants α, C_3, C_4 and ρ_*, such that for all $\rho < \rho_*$ and all $n \geq 1$*

$$P(|\mathscr{K}| \leq \rho(\log n)^{3/4} | U_n = 0) \leq C_3 e^{-C_4(\log n)^{\alpha}}.$$

Choosing $\rho = \rho_0 \leq 1$ small enough, by Proposition 2, we have the following bound for the second term on the right-hand side of (9),

$$E[P_U(M_n = 0)1\{|\mathscr{K}| < \rho_0(\log n)^{3/4}, U_n = 0\}] \leq C_3 C_5 \frac{1}{n} \exp\left(-C_4(\log n)^{\alpha}\right), \tag{10}$$

where we have used the fact that $P(U_n = 0) \leq \frac{C_5}{n}$ for some constant $C_5 > 0$.

To bound the first term on the right-hand side of (9), we will use Proposition 1 with $a = 1/2$ and $\rho = \rho_0/4$. Let us first show that (6) is satisfied. Indeed, note that for each $n \geq 0$ when U_n is at a new site, $E_U[(M_{n+1} - M_n)^2 | \mathscr{G}_n] \geq 1/2$. Therefore, for all $k \in \mathscr{K}$, with $\rho = \rho_0$, one has that for n large enough

$$V_{t_{k+1}} - V_{t_k} = \sum_{n=t_k+1}^{t_{k+1}} E_U[(M_n - M_{n-1})^2 | \mathscr{G}_{n-1}]$$
$$\geq \big(r_U(t_{k+1}-1) - r_U(t_k - 1)\big)/2 \geq \big(r_U(t_{k+1}) - r_U(t_k) - 1\big)/2$$
$$\geq (\rho_0/4)(t_{k+1} - t_k)/(\log n)^{2a}.$$

Hence, on the event $|\mathscr{K}| \geq \rho_0(\log n)^{3/4}$, we have that there exist $k_1, \ldots, k_N \in \mathscr{K}$ with $N = [\rho_0(\log n)^{3/4}]$ such that

$$P_U\left(\cap_{i=1}^N A_{k_i}\right) = 1.$$

Let us now show that there is a sequence of i.i.d. random variables $(G_k : k \geq 0)$ such that (5) is satisfied with $C_1 = 24$ and $C_2 = 3$. Indeed, note that

$$|M_{n+1} - M_n| = |X_{\tau_{n+1}} - X_{\tau_n}| \leq \sum_{k=\tau_n}^{\infty} |X_{(k+1)\wedge\tau_{n+1}} - X_{k\wedge\tau_{n+1}}|. \tag{11}$$

Note that the right-hand side of (11) is the number of steps of X between times τ_n and τ_{n+1}. Now, at each time k (with k starting at τ_n) that a step in X is made there is a probability of at least $\frac{1}{4^2} \times \frac{2}{3} = \frac{1}{24}$ that the random walk S makes three succesive steps at times $k+1$, $k+2$ and $k+3$, in such a way that in one of them a step in U is made and at most two of these steps are of the X random walk: if the random walk is at a site previously visited at time k, with probability $2/3$ at time $k+1$ the U random walk will move; if the random walk is at a site which it had never visited before at time k, with probability $\frac{1}{4^2} \times \frac{2}{3} = \frac{1}{24}$, there will be 3 succesive steps of S at times $k+1$, $k+2$ and $k+3$, with the first 2 steps being of the X random walk and the third step of U (we just need to move in the e_1 direction using X at time $k+1$, immediately follow it at time $k+2$ by a reverse step in the $-e_1$ direction using X again, and then immediately at time $k+3$ do a step in U). Since this happens independently each 3 steps in the time scale of X (time increases by one unit whenever X moves), we see that we can bound the martingale increments choosing i.i.d. geometric random variables $(G_k : k \geq 0)$ of parameter $1/24$ in (5) multiplied by 3.

Remark 2 The sequence of i.i.d. geometric random variables constructed above is not the optimal one, in the sense that it is possible to construct other sequences of i.i.d. geometric random variables of parameter larger than $1/24$.

Since now we know that (6) and (5) are satisfied, by Proposition 1, there exist $n_0 \geq 1$ and $C_7 > 0$ such that on the event $|\mathscr{K}| \geq \rho_0(\log n)^{3/4}$ we have that for $n \geq n_0$,

$$P_U(M_n = 0) \leq e^{-C_7\rho_0 \frac{(\log n)^{3/4}}{(\log n)^{1/2}}}.$$

Hence, for $n \geq n_0$ we have

$$E[P_U(M_n = 0)1\{|\mathscr{K}| \geq \rho_0(\log n)^{3/4}, U_n = 0\}] \leq C_5 \frac{1}{n} e^{-C_7 \rho_0 \frac{(\log n)^{3/4}}{(\log n)^{1/2}}}. \tag{12}$$

Using the bounds (10) and (12) back in (9) gives us that there exist constants $C_8 > 0$, $C_9 > 0$ and some $\beta > 0$, such that

$$P(M_n = U_n = 0) \leq \frac{1}{n} C_8 e^{-C_9 (\log n)^\beta}$$

By the Borel–Cantelli lemma, we conclude that the process (M, U) is transient, which gives the transience of S.

Acknowledgments Daniel Camarena and Gonzalo Panizo thank the support of Fondo Nacional de Desarrollo Científico, Tecnológico y de Innovación Tecnológica CG-176-2015. Alejandro Ramírez thanks the support of Iniciativa Científica Milenio and of Fondo Nacional de Desarrollo Científico y Tecnológico grant 1180259

References

1. Benjamini, I., Kozma, G., Schapira, B.: A balanced excited random walk. C. R. Acad. Sci. Paris Ser. I **349**, 459–462 (2011)
2. Benjamini, I., Wilson, D.: Excited random walk. Electron. Commun. Probab. **9**, 86–92 (2003)
3. Bérard, J., Ramírez, A.F.: Central limit theorem for the excited random walk in dimension $d \geq 2$. Electron. Commun. Probab. **12**, 303–314 (2007)
4. Kozma, G.: Non-classical interacting random walks. Problem session, Oberwolfach Rep., 4(2007), No. 2, 1552. Abstracts from the workshop held May 20–26, 2007, organized by F. Comets and M. Zerner
5. Kozygina, E., Zerner, M.: Excited random walks: results, methods, open problems. Bull. Inst. Math. Acad. Sin. **1**, 105–157 (2013)
6. Peres, Y., Popov, S., Sousi, P.: On recurrence and transience of self-interacting random walks. Bull. Braz. Math. Soc. **44**, 841–867 (2013)
7. Peres, Y., Schapira, B., Sousi, P.: Martingale defocusing and transience of a self-interacting random walk. Ann. Inst. Henri Poincaré Probab. Stat. **52**, 1009–1022 (2016)

Limit Theorems for Loop Soup Random Variables

Federico Camia, Yves Le Jan, and Tulasi Ram Reddy

This article is dedicated to the memory of Vladas Sidoravicius, colleague and friend.

Abstract This article deals with limit theorems for certain loop variables for loop soups whose intensity approaches infinity. We first consider random walk loop soups on finite graphs and obtain a central limit theorem when the loop variable is the sum over all loops of the integral of each loop against a given one-form on the graph. An extension of this result to the noncommutative case of loop holonomies is also discussed. As an application of the first result, we derive a central limit theorem for windings of loops around the faces of a planar graph. More precisely, we show that the winding field generated by a random walk loop soup, when appropriately normalized, has a Gaussian limit as the loop soup intensity tends to ∞, and we give an explicit formula for the covariance kernel of the limiting field. We also derive a Spitzer-type law for windings of the Brownian loop soup, i.e., we show that the total winding around a point of all loops of diameter larger than δ, when multiplied by $1/\log\delta$, converges in distribution to a Cauchy random variable as $\delta \to 0$. The random variables analyzed in this work have various interpretations, which we highlight throughout the paper.

F. Camia (✉)
Division of Science, NYU Abu Dhabi, Saadiyat Island, Abu Dhabi, UAE

Department of Mathematics, VU University - Faculty of Sciences, Amsterdam, The Netherlands
e-mail: federico.camia@nyu.edu

Y. Le Jan
Division of Science, NYU Abu Dhabi, Saadiyat Island, Abu Dhabi, UAE

NYU-ECNU Institute of Mathematical Sciences at NYU Shanghai, Shanghai, China
e-mail: yl57@nyu.edu

T. R. Reddy
Deutsche Bank, Mumbai, India
e-mail: tulasi.math@gmail.com

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_11

Keywords Loop soups · Limit theorems · Winding number · Winding field · Spitzer's law · Loop holonomies

1 Introduction

Windings of Brownian paths have been of interest since Spitzer's classic result [16] on their asymptotic behavior which states that, if $\theta(t)$ is the winding angle of a planar Brownian path about a point, then $2(\log t)^{-1}\theta(t)$ converges weakly to a Cauchy random variable as $t \to \infty$. The probability mass function for windings of any planar Brownian loop was computed in [19] (see also [12]). Similar results for random walks were obtained by Belisle [1] and Schapira [15]. Windings of simple random walks on the square lattice were more recently studied in [3, 4].

Symanzik, in his seminal work on Euclidean quantum field theories [17], introduced a representation of a Euclidean field as a "gas" of (interacting) random paths. The noninteracting case gives rise to a Poissonian ensemble of Brownian loops, independently introduced by Lawler and Werner [10] who called it the *Brownian loop soup*. Its discrete version, the *random walk loop soup* was introduced in [9]. These models have attracted a great deal of attention recently because of their connections to the Gaussian free field, the Schramm-Loewner evolution and various models of statistical mechanics (see, e.g., Chapter 9 of [8] and [5, 7, 11, 18]).

Besides the intrinsic interest in the asymptotic behavior of windings of planar Brownian motion and random walks, random variables based on such windings appear naturally in different contexts and are related to various models of current interest in statistical mechanics and conformal field theory, as briefly discussed below.

Integrals over one-forms for loop ensembles, which are generalizations of windings, were considered in [11, Chapter 6]. Various topological aspects of loop soups, such as homotopy and homology, were studied in [13, 14]. In [6], the n-point functions of fields constructed taking the exponential of the winding numbers of loops from a Brownian loop soup are considered. The fields themselves are, a priori, only well-defined when a cutoff that removes small loops is applied, but the n-point functions are shown to converge to conformally covariant functions when the cutoff is sent to zero. A discrete version of these winding fields, based on the random walk loop soup, was considered in [2]. In that paper, the n-point functions of these discrete winding fields are shown to converge, in the scaling limit, to the continuum n-point functions studied in [6]. The same paper contains a result showing that, for a certain range of parameters, the cutoff fields considered in [6] converge to random generalized functions with finite second moments when the cutoff is sent to zero. A similar result was established later in [12] using a different normalization and a different proof.

In this article, with the exception of Sect. 5, we focus on the high intensity limit of loop ensembles on graphs (see [11] for an introduction and various results on this topic). In Sect. 2, we establish a central limit theorem for random variables that are essentially sums of integrals of a one-form over loops of a random walk loop soup, as the intensity of the loop soup tends to infinity. In Sect. 3 we apply the results of

Sect. 2 to the winding field generated by a random walk loop soup on a finite graph and on the infinite square lattice. In Sect. 4, we discuss an extension of the results of Sect. 2 to the noncommutative case of loop holonomies. In the final Sect. 5, we obtain a non-central limit theorem for the windings of the Brownian loop soup.

2 A Central Limit Theorem for Loop Variables

Let $\mathscr{G} = (X, E)$ be a finite connected graph and, for any vertices $x, y \in X$, let $d(x, y)$ denote the graph distance between x and y and d_x the degree of x. The transition matrix P for the random walk on the graph $\mathscr{G}$ with killing function $\kappa : X \to [0, \infty)$ is given by

$$P_{xy} = \begin{cases} \frac{1}{\kappa_x + d_x} & \text{if } d(x, y) = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Let $G = (I - P)^{-1}$ denote the Green's function corresponding to P. G is well defined as long as κ is not identically zero.

We call a sequence $\{x_0, x_1, \ldots, x_n, x_{n+1}\}$ of vertices of $\mathscr{G}$ with $d(x_i, x_{i+1}) = 1$ for every $i = 0, \ldots, n$ and with $x_{n+1} = x_0$ a *rooted loop* with root x_0 and denote it by γ_r. To each γ_r we associate a weight $w_r(\gamma_r) = \frac{1}{n+1} P_{x_0 x_1} \ldots P_{x_n x_0}$. For a rooted loop $\gamma_r = \{x_i\}$, we interpret the index i as time and define an *unrooted loop* as an equivalence class of rooted loops in which two rooted loops belong to the same class if they are the same up to a time translation. To an unrooted loop γ we associate a weight $\mu(\gamma) = \sum_{\gamma_r \in \gamma} w_r(\gamma_r)$. The *random walk loop soup* $\mathscr{L}_\lambda$ with intensity $\lambda > 0$ is a Poissonian collection of unrooted loops with intensity measure $\lambda\mu$.

A *one-form* on $\mathscr{G}$ is a skew-symmetric matrix A with entries $A_{xy} = -A_{yx}$ if $d(x, y) = 1$ and $A_{xy} = 0$ otherwise. A special case of A is illustrated in Fig. 1. For any (rooted/unrooted) loop $\gamma = \{x_0, x_1, \ldots, x_n, x_0\}$, denote

$$\int_\gamma A = A_{x_0, x_1} + A_{x_1, x_2} + \cdots + A_{x_n, x_0}.$$

Given a one-form A and a parameter $\beta \in \mathbb{R}$, we define a 'perturbed transition matrix' P^β with entries

$$P^\beta_{xy} = \begin{cases} \frac{e^{i\beta A_{xy}}}{\kappa_x + d_x} & \text{if } d(x, y) = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $P^\beta = P$ when $\beta = 0$.

Our aim is to derive a central limit theorem for the loop soup random variable

$$\int_{\mathscr{L}_\lambda} A = \sum_{\gamma \in \mathscr{L}_\lambda} \int_\gamma A$$

as the intensity λ of the loop soup increases to infinity. The key to prove such a result is the following representation of the characteristic function of $\int_{\mathscr{L}_\lambda} A$. The result is not new (see, e.g., Section 6.2 of [11]) but we provide a proof for completeness. Below we give a second interpretation of the same result, which provides an alternative proof.

Lemma 1 *With the above notation, assuming that κ is not identically zero, we have that*

$$\mathbb{E}_{\mathscr{L}_\lambda}\left[e^{\left(i\beta \int_{\mathscr{L}_\lambda} A\right)}\right] = \left(\frac{\det(I - P^\beta)}{\det(I - P)}\right)^{-\lambda}.$$

Proof Note that $\det(I - P^\beta)^\lambda$ is well defined and can be written as $e^{\lambda \log\det(I-P^\beta)} = e^{\lambda \operatorname{Tr}\log(I-P^\beta)}$. Since κ is not identically 0, the spectral radius of P^β is strictly less than 1, which implies that

$$-\log(I - P^\beta) = \sum_{k=1}^{\infty} \frac{(P^\beta)^k}{k},$$

where the series in the above expression is convergent.

The weight of all loops of length $k \geq 2$ is given by $\frac{1}{k}\operatorname{Tr}(P^k)$. Therefore the measure of all loops of arbitrary length is

$$\sum_{k=2}^{\infty} \frac{1}{k}\operatorname{Tr}(P^k) = -\operatorname{Tr}\log(I - P) = -\log\det(I - P).$$

Similarly, we have

$$\int e^{i\beta\int_\gamma A} d\mu(\gamma) = \sum_{k=2}^{\infty} \frac{1}{k}\operatorname{Tr}((P^\beta)^k) = -\log\det(I - P^\beta).$$

Therefore, invoking Campbell's theorem for point processes, we have that

$$\mathbb{E}_{\mathscr{L}_\lambda}\left[e^{\left(i\beta \sum_{\gamma\in\mathscr{L}_\lambda}\int_\gamma A\right)}\right] = e^{\left(\lambda\int[\exp(i\beta\int_\gamma A)-1]d\mu(\gamma)\right)} = \frac{\det(I - P^\beta)^{-\lambda}}{\det(I - P)^{-\lambda}},$$

which concludes the proof.

A way to interpret the lemma, which also provides an alternative proof, is to notice that $\det(I-P)^{-\lambda}$ is the partition function Z_λ of the random walk loop soup on $\mathscr{G}$ with transition matrix P and intensity λ, while $\det(I-P^\beta)^{-\lambda}$ is the partition function Z_λ^β of a modified random walk loop soup on $\mathscr{G}$ whose transition matrix is given by P^β. The expectation in Lemma 1 is given by $1/Z_\lambda = \det(I-P)^\lambda$ times the sum over all loop soup configurations $\mathscr{L}_\lambda$ of $\exp\left(i\beta \sum_{\gamma\in\mathscr{L}_\lambda}\int_\gamma A\right)$ times the weight of $\mathscr{L}_\lambda$. The factor $\exp\left(i\beta \sum_{\gamma\in\mathscr{L}_\lambda}\int_\gamma A\right)$ can be absorbed into the weight of $\mathscr{L}_\lambda$ to produce a modified weight corresponding to a loop soup with transition matrix P^β. Therefore the sum mentioned above gives the partition function $Z_\lambda^\beta = \det(I-P^\beta)^{-\lambda}$. Other interpretations of the quantity in Lemma 1 will be discussed in the next section, after the proof of Lemma 2.

To state our next result, we introduce the Hadamard and wedge matrix product operations denoted by $\odot$ and $\wedge$, respectively. For any two matrices U and V of same size, the Hadamard product between them (denoted $U\odot V$) is given by the matrix (of the same size as U and V) whose entries are the products of the corresponding entries in U and V. The following is the only property of matrix wedge products that will be used in this article: If $\lambda_1,\dots,\lambda_n$ are the eigenvalues of an $n\times n$ matrix U, then $\mathrm{Tr}(U^{\wedge k}) = \sum_{i_1<\dots<i_k}\lambda_{i_1}\dots\lambda_{i_k}$ for all $k\le n$. Recall also that $G=(I-P)^{-1}$ denotes the Green's function corresponding to P, which is well defined as long as the killing function κ is not identically zero.

Theorem 1 *With the above notation, assuming that κ is not identically zero, the distribution of the random variable $\frac{1}{\sqrt{\lambda}}\int_{\mathscr{L}_\lambda} A = \frac{1}{\sqrt{\lambda}}\sum_{\gamma\in\mathscr{L}_\lambda}\int_\gamma A$ tends to a Gaussian distribution as $\lambda\to\infty$. More precisely,*

$$\lim_{\lambda\to\infty}\mathbb{E}_{\mathscr{L}_\lambda}\left[e^{\left(\frac{is}{\sqrt{\lambda}}\int_{\mathscr{L}_\lambda} A\right)}\right] = \exp\left[\frac{-s^2}{2}\Big(\mathrm{Tr}\left((P\odot A^{\odot 2})G\right) - \mathrm{Tr}\left((P\odot A)G(P\odot A)G\right)\Big)\right].$$

Proof Let $E^\beta = P - P^\beta$, then

$$E^\beta_{xy} = \begin{cases} \frac{1-e^{i\beta A_{xy}}}{\kappa_x + d_x} & \text{if } d(x,y)=1,\\ 0 & \text{otherwise.}\end{cases}$$

Invoking Lemma 1, we have that

$$\mathbb{E}_{\mathscr{L}_\lambda}\left[e^{\left(i\beta\sum_{\gamma\in\mathscr{L}_\lambda}\int_\gamma A\right)}\right] = \left(\frac{\det(I-P^\beta)}{\det(I-P)}\right)^{-\lambda}$$
$$= \left(\frac{\det(I-P+E^\beta)}{\det(I-P)}\right)^{-\lambda},$$
$$= \left(\det(I-P)^{-1}(I-P+E^\beta)\right)^{-\lambda},$$
$$= \left(\det(I+(I-P)^{-1}E^\beta)\right)^{-\lambda},$$
$$= \left(\det(I+GE^\beta)\right)^{-\lambda}.$$

Let M be a square matrix of dimension n and $\|M\|$ denote the operator norm of M. Then,

$$\det(I+M) = 1 + \mathrm{Tr}(M) + \mathrm{Tr}(M\wedge M) + \cdots + \mathrm{Tr}(M^{\wedge n})$$

and

$$|\,\mathrm{Tr}(M^{\wedge k})| \le \binom{n}{k}\|M\|^k.$$

Using this, we can write

$$\left(\det(I+GE^\beta)\right)^{-\lambda} = (1+\mathrm{Tr}(GE^\beta)+\mathrm{Tr}(GE^\beta\wedge GE^\beta)+O(\beta^3\|A\|^3))^{-\lambda},$$

which leads to

$$\lim_{\lambda\to\infty}\mathbb{E}_{\mathscr{L}_\lambda}\left[\exp\left(i\frac{s}{\sqrt{\lambda}}\sum_{\gamma\in\mathscr{L}_\lambda}\int_\gamma A\right)\right]$$
$$= \lim_{\lambda\to\infty}\left(1+\mathrm{Tr}(GE^{\frac{s}{\sqrt{\lambda}}})+\mathrm{Tr}(GE^{\frac{s}{\sqrt{\lambda}}}\wedge GE^{\frac{s}{\sqrt{\lambda}}})+O\left(\lambda^{-\frac{3}{2}}\right)\right)^{-\lambda}$$
$$= \lim_{\lambda\to\infty}\exp\left(-\lambda\log\left(1+\mathrm{Tr}(GE^{\frac{s}{\sqrt{\lambda}}})+\mathrm{Tr}(GE^{\frac{s}{\sqrt{\lambda}}}\wedge GE^{\frac{s}{\sqrt{\lambda}}})+O\left(\lambda^{-\frac{3}{2}}\right)\right)\right).$$

To proceed, note that $\mathrm{Tr}(GE^\beta\wedge GE^\beta) = \frac{1}{2}(\mathrm{Tr}^2(GE^\beta)-\mathrm{Tr}(GE^\beta GE^\beta))$. Moreover, using the identities $G_{xy}P_{yx} = G_{yx}P_{xy}$ and $A_{xy} = -A_{yx}$ several times, one gets

$$\begin{aligned}
\mathrm{Tr}(GE^{\beta}) &= \sum_{x\sim y} G_{xy}E^{\beta}_{yx} \\
&= \frac{1}{2}\sum_{x\sim y}(G_{xy}E^{\beta}_{yx} + G_{yx}E^{\beta}_{xy}), \\
&= \frac{1}{2}\sum_{x\sim y} G_{xy}P_{yx}(1-e^{i\beta A_{yx}}) + G_{yx}P_{xy}(1-e^{i\beta A_{xy}}), \\
&= \sum_{x\sim y} G_{xy}P_{yx}(1-\cos(\beta A_{yx})).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathrm{Tr}(GE^{\frac{s}{\sqrt{\lambda}}A}) &= \frac{s^2}{2\lambda}\sum_{x\sim y} G_{xy}P_{yx}A^2_{yx} + O\left(\frac{1}{\lambda^{\frac{3}{2}}}\|A\|^3\right) \\
&= \frac{s^2}{2\lambda}\mathrm{Tr}(G(P\odot A^{\odot 2})) + O\left(\frac{1}{\lambda^{\frac{3}{2}}}\|A\|^3\right).
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathrm{Tr}(GE^{\beta}GE^{\beta}) &= \sum_{\substack{x_0,x_1,x_2,x_3\\ x_0\sim x_1;x_2\sim x_3}} E^{\beta}_{x_0,x_1}G_{x_1x_2}E^{\beta}_{x_2,x_3}G_{x_3x_0}, \\
&= \sum_{\substack{x_0,x_1,x_2,x_3\\ x_0\sim x_1;x_2\sim x_3}} (1-e^{i\beta A_{x_0x_1}})(1-e^{i\beta A_{x_2x_3}})P_{x_0x_1}P_{x_2x_3}G_{x_3x_0}G_{x_1x_2}, \\
&= \sum_{\substack{x_0,x_1,x_2,x_3\\ x_0\sim x_1;x_2\sim x_3}} \big(i\beta A_{x_0x_1} + O(\beta^2\|A\|^2)\big)\big(i\beta A_{x_2x_3} + O(\beta^2\|A\|^2)\big) \\
&\qquad P_{x_0x_1}P_{x_2x_3}G_{x_3x_0}G_{x_1x_2}, \\
&= \sum_{\substack{x_0,x_1,x_2,x_3\\ x_0\sim x_1;x_2\sim x_3}} -\beta^2 A_{x_0x_1}P_{x_0x_1}G_{x_1x_2}A_{x_2x_3}P_{x_2x_3}G_{x_3x_0} + O(\beta^3\|A\|^3), \\
&= -\beta^2\,\mathrm{Tr}[(P\odot A)G(P\odot A)G] + O(\beta^3\|A\|^3).
\end{aligned}$$

Note that the expressions $\mathrm{Tr}(GE^{\frac{1}{\sqrt{\lambda}}})$ and $\mathrm{Tr}(GE^{\frac{1}{\sqrt{\lambda}}}\wedge GE^{\frac{1}{\sqrt{\lambda}}})$ are of the order $\frac{1}{\lambda}$. Using this fact, and expanding the logarithm in power series, the above computations give

$$\begin{aligned}&\lim_{\lambda\to\infty} -\lambda \log\left(1+\mathrm{Tr}(GE^{\frac{s}{\sqrt{\lambda}}})+\mathrm{Tr}(GE^{\frac{s}{\sqrt{\lambda}}}\wedge GE^{\frac{1}{\sqrt{\lambda}}})+O\left(\lambda^{\frac{-3}{2}}\right)\right)\\ &=\lim_{\lambda\to\infty}\left(-\tfrac{s^2}{2}\,\mathrm{Tr}(G(P\odot A^{\odot 2}))-\tfrac{s^2}{2}\,\mathrm{Tr}[(P\odot A)G(P\odot A)G]+O(\lambda^{\frac{-1}{2}})\right)\\ &=-\tfrac{s^2}{2}\,\mathrm{Tr}(G(P\odot A^{\odot 2}))-\tfrac{s^2}{2}\,\mathrm{Tr}[(P\odot A)G(P\odot A)G],\end{aligned}$$

which concludes the proof.

Remark 1 It may be useful to note the following identity, which holds when A is skew-symmetric and P is symmetric:

$$\frac{1}{2}\sum_{\substack{x_0\sim x_1\\ x_2\sim x_3}} P_{x_0x_1}P_{x_2x_3}A_{x_0x_1}A_{x_2x_3}[G_{x_0x_3}G_{x_1x_2}-G_{x_0x_2}G_{x_1x_3}]=\mathrm{Tr}[(P\odot A)G(P\odot A)G].$$

We will use this identity in the Proof of Theorem 2 in the next section.

3 Central Limit Theorem for the Loop Soup Winding Field at High Intensity

The winding field generated by a loop soup on a planar graph $\mathscr{G}=(X,E)$ is defined on the faces f of $\mathscr{G}$, which we identify with the vertices of the dual graph $\mathscr{G}^*=(X^*,E^*)$ (i.e., $f\in X^*$). Fix any face $f\in X^*$ and let $f_0=f, f_1,\dots,f_n$ be a sequence of distinct faces of $\mathscr{G}$ that are nearest-neighbors in $\mathscr{G}^*$, with f_n the infinite face. The sequence $f_0,f_1,\dots,f_n$ determines a directed path p from f to the infinite face. Let e_i^p denote the edge between f_i and f_{i+1} oriented in such a way that it crosses p from right to left. We let cut(f) denote the collection of oriented edges $\{e_i^f\}_{i=0}^{n-1}$. (See Fig. 1 for an example.) Note that cut(f) depends on the choice of p, but since all p's connecting f to the infinite face are equivalent for our purposes, we don't include p in the notation.

Now take an oriented loop ℓ in $\mathscr{G}$ and assume that ℓ crosses p. In this case, we say that ℓ crosses cut(f) and we call the crossing *positive* if ℓ crosses p from right to left and *negative* otherwise. For an oriented loop ℓ in $\mathscr{G}$ and a face $f\in X^*$, we define the *winding number* of ℓ about f to be

$$\begin{aligned}W_\ell(f)=&\text{ number of positive crossings of cut}(f)\text{ by }\ell\\ &-\text{number of negative crossings of cut}(f)\text{ by }\ell\end{aligned}$$

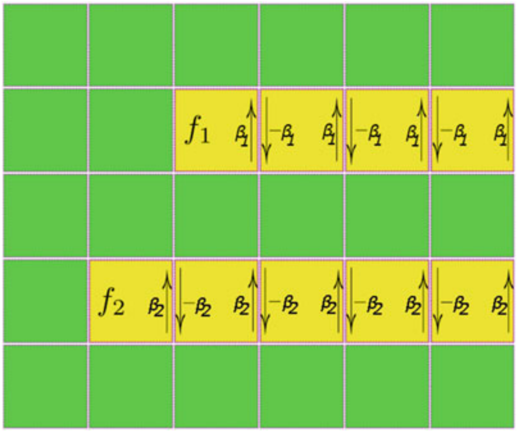


Fig. 1 The figure displays a choice of cuts for faces f_1 and f_2 in a rectangular grid graph

for any choice of cut(f). We note that $W_\ell(f)$ is well defined because the difference above is independent of the choice of cut(f). (This is easy to verify and is left as an exercise for the interested reader.)

For a loop soup $\mathscr{L}_\lambda$, we define

$$W_\lambda = \{W_\lambda(f)\}_{f \in \mathscr{G}^*} = \Big\{ \sum_{\ell \in \mathscr{L}_\lambda} W_\ell(f) \Big\}_{f \in \mathscr{G}^*}$$

to be the *winding field* generated by $\mathscr{L}_\lambda$.

Theorem 1 can be used to prove a CLT for the winding field W_λ, when properly normalized, as $\lambda \to \infty$. In order to use Theorem 1, we need a definition and a lemma. For any collection of faces $f_1, \ldots, f_n$ of $\mathscr{G}$ and any vector $\bar{t} = (t_1, \ldots, t_n)$, define a skew-symmetric matrix $A^{\bar{t}}$ as follows. For each $i = 1, \ldots, n$, choose a cut from f_i to the infinite face as described above and denote it cut(f_i). If $e = (x, y)$ is an edge of cut(f_i) with positive orientation set $A^{\bar{t}}_{xy} = t_i$; if $e = (x, y)$ is an edge of cut(f_i) with negative orientation set $A^{\bar{t}}_{xy} = -t_i$; otherwise set $A^{\bar{t}}_{xy} = 0$. Note that one can write $A^{\bar{t}}$ as $A^{t_1}_{f_1} + \ldots + A^{t_n}_{f_n}$ where A^{t_i} is a matrix such that $(A^{t_i}_{f_i})_{xy} = t_i$ if (x, y) is in cut(f_i) and has positive orientation, $(A^{t_i}_{f_i})_{xy} = -t_i$ if (x, y) is in cut(f_i) and has negative orientation, and $(A^{t_i}_{f_i})_{xy} = 0$ if $(x, y) \notin$ cut(f_i).

Lemma 2 *For any collection of faces $f_1, \ldots, f_n$ of $\mathscr{G}$, there exists a skew-Hermitian matrix $A^{\bar{t}}$ such that the characteristic function of the n-dimensional random vector $(W_\lambda(f_1), \ldots, W_\lambda(f_n))$ is given by*

$$\mathbb{E}_{\mathscr{L}_\lambda}[e^{i\beta(t_1 W_\lambda(f_1) + \cdots + t_n W_\lambda(f_n))}] = \mathbb{E}_{\mathscr{L}_\lambda}\left[e^{\left(i\beta \sum_{\gamma \in \mathscr{L}_\lambda} \int_\gamma A^{\bar{t}} \right)} \right].$$

Proof Using the matrices $A^{\bar{t}}$ describe above, the result follows immediately from the definition of winding number.

The quantity in the lemma has several interpretations. Besides being the characteristic function of the random vector $(W_\lambda(f_1), \ldots, W_\lambda(f_n))$, it can be seen as the n-point function of a winding field of the type studied in [2] (see also [6] for a continuum version). Moreover, by an application of Lemma 1,

$$\mathbb{E}_{\mathscr{L}_\lambda}[e^{i(t_1 W_\lambda(f_1)+\cdots+t_n W_\lambda(f_n))}] = \left(\frac{\det(I - P^{\bar{t}})}{\det(I - P)}\right)^{-\lambda},$$

where

$$P^{\bar{t}}_{xy} = \begin{cases} \frac{e^{iA^{\bar{t}}_{xy}}}{\kappa_x + d_x} & \text{if } d(x, y) = 1 \\ 0 & \text{otherwise} \end{cases}$$

and $A^{\bar{t}}$ is one of the matrices described above. A standard calculation using Gaussian integrals shows that

$$Z^{\bar{t}}_{GFF} = \Pi_{x \in X}\Big(\frac{2\pi}{\kappa_x + d_x}\Big)^{1/2} \det(I - P^{\bar{t}})^{-1/2},$$

where $Z^{\bar{t}}_{GFF}$ is the partition function of the Gaussian Free Field (GFF) on $\mathscr{G}$ with Hamiltonian

$$H^{\bar{t}}(\varphi) = -\frac{1}{2} \sum_{(x,y)\in E} e^{iA^{\bar{t}}_{xy}} \varphi_x \varphi_y + \frac{1}{2} \sum_{x \in X} (\kappa_x + d_x)\varphi_x^2. \tag{1}$$

Hence, $\mathbb{E}_{\mathscr{L}_\lambda}[e^{i(t_1 W_\lambda(f_1)+\cdots+t_n W_\lambda(f_n))}]$ can be written as a ratio of partition functions, namely,

$$\mathbb{E}_{\mathscr{L}_\lambda}[e^{i(t_1 W_\lambda(f_1)+\cdots+t_n W_\lambda(f_n))}] = \Big(\frac{Z^{\bar{t}}_{GFF}}{Z_{GFF}}\Big)^{2\lambda},$$

where Z_{GFF} is the partition function of the 'standard' GFF obtained from (1) by setting $t_1 = \ldots = t_n = 0$.

To state the next theorem we need some additional notation. For any directed edge $e \in \mathrm{cut}(f)$, let e^- and e^+ denote the starting and ending vertices of e, respectively.

Theorem 2 *Consider a random walk loop soup on a finite graph $\mathscr{G}$ with symmetric transition matrix P and the corresponding winding field W_λ. As $\lambda \to \infty$, $\frac{1}{\sqrt{\lambda}} W_\lambda$*

converges to a Gaussian field whose covariance kernel is given by

$$K(f,g) = \sum_{e\in cut(f)} P_{e^+e^-}G_{e^+e^-}\mathbb{1}_{f=g} + 2\sum_{\substack{e_1\in cut(f)\\ e_2\in cut(g)}} P_{e_1^+e_1^-}P_{e_2^+e_2^-}\left(G_{e_1^+e_2^-}G_{e_1^-e_2^+} - G_{e_1^+e_2^+}G_{e_1^-e_2^-}\right)\mathbb{1}_{f\neq g}.$$

Proof Combining Lemma 2 and Theorem 1 shows that the winding field has a Gaussian limit as $\lambda \to \infty$:

$$\left\{\frac{1}{\sqrt{\lambda}}W_\lambda(f) : f \text{ is a face of } \mathcal{G}\right\} \xrightarrow[weakly]{\lambda\uparrow\infty} \left\{W(f) : f \text{ is a face of } \mathcal{G}\right\}$$

where $W(\cdot)$ is a Gaussian process on the faces of G.

Next, we compute the covariance kernel of the limiting Gaussian process. Choose two faces f and g and let $A^{\bar{t}} = A_f^{t_1} + A_g^{t_2}$, where $A_f^{t_1}$ has nonzero entries only along cut(f) and $A_g^{t_2}$ has nonzero entries along cut(g), as described above. Using Theorem 1 we obtain

$$\begin{aligned}
&\lim_{\lambda\to\infty} \log \mathbb{E}_{\mathcal{L}_\lambda}[e^{i\frac{1}{\sqrt{\lambda}}(t_1W_\lambda(f_1)+t_2W_\lambda(f_2))}] \\
&= -\frac{1}{2}\left[\mathrm{Tr}((P\odot(A^{\bar{t}})^{\odot 2})G) + \mathrm{Tr}((P\odot A^{\bar{t}})G(P\odot A^{\bar{t}})G)\right] \\
&= -\frac{1}{2}\left[\mathrm{Tr}((P\odot(A^{t_1}+A^{t_2})^{\odot 2})G) + \mathrm{Tr}((P\odot(A^{t_1}+A^{t_2}))G(P\odot(A^{t_1}+A^{t_2})G)\right] \\
&= -\frac{1}{2}\left[t_1^2K(f,f) + t_2^2K(g,g) - 2t_1t_2K(f,g)\right].
\end{aligned}$$

The variance of $W(f)$ is obtained by setting $t_1 = t$ and $t_2 = 0$. In this case,

$$\begin{aligned}
K(f,f) &= \frac{1}{t^2}\left[\mathrm{Tr}((P\odot(A^t)^{\odot 2})G) + \mathrm{Tr}((P\odot A^t)G(P\odot A^t)G)\right] \\
&= \sum_{e\in \mathrm{cut}(f)} P_{e^-e^+}G_{e^+e^-} + \frac{1}{t^2}(\mathrm{Tr}((P\odot A^t)G(P\odot A^t)G)).
\end{aligned}$$

Since P is assumed to be symmetric, the term $\mathrm{Tr}((P\odot A^t)G(P\odot A^t)G)$ vanishes. Therefore,

$$K(f,f) = \sum_{e\in \mathrm{cut}(f)} P_{e^-e^+}G_{e^-e^+}.$$

A similar calculation, with $A^{\bar{t}} = A_f^{t_1} + A_g^{t_2}$, where $A_f^{t_1}$, gives the covariance:

$$K(f,g) = \frac{-1}{2t_1t_2}\Big[\operatorname{Tr}((P\odot(A^{t_1}+A^{t_2})^{\odot 2})G) + \operatorname{Tr}(P\odot(A^{t_1}+A^{t_2})G(P\odot(A^{t_1}+A^{t_2})G))$$
$$-t_1^2K(f,f) - t_2^2K(g,g)\Big]$$
$$= \frac{-1}{2t_1t_2}\Big[\operatorname{Tr}((P\odot(A^{t_1}+A^{t_2})^{\odot 2})G) + \operatorname{Tr}(P\odot(A^{t_1}+A^{t_2})G(P\odot(A^{t_1}+A^{t_2})G))$$
$$-\operatorname{Tr}((P\odot(A^{t_1})^{\odot 2})G) - \operatorname{Tr}((P\odot A^{t_1})G(P\odot A^{t_1})G)$$
$$-\operatorname{Tr}((P\odot(A^{t_2})^{\odot 2})G) - \operatorname{Tr}((P\odot A^{t_2})G(P\odot A^{t_2})G)\Big]$$
$$= \frac{-1}{t_1t_2}\Big[\operatorname{Tr}((P\odot(A^{t_1}\odot A^{t_2}))G) + \operatorname{Tr}((P\odot A^{t_1})G(P\odot A^{t_2})G)\Big].$$

Since P is symmetric, $\operatorname{Tr}((P\odot(A^{t_1}\odot A^{t_2}))G) = 0$ and, using Remark 1, we obtain

$$K(f,g) = 2\sum_{\substack{e_1\in\operatorname{cut}(f)\\ e_2\in\operatorname{cut}(g)}} P_{e_1^+e_1^-}P_{e_2^+e_2^-}\left(G_{e_1^+e_2^-}G_{e_1^-e_2^+} - G_{e_1^+e_2^+}G_{e_1^-e_2^-}\right),$$

which concludes the proof.

Remark 2 We provide here an alternative, more direct but more specific, Proof of Theorem 2. For any directed edge e, let e^- and e^+ to denote the starting and ending vertices respectively. Moreover, let N_e^+ be the number of positive crossings of e (i.e., from e^- to e^+) by a loop from the loops soup and let N_e^- be the number of negative crossings of e (i.e., from e^+ to e^-) by a loop from the loops soup. The winding number about a face f can be defined as,

$$W_\lambda(f) = \sum_{e\in\operatorname{cut}(f)} (N_e^+ - N_e^-).$$

Therefore the two point function for the winding numbers is given by

$$\mathbb{E}_{\mathscr{L}_1}(W_1(f)W_1(g)) = \mathbb{E}_{\mathscr{L}_1}\Big(\sum_{\substack{e_1\in\operatorname{cut}(f)\\ e_2\in\operatorname{cut}(g)}} (N_{e_1}^+ - N_{e_1}^-)(N_{e_2}^+ - N_{e_2}^-)\Big)$$
$$= \mathbb{E}_{\mathscr{L}_1}\Big(\sum_{\substack{e_1\in\operatorname{cut}(f)\\ e_2\in\operatorname{cut}(g)}} (N_{e_1}^+N_{e_2}^+ + N_{e_1}^-N_{e_2}^- - N_{e_1}N_{e_2}^- - N_{e_1}^-N_{e_2}^+)\Big)$$
$$= 2\sum_{\substack{e_1\in\operatorname{cut}(f)\\ e_2\in\operatorname{cut}(g)}} \left(\mathbb{E}_{\mathscr{L}_1}(N_{e_1}^+N_{e_2}^+) - \mathbb{E}_{\mathscr{L}_1}(N_{e_1}^+N_{e_2}^-)\right)$$

Using this expression and a result from [11] (see [11, Exercise 10, Chapter 2], but note that in [11] the Green's function is defined to be $[(I-P)^{-1}]_{x,y}/(\kappa_y+d_y) = P_{yx}G_{xy} = P_{xy}G_{xy}$, when P is symmetric), we obtain

$$\begin{aligned}
&\mathbb{E}_{\mathscr{L}_1}(W_1(f)W_1(g)) \\
&= \sum_{\substack{e_1\in\text{cut}(f)\\ e_2\in\text{cut}(g)}} \Big(P_{e_2^+e_2^-}P_{e_1^+e_1^-}G_{e_1^+e_2^-}G_{e_2^+e_1^-} + P_{e_2^-e_2^+}P_{e_1^-e_1^+}G_{e_1^-e_2^+}G_{e_2^-e_1^+} \\
&\qquad - P_{e_2^-e_2^+}P_{e_1^+e_1^-}G_{e_1^+e_2^+}G_{e_2^-e_1^-} - P_{e_2^+e_2^-}P_{e_1^-e_1^+}G_{e_1^-e_2^-}G_{e_2^+e_1^+}\Big) \\
&= 2\sum_{\substack{e_1\in\text{cut}(f)\\ e_2\in\text{cut}(g)}} P_{e_1^+e_1^-}P_{e_2^+e_2^-}\Big(G_{e_1^+e_2^-}G_{e_2^+e_1^-} - G_{e_1^+e_2^+}G_{e_2^-e_1^-}\Big).
\end{aligned}$$

Now take $\lambda = n \in \mathbb{N}$ and note that, for any face f, $W_n(f)$ is distributed like $\sum_{i=1}^n W_1^i(f)$, where $\{W_1^i(f)\}_{i=1,\dots,n}$ are n i.i.d. copies of $W_1(f)$. Therefore, for any collection of faces $f_1,\dots,f_m$, the central limit theorem implies that, as $\lambda = n \to \infty$, the random vector $\frac{1}{\sqrt{n}}(W_n(f_1),\dots,W_n(f_m))$ converges to a multivariate Gaussian with covariance kernel given by the two-point function $\mathbb{E}_{\mathscr{L}_1}(W_1(f)W_1(g))$ calculated above.

Remark 3 Theorem 2 can be extended to infinite graphs, as we now explain. For concreteness and simplicity, we focus on the square lattice and consider a random walk loop soup with constant killing function: $\kappa_x = \kappa > 0$ for all $x \in \mathbb{Z}^2$. Note that in this case the transition matrix P and the Green's function G are symmetric. Moreover, contrary to the case $\kappa = 0$, the winding field of the random walk loop soup on $\mathbb{Z}^2$ is well defined when $\kappa > 0$. To see this, note that, since the loop soup is a Poisson process, we can bound the expected number of loops intersecting $(-a,0),(b,0)\in\mathbb{Z}^2$ as follows:

$$\begin{aligned}
\mathbb{E}_{\mathscr{L}_\lambda}\big(\#\text{ loops joining }(-a,0)\text{ and }(b,0)\big) &= \lambda\mu\big(\gamma : \gamma(-a,0),(b,0)\in\gamma\big) \\
&\le \lambda \sum_{m\ge 2(a+b)} \Big(\frac{4}{\kappa+4}\Big)^m \\
&= \frac{\lambda\kappa}{4}\Big(1+\frac{\kappa}{4}\Big)^{-2(a+b)}.
\end{aligned}$$

Hence, the expected number of loops winding around the origin is bounded above by $\frac{\lambda\kappa}{4}\sum_{a=1}^\infty\sum_{b=1}^\infty (1+\kappa/4)^{-2(a+b)} < \infty$ for any $\kappa > 0$. This means that, with probability one, the number of loops winding around any vertex is finite. Because of this, one can obtain the winding field on $\mathbb{Z}^2$ as the weak limit of winding fields in large finite graphs $\mathscr{G}_n = [-n,n]^2 \cap \mathbb{Z}^2$ as $n \to \infty$. It is now clear that one can apply the arguments in Remark 2 to the case of the winding field on $\mathbb{Z}^2$.

4 Holonomies of Loop Ensembles

Theorem 1 can be generalized to loop holonomies, a task we accomplish in this section. Assume that the transition matrix P introduced at the beginning of Sect. 2 is symmetric and hence the Green's function is also symmetric. We consider a connection on the graph $\mathscr{G}$, given by assigning to each oriented edge (x, y) a $d \times d$ unitary matrix $\mathbf{U_{xy}}$ of the form $U_{xy} = e^{iA_{xy}}$ for some Hermitian matrix A_{xy}. For any closed loop $\gamma = \{x_0, x_1, \dots, x_n, x_0\}$, we denote

$$\prod_{\gamma} \mathbf{U} = \mathbf{U_{x_0x_1}} \mathbf{U_{x_1x_2}} \dots \mathbf{U_{x_nx_0}}.$$

We also write $\mathrm{Tr}_\gamma\,[\mathbf{U}]$ for $\mathrm{Tr}[\mathbf{U_{x_0x_1}} \mathbf{U_{x_1x_2}} \dots \mathbf{U_{x_nx_0}}]$, which is well defined as the expression inside $\mathrm{Tr}[\cdot]$ is shift invariant. We will re-do the computations leading to Theorem 1, in this case by invoking block matrices. Note that since $\mathbf{U}_{xy} = \mathbf{U}_{yx}^{-1}$, we assume $\mathbf{A}_{xy} = -\mathbf{A}_{yx}$. Denote the corresponding block matrix whose blocks are $\mathbf{A_{xy}}$ with $\mathbf{A}$. Similarly denote $\mathrm{Tr}_\gamma\left[e^{i\beta\mathbf{A}}\right] := \mathrm{Tr}\left[e^{i\beta\mathbf{A_{x_0x_1}}} \dots e^{i\beta\mathbf{A_{x_nx_0}}}\right]$. We denote the tensor product between two matrices A and B to be $A \otimes B$ and the Hadamard product to be $A \odot B$.

In this context, the quantity $\exp\left(i \sum_{\gamma\in\mathscr{L}_\lambda} \frac{1}{\sqrt{\lambda}} \int_\gamma A\right) = \prod_{\gamma\in\mathscr{L}_\lambda} e^{\frac{i}{\sqrt{\lambda}}\int_\gamma A}$ that appears in Theorem 1 will be replaced by $\prod_{\gamma\in\mathscr{L}_\lambda} \frac{1}{d}\,\mathrm{Tr}_\gamma\left(e^{i\frac{1}{\sqrt{\lambda}}\mathbf{A}}\right)$. Two observations are in order: (1) The expectation of this quantity cannot be interpreted as a characteristic function, but other interpretations such as those discussed after Lemmas 1 and 2 are still available. (2) The presence of Tr_γ means that the Proof of Theorem 1 doesn't apply directly to this case; indeed, a careful decomposition of traces is needed to obtain the non-Abelian extension presented in Theorem 3 below.

The first step towards the main result of this section is the following lemma.

Lemma 3 *With the above notation we have*

$$\mathbb{E}_{\mathscr{L}_\lambda}\left[\prod_{\gamma\in\mathscr{L}_\lambda} \frac{1}{d}\,\mathrm{Tr}_\gamma\left(e^{i\beta\mathbf{A}}\right)\right] = \left(\frac{\det(I_{nd} - (P\otimes J_d)\odot \mathbf{U}_\beta)}{\det(I_{nd} - P\otimes I_d)}\right)^{-\lambda},$$

where $(P \otimes J_d) \odot \mathbf{U}$ and $P \otimes I_d$ are block matrices whose blocks are $P_{ij}\mathbf{U_{ij}}$ and $P_{ij} I_d$ respectively, J_d is $d \times d$ matrix whose entries are all 1 and for any k, I_k is the $k \times k$ identity matrix.

Proof The statement follows from a computation similar to that in the proof of Lemma 1, namely

$$\mathbb{E}_{\mathscr{L}_\lambda}\left[\prod_{\gamma\in\mathscr{L}_\lambda}\frac{1}{d}\operatorname{Tr}_\gamma\left(e^{i\beta\mathbf{A}}\right)\right]=\frac{\exp(-\lambda\sum_{k=1}^{\infty}\frac{1}{k}\operatorname{Tr}((P\otimes J_d)\odot\mathbf{U}_\beta)^k)}{\exp(-\lambda\sum_{k=1}^{\infty}\frac{1}{k}\operatorname{Tr}(P\otimes I_d)^k)}$$

$$=\frac{e^{\lambda\operatorname{Tr}\log(I-(P\otimes J_d)\odot\mathbf{U}_\beta)}}{e^{\lambda\operatorname{Tr}\log(I-P\otimes I_d)}}$$

$$=\frac{(\det(I-(P\otimes J_d)\odot\mathbf{U}_\beta))^{-\lambda}}{(\det(I-P\otimes I_d))^{-\lambda}}.$$

We note that a similar computation can be found in [11, Proposition 23].

We are now ready to state and prove the main result of this section.

Theorem 3 *With the notation above we have*

$$\lim_{\lambda\to\infty}\mathbb{E}_{\mathscr{L}_\lambda}\left[\prod_{\gamma\in\mathscr{L}_\lambda}\frac{1}{d}\operatorname{Tr}_\gamma\left(e^{i\frac{1}{\sqrt{\lambda}}\mathbf{A}}\right)\right]=$$

$$\exp\Big(-\frac{1}{2}\Big[\sum_{x\sim y}G_{xy}P_{xy}\operatorname{Tr}(\mathbf{A}_{xy}^2)$$

$$+\sum_{\substack{x_0\sim x_1\\ x_2\sim x_3}}P_{x_0x_1}P_{x_2x_3}\operatorname{Tr}[\mathbf{A}_{x_0x_1}\mathbf{A}_{x_2x_3}](G_{x_0x_3}G_{x_1x_2}-G_{x_0x_2}G_{x_1x_3})\Big]\Big).$$

Proof We follow the computation in the Proof of Theorem 1. From Lemma 3, defining $\mathbf{E}^{\beta\mathbf{A}}=P\otimes I_d-(P\otimes J_d)\odot\mathbf{U}_\beta$, we have

$$\mathbb{E}_{\mathscr{L}_\lambda}\left[\prod_{\gamma\in\mathscr{L}_\lambda}\frac{1}{d}\operatorname{Tr}_\gamma\left(e^{i\beta\mathbf{A}}\right)\right]=\left(\frac{\det(I_{nd}-(P\otimes J_d)\odot\mathbf{U}_\beta)}{\det(I_{nd}-P\otimes I_d)}\right)^{-\lambda}$$

$$=(\det(I_{nd}+(I_{nd}-P\otimes I_d)^{-1}\mathbf{E}^{\beta A}))^{-\lambda}.$$

Therefore,

$$\lim_{\lambda\to\infty}\left(\frac{\det(I_{nd}-(P\otimes J_d)\odot\mathbf{U}_{\frac{1}{\sqrt{\lambda}}})}{\det(I_{nd}-P\otimes I_d)}\right)^{\lambda}$$

$$=\lim_{\lambda\to\infty}\Big(1+\operatorname{Tr}((I_{nd}-P\otimes I_d)^{-1}\mathbf{E}^{\frac{1}{\sqrt{\lambda}}\mathbf{A}})$$

$$
\begin{aligned}
&\quad + \operatorname{Tr}((I_{nd} - P \otimes I_d)^{-1}\mathbf{E}^{\frac{1}{\sqrt{\lambda}}\mathbf{A}} \wedge (I_{nd} - P \otimes I_d)^{-1}\mathbf{E}^{\frac{1}{\sqrt{\lambda}}\mathbf{A}}) \\
&\quad + O\left(\frac{1}{\lambda^{\frac{3}{2}}}\|\mathbf{A}\|^3\right)\Bigg)^{\lambda} \\
&= \lim_{\lambda\to\infty} \exp\Big[\lambda\Big(\operatorname{Tr}\big((I_{nd} - P \otimes I_d)^{-1}\mathbf{E}^{\frac{1}{\sqrt{\lambda}}\mathbf{A}}\big) \\
&\quad + \operatorname{Tr}\big((I_{nd} - P \otimes I_d)^{-1}\mathbf{E}^{\frac{1}{\sqrt{\lambda}}\mathbf{A}} \wedge (I_{nd} - P \otimes I_d)^{-1}\mathbf{E}^{\frac{1}{\sqrt{\lambda}}\mathbf{A}}\big)\Big)\Big].
\end{aligned}
$$

Note that $(I_{nd} - P \otimes I_d)^{-1} = ((I_n - P) \otimes I_d)^{-1} = G \otimes I_d$. Moreover, expanding the traces of block matrices in terms of traces of blocks, we have

$$
\begin{aligned}
\operatorname{Tr}((G \otimes I_d)\mathbf{E}^{\beta\mathbf{A}}) &= \sum_{x\sim y} G_{xy} P_{yx} \operatorname{Tr}(I_d - e^{i\beta\mathbf{A}_{yx}}) \\
&= -\frac{\beta^2}{2}\sum_{x\sim y} G_{xy} P_{xy} \operatorname{Tr}(\mathbf{A}_{xy}^2) + O(\beta^3\|\mathbf{A}\|_\infty^3).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&\operatorname{Tr}(\mathbf{E}^{\beta\mathbf{A}}(G \otimes I_d)\mathbf{E}^{\beta\mathbf{A}}(G \otimes I_d)) \\
&\quad = \sum_{\substack{x_0\sim x_1\\ x_2\sim x_3}} \operatorname{Tr}(E^{\beta\mathbf{A}_{x_0,x_1}} G_{x_1x_2} E^{\beta\mathbf{A}_{x_2,x_3}} G_{x_3x_0}) \\
&\quad = \sum_{\substack{x_0\sim x_1\\ x_2\sim x_3}} \operatorname{Tr}[E^{\beta\mathbf{A}_{x_0x_1}} G_{x_1x_2} E^{\beta\mathbf{A}_{x_2x_3}} G_{x_3x_0}] \\
&\quad = \sum_{\substack{x_0\sim x_1\\ x_2\sim x_3}} \operatorname{Tr}[(I_d - e^{i\beta\mathbf{A}_{x_0x_1}})(I_d - e^{i\beta\mathbf{A}_{x_2x_3}}) P_{x_0x_1} P_{x_2x_3} G_{x_0x_3} G_{x_1x_2}] \\
&\quad = \sum_{\substack{x_0\sim x_1\\ x_2\sim x_3}} P_{x_0x_1} P_{x_2x_3} \operatorname{Tr}[-\beta^2\mathbf{A}_{x_0x_1}\mathbf{A}_{x_2x_3} G_{x_0x_3} G_{x_2x_1}] + O(\beta^3\|\mathbf{A}\|_\infty^3) \\
&\quad = -\frac{\beta^2}{2}\sum_{\substack{x_0\sim x_1\\ x_2\sim x_3}} P_{x_0x_1} P_{x_2x_3} \operatorname{Tr}[\mathbf{A}_{x_0x_1}\mathbf{A}_{x_2x_3}](G_{x_0x_3} G_{x_1x_2} - G_{x_0x_2} G_{x_1x_3}) \\
&\qquad + O(\beta^3\|\mathbf{A}\|_\infty^3).
\end{aligned}
$$

Invoking the identity $\mathrm{Tr}\left(M \wedge M\right) = \frac{1}{2}(\mathrm{Tr}(M)^2 - \mathrm{Tr}(M^2))$ and the computation above, we have that

$$\lim_{\lambda\to\infty} \log \mathbb{E}_{\mathscr{L}_\lambda}\left[\prod_{\gamma\in\mathscr{L}_\lambda} \frac{1}{d}\,\mathrm{Tr}_\gamma\left(e^{i\frac{1}{\sqrt{\lambda}}\mathbf{A}}\right)\right]$$

$$= -\frac{1}{2}\sum_{x\sim y} G_{xy}P_{xy}\,\mathrm{Tr}(\mathbf{A}_{xy}^2)$$

$$-\frac{1}{2}\sum_{\substack{x_0\sim x_1\\ x_2\sim x_3}} P_{x_0x_1}P_{x_2x_3}\,\mathrm{Tr}[\mathbf{A}_{x_0x_1}\mathbf{A}_{x_2x_3}](G_{x_0x_3}G_{x_1x_2} - G_{x_0x_2}G_{x_1x_3}),$$

which concludes the proof.

5 A Spitzer-Type Law for Windings of the Brownian Loop Soup

In this last section we present a result of a different nature but related to those discussed in the previous sections. Here we keep the intensity λ fixed but we take a continuum scaling limit, letting the lattice mesh size go to zero, which leads to the Brownian loop soup [9]. The scale invariance of the Brownian loop soup immediately implies that the winding around any deterministic point is infinite even for fixed intensity $\lambda < \infty$. Hence, in order to study the law of windings, a renormalization procedure is needed. In this case we will obtain a non-central limit theorem (Theorem 4 below). The situation is similar to the case of a single Brownian path. Spitzer showed [16] that the winding of Brownian motion about a given point up to time t, when scaled by $1/(2\log t)$, converges in distribution to a Cauchy random variable as $t \to \infty$. Belisle [1] proved an analog of Spitzer's law for windings of planar random walks. (Subsequent results on the asymptotic law of windings of Brownian motion are surveyed in the book by Yor [20]). In this section we prove a similar result for the Brownian loop soup in a bounded domain. Our renormalization procedure involves a cutoff $\delta > 0$ on the diameter of loops.

Recall that the Brownian loop soup in a planar domain $D \subset \mathbb{C}$ is defined as a Poisson process of loops with intensity measure μ^{loop} given by

$$\mu^{\mathrm{loop}}(\cdot) = \int_D \int_0^\infty \frac{1}{2\pi t^2}\mu_{\mathrm{BB}}^{z,t}(\cdot)\,dt\,dA(z),$$

where $\mu_{\mathrm{BB}}^{z,t}$ is the Brownian Bridge measure of time length t starting at z and dA is the area measure on the complex plane (see [10] for a precise definition).

For any $z \in D$, we let $W_\lambda^\delta(z)$ denote the sum of the winding numbers about z of all Brownian loops contained in D with diameter at least δ for some $\delta > 0$.

Theorem 4 *Consider a bounded domain $D \subset \mathbb{C}$. For any $z \in D$, as $\delta \downarrow 0$, $\frac{W_\lambda^\delta(z)}{\log \delta}$ converges weakly to a Cauchy random variable with location parameter* 0 *and scale parameter* $\frac{\lambda}{2\pi}$.

Proof Let d_z denote the distance between z and the boundary of D and, for $\delta < d_z$, let $W_\lambda^{\delta,d_z}(z)$ denote the sum of the winding numbers about z of all Brownian loops with diameter between δ and d_z. Note that, because of the Poissonian nature of the Brownian loop soup, the random variables $W_\lambda^{\delta,d_z}(z)$ and $W_\lambda^{d_z}(z)$ are independent.

The key ingredient in the proof is Lemma 3.2 of [6], which states, in our notation, that

$$\mathbb{E}\big(e^{i\beta W_\lambda^{\delta,d_z}(z)}\big) = \Big(\frac{d_z}{\delta}\Big)^{-\lambda\frac{\beta(2\pi-\beta)}{4\pi^2}} = d_z^{-\lambda\frac{\beta(2\pi-\beta)}{4\pi^2}} e^{\lambda\frac{\beta(2\pi-\beta)}{4\pi^2}\log\delta}$$

when $\beta \in [0, 2\pi)$, and that the same expression holds with β replaced by (β mod 2π) when $\beta \notin [0, 2\pi)$. With this result, choosing $\beta = s/\log\delta$, the limit as $\delta \to 0$ of the characteristic function of $\frac{W_\lambda^\delta(z)}{\log\delta}$ can be computed as follows:

$$\lim_{\delta\to 0}\mathbb{E}\big(e^{i\frac{s}{\log\delta}W_\lambda^\delta(z)}\big) = \lim_{\delta\to 0}\mathbb{E}\big(e^{i\frac{s}{\log\delta}W_\lambda^{\delta,d_z}(z)}\big)\mathbb{E}\big(e^{i\frac{s}{\log\delta}W_\lambda^{d_z}(z)}\big) = e^{-\frac{\lambda}{2\pi}|s|},$$

where the right hand side is the characteristic function of a Cauchy random variable with location parameter 0 and scale parameter $\frac{\lambda}{2\pi}$.

Remark 4 One can also consider the joint distribution of n random variables, namely $\frac{W_\lambda^\delta(z_1)}{\log\delta}, \dots, \frac{W_\lambda^\delta(z_n)}{\log\delta}$ for n distinct points $z_1, \dots, z_n \in D$. If one takes $d < \min_{i<j}\Big(\frac{|z_i - z_j|}{2}\Big)$, the random variables $W_\lambda^{\delta,d}(z_1), \dots, W_\lambda^{\delta,d}(z_n)$ defined in the Proof of Theorem 4 are independent. This observation shows that the convergence in distribution of Theorem 4 holds also for multiple points, namely, as $\delta \downarrow 0$, the random variables $\frac{W_\lambda^\delta(z_1)}{\log\delta}, \dots, \frac{W_\lambda^\delta(z_n)}{\log\delta}$ converge jointly in distribution to n independent Cauchy random variables.

Acknowledgments The first author thanks David Brydges for an enlightening discussion during the workshop "Random Structures in High Dimensions" held in June–July 2016 at the Casa Matemática Oaxaca (CMO) in Oaxaca, Mexico. All authors thank an anonymous referee for a careful reading of the manuscript and useful suggestions. The research presented in this paper was carried out while the third author was a postdoctoral associate in the Division of Science of NYU Abu Dhabi. The second author acknowledges the support of NYU-ECNU Institute of Mathematical Sciences at NYU Shanghai.

References

1. Belisle, C.: Windings of random walks. Ann. Probab. **17**, 1377–1402 (1989)
2. van de Brug, T., Camia, F., Lis, M.: Spin systems from loop soups. Electron. J. Probab. **23**, 17 pp. (2018)
3. Budd, T.: The peeling process in random planar maps coupled to an O(n) loop model (with an appendix by Linxiao Chen) (2018). Preprint arXiv:1809.02012
4. Budd, T.: Winding of simple walks on the square lattice. J. Comb. Theory Ser. A. **172**, 105191 (2019)
5. Camia, F.: Scaling limits, Brownian loops, and conformal fields. In: Contucci, P., Giardinà, C. (eds.) Advances in Disordered Systems, Random Processes and Some Applications, pp. 205–269. Cambridge University Press, Cambridge (2017)
6. Camia, F., Gandolfi, A., Kleban, M.: Conformal correlation functions in the Brownian loop soup. Nuclear Phys. B **902**, 483–507 (2016)
7. Lawler, G.F.: Topics in loop measures and the loop-erased walk. Probab. Surveys **15**, 28–101 (2018)
8. Lawler, G.F., Limic V.: Random Walk: A Modern Introduction. Cambridge University Press, Cambridge (2010)
9. Lawler, G.F., Trujillo Ferreras, J.A.: Random walk loop soup. Trans. Amer. Math. Soc. **359**, 767–787 (2007)
10. Lawler, G.F., Werner, W.: The Brownian loop soup. Probab. Theory Relat. Fields **128**, 565–588 (2004)
11. Le Jan, Y.: Markov Paths, Loops and Fields. Lecture Notes in Mathematics, vol. 206. Springer, Heidelberg (2011)
12. Le Jan, Y.: Brownian winding fields. In: Donati-Martin, C., Lejay, A., Rouault, A. (eds.) Séminaire de Probabilités L. Lecture Notes in Mathematics, pp. 487–492, vol. 2252. Springer, Cham (2019)
13. Le Jan, Y.: Brownian loops topology. Potential Analy. **53**, 223–229 (2019)
14. Lupu, T.: Topological expansion in isomorphisms with random walks for matrix valued fields (2019). Preprint arXiv:1908.06732
15. Schapira, B., Young, R.: Windings of planar random walks and averaged Dehn function. Ann. Inst. H. Poincaré Probab. Statist. **47**, 130–147 (2011)
16. Spitzer, F.: Some theorems concerning 2-dimensional Brownian motion. Trans. Amer. Math. Soc. **87**, 187–197 (1958)
17. Symanzik, K.: Euclidean quantum field theory. Rend. Scu. Int. Fis. Enrico. Fermi. **45**, 152–226 (1969)
18. Sznitman, A.-S.: Topics in Occupation Times and Gaussian Free Fields. European Mathematical Society, Zürich (2012)
19. Yor, M.: Loi de l'indice du lacet Brownien, et distribution de Hartman-Watson. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **53**, 71–95 (1980)
20. Yor, M.: Some Aspects of Brownian Motion. Part I: Some Special Functionals. Birkhäuser, Basel (1992)

The Stable Derrida–Retaux System at Criticality

Xinxing Chen and Zhan Shi

Dedicated to the memory of Vladas Sidoravicius

Abstract The Derrida–Retaux recursive system was investigated by Derrida and Retaux (J Stat Phys 156:268–290, 2014) as a hierarchical renormalization model in statistical physics. A prediction of Derrida and Retaux (J Stat Phys 156:268–290, 2014) on the free energy has recently been rigorously proved (Chen et al., The Derrida–Retaux conjecture on recursive models. https://arxiv.org/abs/1907.01601), confirming the Berezinskii–Kosterlitz–Thouless-type phase transition in the system. Interestingly, it has been established in the paper by Chen et al. that the prediction is valid only under a certain integrability assumption on the initial distribution, and a new type of universality result has been shown when this integrability assumption is not satisfied. We present a unified approach for systems satisfying a certain domination condition, and give an upper bound for derivatives of all orders of the moment generating function. When the integrability assumption is not satisfied, our result allows to identify the large-time order of magnitude of the product of the moment generating functions at criticality, confirming and completing a previous result in Collet et al. (Commun Math Phys 94:353–370, 1984).

Keywords Derrida-Retaux recursive system · Moment generating function

2020 Mathematics Subject Classification 60J80, 82B27

X. Chen (✉)
School of Mathematical Sciences, Shanghai Jiaotong University, Shanghai, China
e-mail: chenxinx@sjtu.edu.cn

Z. Shi
LPSM, Sorbonne Université Paris VI, Paris Cedex 05, France
e-mail: zhan.shi@sorbonne-universite.fr

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_12

1 Introduction

Fix an integer $m \geq 2$. Let X_0 be a random variable taking values in $\mathbb{Z}_+ := \{0, 1, 2, \ldots\}$. To avoid trivial discussion, it is assumed, throughout the paper, that $\mathbf{P}(X_0 \geq 2) > 0$. Let us consider the Derrida–Retaux recursive system $(X_n, n \geq 0)$ defined as follows: for all $n \geq 0$,

$$X_{n+1} = (X_{n,1} + \cdots + X_{n,m} - 1)^+, \tag{1}$$

where $X_{n,i}$, $i \geq 1$, are independent copies of X_n. This was investigated by Derrida and Retaux [6] as a toy model to study depinning in presence of impurities [8–12, 14, 15]. We refer to [7] for an overview on rigorous results and predictions about the Derrida–Retaux system.

Assuming $\mathbf{E}(X_0) < \infty$, it is immediate from (1) that $\mathbf{E}(X_{n+1}) \leq m\,\mathbf{E}(X_n)$, so the free energy

$$F_\infty := \lim_{n\to\infty} \downarrow \frac{\mathbf{E}(X_n)}{m^n} \in [0, \infty),$$

is well-defined. A remarkable result by Collet et al. [5] tells us that assuming $\mathbf{E}(X_0\, m^{X_0}) < \infty$ (which we take for granted throughout the paper) and writing $\eta := (m-1)\mathbf{E}(X_0\, m^{X_0}) - \mathbf{E}(m^{X_0})$, then $F_\infty > 0$ if $\eta > 0$, and $F_\infty = 0$ if $\eta \leq 0$.

As such, it is natural to say that the system $(X_n, n \geq 0)$ is supercritical if $\eta > 0$, is critical if $\eta = 0$, and is subcritical if $\eta < 0$.

It has been conjectured by Derrida and Retaux [6] that if $\eta > 0$, then we would have

$$F_\infty = \exp\Big(- \frac{C + o(1)}{\eta^{1/2}}\Big), \qquad \eta \to 0+, \tag{2}$$

for some constant $C \in (0, \infty)$ possibly depending on the law of X_0. A (somehow weak) result has been proved in [1]: assuming $\mathbf{E}(X_0^3 m^{X_0}) < \infty$,

$$F_\infty = \exp\Big(- \frac{1}{\eta^{1/2+o(1)}}\Big), \qquad \eta \to 0+ \, .$$

This confirms that the Derrida–Retaux system has a Berezinskii–Kosterlitz–Thouless-type phase transition of infinite order. The integrability assumption $\mathbf{E}(X_0^3 m^{X_0}) < \infty$ might look exotic, but it is optimal. [We believe that there should be a change-of-measures argument, and that the assumption is equivalent to saying that X_0 has a finite second moment under a new probability measure; however, we have not succeeded in making this idea into a rigorous argument.] In fact, it has also been proved in [1] that if $\mathbf{P}(X_0 = k) \sim c\, m^{-k} k^{-\alpha}$, $k \to \infty$, for some $2 < \alpha < 4$ and $c > 0$,[1] then

$$F_\infty = \exp\Big(- \frac{1}{\eta^{\nu + o(1)}}\Big), \qquad \eta \to 0+ , \tag{3}$$

where $\nu = \nu(\alpha) := \frac{1}{\alpha - 2}$. In other words, (2) predicts only a small part of universalities, under the assumption $\mathbf{E}(X_0^3 m^{X_0}) < \infty$, while other universality phenomena are described by (3). We expect many other universality results in the latter setting (for example, corresponding to those in [2] for an analogous continuous-time model); unfortunately, they are currently only on a heuristic level.

It is well-known that sum of i.i.d. random variables, after an appropriate normalization, converges to a Gaussian limiting law under the condition of finiteness of second moment, and to a stable limiting law under a weaker integrability condition. We say that the Derrida–Retaux system has a "finite variance" if $\mathbf{E}(X_0^3 m^{X_0}) < \infty$, and that it is a stable system if integrability condition holds for lower orders. In this paper, we are interested in the stable system when it is critical, i.e., when $(m-1)\mathbf{E}(X_0\, m^{X_0}) = \mathbf{E}(m^{X_0})$. [We are going to see in Sect. 2, quite easily, that this implies $(m-1)\mathbf{E}(X_n\, m^{X_n}) = \mathbf{E}(m^{X_n})$ for all $n \ge 0$.] We write $(Y_n,\ n \ge 0)$ instead of $(X_n,\ n \ge 0)$ in order to insist on criticality. From now on, we assume $(Y_n,\ n \ge 0)$ to be a Derrida–Retaux system satisfying $(m-1)\mathbf{E}(Y_0\, m^{Y_0}) = \mathbf{E}(m^{Y_0}) < \infty$, such that

$$\mathbf{P}(Y_0 = k) \sim c_0\, m^{-k} k^{-\alpha}, \qquad k \to \infty, \tag{4}$$

for some $2 < \alpha < 4$ and $c_0 > 0$. We intend to prove the following result.

Theorem 1 *Let* $(Y_n,\ n \ge 0)$ *be such that* $(m-1)\mathbf{E}(Y_0\, m^{Y_0}) = \mathbf{E}(m^{Y_0}) < \infty$. *Under assumption* (4)*, there exist constants* $c_2 \ge c_1 > 0$ *such that for all* $n \ge 1$*,*

$$c_1\, n^{\alpha-2} \le \prod_{i=0}^{n-1} [\mathbf{E}(m^{Y_i})]^{m-1} \le c_2\, n^{\alpha-2}.$$

When the system is of "finite-variance" (i.e., $\mathbf{E}(Y_0^3 m^{Y_0}) < \infty$), the analogue of Theorem 1 was known [3, 5], and has played an important role in the study of the asymptotics of $\mathbf{P}(Y_n > 0)$ and $\mathbf{E}(Y_n)$ in [4]. It would be tempting to believe that Theorem 1 could play an equally important role in the study of the same problems for the stable system.

Just like the usual random walk has a nice continuous-time analogue which is Brownian motion, the Derrida–Retaux system has analogues in continuous time (Derrida and Retaux [6], Hu et al. [13]), defined via appropriate integro-differential equations. For the continuous-time analogue of the stable Derrida–Retaux system,

[1]Notation: $a_k \sim b_k$, $k \to \infty$, means $\lim_{k\to\infty} \frac{a_k}{b_k} = 1$.

see [2]. These continuous-time models have been studied in depth in [6, 13] and [2], while most of the corresponding problems remain open for the original Derrida–Retaux system.

With the exception of the case $\alpha = 3$, Theorem 1 was already stated in Collet et al. [5]: its proof in case $2 < \alpha < 3$ was indicated, whereas the proof in case $3 < \alpha < 4$ was only summarized in a "very succinct account". By means of the notion of dominability (see the forthcoming Definition 1), we give a unified approach to the system in both situations, i.e., either it is stable (no need for discussions separately on the cases $2 < \alpha < 3$ and $3 < \alpha < 4$) or is of "finite variance". Concretely, in both situations, we use a truncating argument by considering a bounded random variable defined by

$$Z_0 = Z_0(M) := Y_0 \, \mathbf{1}_{\{Y_0 \le a(M)\}},$$

where $a(M) \in [1, \infty]$ can be possibly infinite (in which case there is no need for truncation), whose value depends on an integer parameter $M \ge 1$. Consider the Derrida–Retaux system $(Z_n, \ n \ge 0)$ whose initial distribution is given by Z_0.[2] We prove, in Theorem 4, that in both situations, it is possible to choose a convenient value of $a(M)$ such that the new system $(Z_n, \ n \ge 0)$ is dominable (in the sense of Definition 1), while it is possible to connect the moment generating functions of Y_n and Z_n. In Theorem 3, we give an upper bound for the moment generating function of any dominable system $(Z_n, \ n \ge 0)$. As such, a combined application of Theorems 4 and 3 will yield information for the moment generating function of the original Derrida–Retaux system, in both situations. In the stable case, it will yield Theorem 1, whereas in the case of "finite variance", under a stronger integrability assumption on the law of Y_0, it will give the following result:

Theorem 2 *Let* $(Y_n, \ n \ge 0)$ *be such that* $(m-1)\mathbf{E}(Y_0 \, m^{Y_0}) = \mathbf{E}(m^{Y_0})$. *If* $\mathbf{E}(s^{Y_0}) < \infty$ *for some* $s > m$*, then there exists a constant* $c_3 > 0$ *such that for all integers* $n \ge 1$ *and* $k \ge 1$*,*

$$\frac{\mathrm{d}^k}{\mathrm{d}u^k} \mathbf{E}(u^{Y_n})\Big|_{u=m} \le k! \, \mathrm{e}^{c_3 k} \, n^{k-1} \, . \tag{5}$$

In the "finite-variance" case $\mathbf{E}(Y_0^3 m^{Y_0}) < \infty$, (5) for $k \in \{1, 2, 3\}$ was known: the case $k = 1$ is simple because by criticality, $\mathbf{E}(Y_n m^{Y_n - 1}) = \frac{1}{m(m-1)} \mathbf{E}(m^{Y_n})$ which is bounded in n [3, 5], the case $k = 3$ was proved in [3], and the case $k = 2$, stated in [1], follows immediately from the cases $k = 1$ and $k = 3$ by means of the Cauchy–Schwarz inequality. More generally, if $\mathbf{E}(Y_0^\ell m^{Y_0}) < \infty$ for some integer $\ell \ge 1$, then for all $k \in [1, \ell] \cap \mathbb{Z}$, it is quite easy to prove [4] by induction in k,

[2]Strictly speaking, it is a *sequence* of Derrida–Retaux systems, indexed by M.

using the recursion (1), that there exists a constant $c > 0$ such that for all integer $n \geq 1$,

$$\frac{\mathrm{d}^k}{\mathrm{d}u^k}\,\mathbf{E}(u^{Y_n})\Big|_{u=m} \leq c\,n^{k-1}\,.$$

Theorem 2 gives information about the dependence in k of the constant c, under the integrability assumption $\mathbf{E}(s^{Y_0}) < \infty$ for some $s > m$.

The rest of the paper is organized as follows. In Sect. 2, we introduce the notion of dominable systems. Theorem 3, which gives an upper bound for the moment generating function of dominable systems, is the main technical result of the paper. The brief Sect. 3 is devoted to the proof of Theorem 2, obtained as a simple consequence of Theorem 3. In Sect. 4, for both "finite-variance" and stable systems, we construct a dominable system $(Z_n,\ n \geq 0)$ such that Z_0 is obtained from an appropriate truncation of Y_0. Finally, Theorem 1 is proved in Sect. 5, also as a consequence of Theorem 3.

2 Dominable Systems

We introduce the notion of dominable systems and prove a general upper bound for the moment generating function of such systems (Theorem 3). As before, we talk about the Derrida–Retaux system $(Z_n,\ n \geq 0)$, while it is, in fact, a sequence of Derrida–Retaux systems $(Z_n(M),\ n \geq 0)$ indexed by the integer-valued parameter M.

Definition 1 Let $\gamma > 0$. The system $(Z_n,\ n \geq 0)$ is said to be γ-dominable if for all sufficiently large integer M, say $M \geq M_0$, $Z_0 = Z_0(M)$ is bounded, and there exists a constant $\vartheta(M) \geq 1$ such that $M \mapsto \vartheta(M)$ is non-decreasing in $M \geq M_0$, and that [3]

$$\mathbf{E}(Z_0^k m^{Z_0}) \leq M^{k-3}(\vartheta(M) + k!), \qquad k \geq 3, \tag{6}$$

$$\vartheta(n \vee M) \prod_{i=0}^{n-1} [\mathbf{E}(m^{Z_i})]^{m-1} \leq \gamma\,(n \vee M)^2, \qquad n \geq 1. \tag{7}$$

Remark 1 Let $(Z_n,\ n \geq 0)$ be γ-dominable. By (7) and the trivial inequality $\mathbf{E}(m^{Z_i}) \geq 1$ ($\forall i \geq 0$), we have $\mathbf{E}(Z_n) \leq \mathbf{E}(m^{Z_n}) \leq [\gamma\,(n \vee M)^2]^{1/(m-1)}$, so the free energy $F_\infty := \lim_{n\to\infty} \frac{\mathbf{E}(Z_n)}{m^n}$ vanishes; by the criterion of Collet et al. [5] recalled in the introduction, the system $(Z_n,\ n \geq 0)$ is subcritical or critical for all $M \geq M_0$: we have $(m-1)\mathbf{E}(Z_0\, m^{Z_0}) \leq \mathbf{E}(m^{Z_0})$.

[3] In (6) and (7), $k \geq 3$ and $n \geq 1$ are integers. Notation: $a \vee b := \max\{a,\ b\}$.

Theorem 3 *Let $\gamma > 0$. Let $(Z_n,\ n \ge 0)$ be a γ-dominable system, and let $H_n(u) := \mathbf{E}(u^{Z_n})$. There exists a constant $c_4 \ge 1$, depending only on $(m,\ \gamma)$, such that for all integers $k \ge 3$, $n \ge 1$ and $M \ge M_0$,*

$$H_n^{(k)}(m) \le k!\, c_4^k\, (n \vee M)^{k-1}\,,$$

where $H_n^{(k)}(\cdot)$ stands for the k-th derivative of $H_n(\cdot)$.

Corollary 1 *Let $\gamma > 0$. Let $(Z_n,\ n \ge 0)$ be a γ-dominable system. There exists a constant $c_5 > 0$, depending only on $(m,\ \gamma)$, such that for $M \ge M_0$, $n \ge 1$ and $v := m + \frac{1}{2c_4(n \vee M)}$,*

$$\mathbf{E}(Z_n^2 v^{Z_n}) \le c_5[\vartheta(n \vee M)]^{1/2} \prod_{i=0}^{n-1} [\mathbf{E}(m^{Z_i})]^{(m-1)/2}; \tag{8}$$

in particular, we have, with $c_6 := \gamma^{1/2} c_5$,

$$\mathbf{E}(Z_n^2 v^{Z_n}) \le c_6(n \vee M). \tag{9}$$

The rest of the section is devoted to the proof of Theorem 3 and Corollary 1. We start by mentioning a general technique going back to Collet et al. [5]. Let $(X_n,\ n \ge 0)$ denote a Derrida–Retaux system satisfying $\mathbf{E}(X_0\, m^{X_0}) < \infty$. Let

$$G_n(u) := \mathbf{E}(u^{X_n}), \qquad n \ge 0.$$

The iteration formula (1) is equivalent to:

$$G_{n+1}(u) = \frac{1}{u}\, [G_n(u)]^m + (1 - \frac{1}{u})\, [G_n(0)]^m, \qquad n \ge 0. \tag{10}$$

A useful trick of Collet et al. [5] consists in observing that this yields

$$(u-1)u\, G'_{n+1}(u) - G_{n+1}(u) = [m(u-1)\, G'_n(u) - G_n(u)]\, [G_n(u)]^{m-1}\,.$$

In particular, taking $u = m$ yields that

$$G_{n+1}(m) - (m-1)m\, G'_{n+1}(m) = [G_n(m) - (m-1)m\, G'_n(m)]\, [G_n(m)]^{m-1}\,.$$

Iterating this formula gives that for $n \ge 1$,

$$G_n(m) - (m-1)m\, G'_n(m) = [G_0(m) - (m-1)m\, G'_0(m)] \prod_{i=0}^{n-1} [G_i(m)]^{m-1}\,. \tag{11}$$

In particular, (11) tells us that the sign of $\mathbf{E}(m^{X_n}) - (m-1)\mathbf{E}(X_n\, m^{X_n})$ remains identical for all $n \ge 0$: it either is always positive (meaning that the system is supercritical), or is always negative (subcritical), or vanishes identically (critical).

A couple of known results which we are going to use for the subcritical or critical system: if $(X_n,\, n \ge 0)$ is a Derrida–Retaux satisfying $\mathbf{E}(m^{X_0}) \le (m-1)\mathbf{E}(X_0\, m^{X_0}) < \infty$, then for all $n \ge 0$,

$$(m-1)\mathbf{E}(X_n\, m^{X_n}) \le \mathbf{E}(m^{X_n}) \le m^{1/(m-1)}, \qquad n \ge 0, \tag{12}$$

$$\prod_{i=0}^{n-1} [\mathbf{E}(m^{X_i})]^{m-1} \le c_7\, n^2, \qquad n \ge 1, \tag{13}$$

where $c_7 > 0$ is a constant depending on m and on the law of X_0. See [1, (3.11)] for the second inequality in (12) (proved in [1] for critical systems, and the same proof valid for subcritical systems as well), and [3, Proposition 1] for (13).

The proof of Theorem 3 relies on the following preliminary result. Recall that $H_n(u) := \mathbf{E}(u^{Z_n})$.

Lemma 1 *Let $(Z_n,\, n \ge 0)$ be a γ-dominable system for some $\gamma > 0$, in the sense of Definition 1. There exist constants $c_8 > 0$ and $c_9 > 0$, depending only on m, such that for $M \ge M_0$ and all integer $n \ge 1$,*

$$H_n''(m) \le c_8\, [\vartheta(M)]^{1/2} \prod_{i=0}^{n-1} H_i(m)^{(m-1)/2}, \tag{14}$$

$$H_n'''(m) \le c_9\, \vartheta(M) \prod_{i=0}^{n-1} H_i(m)^{m-1}. \tag{15}$$

Proof Let

$$D_n(u) := (m-1)u^3 H_n'''(u) + (4m-5)u^2 H_n''(u) + 2(m-2)u H_n'(u), \qquad n \ge 0.$$

Recall from Remark 1 that $m(m-1)H_0'(m) \le H_0(m)$. By Chen et al. [3, Equation (19)], this yields $D_{n+1}(m) \le D_n(m) H_n(m)^{m-1}$ for all $n \ge 0$. [In [3], it was proved that if $m(m-1)H_0'(m) = H_0(m)$, then $D_{n+1}(m) = D_n(m)H_n(m)^{m-1}$, but the proof is valid for inequalities in place of equalities.] Accordingly, for all $n \ge 1$,

$$D_n(m) \le D_0(m) \prod_{i=0}^{n-1} H_i(m)^{m-1}.$$

By definition, $D_n(m) \ge m^3(m-1)H_n'''(m)$, so

$$H_n'''(m) \le \frac{D_0(m)}{m^3(m-1)} \prod_{i=0}^{n-1} H_i(m)^{m-1}\,. \tag{16}$$

For $j \geq 0$, writing $Z_j^3 m^{Z_j} \leq \frac{9m^3}{2} Z_j(Z_j-1)(Z_j-2)m^{Z_j-3}\, \mathbf{1}_{\{Z_j \geq 3\}} + 8m^2$, and $Z_j^\ell \leq Z_j^3$ for $\ell \in \{1, 2, 3\}$, we obtain

$$\max_{\ell \in \{1,2,3\}} \mathbf{E}(Z_j^\ell m^{Z_j}) \leq \frac{9m^3}{2}\, H_j'''(m) + 8m^2 . \tag{17}$$

Considering $j = 0$ and $\ell \in \{1, 2\}$, this gives $\max\{H_0'(m), H_0''(m)\} \leq \frac{9m^3}{2} H_0'''(m) + 8m^2$. Consequently, with $c_{10} := m^3(m-1)$, $c_{11} := m^2(4m-5)$ and $c_{12} := 2m(m-2)$,

$$D_0(m) = c_{10}\, H_0'''(m) + c_{11}\, H_0''(m) + c_{12}\, H_0'(m) \leq c_{13}\, H_0'''(m) + c_{14}\,,$$

where $c_{13} := c_{10} + \frac{9m^3}{2}(c_{11} + c_{12})$ and $c_{14} := 8m^2(c_{11} + c_{12})$. Since $H_0'''(m) \leq \vartheta(M) + 6 \leq 7\vartheta(M)$ by assumption (6) (applied to $k = 3$; recalling that $\vartheta(M) \geq 1$), we get $D_0(m) \leq 7c_{13}\, \vartheta(M) + c_{14} \leq c_{15}\, \vartheta(M)$ with $c_{15} := 7c_{13} + c_{14}$. Going back to (16), we have

$$H_n'''(m) \leq \frac{c_{15}}{m^3(m-1)}\, \vartheta(M) \prod_{i=0}^{n-1} H_i(m)^{m-1},$$

proving (15) with $c_9 := \frac{c_{15}}{m^3(m-1)}$.

It remains to prove (14). By (17) (applied to $j = n$),

$$\mathbf{E}(Z_n^3 m^{Z_n}) \leq \frac{9m^3}{2}\, H_n'''(m) + 8m^2\,,$$

whereas by (12), $\mathbf{E}(Z_n m^{Z_n}) \leq \frac{m^{1/(m-1)}}{m-1} =: c_{16}$, it follows from the Cauchy–Schwarz inequality that

$$\mathbf{E}(Z_n^2 m^{Z_n}) \leq c_{16}^{1/2} \Big(\frac{9m^3}{2}\, H_n'''(m) + 8m^2\Big)^{1/2}.$$

Recall from (15), which we have just proved, that $H_n'''(m) \leq c_9 \vartheta(M) \prod_{i=0}^{n-1} H_i(m)^{m-1}$. Writing $8m^2 \leq 8m^2 \vartheta(M) \prod_{i=0}^{n-1} H_i(m)^{m-1}$ (because $\vartheta(M) \geq 1$ and $H_i(m) \geq 1$), this yields

$$\mathbf{E}(Z_n^2 m^{Z_n}) \leq c_{17}\, [\vartheta(M)]^{1/2} \prod_{i=0}^{n-1} H_i(m)^{(m-1)/2},$$

with $c_{17} := c_{16}^{1/2}(\frac{9m^3}{2}\, c_9 + 8m^2)^{1/2}$. Since $H_n''(m) \leq \mathbf{E}(Z_n^2 m^{Z_n})$, we obtain (14) with $c_8 := c_{17}$. □

Remark 2 We often use the following inequalities for dominable systems:

$$H_n''(m) \le c_8 \, [\vartheta(n \vee M)]^{1/2} \prod_{i=0}^{n-1} H_i(m)^{(m-1)/2}, \tag{18}$$

$$H_n'''(m) \le c_9 \, \vartheta(n \vee M) \prod_{i=0}^{n-1} H_i(m)^{m-1}, \tag{19}$$

They are immediate consequences of Lemma 1 and the monotonicity of $M \mapsto \vartheta(M)$.

We also need an elementary inequality.

Lemma 2 *Let $\ell \ge 4$ be an integer, and let*

$$B_\ell := \{\mathbf{u} := (u_1, \dots, u_m) \in ([0, \ \ell - 1] \cap \mathbb{Z})^m : \ u_1 + \dots + u_m = \ell\}. \tag{20}$$

There exists a constant $c_{18} > 0$, depending only on m, such that for all $y \ge 3m$,[4]

$$\sum_{\mathbf{u}:=(u_1,\dots,u_m)\in B_\ell} y^{(\ell-\eta(\mathbf{u})-2)^+} \prod_{i:\, u_i \ge 3} \frac{1}{u_i(u_i-1)} \le c_{18} \, \frac{y^{\ell-4}}{\ell^2}, \tag{21}$$

where $a^+ := \max\{a, \ 0\}$ as before, and

$$\eta(\mathbf{u}) := \sum_{i=1}^m \mathbf{1}_{\{u_i \ge 1\}} \ge 2. \tag{22}$$

Proof The sum over $\mathbf{u} := (u_1, \dots, u_m) \in B_\ell$ satisfying $u_{\max} := \max_{1 \le i \le m} u_i \le 2$ is very simple: in this case, $\ell \le 2m$; since $\eta(\mathbf{u}) \ge 2$, we have $(\ell - \eta(\mathbf{u}) - 2)^+ \le \ell - 4$. The number of such $\mathbf{u}$ being smaller than 3^m, we get

$$\sum_{\mathbf{u}\in B_\ell:\, u_{\max}\le 2} y^{(\ell-\eta(\mathbf{u})-2)^+} \prod_{i:\, u_i \ge 3} \frac{1}{u_i(u_i-1)} \le 3^m y^{\ell-4} \le c_{19} \, \frac{y^{\ell-4}}{\ell^2},$$

with $c_{19} := 3^m (2m)^2$. [Notation: $\sum_\varnothing := 0$.]

Let LHS $_{(21)}$ denote the expression on the left-hand side of (21). Then

$$\text{LHS}_{(21)} \le c_{19} \, \frac{y^{\ell-4}}{\ell^2} + \sum_{j=2}^{\ell \wedge m} \binom{m}{j} j! \, y^{(\ell-j-2)^+} \sum_{(u_1,\dots,u_j)} \prod_{i:\, 1 \le i \le j, \, u_i \ge 3} \frac{1}{u_i(u_i-1)},$$

[4] Strictly speaking, we should write $\prod_{i:\,1\le i\le m,\, u_i\ge 3} \frac{1}{u_i(u_i-1)}$ for $\prod_{i:\, u_i\ge 3} \frac{1}{u_i(u_i-1)}$. Notation: $\prod_\varnothing := 1$.

where, on the right-hand side, $\sum_{(u_1,\ldots,u_j)}$ sums over all $(u_1,\ldots,u_j) \in \mathbb{Z}_+^j$ with $1 \le u_1 \le u_2 \le \cdots \le u_j$ such that $u_1 + \cdots + u_j = \ell$ and that $u_j \ge 3$. Note that $u_j \ge 3$ implies $j \le \ell - 2$, thus $(\ell - j - 2)^+ = \ell - j - 2$. Moreover, we have $u_j \ge \frac{\ell}{j} \ge \frac{\ell}{m}$, thus $u_j(u_j - 1) \ge \frac{1}{2}u_j^2 \ge \frac{\ell^2}{2m^2}$. Consequently, $\prod_{i \le j:\, u_i \ge 3} \frac{1}{u_i(u_i-1)}$ is bounded by $\frac{2m^2}{\ell^2} \prod_{i \le j-1:\, u_i \ge 3} \frac{1}{u_i(u_i-1)}$. This leads to (using $\binom{m}{j} j! \le m^j$):

$$\begin{aligned}
\text{LHS}_{(21)} &\le c_{19}\, \frac{y^{\ell-4}}{\ell^2} + \frac{2m^2}{\ell^2} \sum_{j=2}^{(\ell-2)\wedge m} m^j y^{\ell-j-2} \sum_{(u_1,\ldots,u_j)} \prod_{i \le j-1:\, u_i \ge 3} \frac{1}{u_i(u_i-1)} \\
&\le c_{19}\, \frac{y^{\ell-4}}{\ell^2} + \frac{2m^2}{\ell^2} \sum_{j=2}^{(\ell-2)\wedge m} m^j y^{\ell-j-2} \prod_{i=1}^{j-1} \Big(1 + \sum_{u=3}^{\infty} \frac{1}{u(u-1)}\Big).
\end{aligned}$$

Of course, $\sum_{u=3}^{\infty} \frac{1}{u(u-1)} = \frac{1}{2}$; also, we bound $\sum_{j=2}^{(\ell-2)\wedge m}$ by $\sum_{j=2}^{\infty}$. This yields that

$$\text{LHS}_{(21)} \le c_{19}\, \frac{y^{\ell-4}}{\ell^2} + \frac{2m^2}{\ell^2} \sum_{j=2}^{\infty} m^j y^{\ell-j-2} (\frac{3}{2})^{j-1}.$$

On the right-hand side, write $\sum_{j=2}^{\infty} m^j y^{\ell-j-2} (\frac{3}{2})^{j-1} = m^2 y^{\ell-4} \sum_{j=2}^{\infty} (\frac{m}{y})^{j-2} (\frac{3}{2})^{j-1}$; in view of our choice $y \ge 3m$, this is bounded by $m^2 y^{\ell-4} \sum_{j=2}^{\infty} (\frac{1}{3})^{j-2} (\frac{3}{2})^{j-1} = 3m^2 y^{\ell-4}$. As a consequence,

$$\text{LHS}_{(21)} \le c_{19}\, \frac{y^{\ell-4}}{\ell^2} + \frac{6m^4 y^{\ell-4}}{\ell^2},$$

yielding (21) with $c_{18} := c_{19} + 6m^4$. □

We have all the ingredients for the proof of Theorem 3.

Proof of Theorem 3 Let $\gamma > 0$ and let $(Z_n, n \ge 0)$ be a γ-dominable system. Write $H_n(u) := \mathbf{E}(u^{Z_n})$ as before. Recall $\vartheta(M) \ge 1$ (for all $M \ge M_0$) from (6) and (7). Write, for brevity,

$$\begin{aligned}
Q_n &:= \prod_{j=0}^{n-1} H_j(m)^{m-1} = \prod_{j=0}^{n-1} [\mathbf{E}(m^{Z_j})]^{m-1}, \\
M_n &:= n \vee M, \\
\vartheta_n &:= \vartheta(M_n) = \vartheta(n \vee M), \qquad n \ge 1,\ M \ge M_0.
\end{aligned}$$

By assumption (7),

$$Q_n \le \vartheta_n Q_n \le \gamma\, M_n^2, \qquad n \ge 1. \tag{23}$$

We claim that for all integer $k \geq 3$,

$$H_n^{(k)}(m) \leq c_4^{k-1}\,(k-2)!\,M_n^{k-3}\,\vartheta_n Q_n, \qquad n \geq 1, \tag{24}$$

where $c_4 := \max\{4 + \frac{c_{18}c_{20}}{m},\ c_9^{1/2},\ \gamma\}$, with $c_{20} := m^{m/(m-1)}(c_8^m \vee 1)(\gamma^{3m/2} \vee 1)$. Since $\vartheta_n Q_n \leq \gamma\, M_n^2$ (see (23)), (24) will imply Theorem 3.

It remains to prove (24), which we do by induction in $k \geq 3$.

By (19), $H_n'''(m) \leq c_9\,\vartheta_n Q_n$ for $n \geq 1$. So (24) holds for $k = 3$ since $c_9 \leq c_4^2$.

Let $\ell \geq 4$ be an integer. Suppose (24) holds for all $k \in \{3, 4, \ldots, \ell - 1\}$. We need to prove (24) for $k = \ell$.

We first prove that the induction assumption yields that for $n \geq 1$ and $\mathbf{u} := (u_1, \ldots, u_m) \in B_\ell$ (defined in (20)), we have, with $c_{20} := m^{m/(m-1)}(c_8^m \vee 1)(\gamma^{3m/2} \vee 1)$ as before,

$$\prod_{i=1}^{m} H_n^{(u_i)}(m) \leq c_{20}c_4^{\ell-2}M_n^{(\ell-\eta(\mathbf{u})-2)^+}\vartheta_n Q_n \prod_{i:\,u_i\geq 3}(u_i-2)!\,, \tag{25}$$

where $\eta(\mathbf{u}) := \sum_{i=1}^m \mathbf{1}_{\{u_i \geq 1\}}$ is as in (22).

To check (25), let $n \geq 1$ and $\mathbf{u} \in B_\ell$. Since $H_n(m) \leq m^{1/(m-1)} =: c_{21}$ (see (12)) and $H_n'(m) \leq \frac{1}{m-1}\,\mathbf{E}(m^{Z_n}) \leq \frac{c_{21}}{m-1} \leq c_{21}$, we have

$$\prod_{i=1}^{m} H_n^{(u_i)}(m) \leq c_{21}^m H_n''(m)^{\lambda_2(\mathbf{u})} \prod_{i:\,u_i\geq 3} H_n^{(u_i)}(m)\,,$$

where $\lambda_2(\mathbf{u}) := \sum_{i=1}^m \mathbf{1}_{\{u_i = 2\}}$. By (18), we have $H_n''(m) \leq c_8\,\vartheta_n^{1/2}\,Q_n^{1/2}$; thus with $c_{22} := c_{21}^m \max\{c_8^m,\ 1\}$,

$$\prod_{i=1}^{m} H_n^{(u_i)}(m) \leq c_{22}\,(\vartheta_n Q_n)^{\lambda_2(\mathbf{u})/2} \prod_{i:\,u_i\geq 3} H_n^{(u_i)}(m)\,.$$

By the induction assumption in (24), $H_n^{(u_i)}(m) \leq c_4^{u_i-1}(u_i-2)!\,M_n^{u_i-3}\,\vartheta_n Q_n$ if $u_i \geq 3$. As such, we have

$$\begin{aligned}\prod_{i=1}^{m} H_n^{(u_i)}(m) &\leq c_{22}\,(\vartheta_n Q_n)^{\lambda_2(\mathbf{u})/2} \prod_{i:\,u_i\geq 3}\Big(c_4^{u_i-1}(u_i-2)!\,M_n^{u_i-3}\vartheta_n Q_n\Big)\\ &= c_{22}\,(\vartheta_n Q_n)^{\lambda_2(\mathbf{u})/2}\,(c_4 M_n)^{\sum_{i:\,u_i\geq 3}(u_i-1)}(M_n^{-2}\vartheta_n Q_n)^{\eta_3(\mathbf{u})} \prod_{i:\,u_i\geq 3}(u_i-2)!\,,\end{aligned}$$

where $\eta_3(\mathbf{u}) := \sum_{i=1}^m \mathbf{1}_{\{u_i \geq 3\}}$. Note that $\sum_{i:\,u_i\geq 3}(u_i - 1) = \sum_{i=1}^m (u_i - 1)^+ - \lambda_2(\mathbf{u}) = \ell - \eta(\mathbf{u}) - \lambda_2(\mathbf{u}) \leq \ell - 2$. So $c_4^{\sum_{i:\,u_i\geq 3}(u_i-1)} \leq c_4^{\ell-2}$ (using $c_4 > 1$). This leads to:

$$\prod_{i=1}^{m} H_n^{(u_i)}(m) \le c_{22}\, c_4^{\ell-2}\, (\vartheta_n Q_n)^{\frac{\lambda_2(\mathbf{u})}{2}+\eta_3(\mathbf{u})}\, M_n^{\ell-\eta(\mathbf{u})-\lambda_2(\mathbf{u})-2\eta_3(\mathbf{u})} \prod_{i:\, u_i\ge 3} (u_i-2)!$$

$$= c_{22}\, c_4^{\ell-2}\, (\vartheta_n Q_n)^{\frac{\lambda_2(\mathbf{u})}{2}+\eta_3(\mathbf{u})-1}\, M_n^{\ell-\eta(\mathbf{u})-\lambda_2(\mathbf{u})-2\eta_3(\mathbf{u})}\, \vartheta_n Q_n \prod_{i:\, u_i\ge 3} (u_i-2)!\, .$$

Assume for the moment $\frac{\lambda_2(\mathbf{u})}{2} + \eta_3(\mathbf{u}) \ge 1$. By (23), $\vartheta_n Q_n \le \gamma\, M_n^2$, so we have $(\vartheta_n Q_n)^{\frac{\lambda_2(\mathbf{u})}{2}+\eta_3(\mathbf{u})-1} \le c_{23}\, M_n^{\lambda_2(\mathbf{u})+2\eta_3(\mathbf{u})-2}$, with $c_{23} := \max\{\gamma^{3m/2},\ 1\}$. Since $c_{20} = c_{22}c_{23}$, we get

$$\prod_{i=1}^{m} H_n^{(u_i)}(m) \le c_{20}\, c_4^{\ell-2}\, M_n^{\ell-\eta(\mathbf{u})-2}\, \vartheta_n Q_n \prod_{i:\, u_i\ge 3} (u_i-2)!,$$

yielding (25). If, on the other hand, $\frac{\lambda_2(\mathbf{u})}{2} + \eta_3(\mathbf{u}) < 1$, then $\lambda_2(\mathbf{u}) \le 1$ and $\eta_3(\mathbf{u}) = 0$, i.e., $\max_{1\le i\le m} u_i \le 2$. This time, $\ell - \eta(\mathbf{u}) = \sum_{i=1}^m (u_i-1)^+ \le 1$. The situation is very simple if we look at $\prod_{i=1}^m H_n^{(u_i)}(m)$ directly: at most one term among $H_n^{(u_i)}(m)$ is $H_n''(m)$ (which is bounded by $c_8\, \vartheta_n^{1/2}\, Q_n^{1/2}$ as we have seen in (18)), while all the rest is either $H_n(m)$ (which is bounded by $m^{1/(m-1)} =: c_{21}$) or $H_n'(m)$ (which is bounded by 1 because by (12), $H_n'(m) = \mathbf{E}(Z_n\, m^{Z_n-1}) \le \frac{m^{1/(m-1)}}{m(m-1)} \le 1$). Hence

$$\prod_{i=1}^{m} H_n^{(u_i)}(m) \le c_{21}^m c_8\, \vartheta_n^{1/2}\, Q_n^{1/2} \le c_{24}\, \vartheta_n Q_n\, ,$$

with $c_{24} := c_{21}^m c_8$. [We have used $\vartheta_n Q_n \ge 1$.] This again gives (25) because $c_{24} \le c_{20}$ and $c_4 > 1$.

Now that (25) is proved, it is painless to complete the proof of Theorem 3. Indeed, by (10),

$$s H_{n+1}(s) = [H_n(s)]^m + (s-1)[H_n(0)]^m\, .$$

On both sides, we differentiate ℓ times with respect to s (recalling that $\ell \ge 4$, so the last term on the right-hand side, being affine in s, makes no contribution to the derivatives), and apply the general Leibniz rule to the first term on the right-hand side; this leads to:

$$
\begin{aligned}
sH_{n+1}^{(\ell)}(s) + \ell H_{n+1}^{(\ell-1)}(s) &= \sum_{(u_1,\dots,u_m)\in\mathbb{Z}_+^m:\, u_1+\cdots+u_m=\ell} \frac{\ell!}{u_1!\cdots u_m!} \prod_{i=1}^m H_n^{(u_i)}(s) \\
&= mH_n^{(\ell)}(s)H_n(s)^{m-1} + \sum_{\mathbf{u}\in B_\ell} \frac{\ell!}{u_1!\cdots u_m!} \prod_{i=1}^m H_n^{(u_i)}(s)\,.
\end{aligned}
$$

Note that the expression on the left-hand side is at least $sH_{n+1}^{(\ell)}(s)$. We take $s = m$ to see that

$$
H_{n+1}^{(\ell)}(m) - H_n^{(\ell)}(m)H_n(m)^{m-1} \le \frac{1}{m} \sum_{\mathbf{u}\in B_\ell} \frac{\ell!}{u_1!\cdots u_m!} \prod_{i=1}^m H_n^{(u_i)}(m).
$$

By (25), we have $\prod_{i=1}^m H_n^{(u_i)}(m) \le c_{20} c_4^{\ell-2} M_n^{(\ell-\eta(\mathbf{u})-2)^+} \vartheta_n Q_n \prod_{i:\,u_i\ge 3}(u_i-2)!$, where $\eta(\mathbf{u}) := \sum_{i=1}^m \mathbf{1}_{\{u_i\ge 1\}}$ is as in (22). Hence

$$
H_{n+1}^{(\ell)}(m) - H_n^{(\ell)}(m)H_n(m)^{m-1} \le \frac{c_{20}}{m} c_4^{\ell-2} \ell!\, \vartheta_n Q_n \sum_{\mathbf{u}\in B_\ell} M_n^{(\ell-\eta(\mathbf{u})-2)^+} \prod_{i:\,u_i\ge 3} \frac{1}{u_i(u_i-1)}\,,
$$

which, in view of Lemma 2 (applied to $y := M_n$), yields that, for $n \ge 1$,

$$
H_{n+1}^{(\ell)}(m) \le H_n^{(\ell)}(m)H_n(m)^{m-1} + c_{18}\frac{c_{20}}{m} c_4^{\ell-2}\,(\ell-2)!\, M_n^{\ell-4} \vartheta_n Q_n.
$$

Recall that $Q_n := \prod_{j=0}^{n-1} H_j(m)^{m-1}$. Iterating this inequality, and by means of the monotonicity of $n \mapsto M_n^{\ell-4}\vartheta_n Q_n$, we get

$$
\begin{aligned}
H_n^{(\ell)}(m) &\le H_0^{(\ell)}(m) Q_n + \sum_{j=0}^{n-1} \frac{c_{18}c_{20}}{m} c_4^{\ell-2} (\ell-2)!\, M_n^{\ell-4}\, \vartheta_n Q_n \\
&= \Big(\frac{H_0^{(\ell)}(m)}{\vartheta_n} + n\frac{c_{18}c_{20}}{m} c_4^{\ell-2}(\ell-2)!\, M_n^{\ell-4} \Big) \vartheta_n Q_n\,.
\end{aligned}
$$

We use $n \le M_n$ so that $n\, M_n^{\ell-4} \le M_n^{\ell-3}$. On the other hand, $H_0^{(\ell)}(m) \le M^{\ell-3}(\vartheta(M) + \ell!)$ (by assumption (6)), which is bounded by $M_n^{\ell-3}\vartheta_n\,(1+\ell!)$. Thus

$$
\begin{aligned}
H_n^{(\ell)}(m) &\le \Big(1 + \ell! + \frac{c_{18}c_{20}}{m} c_4^{\ell-2} (\ell-2)!\Big) M_n^{\ell-3}\, \vartheta_n Q_n \\
&\le (1 + \ell(\ell-1) + \frac{c_{18}c_{20}}{m} c_4^{\ell-2})\, (\ell-2)!\, M_n^{\ell-3}\, \vartheta_n Q_n\,.
\end{aligned}
$$

Since $\ell \geq 4$, we have $1 + \ell(\ell - 1) \leq \ell^2 \leq 2^\ell \leq 4\,c_4^{\ell-2}$ (because $c_4 \geq 2$), so $1 + \ell(\ell - 1) + \frac{c_{18}c_{20}}{m} c_4^{\ell-2} \leq 4\,c_4^{\ell-2} + \frac{c_{18}c_{20}}{m} c_4^{\ell-2} \leq c_4^{\ell-1}$ by means of the fact that $c_4 := \max\{4 + \frac{c_{18}c_{20}}{m},\ c_9^{1/2},\ \gamma\}$. Consequently,

$$H_n^{(\ell)}(m) \leq c_4^{\ell-1}(\ell - 2)!\, M_n^{\ell-3}\, \vartheta_n Q_n\,,$$

implying (24) for $k = \ell$, and completing the proof of Theorem 3. □

Proof of Corollary 1. Only (8) needs proving because (9) will follow immediately from (8) and assumption (7).

Let $n \geq 1$ and $s \in [m,\ m + \frac{1}{2c_4 M_n}]$, where we keep using the notation

$$M_n := n \vee M\,.$$

Write $H_n(u) := \mathbf{E}(u^{Z_n})$ as before. Then

$$\mathbf{E}(Z_n^2 s^{Z_n}) = s H_n'(s) + s^2 H_n''(s)\,.$$

Since $u \mapsto H_n''(u)$ is non-decreasing, we have $H_n'(s) \leq H_n'(m) + (s - m)H_n''(s)$; hence

$$\mathbf{E}(Z_n^2 s^{Z_n}) \leq s H_n'(m) + s(s - m)H_n''(s) + s^2 H_n''(s) = s H_n'(m) + s(2s - m)H_n''(s)\,.$$

On the right-hand side, we use $H_n'(m) \leq 1$ (which has already been observed as a consequence of (12)), and $m \leq s \leq m + 1$ (so $s(2s - m) \leq (m + 1)(m + 2)$), to see that

$$\mathbf{E}(Z_n^2 s^{Z_n}) \leq m + 1 + (m + 1)(m + 2)H_n''(m + \frac{1}{2c_4 M_n})\,.$$

To bound $H_n''(m + \frac{1}{2c_4 M_n})$, we recall that Z_n is bounded for each n (which is a consequence of the boundedness of Z_0), so by Taylor expansion,

$$H_n''(m + \frac{1}{2c_4 M_n}) - H_n''(m) = \sum_{k=3}^{\infty} \frac{(2c_4 M_n)^{-(k-2)}}{(k - 2)!} H_n^{(k)}(m)\,.$$

By (24), we have $H_n^{(k)}(m) \leq c_4^{k-1}\,(k - 2)!\, M_n^{k-3}\, \vartheta_n Q_n$ (for $k \geq 3$). Hence

$$H_n''(m+\frac{1}{2c_4 M_n})-H_n''(m) \leq \sum_{k=3}^{\infty} \frac{(2c_4 M_n)^{-(k-2)}}{(k - 2)!} c_4^{k-1}\,(k-2)!\, M_n^{k-3}\, \vartheta_n Q_n = c_4\, \frac{\vartheta_n Q_n}{M_n}\,.$$

As such, we arrive at:

$$\mathbf{E}(Z_n^2 s^{Z_n}) \le m + 1 + (m+1)(m+2)\Big(H_n''(m) + c_4\, \frac{\vartheta_n Q_n}{M_n}\Big).$$

By (18), we have $H_n''(m) \le c_8\, \vartheta_n^{1/2} Q_n^{1/2}$, whereas according to assumption (7), we have $M_n \ge \frac{1}{\gamma^{1/2}}\, \vartheta_n^{1/2} Q_n^{1/2}$, this readily yields (8) (recalling that $\vartheta_n Q_n \ge 1$). Corollary 1 is proved. □

3 Proof of Theorem 2

Let $(Y_n,\, n \ge 0)$ be a critical system such that $\mathbf{E}(s^{Y_0}) < \infty$ for some $s > m$. We claim that with $c_{25} := (\sup_{x>0} x\, (\frac{s}{m})^{-x/\mathrm{e}}) \vee 1 \in [1,\, \infty)$, we have, for all integer $k \ge 1$,

$$\mathbf{E}(Y_0^k m^{Y_0}) \le c_{25}^k\, \mathbf{E}(s^{Y_0})\, k!\,. \tag{26}$$

Indeed, by definition,

$$x \le c_{25}\, (\frac{s}{m})^{x/\mathrm{e}}\,,$$

for all $x > 0$. Taking to the power k on both sides and with $x := \frac{\mathrm{e}\,\ell}{k}$, we see that for all integers $k \ge 1$ and $\ell \ge 1$,

$$(\frac{\mathrm{e}\,\ell}{k})^k \le c_{25}^k\, (\frac{s}{m})^{\ell}\,.$$

Since $k! \ge (\frac{k}{\mathrm{e}})^k$ by Stirling's formula, this yields $\ell^k m^{\ell} \le c_{25}^k\, s^{\ell}\, k!$, from which (26) follows.

Let $L \ge 1$ be an integer, and let $Z_0 = Z_0(M,\, L) := Y_0\, \mathbf{1}_{\{Y_0 \le L\}}$. Then Z_0 is bounded, and does not depend on M, though we still treat it as indexed by M. Let us check that assumptions (6) and (7) in Definition 1 are satisfied.

Since $Z_0 \le Y_0$, it follows from (26) that $\mathbf{E}(Z_0^k m^{Z_0}) \le c_{26} c_{25}^k\, k!$ for $k \ge 1$, where $c_{26} = c_{26}(s) := \mathbf{E}(s^{Y_0}) \ge 1$. Let $M \ge c_{26} c_{25}^4$. Then $M^{k-3} \ge (c_{26} c_{25}^3)^{k-3} c_{25}^{k-3} \ge c_{26} c_{25}^3 c_{25}^{k-3} = c_{26} c_{25}^k$ for all integer $k \ge 4$, so $c_{26} c_{25}^k\, k! \le M^{k-3} k!$ for $k \ge 4$. For $k = 3$, we have $c_{26} c_{25}^3\, 3! \le c_{27} + 3!$ with $c_{27} := 6(c_{26} c_{25}^3 - 1) \vee 1 \in [1,\, \infty)$. As such, we see that for all integers $k \ge 3$ and $M \ge c_{26} c_{25}^4$,

$$\mathbf{E}(Z_0^k m^{Z_0}) \le M^{k-3}(c_{27} + k!);$$

in words, assumption (6) is satisfied with $\vartheta(M) := c_{27}$ for all $M \ge M_0 := \lceil c_{26} c_{25}^4 \rceil$.

Assumption (7) is easily seen to be satisfied: by (13), $\prod_{i=0}^{n-1}[\mathbf{E}(m^{Y_i})]^{m-1} \le c_7 n^2$ for all $n \ge 1$, so (7) holds with $\gamma := c_7 c_{27}$.

So we are entitled to apply Theorem 3 to see that for $k \ge 3$, $n \ge 1$ and $M \ge M_0 := \lceil c_{26} c_{25}^4 \rceil$,

$$\mathbf{E}[Z_n(Z_n-1)\cdots(Z_n-k+1)\, m^{Z_n-k}] \le k!\, c_4^k\, (n \vee M)^{k-1}\,.$$

We take $M := M_0$. Recall that $Z_0 = Z_0(M_0,\, L) := Y_0\, \mathbf{1}_{\{Y_0 \le L\}}$. Letting $L \to \infty$, and applying the monotone convergence theorem, we get, for $k \ge 3$ and $n \ge 1$,

$$\mathbf{E}[Y_n(Y_n-1)\cdots(Y_n-k+1)\, m^{Y_n-k}] \le k!\, c_4^k\, (n \vee M_0)^{k-1}\,.$$

This implies the desired inequality for $k \ge 3$. The case $k = 1$ has already been implicitly treated in the proof of Corollary 1 in Sect. 2: $\mathbf{E}(Y_n\, m^{Y_n-1}) = \frac{\mathbf{E}(m^{Y_n})}{m(m-1)} \le \frac{m^{1/(m-1)}}{m(m-1)}$ (see (12)). The case $k = 2$ follows from the cases $k = 1$ and $k = 3$ by the Cauchy–Schwarz inequality. □

4 Truncating the Critical System

Let $(Y_n,\, n \ge 0)$ be a Derrida–Retaux system satisfying $\mathbf{E}(m^{Y_0}) = (m-1)\mathbf{E}(Y_0\, m^{Y_0}) < \infty$ (so the system is critical). Recall that the system $(Y_n,\, n \ge 0)$ is of "finite variance" if $\mathbf{E}(Y_0^3\, m^{Y_0}) < \infty$, and is stable if $\mathbf{P}(Y_0 = j) \sim c_0\, m^{-j}\, j^{-\alpha}$, $j \to \infty$, for some $c_0 > 0$ and $2 < \alpha < 4$ as in (4). The following theorem tells that in either case, we can define

$$Z_0 := Y_0\, \mathbf{1}_{\{Y_0 \le a(M)\}},$$

for some appropriate $a(M) \in [1,\, \infty]$ such that $(Z_n,\, n \ge 0)$ is dominable in the sense of Definition 1; the values of $a(M)$ and $\vartheta(M)$ (as defined in Definition 1) are also given as they are often useful in the applications.

Theorem 4 *Let* $(Y_n,\, n \ge 0)$ *be a Derrida–Retaux system satisfying* $\mathbf{E}(m^{Y_0}) = (m-1)\mathbf{E}(Y_0\, m^{Y_0}) < \infty$. *If it is either of "finite variance" or stable, then there is a dominable system* $(Z_n,\, n \ge 0)$ *such that* $Z_0 \le Y_0$ a.s. *More precisely,*

(i) *if the system is of "finite variance", we can choose* $Z_0 := Y_0\, \mathbf{1}_{\{Y_0 \le M\zeta(M)\}}$ *where* $\zeta(M) := -\log \mathbf{E}(Y_0^3 m^{Y_0}\, \mathbf{1}_{\{Y_0 > M\}}) \le \infty$,[5] *with* $\vartheta(M) := \max\{\mathbf{E}(Y_0^3 m^{Y_0}),\, 1\}$;
(ii) *if the system is stable, we can choose* $Z_0 := Y_0\, \mathbf{1}_{\{Y_0 \le M\}}$, *with* $\vartheta(M) := c_{30}\, M^{4-\alpha}$, *where* c_{30} *is the constant in (31) below.*

[5] So $\lim_{M\to\infty} \zeta(M) = \infty$ by the "finite variance" assumption.

For the sake of clarity, the two situations ("finite variance", stable) are discussed in distinct parts.

4.1 Proof of Theorem 4: The "Finite Variance" Case

Assume $\mathbf{E}(m^{Y_0}) = (m-1)\mathbf{E}(Y_0\, m^{Y_0}) < \infty$ and $\mathbf{E}(Y_0^3\, m^{Y_0}) < \infty$. Let

$$\zeta(M) := -\log \mathbf{E}(Y_0^3 m^{Y_0}\, \mathbf{1}_{\{Y_0>M\}}) \le \infty\,.$$

Since $\lim_{M\to\infty} \zeta(M) \to \infty$, we can choose M sufficiently large so that $M\,\zeta(M) \ge 2$. Let $Z_0 := Y_0\, \mathbf{1}_{\{Y_0 \le M\zeta(M)\}}$.

We claim that assumption (6) is satisfied with $\vartheta(M) := \max\{\mathbf{E}(Y_0^3 m^{Y_0}),\ 1\}$ and that Z_0 is bounded. For any integer $k \ge 3$, we write

$$\begin{aligned}\mathbf{E}(Z_0^k m^{Z_0}) &= \mathbf{E}(Y_0^k m^{Y_0}\, \mathbf{1}_{\{Y_0\le M\}}) + \mathbf{E}(Y_0^k m^{Y_0}\, \mathbf{1}_{\{M<Y_0\le M\zeta(M)\}})\\ &\le M^{k-3}\mathbf{E}(Y_0^3 m^{Y_0}\, \mathbf{1}_{\{Y_0\le M\}}) + \mathbf{E}(Y_0^k m^{Y_0}\, \mathbf{1}_{\{M<Y_0\le M\zeta(M)\}})\,.\end{aligned}$$

The first term on the right-hand side is easy to handle: we have $\mathbf{E}(Y_0^k m^{Y_0}\, \mathbf{1}_{\{Y_0\le M\}}) \le M^{k-3}\mathbf{E}(Y_0^3 m^{Y_0}\, \mathbf{1}_{\{Y_0\le M\}}) \le \mathbf{E}(Y_0^3 m^{Y_0})\, M^{k-3}$. In case $\zeta(M) = \infty$, we have $Y_0 \le M$ a.s., so Z_0 is bounded and the second term on the right-hand side vanishes, which yields (6) with $\vartheta(M) := \max\{\mathbf{E}(Y_0^3 m^{Y_0}),\ 1\}$.

To treat the case $\zeta(M) < \infty$ (in which case Z_0 is obviously bounded), let us look at the second term on the right-hand side: since $\mathbf{E}(Y_0^3 m^{Y_0}\, \mathbf{1}_{\{Y_0>M\}}) = \mathrm{e}^{-\zeta(M)}$, we have

$$\begin{aligned}\mathbf{E}(Y_0^k m^{Y_0}\, \mathbf{1}_{\{M<Y_0\le M\zeta(M)\}}) &\le M^{k-3}\zeta(M)^{k-3}\mathbf{E}(Y_0^3 m^{Y_0}\, \mathbf{1}_{\{Y_0>M\}})\\ &= M^{k-3}\zeta(M)^{k-3}\, \mathrm{e}^{-\zeta(M)}\,.\end{aligned}$$

Applying the inequality $\mathrm{e}^x \ge \frac{x^{k-3}}{(k-3)!}$ (for $x \ge 0$ and $k \ge 3$) to $x := \zeta(M)$ yields (6) again with $\vartheta(M) := \max\{\mathbf{E}(Y_0^3 m^{Y_0}),\ 1\}$.

Consequently, regardless of whether $\zeta(M)$ is finite or infinite, Z_0 is bounded, and assumption (6) is satisfied with $\vartheta(M) := \max\{\mathbf{E}(Y_0^3 m^{Y_0}),\ 1\}$. Note that $\vartheta(M)$ does not depend on M.

Assumption (7) is also satisfied: by (13), $\prod_{i=0}^{n-1}[\mathbf{E}(m^{Y_i})]^{m-1} \le c_7 n^2$ for all $n \ge 1$; since $\mathbf{E}(m^{Z_i}) \le \mathbf{E}(m^{Y_i})$ for all $i \ge 0$, (7) is satisfied with $\gamma := c_7 \max\{\mathbf{E}(Y_0^3 m^{Y_0}),\ 1\}$. □

4.2 Proof of Theorem 4: The Stable Case

We start with a simple inequality.

Lemma 3 *Let* $(X_n,\ n \ge 0)$ *be a Derrida–Retaux system. If* $(m-1)\mathbf{E}(X_0\, m^{X_0}) < \mathbf{E}(m^{X_0}) < \infty$, *then*

$$\prod_{i=0}^{\infty} [\mathbf{E}(m^{X_i})]^{m-1} \le \frac{1}{\mathbf{E}(m^{X_0}) - (m-1)\mathbf{E}(X_0\, m^{X_0})}\,.$$

Proof For the moment, let $(X_n,\ n \ge 0)$ be an arbitrary Derrida–Retaux system satisfying $\mathbf{E}(X_0\, m^{X_0}) < \infty$, and such that $\mathbf{E}(m^{X_0}) \ne (m-1)\mathbf{E}(X_0\, m^{X_0})$. By (11),

$$\prod_{i=0}^{n-1} [\mathbf{E}(m^{X_i})]^{m-1} = \frac{\mathbf{E}(m^{X_n}) - (m-1)\mathbf{E}(X_n\, m^{X_n})}{\mathbf{E}(m^{X_0}) - (m-1)\mathbf{E}(X_0\, m^{X_0})}\,. \tag{27}$$

For the nominator in (27), we observe that $\mathbf{E}(m^{X_n}) - (m-1)\mathbf{E}(X_n\, m^{X_n}) = \mathbf{E}[(1-(m-1)X_n)m^{X_n}] \le \mathbf{P}(X_n = 0) \le 1$. If we assume $(m-1)\mathbf{E}(X_0\, m^{X_0}) < \mathbf{E}(m^{X_0})$, then the denominator in (27) is positive; the lemma follows immediately from the monotone convergence theorem by letting $n \to \infty$. □

We now proceed to the proof of Theorem 4 for stable systems. Assume $\mathbf{E}(m^{Y_0}) = (m-1)\mathbf{E}(Y_0\, m^{Y_0}) < \infty$ and $\mathbf{P}(Y_0 = j) \sim c_0\, m^{-j} j^{-\alpha}$, $j \to \infty$, for some $c_0 > 0$ and $2 < \alpha < 4$ as in (4). This yields the existence of constants $c_{28} \ge c_{29} > 0$ and an integer $j_0 \ge 1$, all depending on m and on the law of Y_0, such that

$$\mathbf{P}(Y_0 = j) \le c_{28}\, m^{-j} j^{-\alpha}, \qquad j \ge 1, \tag{28}$$

$$\mathbf{P}(Y_0 = j) \ge c_{29}\, m^{-j} j^{-\alpha}. \qquad j \ge j_0, \tag{29}$$

Let $M \ge j_0$ be an integer and let

$$Z_0 := Y_0\, \mathbf{1}_{\{Y_0 \le M\}}, \tag{30}$$

which is a bounded random variable. For integer $k \ge 3$, we write

$$\mathbf{E}(Z_0^k m^{Z_0}) = \mathbf{E}(Y_0^k m^{Y_0}\, \mathbf{1}_{\{Y_0 \le M\}}) \le M^{k-3}\, \mathbf{E}(Y_0^3 m^{Y_0}\, \mathbf{1}_{\{Y_0 \le M\}}),$$

so by (28), we have

$$\begin{aligned}\mathbf{E}(Z_0^k m^{Z_0}) &\le c_{28} M^{k-3} \sum_{j=1}^{M} j^{3-\alpha} \le c_{28} M^{k-3} \int_0^{M+1} x^{3-\alpha}\, \mathrm{d}x \\ &= c_{28} M^{k-3} \frac{(M+1)^{4-\alpha}}{4-\alpha} \le c_{28} M^{k-3} \frac{2^{4-\alpha} M^{4-\alpha}}{4-\alpha}\,.\end{aligned}$$

As such, assumption (6) is satisfied with

$$\vartheta(M) := c_{30}\, M^{4-\alpha}, \tag{31}$$

where $c_{30} := \max\{c_{28}\frac{2^{4-\alpha}}{4-\alpha},\ 1\}$. In particular, $M \mapsto \vartheta(M)$ is non-decreasing.

It remains to check assumption (7), which in this case states that for some constant $c_{31} > 0$,

$$\prod_{i=0}^{n-1} [\mathbf{E}(m^{Z_i})]^{m-1} \le c_{31}\,(n \vee M)^{\alpha-2}, \qquad M \ge j_0,\ n \ge 1. \tag{32}$$

Since $\mathbf{E}(m^{Z_i}) \le \mathbf{E}(m^{Y_i})$, there is nothing to prove if $n \le j_0$: it suffices to take c_{31} such that $c_{31} \ge \prod_{i=0}^{j_0-1}[\mathbf{E}(m^{Y_i})]^{m-1}$. Let us assume $n > j_0$. We have $\prod_{i=0}^{n-1}[\mathbf{E}(m^{Z_i})]^{m-1} \le \prod_{i=0}^{\infty}[\mathbf{E}(m^{Z_i})]^{m-1}$, so by Lemma 3,

$$\prod_{i=0}^{n-1} [\mathbf{E}(m^{Z_i})]^{m-1} \le \frac{1}{\mathbf{E}(m^{Z_0}) - (m-1)\mathbf{E}(Z_0\, m^{Z_0})}\,.$$

By definition, $Z_0 = Y_0\, \mathbf{1}_{\{Y_0 \le M\}}$, and by assumption, $\mathbf{E}(m^{Y_0}) = (m-1)\mathbf{E}(Y_0\, m^{Y_0})$. So

$$\begin{aligned} &\mathbf{E}(m^{Z_0}) - (m-1)\mathbf{E}(Z_0\, m^{Z_0}) \\ &= \mathbf{E}[((m-1)Y_0 - 1)\, m^{Y_0}\, \mathbf{1}_{\{Y_0 > M\}}) + \mathbf{P}(Y_0 > M) \\ &\ge \mathbf{E}[((m-1)Y_0 - 1)\, m^{Y_0}\, \mathbf{1}_{\{Y_0 > M\}})\,, \end{aligned} \tag{33}$$

which, by (29), is $\ge c_{29}\sum_{j=M+1}^{\infty}((m-1)\,j - 1)\,j^{-\alpha} \ge \frac{c_{32}}{M^{\alpha-2}}$ for some constant $c_{32} > 0$ and all $M \ge j_0$. Consequently,

$$\prod_{i=0}^{n-1} [\mathbf{E}(m^{Z_i})]^{m-1} \le \frac{M^{\alpha-2}}{c_{32}}\,, \tag{34}$$

which is bounded by $\frac{(n \vee M)^{\alpha-2}}{c_{32}}$ as $\alpha > 2$. This yields (32). □

Remark 3 For further use, let us note that in the stable case, $\vartheta(M) := c_{30}\, M^{4-\alpha}$ for all $M \ge M_0$ (by (31)), whereas $\prod_{i=0}^{n-1}[\mathbf{E}(m^{Z_i})]^{m-1} \le \frac{M^{\alpha-2}}{c_{32}}$ (by (34)), so inequality (8) in Corollary 1 implies that for $M \ge M_0$, $n \ge 1$ and $v := m + \frac{1}{2c_4(n \vee M)}$ ($c_4 \ge 1$ being as before the constant in Theorem 3),

$$\mathbf{E}(Z_n^2 v^{Z_n}) \le c_{33}(n \vee M)^{(4-\alpha)/2} M^{(\alpha-2)/2}\,, \tag{35}$$

where $c_{33} := \frac{c_5\, c_{30}^{1/2}}{c_{32}^{1/2}}$.

5 Proof of Theorem 1

Let $(Y_n,\, n \ge 0)$ be a Derrida–Retaux system such that $\mathbf{E}(m^{Y_0}) = (m-1)\,\mathbf{E}(Y_0\, m^{Y_0}) < \infty$. We assume that $\mathbf{P}(Y_0 = j) \sim c_0\, m^{-j} j^{-\alpha}$, $j \to \infty$, for some $c_0 > 0$ and $2 < \alpha < 4$ as in (4).

5.1 Upper Bound

We start with a lemma.

Lemma 4 *Let $(X_n,\, n \ge 0)$ be a Derrida–Retaux system satisfying $\mathbf{E}(X_0\, m^{X_0}) < \infty$. There exists a constant $c_{34} > 0$, depending only on m, such that ($\frac{n^2}{0} := \infty$)*

$$\prod_{i=0}^{n-1} [\mathbf{E}(m^{X_i})]^{m-1} \le c_{34}\, \frac{n^2}{\mathbf{E}(X_0^3 m^{X_0} \mathbf{1}_{\{2\le X_0 \le 3n\}})}, \qquad n \ge 1\,.$$

Proof The lemma was known in various forms. Recall from [3] that for $s \in (\frac{m}{2},\, m)$,

$$\prod_{i=0}^{n-1} [\mathbf{E}(m^{X_i})]^{m-1} \le \Big(\frac{m}{2s-m}\Big)^n \frac{1}{\Delta_0(s)}, \tag{36}$$

where

$$\Delta_0(s) := \sum_{k=1}^{\infty} m^k ((m-1)k - 1)(1 - (k+1)x^k + kx^{k+1}) \mathbf{P}(X_0 = k),$$

with $x := \frac{s}{m} \in (\frac{1}{2},\, 1)$. We now reproduce some elementary computations from [1]. For $k \ge 1$, we have $x^k \le \mathrm{e}^{-(1-x)k}$, so $1-(1+k)x^k + kx^{1+k} \ge 1-(1+u)\,\mathrm{e}^{-u}$, where $u := (1-x)k > 0$. Since $1-(1+v)\,\mathrm{e}^{-v} \ge 1-\frac{2}{\mathrm{e}}$ for $v \ge 1$ (because $v \mapsto 1-(1+v)\,\mathrm{e}^{-v}$ is increasing on $(0,\, \infty)$) and $1-(1+v)\,\mathrm{e}^{-v} \ge \frac{v^2}{2\mathrm{e}}$ for $v \in (0,\, 1]$ (because $v \mapsto 1-(1+v)\,\mathrm{e}^{-v} - \frac{v^2}{2\mathrm{e}}$ is increasing on $(0,\, 1]$), we get, for $k \ge 1$,

$$1-(k+1)x^k + kx^{k+1} \ge c_{35}\, \min\{(1-x)^2 k^2,\, 1\}\,,$$

where $c_{35} := \min\{1 - \frac{2}{e}, \frac{1}{2e}\} > 0$. Let $n \geq 1$. We take $s = s_n := (1 - \frac{1}{3n})m$ (so $x = 1 - \frac{1}{3n}$), to see that

$$\Delta_0(s_n) \geq c_{35} \sum_{k=1}^{\infty} m^k((m-1)k - 1) \min\{\frac{k^2}{(3n)^2}, 1\} \mathbf{P}(X_0 = k)$$

$$\geq \frac{c_{35}}{(3n)^2} \sum_{k=1}^{3n} k^2 m^k ((m-1)k - 1) \mathbf{P}(X_0 = k).$$

We use $(m-1)k - 1 \geq \frac{m-1}{2} k$ for $k \geq 2$, so that with $c_{36} := c_{35} \frac{m-1}{18} > 0$

$$\Delta_0(s_n) \geq \frac{c_{36}}{n^2} \sum_{k=2}^{3n} k^3 m^k \mathbf{P}(X_0 = k) = \frac{c_{36}}{n^2} \mathbf{E}(X_0^3 m^{X_0} \mathbf{1}_{\{2 \leq X_0 \leq 3n\}}).$$

This, in view of (36) (applied to $s = s_n$), yields the lemma. □

Proof of the Upper Bound in Theorem 1 By Lemma 4, for all $n \geq 1$,

$$\prod_{i=0}^{n-1} [\mathbf{E}(m^{Y_i})]^{m-1} \leq c_{34} \frac{n^2}{\mathbf{E}(Y_0^3 m^{X_0} \mathbf{1}_{\{2 \leq Y_0 \leq 3n\}})}.$$

Since $\mathbf{P}(Y_0 = j) \sim c_0 m^{-j} j^{-\alpha}$, $j \to \infty$, and $2 < \alpha < 4$, we have $\mathbf{E}(Y_0^3 m^{Y_0} \mathbf{1}_{\{2 \leq Y_0 \leq 3n\}}) \geq c_{37} n^{4-\alpha}$ for some constant $c_{37} > 0$ and all sufficiently large n; this implies the upper bound in Theorem 1. □

5.2 *Lower Bound in Theorem 1*

The proof of the lower bound in Theorem 1 needs some preparation.

Lemma 5 *There exists a constant $c_{38} > 0$ such that for all sufficiently large integer n, say $n \geq n_0$, we have*

$$\textit{either} \prod_{i \in (\frac{n}{2}, n] \cap \mathbb{Z}} [\mathbf{E}(m^{Y_i})]^{m-1} \geq 8, \quad \textit{or} \quad \prod_{i=0}^{n} [\mathbf{E}(m^{Y_i})]^{m-1} \geq c_{38} n^{\alpha-2}.$$

Proof Let $M_0 \geq 1$ and $c_{33} \geq 1$ be the constants in (35) in Remark 3. Let $c_{39} := [120(m-1)c_{33}]^{2/(\alpha-2)} \geq 120$. Let $n \geq \lceil 3c_{39} M_0 \rceil =: n_0$ be an integer, and let $M = M(n) := \lfloor \frac{n}{c_{39}} \rfloor$. So $\frac{n}{2c_{39}} \leq M \leq \frac{n}{120}$. Let $u_n := m - \frac{c_{40}}{n}$, where $c_{40} := 30m$. We can enlarge the value of M_0 if necessary to ensure that $u_n > \frac{m}{2}$.

We discuss on two possible situations, each leading to one of the inequalities stated in the lemma.

First situation: $\mathbf{E}[(1-(m-1)Y_i)u_n^{Y_i}] < \frac{1}{2}$ *for all* $i \in (\frac{n}{2}, n] \cap \mathbb{Z}$. In this situation, we have, for all $i \in (\frac{n}{2}, n] \cap \mathbb{Z}$, $(m-1)\mathbf{E}(Y_i\, u_n^{Y_i}) \ge \mathbf{E}(u_n^{Y_i}) - \frac{1}{2} \ge \frac{1}{2}$, thus

$$\mathbf{E}(Y_i\, u_n^{Y_i}) \ge \frac{1}{2(m-1)}\,.$$

Consider the function $s \mapsto f_i(s) := \mathbf{E}(s^{Y_i})$, $s \in [0, m]$. We have

$$\mathbf{E}(m^{Y_i}) = f_i(m) \ge f_i(u_n) + (m-u_n) f_i'(u_n) \ge 1 + (m-u_n) f_i'(u_n).$$

Since $m - u_n = \frac{30m}{n}$ and $f_i'(u_n) = \frac{1}{u_n}\mathbf{E}(Y_i\, u_n^{Y_i}) \ge \frac{1}{m}\frac{1}{2(m-1)} = \frac{1}{2m(m-1)}$, we get, for all $i \in (\frac{n}{2}, n] \cap \mathbb{Z}$,

$$\mathbf{E}(m^{Y_i}) \ge 1 + \frac{15}{(m-1)n}\,.$$

Consequently,

$$\prod_{i \in (\frac{n}{2}, n]\cap\mathbb{Z}} [\mathbf{E}(m^{Y_i})]^{m-1} \ge \Big(1 + \frac{15}{(m-1)n}\Big)^{(m-1)n/2} \ge 8\,.$$

Second (and last) situation: $\mathbf{E}[(1-(m-1)Y_\ell)u_n^{Y_\ell}] \ge \frac{1}{2}$ *for some* $\ell = \ell(n) \in (\frac{n}{2}, n] \cap \mathbb{Z}$. We will be working with this particular ℓ in the rest of the proof. Let $Z_0 := Y_0\, \mathbf{1}_{\{Y_0 \le M\}}$ as in (30). Let $(Z_n,\, n \ge 0)$ be a Derrida–Retaux system whose initial distribution is given by Z_0.

Since $u_n \in [1, m]$, the function $x \mapsto (1-(m-1)x)u_n^x$ is decreasing on $[0, \infty)$, so

$$\mathbf{E}[(1-(m-1)Z_\ell)u_n^{Z_\ell}] \ge \mathbf{E}[(1-(m-1)Y_\ell)u_n^{Y_\ell}] \ge \frac{1}{2}\,. \tag{37}$$

Consider the function

$$\varphi(s) := \mathbf{E}[(1-(m-1)Z_\ell)s^{Z_\ell}], \qquad s \ge 0,$$

which is well-defined because Z_ℓ is bounded. We have just proved that $\varphi(u_n) \ge \frac{1}{2}$. Let $v_n := m + \frac{1}{2c_4 n}$, where $c_4 \ge 1$ is the constant in Theorem 3. Since $v_n > m$, we have, by concavity of $\varphi(\cdot)$,

$$\varphi(v_n) \ge \varphi(m) + (v_n - m)\varphi'(v_n)\,.$$

By assumption, the system $(Y_n,\ n \ge 0)$ is critical, so $(Z_n,\ n \ge 0)$ is subcritical (or critical in case $Y_0 \le M$ a.s.), which implies that $(m-1)\mathbf{E}(Z_\ell\, m^{Z_\ell}) \le \mathbf{E}(m^{Z_\ell})$. This means that $\varphi(m) \ge 0$. On the other hand, $v_n - m = \frac{1}{2c_4 n}$, whereas

$$\varphi'(v_n) = \mathbf{E}[(1-(m-1)Z_\ell)Z_\ell v_n^{Z_\ell - 1}] \ge -(m-1)\mathbf{E}(Z_\ell^2 v_n^{Z_\ell - 1}),$$

which is $= -\frac{m-1}{v_n}\mathbf{E}(Z_\ell^2 v_n^{Z_\ell}) \ge -\frac{m-1}{m}\mathbf{E}(Z_\ell^2 v_n^{Z_\ell})$. Assembling these pieces together yields that

$$\varphi(v_n) \ge -\frac{1}{2c_4 n}\,\frac{m-1}{m}\,\mathbf{E}(Z_\ell^2 v_n^{Z_\ell})\,.$$

Since $v_n = m + \frac{1}{2c_4 n} \le m + \frac{1}{2c_4 \ell} = m + \frac{1}{2c_4(\ell \vee M)}$ (to obtain the last equality, we have used the fact that $\ell \ge \frac{n}{2} \ge M$), we are entitled to apply inequality (35) in Remark 3 to see that

$$\mathbf{E}(Z_\ell^2 v_n^{Z_\ell}) \le c_{33}(\ell \vee M)^{(4-\alpha)/2} M^{(\alpha-2)/2}\,.$$

Since $(\ell \vee M) = \ell \le n$ and $M \le \frac{n}{c_{39}}$, this yields

$$\mathbf{E}(Z_\ell^2 v_n^{Z_\ell}) \le c_{33} n^{(4-\alpha)/2} (\frac{n}{c_{39}})^{(\alpha-2)/2} = \frac{c_{33}}{c_{39}^{(\alpha-2)/2}}\, n\,.$$

Accordingly,

$$\varphi(v_n) \ge -\frac{1}{2c_4 n}\,\frac{m-1}{m}\,\frac{c_{33}}{c_{39}^{(\alpha-2)/2}}\, n = -\frac{c_{41}}{c_{39}^{(\alpha-2)/2}}\,,$$

where $c_{41} := \frac{c_{33}}{2c_4}\,\frac{m-1}{m}$. On the other hand, we have $\varphi(u_n) \ge \frac{1}{2}$ (see (37)). Since $m = \beta u_n + (1-\beta) v_n$ with $\beta := \frac{1}{1+2c_4 c_{40}} \in (0,\ 1)$, it follows from concavity of $\varphi(\cdot)$ that

$$\varphi(m) \ge \beta\, \varphi(u_n) + (1-\beta)\, \varphi(v_n) \ge \frac{\beta}{2} - (1-\beta)\frac{c_{41}}{c_{39}^{(\alpha-2)/2}} = \frac{1 - \frac{4c_4 c_{40} c_{41}}{c_{39}^{(\alpha-2)/2}}}{2(1+2c_4 c_{40})}\,.$$

Our choice of the constant c_{39} ensures $\frac{4c_4 c_{40} c_{41}}{c_{39}^{(\alpha-2)/2}} = \frac{1}{2}$; hence $\varphi(m) \ge \frac{1}{4(1+2c_4 c_{40})} =: c_{42}$, i.e.,

$$\mathbf{E}(m^{Z_\ell}) - (m-1)\mathbf{E}(Z_\ell\, m^{Z_\ell}) \ge c_{42}\,.$$

Recall from (27) that

$$\prod_{i=0}^{\ell-1}[\mathbf{E}(m^{Z_i})]^{m-1} = \frac{\mathbf{E}(m^{Z_\ell}) - (m-1)\mathbf{E}(Z_\ell\, m^{Z_\ell})}{\mathbf{E}(m^{Z_0}) - (m-1)\mathbf{E}(Z_0\, m^{Z_0})}\,.$$

On the right-hand side, the numerator is at least c_{42}, whereas the denominator has already appeared in (33):

$$\begin{aligned}\mathbf{E}(m^{Z_0}) - (m-1)\mathbf{E}(Z_0\, m^{Z_0}) &= \mathbf{E}[((m-1)Y_0 - 1)\, m^{Y_0}\, \mathbf{1}_{\{Y_0>M\}}) + \mathbf{P}(Y_0 > M)\\ &\le (m-1)\mathbf{E}(Y_0\, m^{Y_0}\, \mathbf{1}_{\{Y_0>M\}})\,.\end{aligned}$$

Our assumption $\mathbf{P}(Y_0 = j) \sim c_0\, m^{-j}\, j^{-\alpha}$ (for $j \to \infty$) yields that $\mathbf{E}(Y_0\, m^{Y_0}\, \mathbf{1}_{\{Y_0>M\}}) \le \frac{c_{43}}{M^{\alpha-2}}$ for some constant $c_{43} > 0$ depending on the law of Y_0. As a consequence,

$$\prod_{i=0}^{\ell-1}[\mathbf{E}(m^{Z_i})]^{m-1} \ge \frac{c_{42}}{c_{43}\, M^{-(\alpha-2)}} = c_{44}\, M^{\alpha-2}\,,$$

where $c_{44} := \frac{c_{42}}{c_{43}}$. This implies that

$$\prod_{i=0}^{n-1}[\mathbf{E}(m^{Y_i})]^{m-1} \ge \prod_{i=0}^{\ell-1}[\mathbf{E}(m^{Y_i})]^{m-1} \ge \prod_{i=0}^{\ell-1}[\mathbf{E}(m^{Z_i})]^{m-1} \ge c_{44}\, M^{\alpha-2}\,.$$

Since $M \ge \frac{n}{2c_{39}}$, this completes the proof of the lemma. □

We have all the ingredients for the proof of the lower bound in Theorem 1.

Proof of the Lower Bound in Theorem 1. It follows quite easily from Lemma 5. Indeed, let n_0 be the integer in Lemma 5, and let $n \ge 2n_0$. According to Lemma 5, there are two possibilities:

- either we have $\prod_{i\in(n/2^{j+1},\, n/2^j]\cap\mathbb{Z}}[\mathbf{E}(m^{Y_i})]^{m-1} \ge 8$ for all $0 \le j \le \lfloor\frac{\log(n/n_0)}{\log 2}\rfloor$, in which case we have

$$\prod_{i=0}^{n}[\mathbf{E}(m^{Y_i})]^{m-1} \ge 8^{\lfloor\frac{\log(n/n_0)}{\log 2}\rfloor} \ge 8^{\frac{\log(n/n_0)}{\log 2}-1} = \frac{n^3}{8n_0^3}\,,$$

  which yields the lower bound in Theorem 1 because $3 > \alpha - 2$.
- or there exists an integer $0 \le j^* = j^*(n) \le \lfloor\frac{\log(n/n_0)}{\log 2}\rfloor$ such that $\prod_{i=0}^{n/2^{j^*}} [\mathbf{E}(m^{Y_i})]^{m-1} \ge c_{38}\, (\frac{n}{2^{j^*}})^{\alpha-2}$ and that $\prod_{i\in(n/2^{j+1},\, n/2^j]\cap\mathbb{Z}}[\mathbf{E}(m^{Y_i})]^{m-1} \ge 8$ for all non-negative integer $j < j^*$ (if there is any); in this case, we have

(recalling notation: $\prod_{\varnothing} := 1$)

$$\prod_{i=0}^{n} [\mathbf{E}(m^{Y_i})]^{m-1} \geq \prod_{i=0}^{n/2^{j^*}} [\mathbf{E}(m^{Y_i})]^{m-1} \times \prod_{j=0}^{j^*-1} \prod_{i \in (n/2^{j+1},\, n/2^j] \cap \mathbb{Z}} [\mathbf{E}(m^{Y_i})]^{m-1}$$
$$\geq c_{38} \, (\frac{n}{2^{j^*}})^{\alpha-2} \times 8^{j^*},$$

which implies that

$$\prod_{i=0}^{n} [\mathbf{E}(m^{Y_i})]^{m-1} \geq c_{38} \, n^{\alpha-2} \, 2^{(5-\alpha) j^*} \geq c_{38} \, n^{\alpha-2} \, .$$

In both situations, the lower bound in Theorem 1 is valid. □

Acknowledgments We are grateful to Victor Dagard, Bernard Derrida, Yueyun Hu and Mikhail Lifshits for stimulating discussions on various aspects of the Derrida–Retaux model throughout the last few years. We also wish to thank two anonymous referees whose comments have led to improvements in the presentation of the paper.

Xinxing Chen partially supported by NSFC grants 11771286 and 11531001 and Zhan Shi Partially supported by ANR MALIN.

References

1. Chen, X., Dagard, V., Derrida, B., Hu, Y., Lifshits, M., Shi, Z.: The Derrida–Retaux conjecture on recursive models. Ann. Probab. (to appear). https://arxiv.org/abs/1907.01601
2. Chen, X., Dagard, V., Derrida, B., Shi, Z.: The critical behaviors and the scaling functions of a coalescence equation. J. Phys. A: Math. Theor. **53**, 195202, 25 (2020)
3. Chen, X., Derrida, B., Hu, Y., Lifshits, M., Shi, Z.: A max-type recursive model: some properties and open questions. In: Sidoravicius, V. (ed.) Sojourns in Probability Theory and Statistical Physics III, pp. 166–186. Springer, Singapore (2019)
4. Chen, X., Hu, Y., Shi, Z.: The sustainability probability for the critical Derrida–Retaux model (in preparation, 2020)
5. Collet, P., Eckmann, J.P., Glaser, V., Martin, A.: Study of the iterations of a mapping associated to a spin-glass model. Commun. Math. Phys. **94**, 353–370 (1984)
6. Derrida, B., Retaux, M.: The depinning transition in presence of disorder: a toy model. J. Stat. Phys. **156**, 268–290 (2014)
7. Derrida, B., Shi, Z.: Results and conjectures on a toy model of depinning. https://arxiv.org/abs/2005.10208
8. Derrida, B., Hakim, V., Vannimenus, J.: Effect of disorder on two-dimensional wetting. J. Stat. Phys. **66**, 1189–1213 (1992)
9. Derrida, B., Giacomin, G., Lacoin, H., Toninelli, F.L.: Fractional moment bounds and disorder relevance for pinning models. Commun. Math. Phys. **287**, 867–887 (2009)
10. Forgacs, G., Luck, J.M., Nieuwenhuizen, T.M., Orland, H.: Wetting of a disordered substrate: exact critical behavior in two dimensions. Phys. Rev. Lett. **57**, 2184 (1986)
11. Giacomin, G.: Random Polymer Models. Imperial College Press, London (2007)

12. Giacomin, G., Toninelli, F.L.: Smoothing effect of quenched disorder on polymer depinning transitions. Commun. Math. Phys. **266**, 1–16 (2006)
13. Hu, Y., Mallein, B., Pain, M.: An exactly solvable continuous-time Derrida–Retaux model. Commun. Math. Phys. **375**, 605–651 (2020)
14. Monthus, C.: Strong disorder renewal approach to DNA denaturation and wetting: typical and large deviation properties of the free energy. J. Stat. Mech. Theory Exp. **2017**(1), 013301 (2017)
15. Tang, L.H., Chaté, H.: Rare-event induced binding transition of heteropolymers. Phys. Rev. Lett. **86**(5), 830 (2001)

A Class of Random Walks on the Hypercube

Andrea Collevecchio and Robert C. Griffiths

Dedicated to the memory of Prof. Vladas Sidoravicius

Abstract We introduce a general class of time inhomogeneous random walks on the N-hypercube. These random walks are reversible with an N-product Bernoulli stationary distribution and have a property of local change of coordinates in a transition. Several types of representations for the transition probabilities are found. The paper studies cut-off for the mixing time. We observe that for a sub-class of these processes with long range (i.e. non-local) there exists a critical value of the range that allows an *almost-perfect* mixing in at most two steps. That is, the total variation distance between the two steps transition and stationary distributions decreases to zero as the dimension of the hypercube N increases. Notice that a well-known result (Theorem 1 in [6]) shows that there does not exist a random walk on Abelian groups (such as the hypercube) which mixes perfectly in exactly two steps.

Keywords Hypercube · Random walk · Markov chain · Mixing

MSC:60J10

1 Introduction

This paper studies a general class $\mathscr{G}$ of reversible Markov chains on the vertices of a hypercube $\{0, 1\}^N$. For each process in this class, there exists a $p \in (0, 1)$ such that the stationary distribution of the chain is a product-measure $\otimes_{i=1}^N \theta_i$, where θ_i are Bernoulli(p) measures. The chains are allowed to have long-range steps. The class

A. Collevecchio (✉) · R. C. Griffiths
School of Mathematics, Monash University, Melbourne, VIC, Australia
e-mail: Andrea.Collevecchio@monash.edu; Bob.Griffiths@monash.edu

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_13

of Markov chains has the following *equivalent* characterizations.

(a) A local changes property (formally introduced as (3)) holds so that the probability that a subset of coordinates change in a transition only depends on the same subset of coordinates at the prior time.
(b) The chain is driven by an acceptance-rejection scheme on a deterministic hypercube random walk which can be described as follows. At each stage a subset of coordinates is chosen according to a certain random rule, independently of the past of the chain. In the acceptance rejection step if a coordinate is chosen then it changes according to a transition matrix (with row and columns labelled by 0, 1)

$$P^{\circ} = \begin{pmatrix} 0 & 1 \\ \frac{q}{p} & 1 - \frac{q}{p} \end{pmatrix} \tag{1}$$

where $q = 1 - p$ and $p \geq q$. Details are in Sect. 5.1.
(c) The spectral decomposition of the transition probability has a diagonal expansion in the tensor product basis, orthogonal on the stationary distribution which is product Bernoulli(p). A characterization of the eigenvalues of the transition distribution solves the *Lancaster problem* for the tensor product basis.
(d) A random walk representation (see Theorem 7).

We emphasize that the processes that we are studying are not necessarily homogeneous in time. Examples from this class of Markov chains are discussed in Sect. 2, and include (1) the classical random walk on the hypercube, (2) long range random walks, (3) walks with a mixture of i.i.d. updates. The type of processes we are studying are intimately connected to the generalized Ehrenfest Urn model studied in [4], by considering the Hamming distance from a vector of all zeroes on the hypercube. The coordinate Markov chain is of course a finer look at how individual balls change colour in an Ehrenfest urn. In the class we study the N coordinates are not necessarily exchangeable, but we often take them to be as this is a natural assumption.

Our contribution can be summarized as follows.

- Spectral expansions for random walks on a hypercube are well studied, however a Lancaster characterization of *all* possible Markov chains with tensor product eigenfunctions and the equivalence of (a), (b), (c), (d) does not seem to be known in the same way as presented here. We provide an explicit, computable formula for the eigenvalues.
- We use the spectral decomposition to study the behaviour of the mixing time of Markov chains in $\mathscr{G}$ as the state space increases, both in terms of total variation and χ^2 distances. Among other things, our results highlight a certain discontinuity of the mixing time in terms of the size of a single step of the random walk. We characterize the cases when the chain mixes *almost-perfectly* in at most two steps. This means that the total variation distance between the distribution of the process at time 2 and the stationary distribution decreases to

zero as the dimension of the hypercube increases (see Definition 5). For example, to illustrate this phenomena, consider the following chain, which is described in detail in Example 2 below. Fix $p > 1/2$ and a parameter $\alpha \in (0, 1]$. At each step exactly $\lfloor \alpha N \rfloor$ coordinates are picked uniformly at random and their value is changed with the following procedure by making transitions according to P° (defined in (1)), which is repeated independently for each coordinate selected. That is, if a coordinate is 0 it changes to 1, while if it is 1 then an independent randomization is used. The 1 becomes 0 with probability q/p, and does not change otherwise. If $\alpha = p$, we prove that the chain we just described mixes almost perfectly in exactly[1] 2 steps.

The value $\alpha = p$ is what we call the critical value. Moreover almost-perfect mixing in at most 2 steps is observed in the window $\alpha \in [p - v/\sqrt{N}, p + v/\sqrt{N}]$ where v is any real number, and can even be random. On the other hand, if $\alpha \neq p$ the chain mixes in the order of $\log N$ steps, and a cut-off is proved in the χ^2 distance.

Notice that p is allowed to depend on N, and we find interesting the case where p_N converges to $1/2$. This can be interpreted as a small perturbation of the case $p = 1/2$. Comparing this result with the existing literature on long-range random walks on the hypercube with $p = 1/2$, we observe a big gap, as the latter process mixes slowly, at least in the χ^2 distance when there is a laziness assumption (see [11]).

The almost-perfect mixing in exactly two steps described above is possibly unexpected also because of a well-known result by Diaconis and Shahshahani [6] which implies that no random walk on an Abelian group reaches perfect stationarity in exactly two steps.

The mixing time in a random walk on the hypercube can be greater than the mixing time for the Hamming distance of the random walk. Example 6 shows how this difference is highly dependent on the inital condition. In Sect. 3 we provide a short literature review.

2 Model and Main Results

We define a class $\mathscr{G}$ of Markov chains as follows. Let $\mathbf{X} = (X_t)_{t \in \mathbb{Z}_+}$ be a reversible Markov chain with state space $\mathscr{V}_N = \{0, 1\}^N$, for some $N \in \mathbb{N}$. The process $\mathbf{X}$ is in the class $\mathscr{G}$ if and only if satisfies the following two conditions.

Condition 1 There exists a parameter $p \geq 1/2$ such that the following is the unique stationary measure for $\mathbf{X}$,

$$\pi_N(\boldsymbol{y}, p) := p^{\|\boldsymbol{y}\|}(1-p)^{N-\|\boldsymbol{y}\|} := p^{\|\boldsymbol{y}\|} q^{N-\|\boldsymbol{y}\|}, \tag{2}$$

[1]It is immediate to see that in 1 step the chain is far away from stationarity. In fact, if the starting point is $(0, 0, \ldots, 0)$ then, after 1 step, there are exactly $\lfloor pN \rfloor$ ones. π gives a very small probability for this to happen.

where $\boldsymbol{y} \in \mathscr{V}_N$, and $\|\boldsymbol{y}\|$ is the sum of ones appearing in $\boldsymbol{y}$, i.e. the Hamming distance between $\boldsymbol{y}$ and $\mathbf{0} = (0, 0, \ldots, 0) \in \mathscr{V}_N$. In many occasions, we drop N from the notation, and simply use $\pi(\boldsymbol{y}, p)$.

Condition 2 For any $\boldsymbol{y} \in \mathscr{V}_N$, denote by $\boldsymbol{y} = (y[1], y[2], \ldots y[N])$ its coordinates. For all $B \subseteq [N] = \{1, 2, \ldots N\}$ let $\boldsymbol{y}(B)$ be the projection from $\mathscr{V}_N$ on B defined as the vector $\boldsymbol{y}(B) = (\boldsymbol{y}[j], j \in B)$. We assume that

$$\mathbb{P}(X_t(B) \in C \mid X_{t-1}) = \mathbb{P}(X_t(B) \in C \mid X_{t-1}(B)). \tag{3}$$

For any pair of probability measures μ and ν defined on a countable space Ω, define the total variation distance

$$\|\mu - \nu\|_{TV} = \max_{A \subseteq \Omega} |\mu(A) - \nu(A)|.$$

Definition 1 Let $\mathbf{X} \in \mathscr{G}$. Define a sequence $(Z_t)_{t \in \mathbb{N}}$ of independent random vectors in $\mathscr{V}_N$, with the following distribution

$$\mathbb{P}(Z_t \in S) = \mathbb{P}(X_t \in S \mid X_{t-1} = \mathbf{0}).$$

From now on, we denote the coordinates of Z_t by $(Z_t[1], Z_t[2], \ldots Z_t[N])$.

Remark 1 In what follows, $P_t(\cdot \mid \boldsymbol{x})$ is the probability mass function of X_t given $X_0 = \boldsymbol{x}$. Moreover, when considering a generic $\mathbf{X} \in \mathscr{G}$ we denote by N the dimension of the corresponding hypercube.

The following representation holds.

Theorem 1 (Spectral Representation) *The process* $\mathbf{X} \in \mathscr{G}$ *if and only if*

$$P_t(\boldsymbol{y} \mid \boldsymbol{x}) = \pi(\boldsymbol{y}, p)\left\{1 + \sum_{A \subseteq [N], A \neq \varnothing} \left(\prod_{m=1}^{t} \rho_{A,m}\right)\left(\frac{p}{q}\right)^{|A|} \prod_{j \in A}\left(1 - \frac{\boldsymbol{x}[j]}{p}\right)\left(1 - \frac{\boldsymbol{y}[j]}{p}\right)\right\}, \tag{4}$$

with an explicit representation for the eigenvalues of

$$\rho_{A,m} = \mathbb{E}\left[\prod_{j \in A}\left(1 - \frac{Z_m[j]}{p}\right)\right]. \tag{5}$$

If $\rho_{A,m}$ only depends on $|A|$ then we denote $\rho_{|A|} = \rho_{A,m}$.

Remark 2 The spectral representation in (4) simplifies when $\mathbf{X}$ is time-homogenous and instead of a product, we simply have ρ_A^t. Notice that the previous representation holds also in cases when the chain is reducible and/or periodic.

Example 1 (Lazy Simple Random Walk on the Hypercube (RWH)) Let X_0 be a vertex of the N-dimensional hypercube, and define the process $(X_t)_{t\in\mathbb{N}}$ recursively as follows. Suppose that at stage $t+1$, a fair coin is flipped. If it shows Head, $X_{t+1} = X_t$. If the coin shows Tail, then a coordinate of X_t is chosen uniformly at random, and it is changed. This process has been studied extensively. In particular it was shown in [7] that it exhibits a cut-off at $(1/4)N \log N$.

Example 2 (Non-Local Random Walk on the Hypercube (NLRWH)) Consider the following random walk. Fix parameters $p_N \geq 1/2$, $q_N = 1 - p_N$, and $z_N \in [N]$. Pick a set of coordinates with cardinality z_N uniformly at random, i.e. each possible choice is picked with probability

$$\binom{N}{z_N}^{-1}.$$

For each coordinate i selected, perform the following *Acceptance-Rejection* procedure with parameter p_N:

(a) If $X_t[i] = 0$ then $X_{t+1}[i] = 1$
(b) If $X_t[i] = 1$ then randomize further, and set

$$X_{t+1}[i] = \begin{cases} 0 & \text{with probability } \frac{q_N}{p_N} \\ 1 & \text{otherwise} \end{cases}.$$

The stationary distribution is unique, and is a product Bernoulli (p_N) measure. The case $p_N \equiv 1/2$, with an additional assumption of lazyness, was studied in [11].

Example 3 (Mixture of i.i.d. Updates for Each Coordinate) Fix $p_N \geq 1/2$. Define a process $(X_t)_{t\in\mathbb{N}}$ recursively. Let $X_0 = \mathbf{0} := (0, 0, \ldots, 0) \in \mathscr{V}_N$. Suppose that we have defined X_t, then we obtain X_{t+1} as follows. Let $I_t^{(N)}$ be a random variable with distribution $\nu_{N,t}$. We assume that for any fixed $N \in \mathbb{N}$, the random variables $(I_t^{(N)})_{t\in\mathbb{N}}$ are independent. Given $I_t^{(N)} = \alpha_{N,t}$, for each coordinate $j \in [N] = \{1, 2, \ldots N\}$ flip an independent coin that has probability $\alpha_{N,t}$ of showing Head. If the coin shows Tail then we set $X_{t+1}[j] = X_t[j]$. The coordinate is selected if and only if the corresponding coin shows Head. For each selected coordinate we repeat the Acceptance Rejection procedure described in the Example 2 with parameter p_N. The stationary distribution is a product Bernoulli (p_N) measure.

Example 4 (Blocks Update) Partition the space $[N]$ into L disjoint subsets $C[1], \ldots, C[L]$. Exactly one group is chosen, each with equal probability. For each coordinate j of this group repeat the Acceptance-Rejection method described in Example 2 with parameter p_N. The Markov chain $\mathbf{X}$ is reversible with respect the measure product measure of i.i.d. Bernoulli's with parameter p. Denote γ as the probability distribution with $\gamma[j] = \|A[j]\|/N$, $j \in [L]$. In the representation (5)

$Z_m[j] = 1$, $j \in C[k]$ and $Z_m[j] = 0$, $j \notin C[k]$ with probability $\gamma[k]$, $k \in [L]$. The eigenvalues in a time homogeneous chain are for $A \subseteq [N]$,

$$\rho_A = \sum_{k=1}^{L} \gamma[k]\Big(-\frac{q}{p}\Big)^{\|A\cap C[k]\|}.$$

Condition 2 implies that marginally $\{X_t(C[k])\}_{t\in\mathbb{Z}_+}$, for $k \in [L]$ are random walks in $\mathscr{G}$ on $\mathscr{V}_{\|C[k]\|}$. In $\{X_t\}_{t\in\mathbb{Z}_+}$ the blocks are linked by making a choice of one to choose in a transition. The stationary distribution is a product Bernoulli (p_N) measure.

Example 5 Nestoridi and Nguyen [12] describe a random walk on the edges of a bipartite graph motivated by the Diaconis–Gangolli random walk over $\mathbb{Z}/q\mathbb{Z}$ on contingency tables [3]. Let U, V be vertex sets in a bipartite graph with $|U| = n$, $|V| = n$. A particular version of their random graphs, when $q = 2$, that preserves the parity of each vertex in U and V is the following. At each step select, independently from the past, distinct $u_i, u_j \in U$, , $v_k, v_l \in V$ with uniform probability. Delete existing edges and add edges to those pairs of vertices which are not joined. It is easy to check that the parity of the vertex degrees in U, V is invariant under such a transition. The mixing time of such a chain is studied in [12].

We show next how to express this Markov chain in terms of the model studied this paper. Let $N = n^2$. The process $\mathbf{X} = \{X_t\}_{t\in\mathbb{Z}_+}$ has coordinates indexed in $[n] \times [n]$ and is defined by

$$X_t[i,k] = \begin{cases} 1 & \text{if } (u_i, v_k) \text{ is an edge} \\ 0 & \text{if } (u_i, v_k) \text{ is not an edge.} \end{cases}$$

Set $p = 1/2$. In this context, for each t the random vector Z_t contains exactly four 1's. The distribution of Z_t is such that four pairs with distinct coordinates $(i,k), (i,l), (j,k), (j,l)$ are chosen uniformly from $[n]\times[n]$ with probability $\binom{n}{2}^{-2}$. The entries of Z_t at these coordinates are taken to be 1 and the other entries are taken to be 0. In a transition entries at the four coordinates are toggled from 0 to 1 or 1 to 0. A stationary distribution of the chain is product Bernoulli (1/2) in the sense that if X_0 has this distribution then so does X_t for all $t \geq 0$. However for a *fixed* X_0 the limiting distribution is not the product measure, but has uniform measure on bipartite graphs with the same degree parity as X_0 determines.

Remark 3 (De Finetti Sequences) In Example 3, when $\mathbf{X}$ is time homogenous, and $\nu_{N,t}$ is a Dirac mass at r

$$\rho_{n,t} = \rho_n = \sum_{k=0}^{n} \binom{n}{k} r^k (1-r)^{n-k}\Big(-\frac{q}{p}\Big)^k = \Big(1 - \frac{r}{p}\Big)^n. \tag{6}$$

If $r = p$ then $\rho_n = 0$ for $n = 1, \dots N$. $\mathbf{X}$ is then a sequence of $i.i.d.$ random variables, which mixes in one step. If $r \neq p$ then $|\rho_1|$ is the maximum value of $|\rho_n|$. More generally, when $\nu_{N,t} \equiv \nu_N$,

$$\rho_{n,t} = \rho_n = \int_{[0,1]} \left(1 - \frac{r}{p}\right)^n \nu_N(dr).$$

The stationary distribution is a product Bernoulli (p_N) measure.

This Remark is continued in Theorem 5 where mixing times are found when ν has a density $h(x) \sim cx^\gamma$, $\gamma > -1$ as $x \to 0$. Beta densities have this property.

Definition 2 We define the collection $\mathscr{C}$ of sequences $(\mathbf{X}^{(N)})_{N\in\mathbb{N}} \in \mathscr{G}^{\mathbb{N}}$ with the following property. For each $N \in \mathbb{N}$,

- the state space of $\mathbf{X}^{(N)}$ is $\mathscr{V}_N$, and
- there exists a sequence $(p_N)_N \in [1/2, 1]^{\mathbb{N}}$ such that $\pi_N(\cdot, p_N)$ is a stationary distribution for $\mathbf{X}^{(N)}$, and $\lim_{N\to\infty} p_N = p$ for some $p \in [1/2, 1]$.

From now on, once an element of $\mathscr{C}$ is fixed, denote by $P_t^{(N)}$ the transition kernel of $\mathbf{X}^{(N)}$.

Define

$$t_{\text{mix}}(\varepsilon, \boldsymbol{x}) = \inf\{t \colon \|P_t^{(N)}(\cdot \mid \boldsymbol{x}) - \pi_N(\cdot)\|_{TV} \leq \varepsilon\}.$$

Let $t_{\text{mix}}(\varepsilon) = \sup_{\boldsymbol{x}\in\mathscr{V}_N} t_{\text{mix}}(\varepsilon, \boldsymbol{x})$.

Theorem 2 is quite general and simple to prove. This result provides almost the correct order for the mixing time, missing a logarithmic factor. It provides bounds that are sharp up to a logarithmic factor in the case of exchangeability (defined in (7) below).

Theorem 2 (General Lower Bound for $\mathbf{t_{mix}}$) *Suppose that* $(\mathbf{X}^{(N)})_N \in \mathscr{C}$ *and each process in the sequence is time homogeneous, i.e.* $(Z_i^{(N)})_{i\in\mathbb{N}}$ *are identically distributed for each* N. *Define*

$$\theta_N = \min_{j\in[N]} \mathbb{P}(Z^{(N)}[j] = 1).$$

There exists $a > 0$ *such that* $t_{mix}(\varepsilon) \geq a\theta_N^{-1}$, *where we set* $a/0 = \infty$. *Notice that* $\theta_N > 0$ *guarantees irreducibility of the Markov chain* $\mathbf{X}^{(N)}$.

Definition 3 A random variable Z which takes values on $\mathscr{V}_N$ is said to be exchangeable if

$$\mathbb{P}(Z = \boldsymbol{x}) = \mathbb{P}(Z = \boldsymbol{y}) \qquad \text{whenever } \|\boldsymbol{x}\| = \|\boldsymbol{y}\|. \tag{7}$$

Definition 4 Let $\mathbb{P}$ be a measure and $\mathbb{Q}$ a positive measure both defined on the subsets of $\mathscr{V}_N$. Define

$$\chi^2(\mathbb{P}\,|\mathbb{Q}) = \sum_{\boldsymbol{y}\in\mathscr{V}_N} \frac{\big(\mathbb{P}(\boldsymbol{y}) - \mathbb{Q}(\boldsymbol{y})\big)^2}{\mathbb{Q}(\boldsymbol{y})}.$$

We set $\chi^2(\boldsymbol{x}, t) = \chi^2(P_t(\cdot \mid \boldsymbol{x})\,|\pi_N)$. Moreover, let

$$t^{(2)}_{mix}(\varepsilon, \boldsymbol{x}) = \inf\{t \colon \chi^2(\boldsymbol{x}, t) \le \varepsilon\},$$

and $t^{(2)}_{\text{mix}}(\varepsilon) = \sup_{\boldsymbol{x}\in\mathscr{V}_N} t_{\text{mix}}(\varepsilon, \boldsymbol{x})$.

Definition 5 Consider a sequence $(\mathbf{X}^{(N)})_N \in \mathscr{C}$. We say that this sequence mixes *almost perfectly* in t_0 steps, where $t_0 \in \mathbb{N}$ is fixed, if

$$\lim_{N\to\infty} \sup_{\boldsymbol{x}\in\mathscr{V}_N} \|P_{t_0}(\cdot \mid \boldsymbol{x}) - \pi_N(\cdot, p_N)\|_{TV} = 0.$$

Theorem 3 (Almost-Perfect Mixing in Two Steps) *Consider a sequence* $(\mathbf{X}^{(N)})_N \in \mathscr{C}$. *We make the following assumptions.*

- $p_N > 1/2$ *for all* $N \in \mathbb{N}$ *and* $p_N \to p \ge 1/2$.
- *For* $N \in \mathbb{N}$ *each of the random variables* $Z_1^{(N)}, Z_2^{(N)}$ *is exchangeable within coordinates in the sense of definition 3.*
- *For* $m = 1, 2$ *let* $\zeta_{Nm} = \|Z_m^{(N)}\|$ *and assume that there exists random variables* V_m *such that*

$$\begin{aligned} &\lim_{N\to\infty} \tfrac{\zeta_{Nm}-Np}{\sqrt{Npq}} = V_m && \textit{(in distribution)} \\ &\sup_N \mathbb{E}\left[\tfrac{(\zeta_{Nm}-Np)^a}{(Npq)^{a/2}}\right] < \infty && \textit{for some } a > 1, \\ &\mathbb{E}[\mathrm{e}^{V_j^2/2}] < \infty,\ j = 1, 2. \end{aligned} \tag{8}$$

Then $(\mathbf{X}^{(N)})_{N\in\mathbb{N}}$ *mixes almost-perfectly in two steps.*

Remark 4 The proof of Theorem 3 provides explicit estimates for the error in the total variation for any fixed $t \in \mathbb{N}$. In particular, under the assumption of Theorem 3, we have that

$$\sup_{\boldsymbol{x}\in\mathscr{V}_N} \|P_t(\cdot \mid \boldsymbol{x}) - \pi_N(\cdot, p_N)\|_{TV} \le C\left(\frac{2q_N}{Np_N}\right)^{t-1}.$$

where C depends on V_1 and V_2 only.

The next result describes a cut-off phenomena for NLRWH (recall Example 2).

Theorem 4 (Cut-off for NLRWH) *Let* $(\mathbf{X}^{(N)}) \in \mathscr{C}$ *and assume that each* $\mathbf{X}^{(N)}$ *are defined as in Example 2, with stationary distribution* $\pi_N(\cdot, p_N)$.

1. *If* $\lim_{N\to\infty} z_N/N = 0$, *then both* $t_{mix}(\varepsilon)$ *and* $t^{(2)}_{mix}(\varepsilon)$ *exhibit a sharp cutoff at* $\frac{Np_N}{2z_N}\log N$. *That is, setting*

$$t_C = \frac{Np_N}{2z_N}(\log N + C),$$

for all $C < 0$ *small enough,* $t_{mix}(\varepsilon), t^{(2)}_{mix}(\varepsilon) > t_C$, *and for all* C *large enough* $t_{mix}(\varepsilon), t^{(2)}_{mix}(\varepsilon) < t_C$.

2. *If* $z_N/N = w \in (0, 1] \setminus \{p\}$, *then* $t^{(2)}_{mix}(\varepsilon)$ *exhibits a sharp cutoff at*

$$t_C = \frac{p_N}{2w}(\log N + C).$$

Let $\overline{P}_t(\cdot\ |\boldsymbol{x})$ be the p.m.f. of $\|X_t\|$ conditional on $X_0 = \boldsymbol{x}$. Let $\mathbb{Q}_N$ be a Binomial with parameters N and p. Define χ^2 for the Hamming distance as $\chi^2_H(\boldsymbol{x}, t) = \chi^2(\overline{P}_t(\cdot\ |\boldsymbol{x}) \mid \mathbb{Q}_N)$.

$$\overline{t}^{(2)}_{\mathrm{m}ix}(\varepsilon) = \inf\{t\colon \max_{\boldsymbol{x}} \chi^2_H(\boldsymbol{x}, t) \le \varepsilon\}.$$

In what follows we use the following notation. For two sequences $(a_N)_N$ and $(b_N)_N$ of real numbers $a_N \sim b_N$ if an only if

$$\lim_{N\to\infty} \frac{a_N}{b_N} = 1.$$

Theorem 5 (De Finetti Sequences—Mixing Time) *Suppose that in the de Finetti case described in Remark 3* $p > q$, *and* $\nu(u)$ *has a density* $h(u)$ *with respect to the Lebesque measure satisfying* $h(u) \sim cu^\gamma$ *as* $u \to 0$ *for* $c > 0$ *and* $\gamma > -1$. *Then*

$$\overline{t}^{(2)}_{mix} = J\frac{N}{\log N}, \ \textit{where}\ J = \frac{-\log q}{2(\gamma + 1)}.$$

This mixing time should be compared with the case were ν_N is a dirac mass at a point $\alpha \in (0, 1) \setminus \{p\}$, i.e. i.i.d. updates (also studied in [15]). In this context, using (6), we have

$$\chi^2_H(\mathbf{0}, t) = \sum_{n=1}^{N} \binom{N}{n} \left(\frac{p}{q}\right)^n \left(1 - \frac{\alpha}{p}\right)^{2tn} = \left(1 + \frac{p}{q}\Big(1 - \frac{\alpha}{p}\Big)^{2t}\right)^N - 1. \tag{9}$$

The latter equation shows a completely different behaviour. In fact, it is shown in Sect. 8.2 that $\chi^2_H(\mathbf{0}, t) = \sup_{\boldsymbol{x}\in\mathscr{V}_N} \chi^2_H(\boldsymbol{x}, t)$. Equation (9) shows that when $\nu_{N,t}$ is a dirac mass at $\alpha \neq p$ then it has a cut-off at $b \ln N$ when we consider the χ^2 distance, with b depending on α only. Hence it mixes much faster than the exchangeable case in Theorem 5.

Theorem 6 (Constant Order Mixing at Critical Initial Conditions) *Fix* $\varepsilon > 0$. *Let* $(\mathbf{X}^{(N)}) \in \mathscr{C}$ *and assume that each* $\mathbf{X}^{(N)}$ *is defined as in Example 2, with stationary distribution* $\pi_N(\cdot, p_N)$. *If* $\|\boldsymbol{x}\|/N = p$ *and* $z_N = wN$, *with* $w > 0$, *then there exists* t_ε *not depending on* N *such that* $\bar{t}^{(2)}_{mix}(\varepsilon) \leq t_\varepsilon$.

The following representation characterizes the process in $\mathscr{G}$ in terms of N (possibly) dependent random walks.

Theorem 7 (Random Walk Representation) *Suppose that* $\mathbf{X} \in \mathscr{G}$. *We have*

$$P_t(\boldsymbol{y} \mid \boldsymbol{x}) = \pi(\boldsymbol{y}, p)\mathbb{E}\Bigg[\prod_{j=1}^{N}\Bigg(1 - \Big(-\frac{q}{p}\Big)^{S^{(t)}[\boldsymbol{x},\boldsymbol{y},j]}\Bigg)\Bigg], \tag{10}$$

where $\boldsymbol{S}^{(t)}[\boldsymbol{x}, \boldsymbol{y}, j] = \boldsymbol{x}[j]+\boldsymbol{y}[j]-1+\sum_{k=1}^{t} Z_k[j]$, *and the parameters* $p \geq q$ *are the same as in* (2). *The vectors* $(Z_j)_{j\in\mathbb{N}}$ *are independent and their distribution is defined in Definition 1. Vice versa, if the random vectors* $(Z_j)_{j\in\mathbb{N}}$ *are independent, then the process with transition functions defined as in* (10) *belongs to* $\mathscr{G}$.

In the following example we discuss some implications of the random walk representation in the case of infinite de Finetti sequences.

Remark 5 (Infinite de Finetti Sequences) A Markov chain $(X_t)_{t\in\mathbb{Z}_+}$ with state space $\mathscr{V}_\infty$ is well defined if $(Z_t)_{t\in\mathbb{Z}_+}$ are homogeneous independent sequences in $\mathscr{V}_\infty$ such that their coordinates are distributed exchangeably. By de Finetti's theorem there exists a sequence of measures $(\nu_t)_{t\in\mathbb{Z}_+}$ such that for any N coordinates of Z_t, with $\boldsymbol{\eta} \in \mathscr{V}_N$.

$$\mathbb{P}\big(Z_t[j_1] = \boldsymbol{\eta}[1], \ldots, Z_t[j_N] = \boldsymbol{\eta}[N]\big) = \int_{[0,1]} r^{\|\boldsymbol{\eta}\|}(1-r)^{N-\|\boldsymbol{\eta}\|}\nu_t(dr).$$

Eigenvalues in the spectral decomposition of the transition distribution from t to $t+1$ for N coordinates are

$$\rho_{n,t} = \int_{[0,1]}\Big(1 - \frac{r}{p}\Big)^n \nu_t(dr), \qquad n \leq N,$$

and eigenvalues in the spectral decomposition of the transition distribution from 0 to $t+1$ are $\prod_{m=1}^{t} \rho_{n,m}$. The finite dimensional distributions of $(X_t)_{t\in\mathbb{Z}_+}$ are well defined and consistent because of the way the process is constructed from $(Z_t)_{t\in\mathbb{Z}_+}$. If in addition, the coordinates of X_0 are exchangeable then the coordinates of X_t are exchangable as well, and there exists a sequence of measures μ_t on [0, 1] such that for the projection of X_t to its first N coordinates $X_t^{(N)}$ satisfies, for $\boldsymbol{y} \in \mathscr{V}_N$,

$$\mathbb{P}\big(X_t^{(N)} = \boldsymbol{y}\big) = \int_{[0,1]} s^{\|\boldsymbol{y}\|}(1-s)^{N-\|\boldsymbol{y}\|}\mu_t(ds).$$

The structure of the measure μ_t is seen from the random walk representation (10). Consider the first N coordinates of X_t. Let the coordinates of X_0 be Bernoulli (γ), where γ is a random variable on [0, 1] with measure ν_0. Also let the coordinates of $Z_t[k]$ be Bernoulli ($\theta[k]$) conditionally on the random variable $\theta_t[k]$ that has distribution ν_t. The distribution of $\theta_t = (\theta_t[1], \dots \theta_t[N])$ is the product measure $\otimes_{i=1}^t \nu_i$. Conditionally on θ, we have that the random variable $\sum_{k=1}^t Z_k[j]$ is a sum of independent Bernoulli ($\theta[k]$) random variables. By exchangeability,

$$\begin{aligned} P_t(\boldsymbol{y} \mid \nu_0) &= \mathbb{E}\Bigg[\bigg(p\Big(1-\big(-\frac{q}{p}\big)^{\sum_{k=1}^t Z_k[1]+X_0[1]}\Big)\bigg)^{\|\boldsymbol{y}\|} \\ &\qquad \times\bigg(q\Big(1-\big(-\frac{q}{p}\big)^{\sum_{k=1}^t Z_k[1]+X_0[1]-1}\Big)\bigg)^{N-\|\boldsymbol{y}\|}\Bigg] \\ &:= \mathbb{E}\Big[\omega_t^{\|\boldsymbol{y}\|}(1-\omega_t)^{N-\|\boldsymbol{y}\|}\Big]. \end{aligned}$$

That is, the de Finetti measure μ_t is that of the random variable

$$\omega_t = p\Big(1-\big(-\frac{q}{p}\big)^{\sum_{k=1}^t Z_k[1]+X_0[1]}\Big).$$

It is straightforward to check that $\omega_t \in [0, 1]$.

3 Literature Review

In general it is difficult to study time inhomogeneous chains, and let alone diagonalize them (see, e.g., the work of Saloff-Coste and Zúñiga [14]).

Bassetti and Diaconis [1] consider (among other Markov chains) a random walk on the hypercube where one coordinate changes in a transition with an importance sampling scheme that takes a Markov chain with stationary distribution product Bernoulli (p') to one with p. This fits into the class $\mathscr{G}$, studied in this paper, when $p' = 1/2$ with these parameter values, but is not as general, since only one coordinate changes in a transition.

Nestoridi [11] considers a lazy random walk on the hypercube with $p = 1/2$, where either z coordinates change in an exchangeable way in a transition or the chain is lazy and does not change. In this context, the random walks were considered to be 'fair', i.e. $p = 1/2$, and 'lazy', i.e. at each stage the process would not change with probability 1/2. The latter assumption is convenient to avoid periodicity, and ensures ergodicity of the process. The critical case $z_N = N/2$ was discussed in Section 6 of [11], where an upper bound for the mixing time of order N was provided. Moreover, a lower bound for the χ^2 distance was also provided, and still of the order N (see Remark 2 on page 1297 of [11]), suggesting that the chain would not mix rapidly. Our contribution, for this particular example, is to show that laziness is the cause of a slowing down in the case of χ^2 distance. If we apply the acceptance rejection method described in the examples above, with $p_N \downarrow 1/2$, and $p_N \neq 1/2$, we can observe an almost-perfect mixing in 2 steps (see Theorem 3 above). Notice that when $p_N \neq 1/2$ the chain is aperiodic, as there is a positive probability for the coordinates not to change. Of course, this is a different model, but we can choose p_N in such a way that the similarity between the two models is quite evident. To see this, we can identify the limiting distribution of $\pi_N(\cdot, 1/2)$ with a Uniform over the interval [0, 1]. In fact, we can identify the vertices of the hypercube with a truncated binary expansion and the stationary measure is a product measure of Bernoulli(1/2). In contrast, if we consider a sequence of i.i.d. $(\xi_n)_{n\in\mathbb{N}}$ of Bernoulli(p) with $p \neq 1/2$, the limit of the $\sum_{n=1}^{\infty} \xi_n 2^{-n}$ has a distribution singular with respect to the Lebesgue measure. The latter, is a consequence of a beautiful Theorem of [8]. Hence it makes sense to consider sequences $p_N \to 1/2$. Fix $\varepsilon > 0$ and choose $p_N = 1/2 + \delta_\varepsilon / N^a$ where $a > 1$ and $\delta_\varepsilon > 0$ only depends on a and ε. From Kakutani's Theorem we have that the limiting distribution of the product Bernoulli (p_N), using the binary expansion trick, is uniformly continuous with respect to the Lebesgue measure. Denote by $\eta_\varepsilon(\cdot)$ this distribution. It is not difficult to prove that we can choose δ_ε such that the total variation distance between $\eta_\varepsilon(\cdot)$ and the uniform measure is less than ε.

There are many interesting examples of random walks on a hypercube in [2, Section 3], extended to a group theory context where cutoff times are calculated.

4 Spectral Representation via Tensor Products

Proposition 1 *If* $\mathbf{X} \in \mathscr{G}$ *then there exists constants* $(\gamma_{A,t})_{A\subseteq[N],t\in\mathbb{N}}$ *such that*

$$P_t(\boldsymbol{y} \mid \boldsymbol{x}) = \pi(\boldsymbol{y}, p)\left\{1 + \sum_{A\subseteq[N],A\neq\varnothing} \left(\prod_{m=1}^{t} \gamma_{A,m}\right)\left(\frac{p}{q}\right)^{|A|} \prod_{j\in A}\left(1 - \frac{\boldsymbol{x}[j]}{p}\right)\left(1 - \frac{\boldsymbol{y}[j]}{p}\right)\right\}. \tag{11}$$

Proof The general form of a 1-step transition density expansion for $\mathbf{X}$ is

$$\mathbb{P}(X_{t+1} = \boldsymbol{y} \mid X_t = \boldsymbol{x}) = \pi(\boldsymbol{y}, p)\left\{1 + \sum_{L,M\subseteq[N],L,M\neq\varnothing} \gamma_{LM}^{(t)} \prod_{i\in L} \frac{p - \boldsymbol{x}[i]}{\sqrt{pq}} \prod_{j\in M} \frac{p - \boldsymbol{y}[j]}{\sqrt{pq}}\right\},$$

with $\gamma_{LM}^{(t)} = \gamma_{ML}^{(t)}$, assuming reversibility with respect to π. This is a well-known expansion, named after [9], for $P_t(\boldsymbol{y} \mid \boldsymbol{x})/\pi(\boldsymbol{x})$ (also known as Fourier-Walsh basis expansion in part of the literature) using the tensor product sets

$$\Big\{\bigotimes_{i=1}^{N}\Big\{1, \frac{p-\boldsymbol{x}[i]}{\sqrt{pq}}\Big\}\Big\} \bigotimes \Big\{\bigotimes_{j=1}^{N}\Big\{1, \frac{p-\boldsymbol{y}[j]}{\sqrt{pq}}\Big\}\Big\}. \tag{12}$$

The following steps are well-known from basis theory, but we include these steps for the sake of completeness. We emphasize that we can compute the eigenvalues explicitly.

Roughly speaking, the *Lancaster* expansion applies to the ratio P_t/π in terms of the two tensor product sets which are complete orthogonal function sets on the Bernoulli product distributions on the sequences. The symmetry $\gamma_{LM}^{(t)} = \gamma_{ML}^{(t)}$ is a consequence of reversibility of $\mathbf{X}$. Moreover, for $L \nsubseteq M$, where the p.m.f. of (X_{t+1}, X_t) is $P_{t+1}(\boldsymbol{x}_{t+1} \mid \boldsymbol{x}_t)\pi(\boldsymbol{x}_t, p)$, we have

$$\begin{aligned}
\gamma_{LM}^{(t)} &= \mathbb{E}\Big[\prod_{i\in L}\frac{p-X_t[i]}{\sqrt{pq}}\prod_{j\in M}\frac{p-X_{t+1}[j]}{\sqrt{pq}}\Big] \\
&= \mathbb{E}\Big[\prod_{i\in L}\frac{p-X_t[i]}{\sqrt{pq}}\mathbb{E}\Big[\prod_{j\in M}\frac{p-X_{t+1}[j]}{\sqrt{pq}} \mid X_t\Big]\Big] \\
&= \mathbb{E}\Big[\prod_{i\in L\setminus M}\frac{p-X_t[i]}{\sqrt{pq}}\mathbb{E}\Big[\prod_{j\in M}\frac{p-X_{t+1}[j]}{\sqrt{pq}}\prod_{k\in L\cap\, M}\frac{p-X_t[k]}{\sqrt{pq}} \mid X_t(M)\Big]\Big] = 0
\end{aligned}$$

if $L \nsubseteq M$, in virtue of Condition 2. Using (11) we get the general representation for $P_t(\boldsymbol{y} \mid \boldsymbol{x})$ in (4), using the orthogonality of the functions (12). □

Also the reverse is true.

Proposition 2 *If* $\mathbf{X}$ *is a reversible Markov process which satisfies Condition 1 and whose transition kernel satisfies* (11)*, then* $\mathbf{X} \in \mathscr{G}$.

Proof It is enough to prove that $\mathbf{X}$ satisfies Condition 2. The marginal distribution of $X_{t+1}(B)$ conditional to the event $X_t = \boldsymbol{x}$, is

$$\begin{aligned}
&\mathbb{P}(X_{t+1}(B) = \boldsymbol{y}(B) \mid X_t = \boldsymbol{x}) = \\
&\quad \pi(\boldsymbol{y}(B), p)\Big\{1 + \sum_{A\subseteq B, A\neq\varnothing}\rho_{A,t}\Big(\frac{p}{q}\Big)^{|A|}\prod_{j\in A}\Big(1-\frac{\boldsymbol{x}[j]}{p}\Big)\Big(1-\frac{\boldsymbol{y}[j]}{p}\Big)\Big\},
\end{aligned} \tag{13}$$

where $\pi(\boldsymbol{y}(B), p)$ is the probability that a vector of $|B|$ independent Bernoulli's(p) equals $\boldsymbol{y}(B)$. The right-hand side of (13) depends on $\boldsymbol{x}(B)$ only, so

$$\mathbb{P}(X_{t+1}(B) = \boldsymbol{y}(B) \mid X_t(B) = \boldsymbol{x}) = \mathbb{P}(X_{t+1}(B) = \boldsymbol{y}(B) \mid X_t(B) = \boldsymbol{x}(B)),$$

and this proves our result. □

5 Proof of Theorem 1

5.1 General Construction of the Process X

In this Section we provide a construction for any reversible $\mathbf{X}$ on $\mathscr{V}_N$ which satisfies Conditions 1 and 2. We explicitly construct a collection of reversible Markov processes $\mathscr{G}'$, using an acceptance/rejection method. Soon after, we prove that $\mathscr{G} = \mathscr{G}'$ (Theorem 8 below).

A process $\mathbf{X} \in \mathscr{G}'$ if and only if it can be constructed as follows. Let q, p as in (2), and recall $q \le p$, and $q + p = 1$. Consider a sequence of independent random variables $(Z_t)_{t\in\mathbb{N}}$ which take values in $\mathscr{V}_N$, not necessarily homogeneous in time. Let $(\xi_{i,t})_{t\in\mathbb{N}, i\in[N]}$ be a sequence of i.i.d. Bernoulli(q/p), i.e. $\mathbb{P}(\xi_{i,t} = 1) = q/p = 1 - \mathbb{P}(\xi_{i,t} = 0)$. Consider the following homogeneous Markov process, $\mathbf{X} = (X_t)_{t\in\mathbb{N}}$, which we define recursively. Suppose $(X_i : i \le t)$ is defined, then define X_{t+1} as follows. For all $i \in [N]$,

- If $Z_t[i] = 0$ then $X_{t+1}[i] = X_t[i]$.
- If $X_t[i] = 0$ and $Z_t[i] = 1$ then $X_{t+1}[i] = 1$.
- If $X_t[i] = 1$ and $Z_t[i] = 1$, then $X_{t+1}[i] = X_t[i] + \xi_{i,t} \mod 2$.

Theorem 8 $\mathscr{G} = \mathscr{G}'$.

Proof We first prove that $\mathscr{G}' \subseteq \mathscr{G}$. It is enough to prove the case $t = 0$. Assume $\mathbf{X} \in \mathscr{G}'$. A coordinate i is chosen if and only if $Z_1[i] = 1$. Recall that conditionally on Z_1, the coordinates that are chosen behave independently. Hence,

$$\begin{aligned}&\mathbb{E}\big[X_1[i] - p \mid X_0 = \boldsymbol{x}, Z_1\big] = (1 - Z_1[i])(\boldsymbol{x}[i] - p)\\ &+Z_1[i]\big((1 - \boldsymbol{x}[i])(1 - p) + \boldsymbol{x}[i](-p(q/p) + (1 - p)(1 - (q/p))\big)\\ &= \left(1 - \frac{Z_1[i]}{p}\right)(\boldsymbol{x}[i] - p).\end{aligned}$$

The right side of (14) only depends on $x(A)$ so Condition 2 is satisfied. Therefore for $A \subseteq [N]$

$$\mathbb{E}\Big[\prod_{j\in A}(X_1[j] - p) \mid X_0 = \boldsymbol{x}\Big] = \mathbb{E}\Big[\prod_{i\in A}\left(1 - \frac{Z_1[i]}{p}\right)\Big]\prod_{j\in A}(\boldsymbol{x}[j] - p).$$

If X_0 has an N-product Bernoulli (p) measure then

$$\mathbb{E}\Big[\mathbb{E}\Big[\prod_{j\in A}(X_1[j]-p)\mid X_0\Big]=\mathbb{E}\Big[\prod_{i\in A}\Big(1-\frac{Z_1[i]}{p}\Big)\Big]\mathbb{E}\Big[\prod_{j\in A}(X_0[j]-p)\Big]=0,$$

showing that $\pi(\boldsymbol{y},p)$ is the stationary distribution of the process defined in $\mathscr{G}'$. That is, Condition 1 holds.

Next we prove that $\mathscr{G}\subseteq\mathscr{G}'$. Take Z_1 to have the distribution of $X_1\mid X_0=\mathbf{0}$. From (14) the transition distribution of $X_1\mid X_0$ is uniquely determined by the distribution of Z_1, and Z_1 is the distribution of $X_1\mid X_0=\mathbf{0}$. Therefore $\mathscr{G}\subseteq\mathscr{G}'$. □

Proof of Theorem 1. The proof follows from a combination of Propositions 1, 2 and Theorem 8. The spectral expansion is uniquely determined by (4) and (5). □

Remark 6 (Convex Set of 2×2 *Transition Matrices)* Instead of the acceptance-rejection scheme P° defined in (1), suppose coordinates change according to 2×2 transition matrices $P_1,\ldots,P_N$ each with stationary distribution (q,p). Let $\mathscr{P}_{2\times 2}$ be the set of all 2×2 transition matrices with stationary distribution (q,p). The set $\mathscr{P}_{2\times 2}$ is convex with extreme points P°, and I. That is, if $P\in\mathscr{P}_{2\times 2}$ then there exists $\lambda\in[0,1]$ with $P=\lambda P^\circ+(1-\lambda)I$. This scheme with transition matrices $P_1,\ldots,P_N$ is *not more general* than the one with identical transition matrices P°. The extreme point representation $P_k=\lambda_k P^\circ+(1-\lambda_k)I$ for $k=1,\ldots,N$ implies that Z_t can be chosen to make the models the same.

$$\begin{aligned}Z_t[k]P_k+(1-Z[k])I&=Z_t[k]\lambda_k P^\circ+(1-Z_t[k]\lambda_k)I\\&=Z_t'[k]P^\circ+(1-Z_t'[k])I,\end{aligned}$$

where $Z_t'[k]=Z_t[k]$ with probability λ_k and otherwise is zero.

Remark 7 (Lancaster Characterization) The Lancaster problem in the context of this paper is to characterize eigenvalues $\{\rho_A\}_{A\subseteq N}$ for which (14) below is non-negative, and therefore a transition kernel. This characterization has been answered by Theorem 1. Non-negativity holds if and only if there exists $Z\in\mathscr{V}_N$ such that for $A\subseteq N$

$$\rho_A=\mathbb{E}\Big[\prod_{j\in A}\Big(1-\frac{Z[j]}{p}\Big)\Big],$$

and this is equivalent to the transition kernel belonging to a process in $\mathscr{G}$.

Proof ***(Theorem 7)*** The transition probability for X_1 given $X_0=\boldsymbol{x}$ is

$$\pi(\boldsymbol{y},p)\Big\{1+\sum_{A\subseteq[N],A\neq\varnothing}\rho_A\Big(\frac{p}{q}\Big)^{|A|}\prod_{j\in A}\Big(1-\frac{\boldsymbol{y}[i]}{p}\Big)\Big(1-\frac{\boldsymbol{x}[i]}{p}\Big)\Big\}. \tag{14}$$

Note the identity that for $\mathbf{u} \in \{0,1\}^N$, $A \subseteq [N]$,

$$\prod_{j\in A}\Big(1-\frac{\mathbf{u}[j]}{p}\Big) = \Big(-\frac{q}{p}\Big)^{\|u(A)\|}.$$

The expression in (14) can therefore be written as

$$\pi(\boldsymbol{y},p)\Big\{1+\sum_{A\subseteq[N],A\neq\varnothing}(-1)^{|A|}\mathbb{E}\Big[\Big(-\frac{q}{p}\Big)^{\|Z_1(A)\|+\|\boldsymbol{y}(A)\|+\|\boldsymbol{x}(A)\|-|A|}\Big]\Big\}$$
$$=\pi(\boldsymbol{y},p)\mathbb{E}\Big[\prod_{j=1}^{N}\Big(1-\Big(-\frac{q}{p}\Big)^{Z_1[j]+\boldsymbol{x}[j]+\boldsymbol{y}[j]-1}\Big)\Big].$$

Eq. (10) follows by replacing ρ_A in (14) with

$$\rho_A^t=\mathbb{E}\Big[\Big(-\frac{q}{p}\Big)^{\sum_{k=1}^t\|Z_k(A)\|}\Big]=\mathbb{E}\Big[\Big(-\frac{q}{p}\Big)^{S^{(t)}(A)}\Big],$$

where we used the i.i.d. assumption on the sequence of vectors $(Z_t)_{t\in\mathbb{N}}$. □

Next consider a family of well-known orthogonal polynomials on the Binomial distribution.

Definition 6 (Krawtchouk Polynomials) Define a class of polynomials $\{Q_n(x;N,p)\colon n,N\in\mathbb{Z}_+, x\in\{0,1,\dots,N\}\}$, using the generating function

$$\sum_{n=0}^{N}\binom{N}{n}Q_n(x;N,p)s^n=(1-(q/p)s)^x(1+s)^{N-x}. \tag{15}$$

Proposition 3 *The family of polynomials* $\{Q_n(x;N,p)\colon n,N\in\mathbb{N}, x\in[N]\}$ *satisfy the following properties.*

1. *They are orthogonal in the following sense:* $\mathbb{E}\big[Q_n(X;N,p)Q_m(X;N,p)\big]=\delta_{m,n}h_n^{-1}$, *where* X *is Binomial* (N,p), $h_n=\binom{N}{n}(p/q)^n$ *and the Kronecker* $\delta_{m,n}\in\{0,1\}$ *equals* 1 *if and only if* $m=n$.
2. *If* $\boldsymbol{x}\in\mathcal{V}_N$ *then the family of polynomials satisfy a symmetric function representation*

$$Q_n(\|\boldsymbol{x}\|;N,p)=\binom{N}{n}^{-1}\sum_{A\subseteq[N],|A|=n}\prod_{j\in A}\Big(1-\frac{\boldsymbol{x}[j]}{p}\Big). \tag{16}$$

Proof of the orthogonality in 1 is straghtforward using generating functions (see, e.g., [4]). The representation (16) is seen by noting that the generating function agrees with (15), since

$$1+\sum_{n=1}^{N}s^n \sum_{A\subseteq[N],|A|=n}\prod_{j\in A}\Big(1-\frac{\boldsymbol{x}[j]}{p}\Big)=\prod_{j=0}^{N}\Big(1+s\Big(1-\frac{\boldsymbol{x}[j]}{p}\Big)\Big)$$
$$=\Big(1-(q/p)s\Big)^{\|\boldsymbol{x}\|}(1+s)^{N-\|\boldsymbol{x}\|}.$$

These polynomials are scaled so that for all $n\in[N]$, $Q_n(0;N,p)=1$. The relationship with the 'usual' Krawtchouk polynomials K_n is that for any $x\in\mathbb{N}$,

$$Q_n(x;N,p)=\frac{K_n(x;N,p)}{\frac{N!}{(N-n)!}(-p)^n}.$$

Proposition 4 *If $(Z_t)_{t\in\mathbb{N}}$ are exchangeable, in the sense of Definition 3, then*

$$\rho_A\equiv\rho_{|A|}=\mathbb{E}\big[Q_{|A|}(\|Z_1\|;N,p)\big]$$

and the transition probabilities are

$$\pi(\boldsymbol{y},p)\Big\{1+\sum_{n=1}^{N}\rho_n\Big(\frac{p}{q}\Big)^n\sum_{A\subseteq[N],|A|=n}\prod_{j\in A}\Big(1-\frac{\boldsymbol{x}[j]}{p}\Big)\Big(1-\frac{\boldsymbol{y}[j]}{p}\Big)\Big\}. \tag{17}$$

Proof This follows from (16) since under exchangeability for any $A\subseteq[N]$ with $|A|=n$,

$$\rho_A=\mathbb{E}\Big[\prod_{j\in A}\Big(1-\frac{Z_1[j]}{p}\Big)\Big]$$
$$=\binom{N}{n}^{-1}\sum_{A\subseteq[N],|A|=n}\prod_{j\in A}\mathbb{E}\Big[\Big(1-\frac{Z_1[j]}{p}\Big)\Big]$$
$$=\mathbb{E}\big[Q_{|A|}(\|Z_1\|;N,p)\big].$$

□

Proposition 5 *Suppose Z_1 is exchangeable. Fix $\boldsymbol{y},\boldsymbol{x}\in\mathscr{V}_N$. We have*

$$\mathbb{P}(\|X_1\|=\|\boldsymbol{y}\|\mid X_0=\|\boldsymbol{x}\|)$$
$$=\binom{N}{\|\boldsymbol{y}\|}p^{\|\boldsymbol{y}\|}q^{N-\|\boldsymbol{y}\|}\Big\{1+\sum_{n=1}^{N}\rho_n h_n Q_n(\|\boldsymbol{x}\|;N,p)Q_n(\|\boldsymbol{y}\|;N,p)\Big\}. \tag{18}$$

Proof We want to find the distribution of $\|\boldsymbol{y}\|$ using (17). It is easiest to consider $\|\boldsymbol{y}\|$ in a permutation distribution of $\boldsymbol{y}$. Let S_N be the symmetric group of order N. We have

$$\frac{1}{N!}\sum_{\sigma\in S_N}\sum_{A\subseteq[N],|A|=n}\prod_{j\in A}\Big(1-\frac{\boldsymbol{x}[j]}{p}\Big)\Big(1-\frac{\boldsymbol{y}[\sigma(j)]}{p}\Big)$$
$$=\frac{1}{N!}\frac{(N-n)!}{n!}\left(\sum_{A\subseteq[N],|A|=n}\prod_{j\in A}\Big(1-\frac{\boldsymbol{y}[j]}{p}\Big)\right)\times\left(\sum_{A\subseteq[N],|A|=n}\prod_{j\in A}\Big(1-\frac{\boldsymbol{x}[j]}{p}\Big)\right)$$
$$=\binom{N}{n}Q_n(\|\boldsymbol{x}\|;N,p)Q_n(\|\boldsymbol{y}\|;N,p).$$

There are $\binom{N}{\|y\|}$ different $\tilde{\boldsymbol{y}}\in\mathscr{V}_N$ with $\|\tilde{\boldsymbol{y}}\|=\|\boldsymbol{y}\|$. Sum over these $\tilde{\boldsymbol{y}}$ to find the distribution of $\|\boldsymbol{y}\|$, then (17) gives (18). □

Proposition 5 is known from a generalized Ehrenfest urn in [4] and in [11] as a lazy random walk with $p=1/2$. In such a lazy random walk $\|Z_1\|$ takes values 0 or a fixed $z\in[N]$, each with probability $1/2$. Then

$$\rho_{|A|}=\frac{1}{2}\Big(1+Q_{|A|}(z;N,1/2)\Big)$$

since $Q_{|A|}(0;N,1/2)=1$.

Proposition 5 is more general than first appears in that Z_1 can be taken to be exchangeable without loss of generality. Since $|X_1|$ is invariant under a permutation distribution of labelling we can replace Z_1 with Z_1' having a permutation distribution of labels. Then

$$\rho_n=\binom{N}{n}^{-1}\sum_{A\subseteq[N],|A|=n}\rho_A.$$

Remark 8 (De Finetti Continued—Hamming Distance) The Hamming distance $\|X\|_{t\in\mathbb{Z}_+}$ behaves as the number of balls labelled 1 in an urn with N balls labelled either 0 or 1. This model falls into the generalized Ehrenfest urn models studied in [4]. In this example each ball in the urn is chosen to change state according to a de Finetti choice, then if 0 changes to 1, or if 1 changes to 0 with probability q/p. Propositions (4) and (5) (or [4]) show that the eigenvectors are Krawtchouk polynomials and the eigenvalues are

$$\rho_n=\mathbb{E}\big[Q_n(\|Z_1\|;N,p)\big]=\int_{[0,1]}\Big(1-\frac{r}{p}\Big)^n\nu_N(dr). \tag{19}$$

There is no need to explicitly do the calculation using the Krawtchouk polynomials in (19). It follows easily because

$$\rho_n = \mathbb{E}\Big[\prod_{j=1}^{n}\Big(1-\frac{Z_1[j]}{p}\Big)\Big] = \int_{[0,1]}\Big(1-\frac{r}{p}\Big)^n \nu_N(dr).$$

6 Proof of Theorem 2.

Fix N and fix a coordinate ℓ such that $\mathbb{P}(Z^{(N)}[\ell] = 1) = \theta_N$. Let $A = \{\mathbf{y} \in \mathscr{V}_N : \mathbf{y}[\ell] = 1\}$. Choose $t \le a/\theta_N$, where a is chosen as follows. The quantity $(1-\theta_N)^{1/\theta_N}$ is bounded away from 0 as long as θ_N is bounded away from 1. Choose a such that

$$(1-\theta_N)^{a/\theta_N} > 1 - p_N/2.$$

We have that

$$P_t(A \mid \mathbf{0}) = 1 \le 1 - (1-\theta_N)^{a/\theta_N} \le \frac{p_N}{2}.$$

We have that $\pi(A) = p_N$. Hence,

$$\|P_t(\cdot \mid \mathbf{0}) - \pi(\cdot, p_N)_N)\|_{TV} \ge \pi(A, p_N) - P_t(A \mid \mathbf{0}) > \frac{p_N}{2} \ge \frac{1}{4}.$$

7 Proof of Theorem 3

Definition 7 Let $(H_n(v) : n \in \mathbb{N}, v \in \mathbb{R})$ be the Hermite polynomials, which are defined through the generating function

$$\sum_{n=0}^{\infty} H_n(v)\frac{\psi^n}{n!} = e^{\psi v - \frac{1}{2}\psi^2}.$$

Notice that the $H_n(v)$ are orthogonal polynomials with respect to the standard normal distribution, i.e. for $n \ne m$, we have

$$\int_{-\infty}^{\infty} H_n(v)H_m(v)\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-v^2/2}\mathrm{d}v = n!\delta_{mn},$$

where δ_{mn} is the Kronecker delta.

Recall that $q_N = 1 - p_N$ and that $\lim_{N\to\infty} p_N = p = 1 - q$.

Proposition 6 *Under the assumptions of Theorem 3 we have*

$$\lim_{N\to\infty} \mathbb{E}\big[Q_n(\zeta_N; N, p_N)\big] = \frac{(-1)^n}{h_n^{1/2}\sqrt{n!}}\mathbb{E}[H_n(V_i)] \qquad \textit{for } i \in \{1, 2\}.$$

Proof It is enough to prove convergence in distribution, as we can use the moment condition (8) to appeal to the dominated convergence theorem. In turn, in order to prove the convergence in distribution, it is enough to prove that for any sequence z_N such that

$$\lim_{N\to\infty} \frac{z_N - Np_N}{\sqrt{Np_N(1-p_N)}} = v$$

for some number v, we have that

$$\lim_{N\to\infty} h_n^{1/2} Q_n(z_N; N, p_N) = \frac{(-1)^n}{(n!)^{1/2}} H_n(v). \tag{20}$$

We prove the convergence in (20) using a generating function approach.
Note that for fixed n, as $N \to \infty$ with $p_N \to p$ we have

$$(n!)^{1/2} h_n^{1/2} = \Big(n!\binom{N}{n}(p_N/q_N)^n\Big)^{1/2} \sim \Big(N(p/q)\Big)^{n/2}.$$

Hence, we get the following estimate, which holds for all $z, N \in \mathbb{N}$ and $p \in [1/2, 1)$,

$$\begin{aligned}\sum_{n=0}^{N}(n!)^{1/2}h_n^{1/2}Q_n(z; N, p_N)\frac{s^n}{n!} &\sim \sum_{n=0}^{N}\binom{N}{n}Q_n(z; N, p)\Big(\sqrt{\frac{p}{Nq}}s\Big)^n \\ &= \Big(1-(q/p)\sqrt{\frac{p}{Nq}}s\Big)^z\Big(1+\sqrt{\frac{p}{Nq}}s\Big)^{N-z}.\end{aligned} \tag{21}$$

Taking the logarithm of both sides of (21) and setting $a = (q/p)\sqrt{(p/q)}s = \sqrt{(q/p)}s$, $b = \sqrt{(p/q)}s$, we have

$$\begin{aligned}&\ln\sum_{n=0}^{n}(n!)^{1/2}h_n^{1/2}Q_n(z; N, p)\frac{s^n}{n!} \\ &\sim z\log\Big(1-\frac{a}{\sqrt{N}}\Big) + (N-z)\log\Big(1+\frac{b}{\sqrt{N}}\Big) \\ &= -z\Big(\frac{a+b}{\sqrt{N}} + \frac{1}{2}\frac{a^2-b^2}{N}\Big) + N\Big(\frac{b}{\sqrt{N}} - \frac{1}{2}\frac{b^2}{N}\Big) + \mathscr{O}(N^{-1/2})\end{aligned}$$

$$= -\sqrt{N}p(a+b) + \sqrt{N}b - v\sqrt{pq}(a+b) - \frac{1}{2}(a^2-b^2)p - \frac{1}{2}b^2$$
$$-\frac{1}{2}(a^2-b^2)\sqrt{pq}vN^{-1/2} + \mathscr{O}(N^{-1/2}). \tag{22}$$

We have the following simplifications in (22)

$$-p(a+b)+b = -pa+qb = -\sqrt{pq}s + \sqrt{pq}s = 0$$
$$-\sqrt{pq}(a+b) = -qs - ps = -s$$
$$-(a^2-b^2)p - b^2 = -\Big(\frac{q}{p} - \frac{p}{q}\Big)ps^2 - \frac{p}{q}s^2 = -s^2$$

so (22) is equal to

$$-vs - \frac{1}{2}s^2 + v\mathscr{O}(N^{-1/2}).$$

That is, the generating function (21) is equal to

$$\exp\big\{-vs - \frac{1}{2}s^2 + v\mathscr{O}(N^{-1/2})\big\},$$

which converges to the generating function of $(-1)^n H_n(v)$. □

Using a Césaro sum argument, we immediately get from Proposition 6 the following result.

Corollary 1 *Suppose $p_N \to p \in [1/2, 1]$ and fix $t > 0$, we have*

$$\lim_{N\to\infty} \frac{\sum_{n=1}^{N}\binom{N}{n}\Big(\frac{p_N}{q_N}\Big)^n \prod_{m=1}^{t}\rho_{nm}^2}{\sum_{n=1}^{N}\frac{1}{N^{n(t-1)}}\Big(q_N/p_N\Big)^{n(t-1)}\frac{1}{n!}(\prod_{m=1}^{t}\mathbb{E}[H_n(V_m)])^2} = 1. \tag{23}$$

Proof (Proof of Theorem 3) It is well-known (e.g. see Lemma 12.16 in [10]) that for a reversible time-homogeneous Markov chain

$$\chi^2(\boldsymbol{x}, t) = \sum_{n\geq 1}\lambda_n^{2t} f_n(x)^2, \tag{24}$$

where $\lambda_0 = 1$ and $\{\lambda_n\}_{n\geq 0}$ are eigenvalues and $f_n(x)$ orthonormal eigenvectors with respect to the stationary distribution. A similar formula to (24) also holds for a chain which is non-homogeneous in time, with the transition kernels having the same eigenfunctions, and eigenvalues λ_{nm}, $m = 1,\dots,t$. Then

$$\chi^2(\boldsymbol{x}, t) = \sum_{n\geq 1}\Big(\prod_{m=1}^{t}\lambda_{nm}^2\Big) f_n(x)^2,$$

In our context with a non-homogeneous chain

$$\begin{aligned}\chi^2(\boldsymbol{x},t) &= \sum_{A\subseteq[N]:\ A\neq\varnothing}\Big(\prod_{m=1}^{t}\rho_{Am}^2\Big)\prod_{i\in A}\Big(\frac{p_N-\boldsymbol{x}[i]}{\sqrt{p_Nq_N}}\Big)^2\\ &= \sum_{A\subseteq[N]:\ A\neq\varnothing}\Big(\prod_{m=1}^{t}\rho_{Am}^2\Big)\Big(\frac{p_N}{q_N}\Big)^{|A|}\prod_{i\in A}\Big(1-\frac{\boldsymbol{x}[i]}{p_N}\Big)^2\\ &\le \sum_{n=1}^{N}\binom{N}{n}\Big(\frac{p_N}{q_N}\Big)^n\prod_{m=1}^{t}\rho_{nm}^2.\end{aligned}$$

Next we focus on the case $t = 2$. Notice that the bound in (25) is sharp, as it is achieved for initial condition $\boldsymbol{x} = \mathbf{0}$. Hence,

$$\max_{\boldsymbol{x}}\chi^2(\boldsymbol{x},2) = \chi^2(\mathbf{0},2) = \sum_{n=1}^{N}\binom{N}{n}\Big(\frac{p_N}{q_N}\Big)^n\prod_{m=1}^{2}\rho_{nm}^2.$$

In order to have an estimate of the χ^2 distance, i.e. the numerator in the right-hand side of (23), we simply need an estimate, of

$$\sum_{n=1}^{N}\frac{1}{N^n}\Big(\frac{q_N}{p_N}\Big)^n\frac{1}{n!}\prod_{m=1}^{2}\mathbb{E}[H_n(V_m)])^2. \tag{25}$$

To this end, use the following well-known formula (see, e.g., [13, 18.10.10 p. 448]) which holds for any $v\in\mathbb{R}$,

$$H_n(v) = \frac{2^{n+1}}{\sqrt{\pi}}e^{v^2/2}\int_0^\infty e^{-\tau^2}\tau^n\cos\big(\sqrt{2}v\tau-\frac{1}{2}n\pi\big)d\tau.$$

Therefore

$$\begin{aligned}|H_n(v)| \le \frac{2^{n+1}}{\sqrt{\pi}}e^{v^2/2}\int_0^\infty e^{-\tau^2}\tau^n d\tau &= \frac{2^{n+1}}{\sqrt{\pi}}e^{v^2/2}\frac{1}{2}\Gamma\Big(\frac{n}{2}+\frac{1}{2}\Big)\\ &= \begin{cases} e^{v^2/2}\frac{(2m)!}{2^m m!} & n=2m\\[2ex] \frac{2^{2m+1}}{\sqrt{\pi}}e^{v^2/2}m! & n=2m+1\end{cases}.\end{aligned}$$

If n is even $|H_n(v)| \leq e^{v^2/2}|H_n(0)|$. We use these estimates to provide bounds for the sum of even terms in (25) as follows

$$\begin{aligned}&\sum_{m=1}^{\lfloor N/2\rfloor} \frac{1}{N^{2m}}\Big(\frac{q_N}{p_N}\Big)^{2m}\frac{1}{(2m)!}\prod_{j=1}^{2}\mathbb{E}[H_{2m}(V_j)])^2\\ &\leq \Big(\max_{i\in\{1,2\}}\mathbb{E}[e^{V_i^2/2}]\Big)^4\sum_{m=1}^{[N/2]}\frac{1}{N^{2m}}\Big(\frac{q_N}{p_N}\Big)^{2m}\frac{1}{(2m)!}\Big(\frac{(2m)!}{2^m m!}\Big)^4.\end{aligned} \tag{26}$$

Denote the terms in the sum in the right-hand side of (26) as b_m. We have

$$\frac{b_{m+1}}{b_m} = \frac{1}{N}\cdot\frac{1}{2(m+1)}\Big(\frac{q_N}{p_N}\Big)^2\Big(\frac{2m+1}{N}\Big)^3 < 1$$

for $m+1 \leq \lfloor N/2\rfloor$. Hence, $\max_{m\leq\lfloor N/2\rfloor} b_m = b_1$, i.e. the first term in the sum. Therefore

$$\begin{aligned}&\sum_{m=1}^{\lfloor N/2\rfloor} \frac{1}{N^{2m}}\Big(\frac{q_N}{p_N}\Big)^{2m}\frac{1}{(2m)!}\prod_{j=1}^{2}\mathbb{E}[H_{2m}(V_j)]^2\\ &\qquad\leq \Big(\max_{i\in\{1,2\}}\mathbb{E}[e^{V_i^2/2}]\Big)^4\frac{N}{2}\frac{1}{\sqrt{\pi}}\frac{1}{N^2}\Big(\frac{q_N}{p_N}\Big)^2\frac{1}{2^4}\end{aligned}$$

which tends to zero as $N\to\infty$.

Next consider the odd terms,

$$\begin{aligned}&\sum_{m=0}^{\lfloor N/2\rfloor} \frac{1}{N^{(2m+1)}}\Big(\frac{q_N}{p_N}\Big)^{(2m+1)}\frac{1}{(2m+1)!}\prod_{j=1}^{2}\mathbb{E}[H_{2m+1}(V_j)]\\ &\leq \frac{1}{\sqrt{\pi}}\Big(\max_{i\in\{1,2\}}\mathbb{E}[e^{V_i^2/2}]\Big)^4\sum_{m=0}^{\lfloor N/2\rfloor}\frac{1}{N^{(2m+1)}}\Big(\frac{2q_N}{p_N}\Big)^{(2m+1)}\frac{1}{(2m+1)!}m!^4.\end{aligned} \tag{27}$$

Writing the terms in the latest sum as c_m, we have

$$\frac{c_{m+1}}{c_m} = \Big(\frac{2q_N}{p_N}\Big)^2\frac{1}{2m+3}\Big(\frac{m+1}{N}\Big)^3 \leq \frac{4}{2m+3}.$$

For $m \geq 1$, we have $4/(2m+3) < 1$ Hence $c_1 = \max_{j\geq 1} c_j$. This implies that the right-hand side of (27) is bounded by

$$c_0 + \frac{N}{2}c_1 \leq \frac{C_1}{N}$$

where C_1 is a constant depending only on V_1, V_2. Thus

$$\sup_{\boldsymbol{x}} \chi^2(\boldsymbol{x}, 2) \le \frac{C}{N},$$

where C is a constant depending only on V_1, V_2. □

8 Proof of Theorem 4

8.1 Lower Bound for t_{mix}

The following Theorem is due to David Wilson (see, e.g., Theorem 13.5, p172 in [10]).

Theorem 9 (Wilson Bound) *Let* $\mathbf{X}$ *be an irreducible aperiodic Markov chain with state space* Ω*. Let* Φ *be an eigenfunction with eigenvalue* λ *satisfying* $1/2 < \lambda < 1$*. Fix* $0 < \varepsilon < 1$ *and let* $R > 0$ *satisfy*

$$\mathbb{E}_{\boldsymbol{x}}\Big[\big|\Phi(X_1) - \Phi(\boldsymbol{x})\big|^2\Big] \le R$$

for all $\boldsymbol{x} \in \Omega$*. Then for any* $\boldsymbol{x} \in \Omega$

$$t_{mix}(\varepsilon) \ge \frac{1}{2\log(1/\lambda)}\left[\log\left(\frac{(1-\lambda)\Phi(\boldsymbol{x})^2}{2R}\right) + \log\left(\frac{1-\varepsilon}{\varepsilon}\right)\right]. \tag{28}$$

Next, consider a sequence $(\mathbf{X}^{(N)})_{N\in\mathbb{N}} \in \mathscr{C}$. We apply Wilson's Lemma to each element of the sequence, with the choice of first eigenvalue and eigenvector pair. Then

$$\Phi_N(\boldsymbol{x}) = \|\boldsymbol{x}\| - Np,\ \lambda_N = 1 - \frac{z_N}{Np}.$$

From (47) in the Appendix,

$$\mathbb{E}_{\boldsymbol{x}}\big[X_1(X_1-1) + X_1\big] = N(N-1)p^2\Big(\rho_2 Q_2(\boldsymbol{x}; N, p_N) - 2\rho_1 Q_1(\boldsymbol{x}; N, p_N) + 1\Big)$$
$$+ Np\Big(-\rho_1 Q_1(\boldsymbol{x}; N, p_N) + 1\Big)$$

In particular

$$\mathbb{E}_0\big[X_1^2\big] \sim N(N-1)p^2\Big(\big(1 - \frac{z_N}{Np_N}\big)^2 - 2\big(1 - \frac{z_N}{Np_N}\big) + 1\Big) + Np_N\Big(-\big(1 - \frac{z_N}{Np_N}\big) + 1\Big)$$

Notice that $1/2 < \lambda_N < 1$ is satisfied if $z_N/N < p_N/2$. We apply (28) with the choice $\boldsymbol{x} = \mathbf{0}$. We also cover the case $\lim_{N\to\infty} z_N/N = 0$. There exists a constant $c > 0$ such that

$$\log \lambda_N \geq -c\frac{z_N}{Np_N}. \tag{29}$$

Using (29) in Wilson's bound (28), we get

$$\begin{aligned} t_{\text{mix}}(\varepsilon) &\geq c\frac{Np_N}{2z_N}\left[\log\left(\frac{z_N}{Np_N}\frac{(Np_N)^2}{2R}\right) + \log\left(\frac{1-\varepsilon}{\varepsilon}\right)\right] \\ &\geq c\frac{Np_N}{2z_N}\left[\log\left(\frac{Np_N z_N}{2R}\right) + \log\left(\frac{1-\varepsilon}{\varepsilon}\right)\right] \\ &\geq c\frac{Np_N}{2z_N}\left[\log N + \log\frac{p_N}{2R} + \log\frac{1-\varepsilon}{\varepsilon}\right]. \end{aligned} \tag{30}$$

The dominant term in (30) together gives that

$$t_{\text{mix}}(\varepsilon) \geq c\frac{Np_N}{2z_N}\log N + \mathscr{O}(N).$$

Notice that the bound in Eq. (29) is required just for all large N. Hence, in the case $\lim_{N\to\infty} z_N/N = 0$, we can choose any $c \in (0, 1)$. Hence, for any $\varepsilon > 0$, we have

$$t_{\text{mix}}(\varepsilon) \geq (1-\varepsilon)\frac{Np_N}{2z_N}\log N + \mathscr{O}(N).$$

Comparison between t_{mix} *and* $t_{mix}^{(2)}$.

The following proposition is well-known in the literature. A proof is added for the sake of clarity and completeness.

Proposition 7 *We have the following useful relation between total variation and* χ^2 *distances*

$$4\|P_t(\boldsymbol{x},\cdot) - \pi_N(\cdot)\|_{TV}^2 \leq \chi^2(\boldsymbol{x}, t), \qquad \text{for all } \boldsymbol{x} \in \Omega.$$

Proof

$$
\begin{aligned}
||P_t(\cdot \mid \boldsymbol{x}) - \pi_N(\cdot)||_{\mathrm{TV}} &= \frac{1}{2}\sum_{\boldsymbol{y}} |P_t(\boldsymbol{y} \mid \boldsymbol{x}) - \pi_N(\boldsymbol{y})| \\
&= \frac{1}{2}\sum_{\boldsymbol{y}} \sqrt{\pi_N(\boldsymbol{y})}|P_t(\boldsymbol{y} \mid \boldsymbol{x}) - \pi_N(\boldsymbol{y})|/\sqrt{\pi_N(\boldsymbol{y})} \\
&\le \frac{1}{2}\sqrt{\chi^2(\boldsymbol{x})},
\end{aligned}
$$

where in the last step, the Cauchy–Schwartz inequality is used □

A by-product is the following corollary.

Corollary 2 $t_{mix}(\varepsilon) \le t^{(2)}_{mix}(\sqrt{\varepsilon}/4)$.

8.2 Digression on which χ^2-Distance to Use

Recall that that in the time homogeneous case, when Z_1 is exchangeable, we have

$$
\rho_n = \mathbb{E}\Big[Q_n(\|Z_1\|; N, p)\Big] \tag{31}
$$

then

$$
\chi^2(\boldsymbol{x}, t) = \sum_{n\ge 1} h_n \binom{N}{n}^{-1} \sum_{A\subseteq[N],|A|=n} \left(\mathbb{E}\Big[\prod_{j\in A}\Big(1 - \frac{Z_1[j]}{p}\Big)\Big]\right)^{2t} \prod_{j\in A}\Big(1 - \frac{\boldsymbol{x}[j]}{p}\Big)^2. \tag{32}
$$

The χ^2 distance (32) simplifies to

$$
\chi^2(\boldsymbol{x}, t) = \sum_{n\ge 1} h_n \rho_n^{2t} \binom{N}{n}^{-1} \sum_{A\subseteq[N],|A|=n} \prod_{j\in A}\Big(1 - \frac{\boldsymbol{x}[j]}{p}\Big)^2. \tag{33}
$$

On the other hand, recall that $\overline{P}_t(\cdot \mid \boldsymbol{x})$ the p.m.f. of $\|X_t\|$ conditional on $X_0 = \boldsymbol{x}$. Let $\mathbb{Q}_N$ be a Binomial with parameters N and p. Recall the difinition of the χ^2 for the Hamming distance as $\chi^2_H(\boldsymbol{x}, t) = \chi^2(\overline{P}_t(\cdot \mid \boldsymbol{x}) \mid \mathbb{Q}_N)$. Then

$$
\chi^2_H(\boldsymbol{x}, t) = \sum_{n\ge 1} h_n \rho_n^{2t} Q_n(\|\boldsymbol{x}\|; N, p)^2. \tag{34}
$$

In general

$$
\chi^2(\boldsymbol{x}, t) \ge \chi^2_H(\boldsymbol{x}, t)
$$

which accords with intuition. This is because

$$\binom{N}{n}^{-1} \sum_{A\subseteq[N],|A|=n} \prod_{j\in A} \left(1-\frac{\boldsymbol{x}[j]}{p}\right)^2 \geq \left(\binom{N}{n}^{-1} \sum_{A\subseteq[N],|A|=n} \prod_{j\in A} \left(1-\frac{\boldsymbol{x}[j]}{p}\right)\right)^2$$
$$= Q_n(\|\boldsymbol{x}\|, N, p)^2.$$

On the other hand, we already proved that the supremum over $\boldsymbol{x}$ of the two distinct χ^2 distances (33) and (34) coincide, and occurs when $\boldsymbol{x} = \mathbf{0}$. In other words, $\sup_{\boldsymbol{x}} \chi_H^2(\boldsymbol{x}, t)) = \chi_H^2(\mathbf{0}, t)$ and

$$\sup_{\boldsymbol{x}\in\mathscr{V}_N} \chi^2(\boldsymbol{x}, t) = \chi^2(\mathbf{0}, t) = \chi_H^2(\mathbf{0}, t) = \sum_{n\geq 1} h_n \rho_n^{2t}.$$

The reasoning above implies that if we look at the worse-case scenario, in terms of initial configurations, χ^2 and χ_H^2 behave in the same way. On the other hand, it is possible to choose initial conditions that make χ_H^2 much smaller than χ^2, resulting in a faster mixing for the Hamming distance. This gives the intuition behind Theorem 6.

Upper Bound for $t_{mix}^{(2)}$.

We assume that $\|\boldsymbol{x}\| \neq Np_N$. In virtue of our reasoning in the previous section, we can use the χ^2 distance for the Hamming distance, as when the supremum over $\boldsymbol{x}$ is taken, it coincides with χ_t^2. Recall that

$$\chi_H^2(\boldsymbol{x}, t) = \sum_{n=1}^{N} \rho_n^{2t} h_n Q_n(\|\boldsymbol{x}\|; N, p_N)^2.$$

Proposition 8 *Under the assumptions of Theorem 4, for any $x \in \{0, 1, \ldots, N\} \setminus \{Np\}$,*

$$Q_n(x, N, p_N) \sim \left(1 - \frac{x}{Np_N}\right)^n. \tag{35}$$

If $x = Np_N$,

$$Q_{2n}(Np_N; N, p_N) \sim (-q/p)^n \frac{(2n)!}{n!} \frac{1}{(2N)^n}.$$

and $Q_{2n+1}(Np_N; N, p_N)$ is of smaller order in N than $Q_{2n}(Np_N; N, p_N)$.

Proof Replace s by s/N in the generating function (15) and take the logarithm of both sides to get

$$\log\left(\sum_{n=0}^{N}\binom{N}{n}Q_n(x;N,p_N)\frac{s^n}{N^n}\right)=x\log\left(1-\frac{s}{N}\frac{q_N}{p_N}\right)+(N-x)\log\left(1+\frac{s}{N}\right). \tag{36}$$

Let $\zeta_N = x/N$ and $\alpha = q_N/p_N$. The right-hand side of (36) becomes

$$\begin{aligned}
&N\zeta_N\log(1-s\alpha_N/N)+N(1-\zeta_N)\log(1+s/N)\\
&=\zeta_N\Big(-\alpha_N s-\frac{1}{2N}\alpha_N^2 s^2\Big)+(1-\zeta_N)\Big(s-\frac{1}{2N}s^2\Big)+\mathcal{O}(N^{-2})\\
&=s-\frac{1}{2N}s^2-\zeta_N\Big(s/p_N-s^2(p_N-q_N)\frac{1}{2Np_N}\Big)+\mathcal{O}(N^{-2}).
\end{aligned} \tag{37}$$

If $\zeta_N \neq p_N$ then (37) is equal to

$$s(1-\zeta_N/p_N)+\mathcal{O}(N^{-1})$$

however if $\zeta_N = p_N$ then (37) is equal to

$$-(q_N/p_N)s^2\frac{1}{2N}+\mathcal{O}(N^{-2}).$$

Therefore for fixed $\lim_N \zeta_N \neq p$, recalling that $p = \lim_{N\to\infty} p_N$, asymptotic values are

$$Q_n(N\zeta_N;N,p_N)\sim\Big(1-\zeta_N/p\Big)^n.$$

If $\lim_N \zeta_N = p$ then

$$Q_{2n}(Np;N,p)\sim(-q/p)^n\frac{(2n)!}{n!}\frac{1}{(2N)^n}.$$

and $Q_{2n+1}(Np_N;N,p_N)$ is of smaller order in N than $Q_{2n}(Np_N;N,p_N)$. If $p=q$ then from the original generating function of $(1-s)^{N/2}(1+s)^{N/2}$

$$Q_{2n}(N/2;N,1/2)=(-1)^n\frac{\binom{N/2}{n}}{\binom{N}{2n}},$$

which agrees with the case above with $p_N \to 1/2$ as $N\to\infty$. □

Combining (31) with (35), we have that

$$\rho_n \sim \left(1 - \frac{z_N}{Np}\right)^n.$$

Hence, using (34), we have

$$\begin{aligned}
\chi_H^2(\boldsymbol{x}, t) &\sim \sum_{n=1}^{N} \left(1 - \frac{z_N}{Np}\right)^{2nt} \binom{N}{n} \left(\frac{p}{q}\right)^n \left(1 - \frac{\|\boldsymbol{x}\|}{Np}\right)^{2n} \\
&= \sum_{n=1}^{N} \frac{N^n}{n!} \left(1 - \frac{z_N}{Np}\right)^{2nt} \left(\frac{p}{q}\right)^n \left(1 - \frac{\|\boldsymbol{x}\|}{Np}\right)^{2n} \\
&\leq \exp\left\{N\left(1 - \frac{z_N}{Np}\right)^{2t} \left(1 - \frac{\|\boldsymbol{x}\|}{Np}\right)^2 \left(\frac{p}{q}\right)\right\} - 1 \\
&\leq \exp\left\{Ne^{-2t\frac{z_N}{Np}} \left(1 - \frac{\|\boldsymbol{x}\|}{Np}\right)^2 \left(\frac{p}{q}\right)\right\} - 1 \\
&\leq \exp\left\{Ne^{-2t\frac{z_N}{Np}} \left(\frac{p}{q}\right)\right\} - 1.
\end{aligned} \tag{38}$$

Choose

$$t_N = \frac{Np}{2z_N}(\log N + C) \tag{39}$$

then the upper bound in (38) is equal to

$$\exp\left\{\exp\{-C\}\left(\frac{p}{q}\right)\right\} - 1.$$

For a lower bound take the first term in the $\chi^2(\mathbf{0}, t)$ expression.

$$\begin{aligned}
\sup_{\boldsymbol{x}} \chi^2(\boldsymbol{x}, t_N) \geq \chi^2(\mathbf{0}, t_N) = \chi_H^2(\mathbf{0}, t_N) \ &\geq \ \left(1 - \frac{z_N}{Np}\right)^{2t_N} N\left(\frac{p}{q}\right) \\
&\to e^{-C}\left(\frac{p}{q}\right).
\end{aligned}$$

This shows that (39) is a cutoff time because if C is large and positive both the upper and lower bounds are small, and if C is large and negative both bounds are large. Calculations here are related to chi-squared cutoff calculations for a multinomial model in [5, Section 4.1].

Example 6 This example illustrates the difference between the chi-squared cutoff for the Hamming distance and the chi-squared cutoff for the sequences depending on the initial $\boldsymbol{x}$. Consider a model where $\|Z\| = N$ and $\|\boldsymbol{x}\| = \lfloor Nw \rfloor$. The factor in $\chi^2(\boldsymbol{x}, t)$ of

$$\binom{N}{n}^{-1} \sum_{A \subset [N], |A|=n} \Big(1 - \frac{\boldsymbol{x}[i]}{p}\Big)^2.$$

is the coefficient of $\binom{N}{n} s^n$ in the generating function

$$\Big(1 + \Big(\frac{q}{p}\Big)^2 s\Big)^{\|\boldsymbol{x}\|} \Big(1 + s\Big)^{N - \|\boldsymbol{x}\|}.$$

Replacing s by s/N

$$\Big(1 + \Big(\frac{q}{p}\Big)^2 \frac{s}{N}\Big)^{\|\boldsymbol{x}\|} \Big(1 + \frac{s}{N}\Big)^{N - \|\boldsymbol{x}\|} \to \exp\Big\{sw\frac{q^2}{p^2} + s(1-w)\Big\}.$$

Then it follows that

$$\chi^2(\boldsymbol{x}, t) \sim \sum_{n=1}^{N} \rho_n^{2t} h_n \Big(w\frac{q^2}{p^2} + 1 - w\Big)^n.$$

If $\|Z\| = N$ then $\rho_n = (-q/p)^n$ and

$$\begin{aligned}
\chi^2(\boldsymbol{x}, t) &\sim \sum_{n=1}^{N} \Big(\frac{q}{p}\Big)^{2nt} \binom{N}{n} \Big(w\frac{q}{p} + (1-w)\frac{p}{q}\Big)^n \\
&= \Big(1 + \Big(w\frac{q}{p} + (1-w)\frac{p}{q}\Big)\Big(\frac{q}{p}\Big)^{2t}\Big)^N - 1.
\end{aligned}$$

A calculation shows then, with

$$t_N = \frac{\log N + C + \log\Big(w\frac{q}{p} + (1-w)\frac{p}{q}\Big)}{-2\log(q/p)} \tag{40}$$

then

$$e^{-C} < \chi^2(\boldsymbol{x}, t) < \exp\{e^{-C}\} - 1$$

so the cutoff time calculated from $\chi^2(\boldsymbol{x}, t)$ is given by (40) compared to the Hamming distance which has a finite mixing time when $\|Z\| = N$. In fact, the Hamming distance mixes in exactly 2 steps if $\|\boldsymbol{x}\| \neq N$, and one step if $\|\boldsymbol{x}\| = N$.

9 Proof of Theorem 5

A lower bound for the mixing time is now obtained. Let $t = aN/(\log N)$ where $0 < a < J$. Using the computation given in Remark 3

$$\begin{aligned}
\chi_H^2(\mathbf{0}, t) &= \sum_{n=1}^{N} h_n \rho_n^{2t} \\
&= \sum_{n=1}^{N} \binom{N}{n} \left(\frac{p}{q}\right)^n \left(\int_{[0,1]} \left(1 - \frac{r}{p}\right)^n h(r) dr \right)^{2t} \\
&\geq q^{-N} \binom{N}{\lfloor pN \rfloor} p^{\lfloor pN \rfloor} q^{N - \lfloor pN \rfloor} N^{-2t} \left(\int_0^N \left(1 - \frac{u}{Np}\right)^{\lfloor pN \rfloor} h(u/N) du \right)^{2t} \\
&\sim q^{-N} N^{-2(\gamma+1)t} \frac{1}{\sqrt{2\pi pqN}} \left(\int_0^\infty e^{-u} c u^\gamma du \right)^{2t} \\
&\geq \frac{1}{\sqrt{2\pi pqN}} \exp\Big\{ -N \log q - 2(\gamma+1)aN + Ba(\log N)/N \Big\}
\end{aligned}$$

for a constant B. Since $a < (-\log q)/(2(\gamma+1))$, $\chi_H^2(\mathbf{0}, t) \to \infty$, so

$$\bar{t}_{\mathrm{mix}}^{(2)} \geq J \frac{N}{\log N}.$$

An upper bound is now obtained. Fix $\varepsilon > 0$ such that $1 - \varepsilon > q/p$ and for all $t \in (1 - \varepsilon, 1)$ then

$$h(p(1-t)) < c_\varepsilon (1-t)^\gamma \tag{41}$$

for some constant $c_\varepsilon > 0$. Moreover notice that,

$$\begin{aligned}
p \int_{-q/p}^{1} z^n h(p(1-z)) dz &= p \int_{-q/p}^{1-\varepsilon} z^n h(p(1-z)) dz + p \int_{1-\varepsilon}^{1} z^n h(p(1-z)) dz \\
&\leq p(1-\varepsilon)^n + \int_{1-\varepsilon}^{1} z^n h(p(1-z)) dz.
\end{aligned} \tag{42}$$

Hence

$$\begin{aligned}
\chi_H^2(\mathbf{0}, t) &= \sum_{n=1}^{N} \binom{N}{n} \left(\frac{p}{q}\right)^n \left(\int_{[0,1]} \left(1 - \frac{r}{p}\right)^n h(r) dr \right)^{2t} \\
&\overset{z=(1-r/p)}{=} \sum_{n=1}^{N} \binom{N}{n} \left(\frac{p}{q}\right)^n \left(p \int_{-q/p}^{1} z^n h(p(1-z)) dz \right)^{2t}
\end{aligned}$$

$$\overset{(42)}{\leq} \quad \sum_{n=1}^{N} \binom{N}{n} \left(\frac{p}{q}\right)^n \left(p(1-\varepsilon)^n + \int_{1-\varepsilon}^{1} z^n h(p(1-z))dz \right)^{2t}$$

$$\overset{(41)}{\leq} \quad \sum_{n=1}^{N} \binom{N}{n} \left(\frac{p}{q}\right)^n \left(p(1-\varepsilon)^n + c_\varepsilon p^\gamma \int_{0}^{1} z^n (1-z)^\gamma dz \right)^{2t}$$

$$\overset{\text{Beta function}}{\leq} \sum_{n=1}^{N} \binom{N}{n} \left(\frac{p}{q}\right)^n \left(p(1-\varepsilon)^n + c_\varepsilon p^\gamma \frac{\Gamma(n+1)\Gamma(\gamma+1)}{\Gamma(n+\gamma+2)} \right)^{2t} \tag{43}$$

Notice that, by a simple Stirling approximation, there exists a constant K such that

$$p(1-\varepsilon)^n + c_\varepsilon p^\gamma \frac{\Gamma(n+1)\Gamma(\gamma+1)}{\Gamma(n+\gamma+2)} \leq K n^{-\gamma-1}.$$

Next, choose $t = AN/\log N$ where A is a large enough parameter. Now approximate the sum appearing in (43) with a Riemann integral, where $\alpha = n/N$, as follows

$$\begin{aligned} (43) \leq C \int_0^1 \exp\Big\{ &- N\alpha \log\alpha - N(1-\alpha)\log(1-\alpha) + \alpha N \log(p/q) \\ &-2A(1+\gamma)N - 2A(1+\gamma)\frac{N}{\log N}(\log\alpha)\Big\} K^{2A\frac{N}{\log N}} \frac{1}{\sqrt{N}} \mathrm{d}\alpha \end{aligned} \tag{44}$$

As the unique maximizer of the function $\alpha \mapsto -\alpha\log\alpha - (1-\alpha)\log(1-\alpha) + \alpha\log(p/q)$ in the interval $[0, 1]$ is $\alpha^* = p$, we have that for all large N

$$\text{RHS of (44)} = \frac{C_1}{\sqrt{N}} \exp\{-N(\log q + 2A(1+\gamma) + o(1))\},$$

which decreases to 0 for

$$A > -\frac{\log q}{2(\gamma+1)}.$$

That is, we have shown

$$\bar{t}^{(2)}_{mix} \leq J \frac{N}{\log N},$$

completing the calculation of $\bar{t}^{(2)}_{mix}$.

10 Proof of Theorem 6

Let $Q_{2n+1}(Np; N, p) = 0$ and $Q_{2n}(Np; N, p) = (-q/p)^n \frac{(2n)!}{n!} \frac{1}{(2N)^n}$. Then

$$\chi_H^2(Np, t) = \sum_{n=1}^{[N/2]} Q_{2n}(z; N, p)^{2t} \binom{N}{n} \left(\frac{p}{q}\right)^n \left(\frac{(2n)!}{n!} \frac{1}{(2N)^n}\right)^2.$$

If $z/N = w$ and $w \neq p$, using (35) we have

$$\chi_H^2(Np, t) \sim \sum_{n=1}^{[N/2]} \left(1 - \frac{w}{p}\right)^{2nt} \binom{N}{n} \left(\frac{p}{q}\right)^n \left(\frac{(2n)!}{n!} \frac{1}{(2N)^n}\right)^2. \tag{45}$$

Let b_n denote the nth term in the sum (45). The ratio of terms is

$$\frac{b_{n+1}}{b_n} = \left(1 - \frac{w}{p}\right)^{2t} \frac{N-n}{n+1} \frac{p}{q} \frac{(2n+1)^2}{N^2} < \left(1 - \frac{w}{p}\right)^{2t} \frac{p}{q} \frac{2(N+1)}{N} < 1,$$

for $t > t_\circ$, where $t_\circ$ is a finite time not depending on N. The first term b_1 is therefore maximal for $t > t_\circ$ and

$$\chi_H^2(Np) < \frac{N}{2} b_1 = \frac{1}{2}\left(1 - \frac{w}{p}\right)^{2t} \frac{p}{q}. \tag{46}$$

Choose $t_\varepsilon > t_\circ$ such that the right-hand side of (46) is less than ε.

Proposition 9 *For any $x \in [N] \cup \{0\}$ we have that*

$$\begin{aligned} x(x-1) &= 2q^2 h_2 Q_2(x; N, p) - 2pq(N-1) h_1 Q_1(x; N, p) + N(N-1)p^2 \\ &= N(N-1)p^2 Q_2(x; N, p) - 2N(N-1)p^2 Q_1(x; N, p) + N(N-1)p^2. \end{aligned} \tag{47}$$

Proof Consider, with expectation in the Binomial (N, p) distribution

$$\begin{aligned} &\mathbb{E}\left[X(X-1)\left(1 - \frac{q}{p}s\right)^X (1+s)^{N-x}\right] \\ &= N(N-1)p^2\left(1 - \frac{q}{p}s\right)^2 \left(p\left(1 - \frac{q}{p}s\right) + q(1+s)\right)^{N-2} \\ &= N(N-1)p^2\left(1 - \frac{q}{p}s\right)^2. \end{aligned}$$

Looking at coefficients of s and s^2,

$$\mathbb{E}\big[X(X-1)Q_1(X;N,p)\big] = \binom{N}{1}^{-1} N(N-1)p^2 \times -2\frac{q}{p} = -2pq(N-1)$$

$$\mathbb{E}\big[X(X-1)Q_2(X;N,p)\big] = \binom{N}{2}^{-1} N(N-1)p^2 \times \left(\frac{q}{p}\right)^2 = 2q^2$$

$$\mathbb{E}\big[X(X-1)\big] = N(N-1)p^2,$$

which proves our result. □

Acknowledgments We thank Persi Diaconis, Tim Garoni, Yvonne Griffiths, and Evita Nestoridi for helpful comments that really improved our results.

References

1. Bassetti, F., Diaconis, P.: Examples comparing importance sampling and the Metropolis algorithm. Illinois J. Math. **50**, 67–91 (2006)
2. Diaconis, P.: Group Representations in Probability and Statistics. Lecture Notes-Monograph Series, vol. 11. Institute of Mathematical Statistics, New York (1998)
3. Diaconis, P., Gangolli, A.: Rectangular arrays with fixed margins. In: Discrete Probability and Algorithms (Minneapolis, MN, 1993). IMA Volume Mathematical Application, vol. 72, pp. 15–41. Springer, New York (1995)
4. Diaconis, P., Griffiths, R.C.: Exchangeable pairs of Bernoulli random variables, Krawtchouk polynomials, and Ehrenfest urns. Aust. N. Z. J. Stat. **54**, 81–101 (2012)
5. Diaconis, P., Griffiths, R.C.: Reproducing kernel polynomials on the multinomial distribution. J. Approx. Theory **242**, 1–30 (2019)
6. Diaconis, P., Shahshahani, M. On square roots of the uniform distribution on compact groups. Proc. Am. Math. Soc. **98**, 341–348 (1986)
7. Diaconis, P. Graham, R.L., Morrison, J.A.: Asymptotic analysis of a random walk on a hypercube with many dimensions. Random Struct. Algoritm. **1**, 51–72 (1990)
8. Kakutani, S.: On equivalence of infinite product measures. Ann. Math. **49**, 214–224 (1948)
9. Lancaster, H.O.: The Chi-Squared Distribution. Wiley, New York
10. Levin, D.A., Peres, Y., Wilmer, E.L.: Markov Chains and Mixing Times. American Mathematical Society, Providence (2009)
11. Nestoridi, E.: A non-local random walk on the hypercube. Adv. Appl. Prob. **49**, 1288–1299 (2017)
12. Nestoridi, E., Nguyen, O.: On the mixing time of the Diaconis–Gangolli random walk on contingency tables over $\mathbb{Z}/q\mathbb{Z}$. Ann. Inst. H. Poincaré Probab. Stat. **56**, 983–1001 (2020)
13. Olver, F.W.J., Lozie, D.W., Ronald, F., Boisvert, R.F., Clark, C.W. (eds.) NIST Handbook of Mathematical Functions. Cambridge University, New York (2010)
14. Saloff-Coste, L., Zúñiga, J.: Time inhomogeneous Markov chains with wave-like behavior. Ann. Appl. Probab. **20**, 1831–1853 (2010)
15. Scoppola, B.: Exact solution for a class of random walks on the hypercube. J. Stat. Phys. **143**, 413–419 (2011)

Non-Optimality of Invaded Geodesics in 2d Critical First-Passage Percolation

Michael Damron and David Harper

Abstract We study the critical case of first-passage percolation in two dimensions. Letting (t_e) be i.i.d. nonnegative weights assigned to the edges of $\mathbb{Z}^2$ with $\mathbb{P}(t_e = 0) = 1/2$, consider the induced pseudometric (passage time) $T(x, y)$ for vertices x, y. It was shown in [4] that the growth of the sequence $\mathbb{E}T(0, \partial B(n))$ (where $B(n) = [-n, n]^2$) has the same order (up to a constant factor) as the sequence $\mathbb{E}T^{\text{inv}}(0, \partial B(n))$. This second passage time is the minimal total weight of any path from 0 to $\partial B(n)$ that resides in a certain embedded invasion percolation cluster. In this paper, we show that this constant factor cannot be taken to be 1. That is, there exists $c > 0$ such that for all n,

$$\mathbb{E}T^{\text{inv}}(0, \partial B(n)) \geq (1 + c)\mathbb{E}T(0, \partial B(n)).$$

This result implies that the time constant for the model is different than that for the related invasion model, and that geodesics in the two models have different structure.

Keywords First-passage percolation · Invasion percolation · Near-critical percolation

1 Introduction

We begin with the definition of first-passage percolation (FPP). Consider the integer lattice $\mathbb{Z}^2$ with $\mathscr{E}^2$ denoting the set of nearest-neighbor edges, and let $(t_e)_{e \in \mathscr{E}^2}$ be an i.i.d. family of nonnegative random variables (edge-weights) with common distribution function F. For $x, y \in \mathbb{Z}^2$, a (vertex self-avoiding) path from x to y is a sequence $(v_0, e_1, v_1, \ldots, e_n, v_n)$, where the v_i's, $i = 1, \ldots, n-1$, are

M. Damron (✉) · D. Harper
School of Mathematics, Georgia Institute of Technology, Atlanta, GA, USA
e-mail: mdamron6@gatech.edu; dharper40@gatech.edu

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_14

distinct vertices in $\mathbb{Z}^2$ which are different from x or y, and $v_0 = x$, $v_n = y$, $e_i = \{v_{i-1}, v_i\} \in \mathcal{E}^2$. If $x = y$, the path is called a (vertex self-avoiding) circuit. We define the passage time of a path γ to be $T(\gamma) = \sum_{i=1}^{n} t_{e_i}$. For any $A, B \subset \mathbb{Z}^2$ we denote

$$T(A, B) = \inf\{T(\gamma) : \gamma \text{ is a path from a vertex in } A \text{ to a vertex in } B\}.$$

For $A = \{x\}$ we write $T(x, B)$ to mean $T(\{x\}, B)$ and similarly for $T(A, x)$. A geodesic from A to B is a path γ from A to B such that $T(\gamma) = T(A, B)$. Note that $T = T(x, y)$ as a function of vertices x, y is a psuedometric, and is a.s. a metric if and only if $F(0) = 0$. Thus $(\mathbb{Z}^2, T)$ can be regarded as a random pseudometric space.

1.1 Background and Main Result

FPP is studied as a model for fluid flow in a porous medium, or of the spread of a stochastic growth, such as a bacterial infection. It was introduced in 1965 by Hammersley and Welsh [5] and since then, researchers have developed some of the basics of the theory including asymptotics for $T(0, x)$ as $x \to \infty$, shape theorems, fluctuations of T, and geometry of geodesics (see [1] for a recent survey). Analysis of the model is quite different depending on the relationship between $F(0)$ and the critical value $p_c = 1/2$ for two-dimensional Bernoulli percolation. In the supercritical case, where $F(0) > 1/2$, there exists an infinite cluster (component) of edges with zero weight, and one can then show that $T(0, x)$ is stochastically bounded in x. (To reach x from 0, just enter the infinite cluster and travel near to x in zero time.) In the subcritical (and most studied) case, where $F(0) < 1/2$, $T(0, x)$ grows linearly in x, and there are many results and conjectures about the precise rate of growth.

In the critical case which we study here, where $F(0) = p_c = 1/2$, the (leading order) rate of growth of $T(0, x)$ is considerably more subtle and is closely related to near-critical and critical bond percolation. There is no infinite cluster of zero-weight edges, but there are large zero-weight clusters on all scales. Here, the usual "time constant," defined as

$$\mu = \lim_{n \to \infty} \frac{T(0, ne_1)}{n}$$

is known to be zero (from Kesten's result [7, Theorem 6.1] that $\mu = 0$ if and only if $F(0) \geq 1/2$), so it is natural to ask for the correct (sublinear) growth rate of T. Instead of $T(0, ne_1)$, it is more convenient to consider $T(0, \partial B(n))$, where $B(n) = [-n, n]^2$, and after important work of Chayes-Chayes-Durrett [2] and Zhang [14], it was shown by Damron-Lam-Wang in [4, Theorem 1.2] that

$$\mathbb{E}T(0, \partial B(n)) \asymp \sum_{k=1}^{\lfloor \log n \rfloor} F^{-1}\left(\frac{1}{2} + \frac{1}{2^k}\right), \tag{1}$$

where $a_n \asymp b_n$ means that b_n/a_n is bounded away from 0 and ∞, and F^{-1} is the following generalized inverse of F:

$$F^{-1}(t) = \inf\{x : F(x) \geq t\} \text{ for } t > 0.$$

To prove this result, the authors introduced an embedded invasion percolation cluster (an infinite connected subgraph I of $\mathbb{Z}^2$ containing the origin which we will define in the next section), and showed that

$$\mathbb{E}T(0, \partial B(n)) \asymp \mathbb{E}T^{\text{inv}}(0, \partial B(n)), \tag{2}$$

where T^{inv} is defined analogously to T, but only using paths which remain in I (see (5)). They then argued that $\mathbb{E}T^{\text{inv}}(0, \partial B(n)) \asymp$ the right side of (1).

The main result of our work implies that the symbol $\asymp$ in the comparison (2) cannot be replaced by the stronger $\sim$. In other words, the ratio of the left and right sides does not converge to 1: the invasion passage time is only a good approximation for the true passage time up to a constant factor. Therefore, local properties of geodesics or the passage time cannot be studied by a comparison to invasion.

Theorem 1 *Suppose that $F(0) = p_c = 1/2$. There exists $c_{1.1} > 0$ such that for all large n,*

$$\mathbb{E}[T^{inv}(0, \partial B(n)) - T(0, \partial B(n))] \geq c_{1.1} \sum_{k=1}^{\lfloor \log n \rfloor} F^{-1}\left(\frac{1}{2} + \frac{1}{2^k}\right).$$

In Sect. 1.2 below, we define the embedded invasion percolation model, and give some important properties of critical and near-critical percolation used in the paper. In Sect. 1.3, we give an outline of the proof of Theorem 1, and in Sect. 2 we give the full proof. Throughout the paper, constants will be denoted by c or C depending on whether they are large or small, and their subscripts refer to the sections in which they are defined.

Remark 1 A referee for this paper asked whether an analogue of Theorem 1 holds in an almost sure sense. Such a result would imply our theorem. We outline here how to prove that a.s.,

$$\liminf_{n\to\infty} \frac{T^{\text{inv}}(0, \partial B(n)) - T(0, \partial B(n))}{\sum_{k=1}^{\lfloor \log n \rfloor} F^{-1}\left(\frac{1}{2} + \frac{1}{2^k}\right)} > 0 \tag{3}$$

in the case that the denominator diverges with n. First, setting

$$G_i(n) = \{k \leq \lfloor \log_3 n \rfloor - 3 : 3k + i \in G\},$$

for $i = 1, 2, 3$ (and G defined as in (13)), one can show that, even though the E_k's are not independent,

$$\liminf_{n\to\infty} \frac{1}{\#G_i(n)} \sum_{k\in G_i(n)} \mathbf{1}_{E_{3k+i}} \geq c \text{ a.s.}, \tag{4}$$

where $c = c_{2.2.1}$ is such that $\mathbb{P}(E_k) \geq c$ for all k (from Proposition 1). To prove this, we can use standard results for weakly dependent random variables. Namely, if we define sigma-algebras

$$\Sigma_k = \sigma\left(\mathbf{1}_{E_k}, \mathbf{1}_{E_{k+1}}, \ldots\right)$$
$$\Sigma^k = \sigma\left(\mathbf{1}_{E_0}, \ldots, \mathbf{1}_{E_k}\right)$$

for $k \geq 0$ and the "strong-mixing coefficient"

$$\alpha_\ell = \sup_{k\geq 0} \sup_{A\in\Sigma^k, B\in\Sigma_{k+\ell}} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)|,$$

then we can prove for some $C, c' > 0$

$$\alpha_\ell \leq Ce^{-c'3^\ell} \text{ for all } \ell \geq 1.$$

This is done by approximating the event $A \in \Sigma^k$ by one in which we replace the variables $\mathbf{1}_{E_0}, \ldots, \mathbf{1}_{E_k}$ by $\mathbf{1}_{E'_0}, \ldots, \mathbf{1}_{E'_k}$, where the E'_j are defined the same way as E_j except the p_{3^j}-open path connecting $B(3^{j+2})$ to infinity is only required to connect to from $B(3^{j+2})$ to $\partial B(3^{k+\ell})$. (In this way, the E'_j are independent of B.) Once the strong-mixing bound is established, we can invoke [9, Theorem 2.10] to show a strong law of large numbers for the variables $\left(\mathbf{1}_{E_{3k+i}}\right)_{k\in\cup_n G_i(n)}$ and deduce (4).

Using (4), a summation by parts argument shows that for the same c and $i = 1, 2, 3$,

$$\liminf_{n\to\infty} \frac{\sum_{k\in G_i(n)} F^{-1}(q_{3k+i+1})\mathbf{1}_{E_{3k+i}}}{\sum_{k\in G_i(n)} F^{-1}(q_{3k+i+1})} \geq c \text{ a.s.}$$

Here, q_k is defined as in (7). Combining the cases $i = 1, 2, 3$,

$$\liminf_{n\to\infty} \frac{\sum_{\substack{k\in G \\ k\leq \lfloor \log_3 n\rfloor - 3}} F^{-1}(q_{k+1})\mathbf{1}_{E_k}}{\sum_{\substack{k\in G \\ k\leq \lfloor \log_3 n\rfloor - 3}} F^{-1}(q_{k+1})} \geq c \text{ a.s.}$$

By Lemma 2, the denominator is at least $-F^{-1}(q_0) + \frac{1}{3}\sum_{k \le \lfloor \log_3 n \rfloor - 3} F^{-1}(q_k)$, so

$$\liminf_{n\to\infty} \frac{\sum_{\substack{k\in G \\ k \le \lfloor \log_3 n \rfloor - 3}} F^{-1}(q_{k+1})\mathbf{1}_{E_k}}{\sum_{k \le \lfloor \log_3 n \rfloor - 3} F^{-1}(q_k)} \ge \frac{c}{3} \text{ a.s.}$$

Combining this with (20) and (8) gives (3).

1.2 Coupled Percolation Models

We will couple the FPP model on $(\mathbb{Z}^2, \mathscr{E}^2)$ with invasion percolation and Bernoulli percolation. To describe the coupling, we consider the probability space $(\Omega, \mathscr{F}, \mathbb{P})$, where $\Omega = [0, 1]^{\mathscr{E}^2}$, $\mathscr{F}$ is the product Borel sigma-field, and $\mathbb{P} = \prod_{e\in\mathscr{E}^2} \mu_e$, where each μ_e is the uniform measure on $[0, 1]$. Write $\omega = (\omega_e)_{e\in\mathscr{E}^2} \in \Omega$ so that the coordinates (ω_e) are i.i.d. uniform $[0, 1]$ random variables, and define the edge weights as $t_e = F^{-1}(\omega_e)$ for $e \in \mathscr{E}^2$, so that the collection (t_e) is i.i.d. with common distribution function F.

The uniform variables (ω_e) will be used for two models: invasion percolation and Bernoulli percolation.

- **Invasion percolation** is a another model for a stochastic growth which, unlike FPP, follows a greedy algorithm. Because of its relation to critical Bernoulli percolation, it is known as a model of self-organized criticality.
  To define the growth, we first define, the edge boundary ΔG of an arbitrary subgraph $G = (V, E)$ of $(\mathbb{Z}^2, \mathscr{E}^2)$ by

  $$\Delta G = \{e \in \mathscr{E}^2 : e \notin E, e \text{ has an endpoint in } V\}.$$

  Next, the invasion proceeds in discrete time, as a sequence $(G_n)_{n=0}^{\infty}$ of subgraphs of $(\mathbb{Z}^2, \mathscr{E}^2)$ as follows. Let $G_0 = (\{0\}, \emptyset)$. Given $G_i = (V_i, E_i)$, we define $E_{i+1} = E_i \cup \{e_{i+1}\}$, where e_{i+1} is the a.s. unique edge with $\omega_{e_{i+1}} = \min\{\omega_e : e \in \Delta G_i\}$, and let G_{i+1} be the subgraph of $\mathbb{Z}^2$ induced by E_{i+1}. The graph $I = \cup_{i=0}^{\infty} G_i$ is called the invasion percolation cluster (at time infinity).
  If A, B are subsets of $\mathbb{Z}^2$, we set

  $$T^{\text{inv}}(A, B) = \inf_{\gamma: A \leftrightarrow_I B} T(\gamma), \quad (5)$$

  where the infimum is over all paths from A to B which remain in the invasion I. (Here, $\inf \emptyset$ is defined as $+\infty$.) This T^{inv} is the passage time to which we compare T in Theorem 1.
- **Bernoulli percolation** is a simple model for a random network. The usual setup for Bernoulli percolation requires us to choose a parameter $p \in [0, 1]$ and

then independently declare each edge in our graph $(\mathbb{Z}^2, \mathscr{E}^2)$ to be open with probability p and closed with probability $1 - p$. Using our uniform variables, we can couple all of these models (for different values of p) to the other models described above. For each $e \in \mathscr{E}^2$ and $p \in [0, 1]$, we say that an edge e is p-open in ω if $\omega_e \leq p$ and otherwise say that e is p-closed. Then for any fixed p, the collection of p-open edges has the same distribution as the set of open edges in Bernoulli percolation with parameter p.

Next we give a couple of definitions that will help us work with these models. For $p \in [0, 1]$, a path (or circuit) is said to be p-open (respectively p-closed) if all its edges are p-open (respectively p-closed). Recall that all paths and circuits are assumed to be vertex self-avoiding. The interior of a circuit is the bounded component of its complement, when the circuit is viewed as a Jordan curve in the plane, and the interior is a subset of $\mathbb{R}^2$. The dual graph of $\mathbb{Z}^2$ is written as $((\mathbb{Z}^2)^*, (\mathscr{E}^2)^*)$, where

$$(\mathbb{Z}^2)^* = \mathbb{Z}^2 + \left(\frac{1}{2}, \frac{1}{2}\right) \text{ and } (\mathscr{E}^2)^* = \mathscr{E}^2 + \left(\frac{1}{2}, \frac{1}{2}\right).$$

For $x \in \mathbb{Z}^2$, the vertex dual to x, written x^*, is defined as $x + (1/2, 1/2)$, and for $e \in \mathscr{E}^2$, the edge dual to e, written e^*, is the unique element of $(\mathscr{E}^2)^*$ which bisects e. The percolation model on the original lattice induces one on the dual lattice in the natural way: a dual edge e^* is said to be p-open (for $p \in [0, 1]$) if e is, and is said to be p-closed otherwise.

One relation between invasion percolation and Bernoulli percolation is the following: if the invasion intersects a p_c-open cluster (maximal connected set of p_c-open edges), it must contain the whole cluster. Indeed, if it were to intersect such a cluster but not contain the entire cluster, then for all large n, there would be a p_c-open edge on the edge boundary of G_n. Due to the invasion's greedy algorithm, it therefore would only invade edges that are p_c-open for all large times, and this implies that there is an infinite p_c-open cluster, a contradiction. As a consequence of this fact, we obtain

$$\text{a.s., } I \text{ contains all } p_c\text{-open circuits around the origin.} \tag{6}$$

A central tool used to study invasion percolation is correlation length and we take its definition from [8, Eq. 1.21]. For $n \in \mathbb{N}$ and $p \in (p_c, 1]$, let

$$\sigma(n, m, p) = \mathbb{P}(\text{there is a } p\text{-open left-right crossing of } [0, n] \times [0, m]),$$

where the term "p-open left-right crossing of $[0, n]\times[0, m]$" means a path in $[0, n]\times[0, m]$ with all edges p-open which joins some vertex in $\{0\} \times [0, m]$ to some vertex in $\{n\} \times [0, m]$. For $\epsilon > 0$ and $p > p_c$, we define

$$L(p, \epsilon) = \min\{n : \sigma(n, n, p) \geq 1 - \epsilon\}.$$

$L(p, \epsilon)$ is called the (finite-size scaling) correlation length. It is known that $L(p) < \infty$ for $p > p_c$, $\lim_{p \downarrow p_c} L(p, \epsilon) = \infty$ for $\epsilon > 0$, and that there exists $\epsilon_1 > 0$ such that for all $0 < \epsilon, \epsilon' \leq \epsilon_1$, one has $L(p, \epsilon) \asymp L(p, \epsilon')$ as $p \downarrow p_c$. We will therefore define $L(p) = L(p, \epsilon_1)$ with this fixed ϵ_1 for simplicity. For $n \geq 1$, let

$$p_n = \min\{p > p_c : L(p) \leq n\},$$

and define

$$q_k = p_{3^k} \text{ for } k \geq 0. \tag{7}$$

We note here that

$$\sum_{k=1}^{n} F^{-1}(q_k) \asymp \sum_{k=1}^{n} F^{-1}\left(\frac{1}{2} + \frac{1}{2^k}\right) \text{ as } n \to \infty. \tag{8}$$

This follows from the fact that $n^{-\delta_0} < p_n - p_c < n^{-\epsilon_0}$ for some $\delta_0, \epsilon_0 > 0$ and $n \geq 2$ (explained in [4, Eq. (2.5)]) and from monotonicity of F^{-1} (for instance, see [4, Lemma 4.1]).

We list the following properties of correlation length, with references to their proofs.

1. There exist $c_{1.2.1}, C_{1.2.1} > 0$ such that

$$c_{1.2.1}\left|\log\frac{m}{n}\right| \leq \left|\log\frac{p_m - p_c}{p_n - p_c}\right| \leq C_{1.2.1}\left|\log\frac{m}{n}\right| \text{ for all } m, n \geq 1. \tag{9}$$

   This is a consequence of [11, Prop. 34].
2. There exists $c_{1.2.2} > 0$ such that for all $n \geq 1$,

$$c_{1.2.2}n \leq L(p_n) \leq n. \tag{10}$$

   This follows from [6, Eq. (2.10)].
3. There exist $c_{1.2.3}, C_{1.2.3} > 0$ such that for all $p > p_c$,

$$c_{1.2.3} \leq L(p)^2 \pi_4(L(p))(p - p_c) \leq C_{1.2.3}, \tag{11}$$

   where $\pi_4(n)$ is the probability that there are two vertex-disjoint (except their initial points) p_c-open paths connecting the origin to $\partial B(n)$, and two vertex-disjoint (except for their initial points) p_c-closed dual paths connecting the point $(1/2, 1/2)$ to $\partial B(n)$. This relation appears as [8, Prop. 34].
4. From [10, Section 4], there exists $c_{1.2.4} > 0$ such that for all $n \geq 1$,

$$\mathbb{P}(B(n) \text{ is connected to } \infty \text{ by a } p_n\text{-open path}) \geq c_{1.2.4}. \tag{12}$$

Here, "... connected to ∞ ..." means that there is an infinite vertex self-avoiding p_n-open path starting in $B(n)$.

1.3 Outline of Proof

The proof of Theorem 1 is split into two cases. At the end of Sect. 2.3, we assume that $\sum_k F^{-1}(q_k) < \infty$, and we explicitly construct an event A (whose definition is below (21)) with positive probability such that on A, for all $n \geq R$,

$$T^{\text{inv}}(0, \partial B(n)) - T(0, \partial B(n)) \geq b.$$

Here b, R are positive constants. This is sufficient to show that for $n \geq R$,

$$\mathbb{E}(T^{\text{inv}}(0, \partial B(n)) - T(0, \partial B(n))) \geq b\mathbb{P}(A) > 0,$$

and this is at least a constant times $\sum_k F^{-1}(q_k)$. The comparison (8) then finishes the proof in this case.

For the rest of the outline, we therefore assume that $\sum_k F^{-1}(q_k) = \infty$. Put $\text{Ann}(m, n) = B(n) \setminus B(m)$ for $0 \leq m \leq n$. For large n, we consider subannuli of $B(n)$ of the form $\text{Ann}(3^k, 3^{k+3})$ for $k = 0, \ldots, \lfloor \log_3 n \rfloor - 3$ and in Sect. 2.1 define events (E_k), which are illustrated in Fig. 1, depending on the state of edges in these annuli. Two of the paths involved in the definition of E_k are a p_c-open circuit around the origin in $\text{Ann}(3^k, 3^{k+1})$ and another p_c-open circuit around the origin in $\text{Ann}(3^{k+2}, 3^{k+3})$ (see γ_1^1 and γ_2^1 in Definition 1). Letting C_k and D_k be the outermost and innermost such circuits respectively, the fact that they have zero total weight and are contained in the invasion (see (6)) implies that the difference $\Delta = T^{\text{inv}} - T$ satisfies

$$\Delta(0, \partial B(n)) \geq \Delta(C_k, D_k)\mathbf{1}_{E_k}$$

and furthermore (see (19))

$$\Delta(0, \partial B(n)) \geq \Delta(C_1, D_1)\mathbf{1}_{E_1} + \Delta(C_4, D_4)\mathbf{1}_{E_4} + \cdots + \Delta(C_r, D_r)\mathbf{1}_{E_{3r+1}},$$

where r is the largest integer with $3^{3r+4} \leq n$. (Here we consider only E_k's with values of k differing by at least 3 to ensure that their associated annuli are disjoint.)

To bound the terms in the sum, we define a set of "good" indices $G = \{k : F^{-1}(q_k) \leq 2F^{-1}(q_{k+1})\}$ and we show in (15) and (16) that for such values of k, if E_k occurs, then the passage time $T^{\text{inv}}(C_k, D_k)$ is at least $3F^{-1}(q_{k+1})$, while the ordinary passage time $T(C_k, D_k)$ is at most $2F^{-1}(q_{k+1})$. This is possible because on E_k, any path in the invasion that crosses $\text{Ann}(3^{k+1}, 3^{k+2})$ must contain the edges

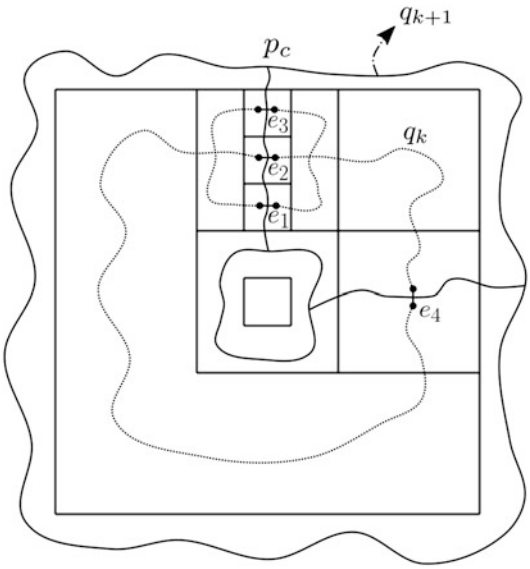


Fig. 1 Illustration of the event E_k. The innermost box is $B(3^k)$ and the outermost is $B(3^{k+2})$. The solid lines represent p_c-open paths, the dotted lines represent q_k-closed dual paths, while the curve with the arrow indicates the a q_{k+1}-open path to infinity. When E_k occurs, any path connecting $B(3^{k+1})$ to $\partial B(3^{k+2})$ remaining in the invasion I must contain the edges e_1, e_2, e_3 (not e_4), whereas a path in the original FPP model may take the edge e_4

e_1, e_2, e_3 (which have weights $\geq q_{k+1}$) shown in Fig. 1, whereas an unrestricted path may simply take edge e_4 (which has weight $\leq q_k$). This implies that

$$\mathrm{e}\Delta(C_k, D_k)\mathbf{1}_{E_k} \geq F^{-1}(q_{k+1})\mathbb{P}(E_k),$$

and combining this with the above inequality,

$$\mathrm{e}\Delta(0, \partial B(n)) \geq \left(\inf_{\ell} \mathbb{P}(E_\ell)\right) \times \sum_{\substack{k:3k+1\in G \\ 3k+3\leq \lfloor \log_3 n \rfloor}} F^{-1}(q_{3k+4}).$$

Similarly, we can obtain

$$\mathrm{e}\Delta(0, \partial B(n)) \geq \frac{1}{3} \left(\inf_{\ell} \mathbb{P}(E_\ell)\right) \times \sum_{\substack{k\in G \\ k+3\leq \lfloor \log_3 n \rfloor}} F^{-1}(q_{k+1}).$$

(Compare to (18).) In Sect. 2.2, we show that the infimum is positive, and so because the definition of G entails that

$$\sum_{\substack{k:k\in G \\ k+3\leq \lfloor \log_3 n \rfloor}} F^{-1}(q_{k+1}) \asymp \sum_{k:k+3\leq \lfloor \log_3 n \rfloor} F^{-1}(q_k)$$

(from Lemma 2), we can finish the proof with another application of (8).

2 Proof of Theorem 1

In this section, we give the proof of Theorem 1. It will be split over three subsections.

2.1 Step 1: Definition of E_k

In this section, we define events $(E_k)_{k\geq 0}$ whose occurrence allows us to give a lower bound for $T^{\text{inv}} - T$. To state this bound precisely, we define C_k to be the outermost p_c-open circuit around the origin in $\text{Ann}(3^k, 3^{k+1})$ (if it exists) and let D_k be the innermost p_c-open circuit around the origin in $\text{Ann}(3^{k+2}, 3^{k+3})$. (On E_k, these circuits will always exist—see the first two bullet points of Definition 1.) The event E_k will be constructed so that for $n \geq 0$ and $k = 0, \ldots, \lfloor \log_3 n \rfloor - 3$, if k is in a certain "good" set of indices

$$G = \{k \geq 0 : F^{-1}(q_k) \leq 2F^{-1}(q_{k+1})\}, \tag{13}$$

then

$$(T^{\text{inv}}(C_k, D_k) - T(C_k, D_k))\mathbf{1}_{E_k} \geq F^{-1}(q_{k+1})\mathbf{1}_{E_k} \text{ a.s.} \tag{14}$$

(Recall that C_k and D_k are contained in the invasion by (6).)

In the following definition, we use the notation $R(N) = [0, N] \times [0, N]$ for $N \geq 0$. Because there are many conditions comprising the event E_k, we encourage the reader to consult Fig. 1 for an illustration.

Definition 1 For $k \geq 0$ and real numbers α, β with $\alpha > 1$ and $\beta \in [0, 1)$, we define the event $E_k = E_k(\alpha, \beta)$ that the following conditions hold.

- There is a p_c-open circuit, γ_1^1, in $\text{Ann}(3^k, 3^{k+1})$, which contains $B(3^k)$ in its interior.
- There is a p_c-open circuit, γ_2^1, in $\text{Ann}(3^{k+2}, 3^{k+3})$, which contains $B(3^{k+2})$ in its interior.

There are edges

- $e_1 \in B_1 := \big(-3^k, 3^{k+1}\big) + R(2 \cdot 3^k)$ with $\omega_{e_1} \in (q_{k+1}, p_c + \alpha(q_{k+1} - p_c))$,
- $e_2 \in B_2 := \big(-3^k, 5 \cdot 3^k\big) + R(2 \cdot 3^k)$ with $\omega_{e_2} \in (q_{k+1}, p_c + \alpha(q_{k+1} - p_c))$,
- $e_3 \in B_3 := \big(-3^k, 7 \cdot 3^k\big) + R(2 \cdot 3^k)$ with $\omega_{e_3} \in (q_{k+1}, p_c + \alpha(q_{k+1} - p_c))$, and
- $e_4 \in B_4 := \big(3^{k+1}, -3^{k+1}\big) + R(2 \cdot 3^{k+1})$ with $\omega_{e_4} \in (p_c + \beta(q_k - p_c), q_k)$,

such that

- there is a p_c-open path which connects γ_1^1 to one endpoint of e_1,
- there is a p_c-open path which connects the other endpoint of e_1 to one endpoint of e_2,

- there is a p_c-open path which connects the other endpoint of e_2 to one endpoint of e_3,
- there is a p_c-open path which connects the other endpoint of e_3 to the p_c-open circuit γ_2^1,
- there is a p_c-open path which connects γ_1^1 to e_4,
- there is a p_c-open path which connects the other endpoint of e_4 to γ_2^1,
- there is a q_k-closed dual path, γ_1^2, which connects one endpoint of e_1^* to one endpoint of e_3^*,
- there is a q_k-closed dual path, γ_2^2, which connects the other endpoint of e_3^* to the other endpoint of e_1^*,
- there is a q_k-closed dual path, γ_1^3, which connects one endpoint of e_4^* to one endpoint of e_2^*,
- there is a q_k-closed dual path, γ_2^3, which connects the other endpoint of e_2^* to the other endpoint of e_4^*, and
- there is a q_{k+1}-open path which connects γ_2^1 to ∞.

Moreover,

- $\{e_4^*\} \cup \gamma_1^3 \cup \{e_2^*\} \cup \gamma_2^3$ forms a dual circuit around zero, and
- $\{e_1^*\} \cup \gamma_2^2 \cup \{e_3^*\} \cup \gamma_1^2$ forms a dual circuit around e_2^*.

From this point forward, we pick α and β satisfying the inequality the next lemma.

Lemma 1 *There exist α, β with $\alpha > 1 > \beta > 0$ such that for all $k \geq 0$,*

$$p_c + \alpha(q_{k+1} - p_c) < p_c + \beta(q_k - p_c).$$

Proof Using (9), we see that

$$3^{c_{1.2.1}} < \frac{q_k - p_c}{q_{k+1} - p_c}.$$

So we choose α and β sufficiently close to 1 that $\alpha < 3^{c_{1.2.1}}\beta$, and this implies

$$p_c + \alpha(q_{k+1} - p_c) < p_c + \beta \cdot 3^{c_{1.2.1}}(q_{k+1} - p_c) < p_c + \beta(q_k - p_c).$$

We will bound the probability $\mathbb{P}(E_k)$ from below in the next step. To finish the current step, we estimate the difference $T^{\text{inv}} - T$ when E_k occurs; that is, we now prove inequality (14). So suppose that $n \geq 0$ and $0 \leq k \leq \lfloor \log_3 n \rfloor - 3$. We will show that

$$T^{\text{inv}}(C_k, D_k)\mathbf{1}_{E_k} \geq 3F^{-1}(q_{k+1})\mathbf{1}_{E_k} \text{ a.s.} \tag{15}$$

and

$$T(C_k, D_k)\mathbf{1}_{E_k} \leq F^{-1}(q_k)\mathbf{1}_{E_k} \text{ a.s.} \tag{16}$$

If we prove these two inequalities, then, under the additional assumption that $k \in G$, we would obtain

$$(T^{\text{inv}}(C_k, D_k) - T(C_k, D_k))\mathbf{1}_{E_k} \geq (3F^{-1}(q_{k+1}) - F^{-1}(q_k))\mathbf{1}_{E_k} \geq F^{-1}(q_{k+1})\mathbf{1}_{E_k},$$

and this would show (14).

We begin by proving (15), and to do this, we show that on the event E_k, any optimal path γ_k^{inv} for $T^{\text{inv}}(C_k, D_k)$ must contain the edges e_1, e_2, and e_3. (In fact, e_1, e_2, e_3 are cut-edges for the invasion, and they separate the circuits C_k and D_k.) Since these edges have weight $t_{e_i} \geq F^{-1}(q_{k+1})$, we would then obtain $T^{\text{inv}}(C_k, D_k) \geq t_{e_1} + t_{e_2} + t_{e_3} \geq 3F^{-1}(q_{k+1})$. The argument is similar for all three edges, so we show that γ_k^{inv} contains e_2. Since γ_k^{inv} crosses $\text{Ann}(3^{k+1}, 3^{k+2})$, by duality it must contain a edge e whose dual is in $\{e_4^*\} \cup \gamma_1^3 \cup \{e_2^*\} \cup \gamma_2^3$. However, after the invasion touches the circuit γ_1^1 for the first time, it has access to infinitely many $p_c + \alpha(q_{k+1} - p_c)$-open edges (through the edges e_1, e_2, e_3). Because C_k does not intersect the interior of γ_1^1 (it lies "on or outside" γ_1^1) all the edges of γ_k^{inv} must then be $p_c + \alpha(q_{k+1} - p_c)$-open. Since γ_1^3 and γ_2^3 are q_k-closed, e_4 is $p_c + \beta(q_k - p_c)$-closed, e_2 is $p_c + \alpha(q_{k+1} - p_c)$-open, and Lemma 1 implies that

$$p_c + \alpha(q_{k+1} - p_c) < p_c + \beta(q_k - p_c) < q_k,$$

it follows that $e_2 \in \gamma_k^{\text{inv}}$. This shows (15).

To complete this step, note that because C_k is "on or outside" γ_1^1 and D_k is "on or inside" γ_2^1, there is a path π (through e_4) from C_k to D_k with passage time equal to $F^{-1}(q_k)$. This implies $T(C_k, D_k) \leq T(\pi) \leq F^{-1}(q_k)$, which is (16).

2.2 Step 2: Lower Bound on $\mathbb{P}(E_k)$

Proposition 1 *There exists* $c_{2.2.1} > 0$ *so that for all* $k \geq 0$, $\mathbb{P}(E_k) \geq c_{2.2.1}$.

Proof To give a lower bound for the probability of E_k, we use several gluing constructions, themselves composed of the RSW theorem, the (generalized) FKG inequality, and Kesten's arms separation method. Because these arguments are now standard, we will confine ourselves to a rough outline of the approach. The interested reader should pay close attention to Fig. 1 throughout the sketch.

For $i = 1, \ldots, 4$, let F_i be the event that in the box B_i in the definition of E_k, there exists an appropriate four-arm edge in the central box of half the size of B_i. Specifically, defining B_i' to be the box with half the sidelength of B_i but with the same center, we let F_i be the event that there is an edge $e_i \in B_i'$ such that $\omega_{e_i} \in I_i$, e_i is connected to the top and bottom sides of B_i by two vertex-disjoint p_c-open paths, and e_i^* is connected to the left and right sides of B_i by two vertex-disjoint q_k-closed paths, where $I_i = (q_{k+1}, p_c + \alpha(q_{k+1} - p_c))$ for $i = 1, 2, 3$. Define F_4 similarly, but with $I_4 = (p_c + \beta(q_k - p_c), q_k)$, the p_c-open paths touching the left and right sides of B_4, and the q_k-closed dual paths touching the top and bottom sides.

We claim that for some $c_{2.2.2} > 0$, one has

$$\mathbb{P}(F_i) \geq c_{2.2.2} \text{ for } i = 1, \ldots, 4, \text{ and for all } k \geq 0. \tag{17}$$

So fix such i and k and first note that for $e \subset B_i'$, if F_e is the event that the edge e satisfies the conditions described in the definition of F_i, then for distinct $e, f \subset B_i'$, the events F_e and F_f are disjoint. Therefore $\mathbb{P}(F_i) = \sum_{e \subset B_i'} \mathbb{P}(F_e)$. Because $L(q_k) \leq 3^k$ (from (10)), one can use [3, Lemma 6.3] to prove that $\mathbb{P}(F_e) \geq c_{2.2.3}\mathbb{P}(F_e')$ for some $c_{2.2.3} > 0$, where F_e' is defined similarly to F_e, but the q_k-closed paths are instead p_c-closed. Last, Kesten's arm separation method (see [11, Theorem 11]) implies that $\mathbb{P}(F_e') \geq c_{2.2.4}|I_i|\pi_4(2^k)$ for some $c_{2.2.4} > 0$, where $|I_i|$ is the length of the interval I_i and π_4 is defined below (11). Putting together these pieces, we obtain

$$\begin{aligned}\mathbb{P}(F_i) = \sum_{e \subset B_i'} \mathbb{P}(F_e) \geq c_{2.2.3} \sum_{e \subset B_i'} \mathbb{P}(F_e') &\geq c_{2.2.5}|I_i|\pi_4(2^k)2^{2k} \\ &\geq c_{2.2.6}(q_{k+1} - p_c)\pi_4(2^{k+1})2^{2k}.\end{aligned}$$

for some $c_{2.2.5}, c_{2.2.6} > 0$. Using the scaling relation stated above in (11), the right side is bounded below by $c_{2.2.7} > 0$. This demonstrates the claim in (17).

Now that we have constructed the four-arm edges in the boxes B_i, we need to create the other macroscopic connections. By the RSW theorem [12, 13], the FKG inequality, and independence, one has

$$\mathbb{P}(J) \geq c_{2.2.8} \text{ for all } k \geq 0,$$

for some $c_{2.2.8} > 0$, where J is the event that the following occur:

1. There is a p_c-open circuit around the origin in $\text{Ann}(3^k, 3^{k+1})$ which is connected by a p_c-open path in this annulus to the bottom side of B_1 and by another p_c-open path in this annulus to the left side of B_4,
2. there is a p_c-open circuit around the origin in $\text{Ann}(3^{k+2}, 3^{k+3})$ which is connected by a p_c-open path in this annulus to the top side of B_3 and by another p_c-open path in this annulus to the right side of B_4, and
3. there are q_k-closed dual paths in the following regions: (a) one connecting the left side of B_2 to the bottom side of B_4, and one connecting the left side of B_1 to the left side of B_3, all in the component of $\text{Ann}(3^{k+1}, 3^{k+2}) \setminus \cup_i B_i$ that contains

the point $(-2\cdot 3^{k+1}, 0)$, and (b) one connecting the right side of B_2 to the top side of B_4, and one connecting the right side of B_1 to the right side of B_3, all in the component of $\mathrm{Ann}(3^{k+1}, 3^{k+2}) \setminus \cup_i B_i$ that contains the point $(2\cdot 3^{k+1}, 2\cdot 3^{k+1})$.

The paths described in J must be "connected" to the four-arm edges described in the events F_i, and this is done with the generalized FKG inequality (see [11, Lem. 13]) and Kesten's arm extension method. Specifically, if $\hat{J}$ is the event that $\cap_i F_i \cap J$ occurs, but with the additional stipulations that:

1. the first open connection in $\mathrm{Ann}(3^k, 3^{k+1})$ described in item 1 of the definition of J is connected in this annulus to the "lower" p_c-open arm in B_1 and the second is connected to the "left" p_c-open arm in B_4,
2. the first open connection in $\mathrm{Ann}(3^{k+2}, 3^{k+3})$ described in item 2 of the definition of J is p_c-connected in this annulus to the "upper" p_c-open arm in B_3 and the second is p_c-connected to the "right" p_c-open arm in B_4,
3. the "upper" p_c-open arm in B_1 is p_c-connected to the "lower" p_c-open arm in B_2, and the "upper" p_c-open arm in B_2 is p_c-connected to the "lower" p_c-open arm in B_3,
4. the first q_k-closed dual path described in item 3(a) of the definition of J is q_k-connected to the "left" q_k-closed arm in B_2 and the "bottom" q_k-closed arm in B_4, and the second is q_k-connected to the "left" q_k-closed arm in B_1 and the "left" q_k-closed arm in B_3, and
5. the first q_k-closed path described in item 3(b) of the definition of J is q_k-connected to the "right" q_k-closed arm in B_2 and the "top" q_k-closed arm in B_4, and the second is q_k-connected to the "right" q_k-closed arm in B_1 and the "right" q_k-closed arm in B_3,

then

$$\mathbb{P}(\hat{J}) \geq c_{2.2.9} \text{ for all } k \geq 0.$$

Finally, we must combine the event $\hat{J}$ with the connection to infinity. Letting H be the event that there is a q_{k+1}-open path connecting $B(3^{k+2})$ to infinity, then by (12), one has $\mathbb{P}(H) \geq c_{2.2.10}$ for all $k \geq 0$. To combine this with $\hat{J}$, we again use the generalized FKG inequality. It implies that

$$\mathbb{P}(\hat{J} \cap H) \geq c_{2.2.11} \text{ for all } k \geq 0.$$

Because $\hat{J} \cap H$ implies the event E_k, this completes the sketch of the proposition.

2.3 Step 3: Good Indices and the End of the Proof

In this last step of the proof, we first prove a lemma which will imply that in the case that $(x_k) = (F^{-1}(q_k))$ is not summable, the sum of $F^{-1}(q_k)$ over all $k \in G$ for $k \leq$

n is comparable to the sum over all $k \leq n$. Recall that G is the "good" set of indices defined in (13) for which the lower bound for $T^{\text{inv}}(C_k, D_k) - T(C_k, D_k)$ from (14) holds. We will use this lemma along with (14) to prove Theorem 1 afterward.

Lemma 2 *Let $(x_k)_{k\geq 0}$ be a nonnegative monotone nonincreasing sequence. Then for all $n \geq 0$,*

$$\sum_{k\leq n} x_k \leq 3x_0 + 3\sum_{\substack{k\in G\\ k\leq n}} x_{k+1},$$

where $G = \{k : x_k/x_{k+1} \leq 2\}$.

Proof If $k, k+1, \ldots, k+m \in G^c$ and $1 \leq \ell \leq m+1$, one has

$$x_{k+\ell} \leq \frac{x_{k+\ell-1}}{2} \leq \cdots \leq \frac{x_k}{2^\ell},$$

and so

$$\sum_{\ell=0}^{m} x_{k+\ell} \leq x_k + x_k \sum_{\ell=1}^{m+1} 2^{-\ell} \leq 2x_k.$$

By partitioning G^c into a collection of maximal disjoint intervals and applying this inequality to each such interval, we obtain

$$\sum_{\substack{k\in G^c\\ k\leq n}} x_{k+1} \leq \sum_{\substack{k\in G^c\\ k\leq n}} x_k \leq 2x_0 + 2\sum_{\substack{k\in G\\ k\leq n}} x_{k+1}.$$

Here, we have used that if k is the first element of an interval in G^c, then $x_{k-1} \in G$, unless $k = 0$. Adding $x_0 + \sum_{k\in G, k\leq n} x_{k+1}$ to both sides completes the proof of the lemma.

Proof (Proof of Theorem 1) Now we prove the main theorem. The main step is to show that for $i = 0, 1, 2$,

$$T^{\text{inv}}(0, \partial B(n)) - T(0, \partial B(n)) \geq \sum_{\substack{k:3k+i\in G\\ 3k+i\leq \lfloor \log_3 n\rfloor - 3}} F^{-1}(q_{3k+i+1})\mathbf{1}_{E_{3k+i}}. \tag{18}$$

To justify this inequality, recall that for a given k such that E_k occurs, C_k and D_k have zero weight and are therefore in the invasion by (6), so one has

$$T(A, B) = T(A, C_k) + T(C_k, D_k) + T(D_k, B)$$

and

$$T^{\mathrm{inv}}(A, B) = T^{\mathrm{inv}}(A, C_k) + T^{\mathrm{inv}}(C_k, D_k) + T^{\mathrm{inv}}(D_k, B)$$

for any $A \subset B(3^k)$ and $B \subset B(3^{k+3})^c$. In this way, we decouple the passage times between circuits. For a given $i = 0, 1, 2$, therefore, if $k_1, \ldots, k_r$ satisfy $k_1 < \cdots < k_r$ and $3k_r + i \le \lfloor \log_3 n \rfloor - 3$, and E_{3k_s+i} occurs for all s, then

$$\begin{aligned} T^{\mathrm{inv}}(0, \partial B(n)) - T(0, \partial B(n)) = {} & T^{\mathrm{inv}}(0, C_{3k_1+i}) - T(0, C_{3k_1+i}) \\ & + T^{\mathrm{inv}}(D_{3k_r+i}, \partial B(n)) - T(D_{3k_r+i}, \partial B(n)) \\ & + \sum_{s=1}^{r} \Big(T^{\mathrm{inv}}(C_{3k_s+i}, D_{3k_s+i}) - T(C_{3k_s+i}, D_{3k_s+i}) \Big) \\ & + \sum_{s=1}^{r-1} \Big(T^{\mathrm{inv}}(D_{3k_s+i}, C_{3k_{s+1}+i}) - T(D_{3k_s+i}, C_{3k_{s+1}+i}) \Big). \end{aligned}$$

(Here we have chosen indices of the form $3k + i$ to ensure that the annuli associated to the events E_{3k+i} are disjoint.) Using the fact that $T^{\mathrm{inv}} \ge T$,

$$T^{\mathrm{inv}}(0, \partial B(n)) - T(0, \partial B(n)) \ge \sum_{s=1}^{r} \Big(T^{\mathrm{inv}}(C_{3k_s+i}, D_{3k_s+i}) - T(C_{3k_s+i}, D_{3k_s+i}) \Big).$$

Applying this idea to all the circuits C_{3k+i}, D_{3k+i} for $3k + i \le \lfloor \log_3 n \rfloor - 3$ with $3k + i \in G$, we obtain

$$\begin{aligned} & T^{\mathrm{inv}}(0, \partial B(n)) - T(0, \partial B(n)) \\ \ge & \sum_{\substack{k:3k+i \in G \\ 3k+i \le \lfloor \log_3 n \rfloor - 3}} (T^{\mathrm{inv}}(C_{3k+i}, D_{3k+i}) - T(C_{3k+i}, D_{3k+i})) \mathbf{1}_{E_{3k+i}}. \end{aligned} \tag{19}$$

Combining this with (14), we obtain (18).

Averaging (18) over $i = 0, 1, 2$ produces

$$T^{\mathrm{inv}}(0, \partial B(n)) - T(0, \partial B(n)) \ge \frac{1}{3} \sum_{\substack{k \in G \\ k \le \lfloor \log_3 n \rfloor - 3}} F^{-1}(q_{k+1}) \mathbf{1}_{E_k}. \tag{20}$$

By Proposition 1, this becomes

$$\mathbb{E}(T^{\mathrm{inv}}(0, \partial B(n)) - T(0, \partial B(n))) \ge \frac{c_{2.2.1}}{3} \sum_{\substack{k \in G \\ k \le \lfloor \log_3 n \rfloor - 3}} F^{-1}(q_{k+1}).$$

Lemma 2 then implies

$$\mathbb{E}(T^{\text{inv}}(0, \partial B(n)) - T(0, \partial B(n))) \geq \frac{c_{2.2.1}}{3}\left[-F^{-1}(q_0) + \frac{1}{3}\sum_{k \leq \lfloor \log_3 n \rfloor - 3} F^{-1}(q_k)\right].$$

Using (8), this implies the inequality of Theorem 1 if $\sum_k F^{-1}(q_k) = \infty$.

If $\sum_k F^{-1}(q_k) < \infty$, then we explicitly construct an event A on which geodesics in I have higher weight than true geodesics. First pick a, b with $0 < a < b$ such that $\mathbb{P}(t_e \in [a/2, a]) > 0$ and $\mathbb{P}(t_e \geq b) > 0$. (If this is impossible, then $\sum_k F^{-1}(q_k) = \infty$.) Fix an integer R which is a multiple of 10 and satisfies

$$R \geq 10b/a, \tag{21}$$

and let Γ be the set of edges of the form $\{(-n-1, 0), (-n, 0)\}$ for $0 \leq n \leq R-1$. Last, define A to be the event that

1. $B(R)$ is connected to infinity by a path of edges e with $t_e \leq a$,
2. all edges e with both endpoints in $\partial B(R)$ have $t_e = 0$,
3. all edges $e \in \Gamma$ have $t_e \in [a/2, a]$,
4. all edges $e \notin \Gamma$ with one endpoint in $B(R/2)$ and one endpoint in $B(R/2)^c$ have $t_e \geq b$
5. all edges $e \notin \Gamma$ with both endpoints within ℓ^∞ distance $R/5$ of $(-R/2, 0)$ have $t_e \geq b$, and
6. all other edges e with both endpoints in $B(R)$ have $t_e = 0$.

(See Fig. 2 for an illustration of the event A.) We claim that

$$\mathbb{E}[T^{\text{inv}}(0, \partial B(n)) - T(0, \partial B(n))] \geq b\mathbb{P}(A) > 0 \text{ for } n \geq R. \tag{22}$$

Assuming this claim, the statement of Theorem 1 follows from (8) if $\sum_k F^{-1}(q_k) < \infty$.

To show (22), we show that the difference of passage times is at least b on the event A. Because A has positive probability (conditions (2)–(6) are clear, and for condition (1), we use that $\mathbb{P}(t_e \leq a) > 1/2$, and so with positive probability, any given vertex on $\partial B(R)$ is connected to infinity by a path outside $B(R)$ all whose edges have weight $\leq a$), this will complete the proof. First note that due to item (2), on A we have

$$T^{\text{inv}}(0, \partial B(n)) - T(0, \partial B(n)) \geq T^{\text{inv}}(0, \partial B(R)) - T(0, \partial B(R)) \text{ for } n \geq R.$$

Next, since there is an infinite edge-self avoiding path starting at 0 whose edges have weight $\leq a$ (just follow Γ to $\partial B(R)$ and then to infinity using item (1)), all edges e in I satisfy $t_e \leq a$. Therefore each path in I connecting 0 to $\partial B(R)$ must

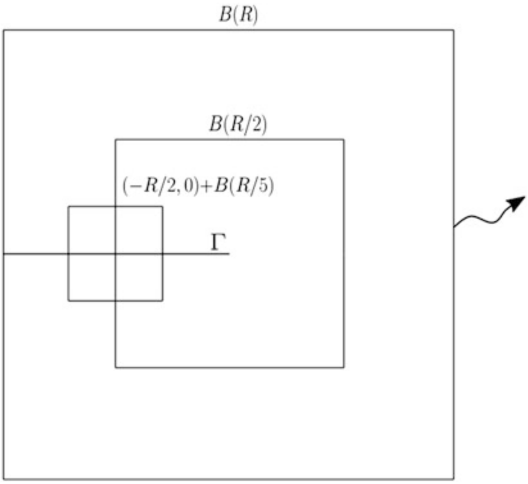


Fig. 2 Illustration of the event A. The arrowed curve emanating from $B(R)$ has edges with weight $\leq a$ and all edges on $\partial B(R)$ have weight zero. The edges in the segment Γ (which starts at the origin) have weight in the interval $[a/2, a]$, and edges touching $\partial B(R/2)$ (and in the box $(-R/2, 0) + B(R/5)$) but not in Γ have weight $\geq b$. All other edges in $B(R)$ have weight zero. On this event, any path from the origin to $\partial B(R)$ in the invasion must contain the segment of Γ passing through $\partial B(R/2)$ on the left, and must therefore pick up at least weight $aR/5$

contain all edges in Γ with both endpoints within ℓ^∞ distance $R/5$ of $(-R/2, 0)$. This implies by item (3) that on A, one has

$$T^{\text{inv}}(0, \partial B(R)) \geq \frac{2R}{5} \cdot \frac{a}{2} = \frac{aR}{5}.$$

On the other hand, there exists a path from 0 to $\partial B(R)$ with passage time equal to b: simply follow the positive e_1-axis. Therefore on A, one has

$$T(0, \partial B(R)) \leq b.$$

By the definition of R in (21), $aR/5 - b \geq b$, and this shows (22), completing the proof of Theorem 1.

Acknowledgments We thank an anonymous referee for useful comments, notably suggesting the almost sure result outlined in Remark 1. The research of M. D. is supported by an NSF CAREER grant.

References

1. Auffinger, A., Damron, M., Hanson, J.: 50 years of first-passage percolation. In: University Lecture Series, vol. 68. American Mathematical Society, Providence (2017)
2. Chayes, J.T., Chayes, L., Durrett, R.: Critical behavior of the two-dimensional first-passage time. J. Stat. Phys. **45**, 933–951 (1986)
3. Damron, M., Sapozhnikov, A., Vágvögyi, B.: Relations between invasion percolation and critical percolation in two dimensions. Ann. Probab. **37**, 2297–2331 (2009)
4. Damron, M., Lam, W-L., Wang, X.: Asymptotics for 2D critical first passage percolation. Ann. Probab. **45**, 2941–2970 (2017)
5. Hammersley, J.M., Welsh, D.J.A.: First-passage percolation, subadditive processes, stochastic networks, and generalized renewal theory. In: Proceedings of the International Research Seminar, Statistical Laboratory, University California, Berkeley, California, pp. 61–110. Springer, New York (1965)
6. Járai, A.A.: Invasion percolation and the incipient infinite cluster in $2D$. Commun. Math. Phys. **236**, 311–334 (2003)
7. Kesten, H.: Aspects of first-passage percolation. In: École d'été de Probabilités de Saint Flour XIV. Lecture Notes in Mathematics, vol. 1180, pp. 125–264 (1986)
8. Kesten, H.: Scaling relations for $2D$-percolation. Commun. Math Phys. **109**, 109–156 (1987)
9. McLeish, D.L.: A maximal inequality and dependent strong laws. Ann. Probab. **3**, 829–839 (1975)
10. Nguyen, B.G.: Correlation Lengths for Percolation Processes, Ph. D. dissertation. University of California, California (1985)
11. Nolin, P.: Near-critical percolation in two dimensions. Electron. J. Probab. **13**, 1562–1623 (2008)
12. Russo, L.: A note on percolation. Z. Warsch. und Verw. Gebiete. **43**, 39–48 (1978)
13. Seymour, P.D., Welsh, D.J.A.: Percolation probabilities on the square lattice. In: Advances in Graph Theory (Cambridge Combinatorial Conference, Trinity College, Cambridge, 1977). Ann. Discrete Math. **3**, 277–245 (1978)
14. Zhang, Y.: Double behavior of critical first-passage percolation. In: Perplexing Problems in Probability. Progress in Probability, vol. 44, pp. 143–158. Birkhäuser, Boston (1999)

Empirical Spectral Distributions of Sparse Random Graphs

Amir Dembo, Eyal Lubetzky, and Yumeng Zhang

Abstract We study the spectrum of a random multigraph with a degree sequence $\mathbf{D}_n = (D_i)_{i=1}^n$ and average degree $1 \ll \omega_n \ll n$, generated by the configuration model, and also the spectrum of the analogous random simple graph. We show that, when the empirical spectral distribution (ESD) of $\omega_n^{-1}\mathbf{D}_n$ converges weakly to a limit ν, under mild moment assumptions (e.g., D_i/ω_n are i.i.d. with a finite second moment), the ESD of the normalized adjacency matrix converges in probability to $\nu \boxtimes \sigma_{\text{SC}}$, the free multiplicative convolution of ν with the semicircle law. Relating this limit with a variant of the Marchenko–Pastur law yields the continuity of its density (away from zero), and an effective procedure for determining its support. Our proof of convergence is based on a coupling between the random simple graph and multigraph with the same degrees, which might be of independent interest. We further construct and rely on a coupling of the multigraph to an inhomogeneous Erdős-Rényi graph with the target ESD, using three intermediate random graphs, with a negligible fraction of edges modified in each step.

Keywords Random matrices · Empirical spectral distribution · Random graphs

MSC Subject Class 05C80, 60B20

A. Dembo
Department of Mathematics, Stanford University, Stanford, CA, USA
e-mail: amir@math.stanford.edu

E. Lubetzky (✉)
Courant Institute, New York University, New York, NY, USA
e-mail: eyal@courant.nyu.edu

Y. Zhang
Department of Statistics, Stanford University, Stanford, CA, USA

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_15

1 Introduction

We study the spectrum of a random multigraph $\mathsf{G}_n = ([n], \mathsf{E}_n)$ of n vertices of degrees $\{D_i^{(n)}\}_{i=1}^n$, constructed by the configuration model, where the *even*

$$\sum_{i=1}^{n} D_i^{(n)} = 2|\mathsf{E}_n| = n\omega_n(1 + o(1)), \quad (1)$$

is assumed to be such that

$$\omega_n \to \infty, \qquad \omega_n = o(n). \quad (2)$$

Specifically, setting $[n] = \{1, 2, \ldots, n\}$, equip each vertex $i \in [n]$ with $D_i^{(n)}$ half-edges, whereby the edge set E_n results from a uniformly chosen perfect matching of the $2|\mathsf{E}_n|$ half-edges. The uniformly chosen *simple* graph $\mathscr{G}_n = ([n], \mathscr{E}_n)$ with the degrees $D_i^{(n)}$—assuming of course that this degree sequence is graphical (i.e., there exist simple graphs with these degrees)—is similarly described via a uniform perfect matching of half-edges, subject to the constraint of having neither self-loops nor multiple edges.

Our study of the spectrum of the adjacency matrix $\mathbf{A}_{\mathsf{G}_n}$ of the multigraph G_n, proceeds through a sequence of couplings, relating it to certain "band" matrices, with independent albeit non-identically-distributed entries (adjacency matrices of Erdős-Rényi *inhomogeneous* random graphs). Various spectral features of the latter will then be derived using the powerful tools that have been developed in the last few decades in random matrix theory and free probability.

In Proposition 1 we further provide a novel coupling of $\mathscr{G}_n$ and G_n, which may be of independent interest. Utilizing this coupling we deduce that the uniformly chosen random *simple* graph $\mathscr{G}_n$, satisfying the same degree assumptions as G_n, will also have the same limiting spectrum.

For random *regular* graphs—the case of $D_i^{(n)} = d_n$ for all i—it was shown by Tran et al. [19] (extending a previous result of [8]) that, whenever $d_n \gg 1$, the empirical spectral distribution (ESD, defined for a symmetric matrix $\mathbf{A}$ with eigenvalues $\lambda_1 \geq \ldots \geq \lambda_n$ as $\mathcal{L}^{\mathbf{A}} = \frac{1}{n}\sum_{i=1}^n \delta_{\lambda_i}$) of the normalized matrix $\hat{\mathbf{A}}_{\mathsf{G}_n} = \frac{1}{\sqrt{d_n}}\mathbf{A}_{\mathsf{G}_n}$ converges weakly, in probability, to σ_{SC}, the standard semicircle law (with support $[-2, 2]$).

The non-regular case with $|\mathsf{E}_n| = O(n)$ has been studied by Bordenave and Lelarge [6] when the graphs G_n converge in the Benjamini–Schramm sense, translating in the above setup to having $\{D_i^{(n)}\}$ that are i.i.d. in i and uniformly integrable in n. The existence and uniqueness of the limiting ESD was obtained in [6] by relating this ESD to a recursive distributional equation—arising from the Galton–Watson trees that correspond to the local neighborhoods in G_n—and showing that this equation has a unique fixed point. See also, e.g., [5, 7, 14] and the references therein, for the analysis of the limiting spectrum at $\lambda = 0$ for Erdős–Rényi graphs of constant average degree. Note that (a) this approach relies on the locally-tree-like structure of the graphs, and is thus tailored for low (at most logarithmic) degrees; and (b) very little is known on this limit, even in seemingly simple settings such as when all degrees are either 3 or 4.

At the other extreme, when $|\mathsf{E}_n|$ diverges polynomially with n (whence the tree approximations are invalid), the trace method—the standard tool for establishing the convergence of the ESD of an Erdős–Rényi random graph to σ_{SC}—faces the obstacle of non-negligible dependencies between edges in the configuration model (the trace method can handle dependencies, but here $n^{-1}\operatorname{tr}((\mathbb{E}\hat{\mathbf{A}}_{\mathsf{G}_n})^{2k}) \asymp \omega_n^k$, thus the precise cancellations of many diverging terms are needed for it to work; such cancellations are very difficult to attain in the presence of dependencies).

1.1 Limiting ESD as a Free Multiplicative Convolution

Our assumptions on the triangular sequence $\{D_i^{(n)}\}$ of degrees are that (2) holds, and in addition, for ω_n satisfying (1), the normalized degrees $\hat{D}_i^{(n)} = D_i^{(n)}/\omega_n$ satisfy that

$$\{\hat{D}_{U_n}^{(n)}\} \quad \text{is uniformly integrable with} \quad \mathbb{E}[(\hat{D}_{U_n}^{(n)})^2] = o(\sqrt{n/\omega_n})\,, \tag{3}$$

where U_n is uniformly chosen in $\{1, \ldots, n\}$. Let

$$\hat{\mathbf{A}}_{\mathsf{G}_n} := \omega_n^{-1/2} A_{\mathsf{G}_n} \qquad \text{and} \qquad \hat{\boldsymbol{\Lambda}}_n := \operatorname{diag}(\hat{D}_1^{(n)}, \ldots, \hat{D}_n^{(n)})\,.$$

Call a degree sequence $\{D_i^{(n)}\}$ *graphical* if for every n there exists a simple graph $\mathscr{G}_n$ with such degrees (equivalently, the criterion of the Erdős–Gallai theorem [9] is met). Our main result derives the limiting ESD under conditions (1)–(3) on the degree sequence (see, e.g., Fig. 1).

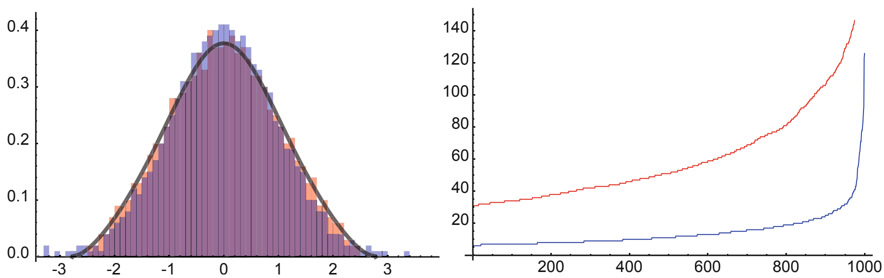


Fig. 1 Spectra of two random multigraphs on $n = 1000$ vertices with different degree sequences $\{D_i\}$. In red, $D_i = [\tau_i\sqrt{n}]$ for all i, and in blue, $D_i = [\tau_i \log n]$ for $i < n - \sqrt{n}$ and $D_i = [\tau_i\sqrt{n}]$ for $i \geq n - \sqrt{n}$, with $\tau_i \sim 1 + \text{Exp}(1)$ i.i.d. (right plot). The limiting law for the ESD, shown by Theorem 1 to be $\nu_{\hat{D}} \boxtimes \sigma_{\text{SC}}$, is plotted in black (left plot)

Theorem 1 *Let $\{D_i^{(n)}\}_{i=1}^n$ be a degree sequence satisfying (1)–(3), and further suppose that the* ESD $\mathcal{L}^{\hat{\mathbf{A}}_n}$ *converges weakly to a limit $\nu_{\hat{D}}$.*

(a) The ESD $\mathcal{L}^{\hat{\mathbf{A}}_{\mathsf{G}_n}}$ *corresponding to the multigraph $\mathsf{G}_n = ([n], \mathsf{E}_n)$ with degrees $\{D_i^{(n)}\}_{i=1}^n$ (generated via the configuration model), converges weakly, in probability, to $\nu_{\hat{D}} \boxtimes \sigma_{\text{SC}}$.*

(b) If $\{D_i^{(n)}\}$ is graphical then the same convergence holds for the ESD $\mathcal{L}^{\hat{\mathbf{A}}_{\mathcal{G}_n}}$ *corresponding to a uniformly chosen simple graph $\mathcal{G}_n = ([n], \mathcal{E}_n)$ with this degree sequence.*

In the above theorem, the *free multiplicative convolution* of a symmetric probability measure ψ and a probability measure φ on $\mathbb{R}_+$ with $\varphi, \psi \neq \delta_0$, denoted $\varphi \boxtimes \psi$, is the unique probability measure such that $S_{\varphi\boxtimes\psi}(z) = S_\varphi(z)S_\psi(z)$ for z in the common domain of the corresponding S-transforms (see [2, Thm. 7], extending the definition of $\varphi \boxtimes \psi$ from [4] and [20] in case both φ, ψ, are of bounded support and non-zero mean). To define the S-transform, recall that the Cauchy–Stieltjes transform of a probability measure μ on $\mathbb{R}$, uniquely determining it, is $G_\mu(z) := \int [t - z]^{-1} d\mu(t)$. For φ as above, the related

$$m_\varphi(z) := z^{-1}G_\varphi(z^{-1}) - 1 = \int \frac{zt}{1 - zt} d\varphi(t)\,, \tag{4}$$

is invertible as a formal power series in $z \in \mathbb{C}_+$, and the S-transform is defined as

$$S_\varphi(w) := (1 + w^{-1})m_\varphi^{-1}(w) \quad \text{for } w \in m_\varphi(\mathbb{C}_+) \tag{5}$$

(cf., e.g., [2, Prop. 1]). Following the extension in [13] of the S-transform to measures of zero mean and finite moments of all order, the S-transform is similarly defined for ψ as above in [2, Thm. 6]. In particular, with σ_{SC} being symmetric and $\nu_{\hat{D}} \neq \delta_0$ supported on $\mathbb{R}_+$, the measure $\nu_{\hat{D}} \boxtimes \sigma_{\text{SC}}$ is well-defined.

Corollary 1 *Let $\{\hat{D}_i^{(n)} : 1 \leq i \leq n\}$ be i.i.d. for each n, such that $\mathbb{E}\hat{D}_1^{(n)} = 1$, $\sup_n \mathbb{E}[(\hat{D}_1^{(n)})^2] < \infty$, and the law of $\hat{D}_1^{(n)}$ converges weakly to some $\nu_{\hat{D}}$. For $\omega_n \to \infty$ such that $\omega_n = o(n)$, let G_n denote the uniform multigraph of degrees $D_i^{(n)} = [\omega_n \hat{D}_i^{(n)}]$ (modifying $D_n^{(n)}$ by one if needed for an even sum). Further, for any integers $\bar{d}_n = o(n)$ with $\omega_n = o(\bar{d}_n)$, the truncated degrees $[\omega_n \hat{D}_i^{(n)}] \wedge \bar{d}_n$ are graphical* WHP *(after increasing the minimal degree by one, if needed, for an even sum).*

Denoting by $\mathcal{G}_n$ the uniform simple graph, both ESDs $\mathcal{L}^{\hat{\mathbf{A}}_{\mathsf{G}_n}}$ *and* $\mathcal{L}^{\hat{\mathbf{A}}_{\mathcal{G}_n}}$ *converge weakly, in probability, to $\nu_{\hat{D}} \boxtimes \sigma_{\text{SC}}$.*

Remark 1 The reason for the appearance of $\nu_{\hat{D}} \boxtimes \sigma_{\text{SC}}$ in our context is due to the fact that it is the limiting ESD of $\mathbf{B}_n := \hat{\boldsymbol{\Lambda}}_n^{1/2}\mathbf{X}_n\hat{\boldsymbol{\Lambda}}_n^{1/2}$ when $\max_i \hat{D}_i^{(n)} = O(1)$ and $\mathbf{X}_n$ is a standard GOE random matrix. Indeed, as its name suggest, the free multiplicative convolution $\varphi \boxtimes \psi$ is the law of the product ab of free, bounded, random non-commutative operators a of law φ and b of law ψ (cf. [1, Defn. 5.2.1, 5.2.3, 5.3.1, 5.3.28] for the precise meaning of all this). This extends to the limiting ESD for the product of *asymptotically free* matrices: two sequences $\mathbf{X}_n, \mathbf{Y}_n$ of random self-adjoint, matrices are asymptotically free if $\mathbb{E}[\mathrm{tr}_n(f_1(\mathbf{X}_n)g_1(\mathbf{Y}_n)\cdots f_k(\mathbf{X}_n)g_k(\mathbf{Y}_n))] = o(1)$ for the normalized trace $\mathrm{tr}_n(\cdot) = \frac{1}{n}\mathrm{tr}(\cdot)$ and any collections $(f_i)_{i=1}^k$ and $(g_i)_{i=1}^k$ of polynomials (with k fixed) that satisfy $\mathbb{E}[\mathrm{tr}_n(f_i(\mathbf{X}_n))] = o(1)$ and $\mathbb{E}[\mathrm{tr}_n(g_i(\mathbf{Y}_n))] = o(1)$ for all $1 \le i \le k$ (see [1, Defn. 5.4.1] or [17, §2.5]). It is known that the GOE $\mathbf{X}_n$ is asymptotically free of any uniformly bounded diagonal $\hat{\boldsymbol{\Lambda}}_n$ (see, e.g., [1, Theorem 5.4.5]), which in turn implies that $\nu_{\hat{D}} \boxtimes \sigma_{\mathrm{SC}}$ is the weak limit of the ESD for the random matrices $\mathbf{B}_n$ (the spectral radius of the GOE $\mathbf{X}_n$ is $O(1)$ with high probability, so by a standard truncation argument we arrive at the bounded case of [1, Corollary 5.4.11(iii)]).

Theorem 1 and Corollary 1 are proved in Sect. 2. This is achieved by first analyzing the ESD of the random multigraph G_n; the move from multigraphs to simple graphs is achieved via the following proposition, which we prove in Sect. 3.

Proposition 1 *Fixing graphical degrees $D_1 \ge D_2 \ge \cdots \ge D_n$, let G_n and $\mathscr{G}_n$ be the corresponding random multigraph and uniform simple graph, respectively. There exists a coupling μ between the matchings which yield G_n and $\mathscr{G}_n$ so the number $\Delta_n \le 2|\mathsf{E}_n|$ of half-edges whose matching links are different between the two graphs, satisfies*

$$\mathbb{E}_\mu[\Delta_n(\Delta_n - 1)] \le 4\sum_{i=1}^{n-1}\sum_{j=i+1}^{i+D_i}(2D_iD_j - D_i - D_j)\,. \tag{6}$$

Remark 2 A crude, yet already useful, upper bound on the RHS of (6) is

$$8\sqrt{2|\mathsf{E}_n|}\sum_{i=1}^{n} D_i^2\,. \tag{7}$$

(Indeed, $\left(\sum_{j=i+1}^{i+D_i} D_j\right)^2 \le D_i\sum_{j=1}^n D_j^2$ by Cauchy–Schwarz for any $i \in [n]$; thus, again by Cauchy–Schwarz, the RHS of (6) is at most $8(\sum_i D_i^2)(\sum_i D_i)^{\frac{1}{2}}$, and $\sum_i D_i = 2|\mathsf{E}_n|$.) In general, the RHS of (6) can be replaced by any bound on the expected number of pairs of half-edges $e \ne f$ on a which a "switch" would yield a non-simple graph.

Remark 3 The Proof of Theorem 1(a) extends to the dense regime, where ω_n/n is bounded below (and above). However, the minimal expected edit distance between G_n and $\mathscr{G}_n$ exceeds the expected number $O(\omega_n^2)$ of parallel edges in G_n, which in the dense regime is already $O(|\mathsf{E}_n|)$, thereby blocking in the dense regime our route to Theorem 1(b) as a consequence of part (a). Further, our assumption (3) allows having a maximal degree that far exceeds n (indeed, prior to truncation this happens in the i.i.d. setting of Corollary 1). Even for specified graphical degrees, the number of simple graphs $\mathscr{G}_n$ oscillates widely as the degrees change, so (3) might not suffice for the statement of Theorem 1(b) to be true in the dense regime. Going back to the sparse regime, assumptions à la (3) have little to do with controlling extreme eigenvalues, or with bringing the corresponding local law to the celebrated GOE-universality class of homogeneous Erdős-Rényi graphs. Indeed, one must further restrict $\hat{\boldsymbol{\Lambda}}_n$, in order to have any hope of transferring the many fine results on extreme eigenvalues and local laws that are available for the GOE, via $\mathbf{B}_n$ of Remark 1 to $\hat{\mathbf{A}}_{\mathsf{G}_n}$.

1.2 Properties of the Limiting ESD

The next two propositions, proved in Sect. 4, relate the limiting measure $\nu_{\hat{D}} \boxtimes \sigma_{\rm SC}$ with a Marchenko–Pastur law, and thereby, via [16], yield its support and density regularity.

Proposition 2 *For the law $\nu_{\hat{D}}$ of a nonnegative random variable $\hat{D}$ with $\mathbb{E}\hat{D} = 1$, let $\mu_{\rm MP}$ be the Marchenko–Pastur limit (on $\mathbb{R}_+$) of the ESD of $n^{-1}\boldsymbol{\Lambda}_n\widetilde{\mathbf{X}}_n\widetilde{\mathbf{X}}_n^{\star}\boldsymbol{\Lambda}_n$, in which the non-symmetric $\widetilde{\mathbf{X}}_n$ has standard i.i.d. complex Gaussian entries and $\mathcal{L}^{\boldsymbol{\Lambda}_n} \Rightarrow \nu$ for non-negative diagonal matrices $\boldsymbol{\Lambda}_n$ and the size-biased ν with $\frac{d\nu}{d\nu_{\hat{D}}}(x) = x$ on $\mathbb{R}_+$. The free multiplicative convolution $\mu = \nu_{\hat{D}} \boxtimes \sigma_{\rm SC}$ has the Cauchy–Stieltjes transform*

$$G_\mu(z) = -z^{-1}\Big[1 + G_{\widetilde{\mu}}(z)^2\Big], \quad \forall z \in \mathbb{C}_+, \tag{8}$$

where $\widetilde{\mu}$ is the law of the symmetric X such that X^2 is distributed according to $\mu_{\rm MP}$.

Recall [16, Lemma 3.1, Lemma 3.2] that $h(z) := G_{\mu_{\rm MP}}(z)$ is uniformly bounded on $\mathbb{C}_+$ away from the imaginary axis, and [16, Theorem 1.1] that $h(z) \to h(x)$ whenever $z \in \mathbb{C}_+$ converges to $x \in \mathbb{R} \setminus \{0\}$. Further, the $\mathbb{C}_+$-valued function $h(x)$ is continuous on $\mathbb{R} \setminus \{0\}$ and the continuous density

$$\boldsymbol{\rho}_{\rm MP}(x) := \frac{d\mu_{\rm MP}}{dx} = \frac{1}{\pi}\Im(h(x)), \tag{9}$$

is real analytic at any $x \neq 0$ where it is positive. The density $\widetilde{\boldsymbol{\rho}}(x) = |x|\boldsymbol{\rho}_{\rm MP}(x^2)$ of $\widetilde{\mu}$ inherits these regularity properties. Bounding $\widetilde{\boldsymbol{\rho}}$ uniformly and analyzing the effect of (8) we next make similar conclusions about the density $\boldsymbol{\rho}(x)$ of μ, now also at $x = 0$, and its support (see Fig. 2).

Proposition 3 *In the setting of Proposition 2, for $x \neq 0$ there is density*

$$\boldsymbol{\rho}(x) := \frac{d\mu}{dx} = -2\Re(h(x^2))\widetilde{\boldsymbol{\rho}}(x), \tag{10}$$

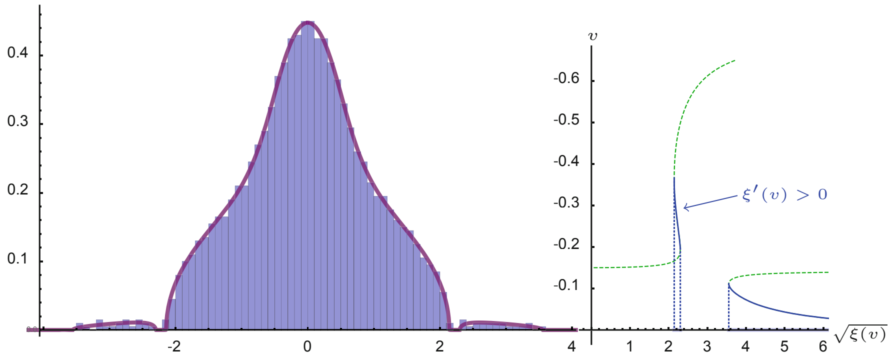


Fig. 2 Recovering the support of the limiting ESD. Left: ESD of the random multigraph on $n = 1000$ vertices with degrees $\sqrt{n}$, $3\sqrt{n}$, $15\sqrt{n}$ in frequencies 0.5, 0.49, 0.01, resp. Right: $\xi(v)$ from Remark 4

which is continuous, symmetric, and moreover real analytic where positive. The support of μ *is* $\mathrm{supp}(\mu) := \{x \in \mathbb{R} : \boldsymbol{\rho}(x) > 0\} = \mathrm{supp}(\widetilde{\mu})$*, which up to the mapping* $x \mapsto x^2$ *further matches* $\mathrm{supp}(\mu_{\mathrm{MP}})$*. In addition* $\pi\widetilde{\boldsymbol{\rho}}(x) \leq 1 \wedge (2/|x|)$*,* $\pi\boldsymbol{\rho}(x) \leq (\mathbb{E}\hat{D}^{-2})^{1/2} \wedge (4/|x|^3)$ *and if* $\nu_{\hat{D}}(\{0\}) = 0$ *then* μ *is absolutely continuous (i.e.,* $\mu(\{0\}) = 0$*).*

Remark 4 Recall the unique inverse of h on $h(\mathbb{C}_+)$ given by

$$\xi(h) := -\frac{1}{h} + \mathbb{E}\left[\frac{\hat{D}^2}{1 + h\hat{D}}\right], \tag{11}$$

namely $\xi(h(z)) = z$ on $\mathbb{C}_+$ (see [16, Eqn. (1.4)]); this inverse extends analytically to a neighborhood of $\mathbb{C}_+ \cup \Gamma$ for $\Gamma := \{h \in \mathbb{R} : h \neq 0, -h^{-1} \in \mathrm{supp}(\hat{\nu}_D)^c\}$ and [16, Theorems 4.1 and 4.2] show that $x \in \mathrm{supp}(\mu_{\mathrm{MP}})^c$ iff $\xi'(v) > 0$ for $v \in \Gamma$, where $v = h(x)$ and $x = \xi(v)$ (thus validating the characterization of $\mathrm{supp}(\mu_{\mathrm{MP}})$ which has been proposed in [12]). We show in Lemma 2 that $\Re(h(x^2)) < 0$ everywhere, hence the behavior of $\boldsymbol{\rho}(x)$ at the soft-edges of $\mathrm{supp}(\mu)$ can be read from the soft-edges of $\mathrm{supp}(\mu_{\mathrm{MP}})$ (as in [11, Prop. 2.3]), depicted in Fig. 3.

Corollary 2 *Suppose* $\nu_{\hat{D}}$ *of mean one is supported on two atoms* $\alpha > \eta > 0$*. The support* $\mathrm{supp}(\mu)$ *of* $\mu = \nu_{\hat{D}} \boxtimes \sigma_{\mathrm{SC}}$ *is then disconnected iff*

$$\alpha > \eta\left[\frac{3}{1 - (1 - \eta)^{1/3}} - 1\right]. \tag{12}$$

Moreover, when (12) *holds,* $\mathrm{supp}(\mu) \cap \mathbb{R}_+$ *consists of exactly two disjoint intervals.*

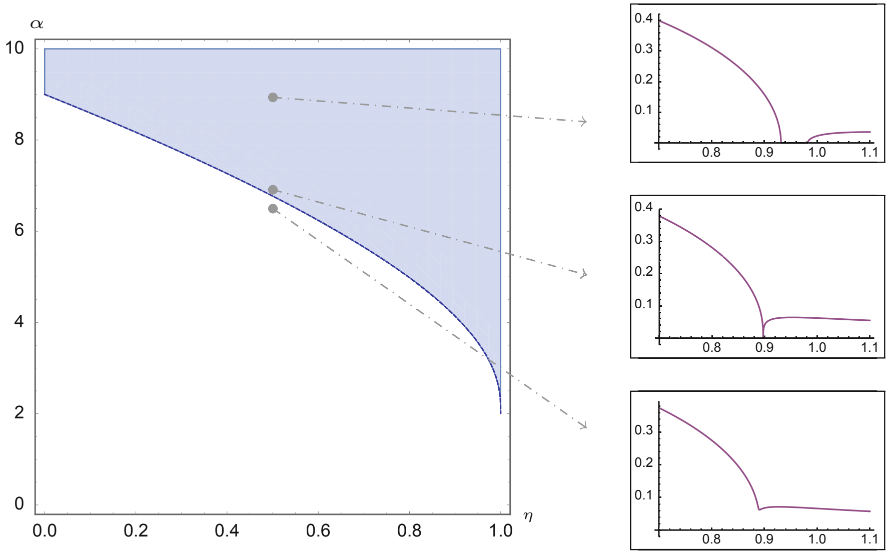


Fig. 3 Phase diagram for the existence of holes in the limiting ESD when $\nu_{\hat{D}}$ is supported on two atoms $\alpha > \eta > 0$ (see Corollary 2). Left: the region (12) (where $\text{supp}(\mu)$ is not connected) highlighted in blue. Right: zooming in on the emergence of a hole as α varies at $\eta = \frac{1}{2}$

2 Convergence of the ESD's

The Proof of Theorem 1 will use the following standard lemma.

Lemma 1 *Let* $\{\mathbf{M}_{n,r}\}_{n,r\in\mathbb{N}}$ *be a family of matrices of order* n*, define* $\mu_{n,r} := \mathcal{L}^{\mathbf{M}_{n,r}}$ *and* $\eta(r) := \limsup_{n\to\infty} \frac{1}{n}\,\text{tr}\left((\mathbf{M}_{n,r} - \mathbf{M}_{n,\infty})^2\right)$*. Let* $\{\mu_r : r \in \mathbb{N}\}$ *denote a family of measures such that*

$$\mu_{n,r} \Rightarrow \mu_r \text{ as } n \to \infty \text{ for every } r \in \mathbb{N}\,, \tag{13}$$

$$\mu_{n,\infty} \text{ is tight}\,, \tag{14}$$

$$\eta(r) \to 0 \quad \text{as } r \to \infty\,. \tag{15}$$

Then the weak limit of $\mu_{n,\infty}$ *as* $n \to \infty$ *exists and equals* $\lim_{r\to\infty}\mu_r$*.*

Proof Let μ_∞ be a limit point of $\mu_{n,\infty}$, the existence of which is guaranteed by the tightness assumption (14). A standard consequence of the Hoffman–Wielandt bound (cf. [1, Lemma 2.1.19]) and Cauchy–Schwarz is that for matrices $\mathbf{A}$ and $\mathbf{B}$ of order n,

$$d_{\text{BL}}\left(\mathcal{L}^{\mathbf{A}}, \mathcal{L}^{\mathbf{B}}\right)^2 \le \frac{1}{n}\,\text{tr}\left((\mathbf{A} - \mathbf{B})^2\right)\,,$$

where d_{BL} is the bounded-Lipchitz metric on the space $M_1(\mathbb{R}_+)$ of probability measures on $\mathbb{R}_+$ (see the proof of [1, Theorem 2.1.21]). Thus, by (13) and the triangle-inequality for d_{BL}, it follows that

$$\eta(r) \geq d_{\mathrm{BL}}(\mu_\infty, \mu_r)^2 .$$

Consequently, $\mu_r \to \mu_\infty$ as $r \to \infty$, from which the uniqueness of μ_∞ also follows.

Proof of Theorem 1 In **Step I** we reduce the proof to treating the *single-adjacency* matrix $\mathbf{A}_n$ of G_n, where multiple copies of an edge/loop are replaced by a single one (that is, $\mathbf{A}_n = \mathbf{A}_{\mathsf{G}_n} \wedge 1$ entry-wise), and further $\{\omega_n^{-1} D_i^{(n)}\} \subseteq \mathcal{S}$ for some fixed finite set $\mathcal{S}$. Scaling $\hat{\mathbf{A}}_n := \omega_n^{-1/2}\mathbf{A}_n$ we rely in **Step II**, on Proposition 4 to replace the limit points of $\mathcal{L}^{\hat{\mathbf{A}}_n}$ by those of $\mathcal{L}^{\omega_n^{-1/2}\widetilde{\mathbf{A}}_n}$ for symmetric matrices $\widetilde{\mathbf{A}}_n$ of independent Bernoulli entries. Using the moment method, **Step III** relates the latter limit points to the limit of $\mathcal{L}^{\mathbf{B}_n}$ for the matrices $\mathbf{B}_n$ of Remark 1.

Step I We claim that if $\mathcal{L}^{\hat{\mathbf{A}}_n} \Rightarrow \mu$ in probability, then the same applies for $\mathcal{L}^{\hat{\mathbf{A}}_{\mathsf{G}_n}}$. This will follow from Lemma 1 with $\mathbf{M}_{n,r} = \hat{\mathbf{A}}_n$ and $\mathbf{M}_{n,\infty} = \hat{\mathbf{A}}_{\mathsf{G}_n}$ upon verifying that

$$\xi_n := \mathbb{E}\left[\tfrac{1}{n}\,\mathrm{tr}\left((\hat{\mathbf{A}}_{\mathsf{G}_n} - \hat{\mathbf{A}}_n)^2\right)\right] \to 0 . \tag{16}$$

Indeed, Condition (13) has been assumed; Condition (14) follows from the fact that

$$\begin{aligned}\frac{1}{2n}\,\mathrm{tr}\left(\hat{\mathbf{A}}_{\mathsf{G}_n}^2\right) &\leq \frac{1}{n}\,\mathrm{tr}\left((\hat{\mathbf{A}}_{\mathsf{G}_n} - \hat{\mathbf{A}}_n)^2\right) + \frac{1}{n}\,\mathrm{tr}(\hat{\mathbf{A}}_n^2)\\ &\leq \frac{1}{n}\,\mathrm{tr}\left((\hat{\mathbf{A}}_{\mathsf{G}_n} - \hat{\mathbf{A}}_n)^2\right) + \frac{2|\mathsf{E}_n|}{n\omega_n},\end{aligned}$$

so in particular $\mathbb{E}[\frac{1}{n}\,\mathrm{tr}(\hat{\mathbf{A}}_{\mathsf{G}_n}^2)] \leq \xi_n + 1 + o(1)$, yielding tightness; and Condition (15) holds in probability by (16) and Markov's inequality. Recall that the stochastic ordering $X \preceq X'$ denotes that $\mathbb{P}(X > x) \leq \mathbb{P}(X' > x)$ for all $x \in \mathbb{R}$, or equivalently, that there exists a coupling of (X, X') such that $\mathbb{P}(X \leq X') = 1$. To establish (16), observe that, for every i and j we have $(\mathbf{A}_{\mathsf{G}_n})_{i,j} \preceq \mathrm{Bin}(m, q)$ for $m = D_i^{(n)}$ and $q = (D_j^{(n)} - 1_{i=j})/(2|\mathsf{E}_n| - 1)$, whereas $\mathrm{Bin}(m, q) \preceq Y_\lambda \sim \mathrm{Po}(\lambda)$ for every m and λ such that $1 - q \geq e^{-\lambda/m}$. Thus,

$$\mathbb{E}\left[(\mathbf{A}_{\mathsf{G}_n} - \mathbf{A}_n)_{i,j}^2\right] \leq \mathbb{E}\left[(Y_\lambda - 1)_+^2\right] \leq \lambda^2 .$$

Since $q \le \frac{1+o(1)}{\omega_n}$ uniformly over i, j, we take WLOG $\lambda = m D_j^{(n)}/|\mathsf{E}_n|$, yielding for n large

$$\xi_n \le \frac{2}{n\omega_n} \sum_{i,j=1}^{n} \Big[\frac{D_i^{(n)} D_j^{(n)}}{|\mathsf{E}_n|}\Big]^2 \le \frac{4\omega_n}{n} \Big[\frac{1}{n}\sum_{i=1}^{n} (\hat{D}_i^{(n)})^2\Big] \to 0\,,$$

by our assumption that $\mathbb{E}[(\hat{D}_{U_n}^{(n)})^2] = o(\sqrt{n/\omega_n})$. Considering hereafter only single-adjacency matrices, we proceed to reduce the problem to the case where the variables $\hat{D}_i^{(n)}$ are all supported on a finite set. To this end, let $\ell = 2r^2$ for $r \in \mathbb{N}$ and

$$\hat{D}_i^{(n,r)} = \Psi_r(\hat{D}_i^{(n)}) \quad \text{for} \quad \Psi_r(x) := \sum_{j=1}^{\ell} d_j^{(r)} \mathbb{1}_{\left[d_j^{(r)}, d_{j+1}^{(r)}\right)}(x)\,,$$

where $0 = d_1^{(r)} < \ldots < d_{\ell+1}^{(r)}$ are continuity points of $\nu_{\hat{D}}$ of interdistances in $[\frac{1}{2r}, \frac{1}{r}]$, which are furthermore in $\epsilon_r \mathbb{Z}$ for some irrational $\epsilon_r > 0$. Let

$$D_i^{(n,r)} = \omega_{n,r} \hat{D}_i^{(n,r)} \in \mathbb{Z}_+ \quad \text{for} \quad \omega_{n,r} := \frac{[\epsilon_r \omega_n]}{\epsilon_r}\,,$$

possibly deleting one half-edge from $D_n^{(n,r)}$ if needed to make $\sum_{i=1}^n D_i^{(n,r)}$ even.

Observation 1 *Let $\{d_i\}_{i=1}^n, \{d_i'\}_{i=1}^n$ be degree sequences with $d_i' \le d_i$, and let G be a random multigraph with degrees $\{d_i\}$ generated by the configuration model. Generate H by (a) marking a subset of d_i' half-edges of vertex i BLUE, chosen independently of the matching that generated G; (b) retaining every edge that has two BLUE endpoints; and (c) adding an independent uniform matching on all other BLUE half-edges. Then H has the law of the random multigraph with degrees $\{d_i'\}$ generated by the configuration model.*

Proof Since the configuration model matches the half-edges in G via a uniformly chosen perfect matching, and the coloring step (a) is performed independently of this matching, it follows that the induced matching on the subset of BLUE half-edges that are matched to BLUE counterparts—namely, the edges retained in step (b)—is uniform.

Using this, and noting $D_i^{(n,r)} \le D_i^{(n)}$ for all i, let $\mathsf{G}_n^{(r)} = ([n], \mathsf{E}_n^{(r)})$ be the following random multigraph with degrees $\{D_i^{(n,r)}\}$, coupled to the already-constructed G_n:

(a) For each i, mark a uniformly chosen subset of $D_i^{(n,r)}$ half-edges incident to vertex i as BLUE in G_n.
(b) Retain in $\mathsf{G}_n^{(r)}$ every edge of E_n where both parts are BLUE.

(c) Complete the construction of $\mathsf{G}_n^{(r)}$ via a uniformly chosen matching of all unmatched half-edges.

Let $\hat{\mathbf{A}}_n^{(r)} = \omega_n^{-1/2}\mathbf{A}_n^{(r)}$ for $A_n^{(r)}$, the single-adjacency matrix of $\mathsf{G}_n^{(r)}$. We next control the difference between $\mathcal{L}^{\hat{\mathbf{A}}_n}$ and $\mathcal{L}^{\hat{\mathbf{A}}_n^{(r)}}$. Indeed, by the definition of the coupling of G_n and $\mathsf{G}_n^{(r)}$, the cardinality of the symmetric $\mathsf{E}_n \triangle \mathsf{E}_n^{(r)}$ is at most twice the number of *unmarked* half-edges in G_n. Thus,

$$\frac{1}{4n}\operatorname{tr}\big((\hat{\mathbf{A}}_n - \hat{\mathbf{A}}_n^{(r)})^2\big) \le \frac{1}{2n\omega_n}\Big|\mathsf{E}_n \triangle \mathsf{E}_n^{(r)}\Big| \le \frac{1}{n\omega_n}\sum_{i=1}^n (D_i^{(n)} - D_i^{(n,r)})$$

$$\le \frac{1+o(1)}{\epsilon_r\omega_n} + \frac{1}{r} + \frac{1}{n}\sum_{i=1}^n \hat{D}_i^{(n)} \mathbb{1}_{\{\hat{D}_i^{(n)} \ge r\}} =: \eta(n,r)\,, \qquad (17)$$

where the first term in $\eta(n,r)$ accounts for the discrepancy between ω_n and $\omega_{n,r}$, the term $1/r$ accounts for the degree quantization, while the last term accounts for degree truncation (since $d_{\ell+1}^{(r)} \ge r$). Thanks to the uniform integrability of $\{\hat{D}_{U_n}^{(n)}\}$ from (3), we have that $\eta(r) := \limsup_{n\to\infty} \eta(n,r)$ satisfies $\eta(r) \to 0$ as $r \to \infty$. Furthermore,

$$\int x^2 d\mathcal{L}^{\hat{\mathbf{A}}_n} = \frac{1}{n}\operatorname{tr}(\hat{\mathbf{A}}_n^2) \le 1 + o(1)$$

by the choice of ω_n in (1), yielding the tightness of $\mu_{n,\infty} := \mathcal{L}^{\hat{\mathbf{A}}_n}$. Altogether, we conclude from Lemma 1 that, if $\mathcal{L}^{\hat{\mathbf{A}}_n^{(r)}} \Rightarrow \mu_r$, then $\mathcal{L}^{\hat{\mathbf{A}}_n} \Rightarrow \lim_{r\to\infty}\mu_r$.

Next, let $\omega_n^{(r)} = 2|\mathsf{E}_n^{(r)}|/n$ (as in (1) but for the multigraph $\mathsf{G}_n^{(r)}$). Since (see (17)),

$$\limsup_{n\to\infty}\left|1 - \frac{\omega_n^{(r)}}{\omega_n}\right| \le \eta(r) \to 0 \quad \text{as } r \to \infty\,,$$

WLOG we replace ω_n by $\omega_n^{(r)}$ in the definition of $\hat{\mathbf{A}}_n^{(r)}$, i.e., starting with

$$\hat{D}_i^{(n,r)} \in \{d_1^{(r)}, \ldots, d_\ell^{(r)}\} =: \mathcal{S}_r\,.$$

Note that the hypothesis $\mathcal{L}^{\hat{\boldsymbol{A}}_n} \Rightarrow \nu_{\hat{D}}$ as $n \to \infty$, together with our choice of $\mathcal{S}_r$, implies that $\mathcal{L}^{\hat{\boldsymbol{A}}_n^{(r)}}$ (corresponding to $\hat{\boldsymbol{\Lambda}}_n^{(r)} = \operatorname{diag}(\hat{D}_1^{(n,r)}, \ldots, \hat{D}_n^{(n,r)})$) converges weakly for each r to some $\nu_{\hat{D}_r} \neq \delta_0$, supported on $\mathbb{R}_+$, and further, $\nu_{\hat{D}_r} \Rightarrow \nu_{\hat{D}} \neq \delta_0$, as $r \to \infty$.

Let $\mu^{(2)}$ denote hereafter the pushforward of the measure μ by the mapping $x \mapsto x^2$ (that is, the measure given by $B \mapsto \mu(f^{-1}(B))$ for $f(x) = x^2$.) It is known that, for probability measures on $\mathbb{R}_+$, the free multiplicative convolution is continuous w.r.t. weak convergence; that is, $\nu_k \boxtimes \nu'_k \Rightarrow \nu \boxtimes \nu'$ provided $\nu_k \Rightarrow \nu \neq \delta_0$,

$\nu'_k \Rightarrow \nu' \neq \delta_0$ all of which are supported on $\mathbb{R}_+$ (see, e.g., [2, Prop. 3]). Applying this twice, we find that

$$\nu_{\hat{D}_r} \boxtimes \sigma_{\mathrm{SC}}^{(2)} \boxtimes \nu_{\hat{D}_r} \Rightarrow \nu_{\hat{D}} \boxtimes \sigma_{\mathrm{SC}}^{(2)} \boxtimes \nu_{\hat{D}} \,. \tag{18}$$

From this we next deduce that $\nu_{\hat{D}_r} \boxtimes \sigma_{\mathrm{SC}} \Rightarrow \nu_{\hat{D}} \boxtimes \sigma_{\mathrm{SC}}$. Indeed, recall [2, Lemma 8] that the LHS of (18) equals $(\nu_{\hat{D}_r} \boxtimes \sigma_{\mathrm{SC}})^{(2)}$, while likewise its RHS equals $(\nu_{\hat{D}} \boxtimes \sigma_{\mathrm{SC}})^{(2)}$. For any $f \in C_b(\mathbb{R})$, the function $g(x) = \frac{1}{2}[f(\sqrt{x}) + f(-\sqrt{x})]$ is in $C_b(\mathbb{R}_+)$. Thus, the weak convergence $(\nu_{\hat{D}_r} \boxtimes \sigma_{\mathrm{SC}})^{(2)} \Rightarrow (\nu_{\hat{D}} \boxtimes \sigma_{\mathrm{SC}})^{(2)}$, implies that $\nu_{\hat{D}_r} \boxtimes \sigma_{\mathrm{SC}} \Rightarrow \nu_{\hat{D}} \boxtimes \sigma_{\mathrm{SC}}$ for the corresponding symmetric source measures of the map $x \mapsto x^2$. In conclusion, it suffices hereafter to prove the theorem for the case where $\hat{D}_i^{(n)} \in \mathcal{S}$, a fixed finite set, for all n.

Step II For $1 \leq a \leq \ell$, let $m_a^{(n)} = |\mathsf{V}_n^a|$ where $\mathsf{V}_n^a = \{v \in [n] : \deg(v) = d_a\omega_n\}$ is the set of vertices of degree $d_a\omega_n$ in G_n. By assumption, $m_a^{(n)}/n \to \nu_a$ for $\nu_a := \nu_{\hat{D}}(\{d_a\})$. (Observe that our choice of ω_n dictates that $\sum_a d_a\nu_a = 1$.) For all $1 \leq a, b \leq \ell$, set

$$q_{a,b} := d_a d_b \nu_b \,.$$

Let $\mathsf{H}_n = \cup_{a\leq b}\mathsf{H}_{a,b}^{(n)}$ for the edge-disjoint multigraphs $\mathsf{H}_{a,b}^{(n)}$ that are generated by the configuration model in the following way.

- For $1 \leq a \leq \ell$, let $\mathsf{H}_{(a,a)}^{(n)}$ be the random $D_{a,a}^{(n)}$-regular multigraph on V_n^a, where $D_{a,a}^{(n)} m_a^{(n)}$ is even and $\hat{D}_{a,a}^{(n)} := D_{a,a}^{(n)}/\omega_n$ converges to $q_{a,a}$ as $n \to \infty$.
- For $1 \leq a < b \leq \ell$, let $\mathsf{H}_{a,b}^{(n)}$ be the random bipartite multigraph with sides $(\mathsf{V}_n^a, \mathsf{V}_n^b)$ and degrees $D_{a,b}^{(n)}$ in V_n^a and $D_{b,a}^{(n)}$ in V_n^b, such that the detailed balance

$$D_{a,b}^{(n)} m_a^{(n)} = D_{b,a}^{(n)} m_b^{(n)}$$

  holds, and $\hat{D}_{a,b}^{(n)} := D_{a,b}^{(n)}/\omega_n$ tends to $q_{a,b}$ as $n \to \infty$ (hence, $\hat{D}_{b,a}^{(n)} \to q_{b,a}$).

Finally, setting

$$\lambda_{a,b}^{(n)} := \frac{\omega_n}{n} d_a d_b \,, \tag{19}$$

let $\widetilde{\mathbf{A}}_n$ denote the single-adjacency matrix of the multigraph $\widetilde{\mathsf{H}}_n = \cup_{a\leq b}\widetilde{\mathsf{H}}_{a,b}^{(n)}$, where the edge-disjoint multigraphs $\widetilde{\mathsf{H}}_{a,b}^{(n)}$ are defined as follows.

- For $1 \leq a \leq b \leq \ell$, mutually independently set the multiplicity of the edge between distinct $i \in \mathsf{V}_a^n$ and $j \in \mathsf{V}_b^n$ in $\widetilde{\mathsf{H}}_{a,b}^{(n)}$ to be a $\mathrm{Po}(\lambda_{a,b}^{(n)})$ random variable.

- For $1 \leq a \leq \ell$, mutually independently set the number of loops incident to $i \in \mathsf{V}_n^a$ to be a $\mathrm{Po}(\frac{1}{2}\lambda_{a,a}^{(n)})$ random variable.

Our next proposition shows that $\mathcal{L}^{\hat{\mathbf{A}}_n} \Rightarrow \nu_{\hat{D}} \boxtimes \sigma_{\mathrm{SC}}$, in probability, whenever

$$\mathcal{L}^{\omega_n^{-1/2}\widetilde{\mathbf{A}}_n} \Rightarrow \nu_{\hat{D}} \boxtimes \sigma_{\mathrm{SC}}\,, \qquad \text{in probability}\,. \tag{20}$$

Proposition 4 *The empirical spectral measures of* $\mathbf{A}_n, \mathbf{A}'_n$ *and* $\widetilde{\mathbf{A}}_n$*, the respective single-adjacency matrices of* $\mathsf{G}_n, \mathsf{H}_n$ *and* $\widetilde{\mathsf{H}}_n$*, satisfy*

$$d_{\mathrm{BL}}\left(\mathcal{L}^{\omega_n^{-1/2}\mathbf{A}_n}, \mathcal{L}^{\omega_n^{-1/2}\mathbf{A}'_n}\right) = o(1) \quad \textit{and} \quad d_{\mathrm{BL}}\left(\mathcal{L}^{\omega_n^{-1/2}\mathbf{A}'_n}, \mathcal{L}^{\omega_n^{-1/2}\widetilde{\mathbf{A}}_n}\right) = o(1)\,,$$

in probability, as $n \to \infty$.

Proof Setting

$$\mathsf{G}_n^{(0)} = \mathsf{G}_n\,, \quad \mathsf{G}_n^{(2)} = \mathsf{H}_n\,, \quad \mathsf{G}_n^{(4)} = \widetilde{\mathsf{H}}_n\,,$$

associate with each multigraph its sub-degrees (accounting for edge multiplicities),

$$D_{i,b}^{(n,k)} := \sum_{j \in \mathsf{V}_n^b} (\mathbf{A}_{\mathsf{G}_n^{(k)}})_{i,j}\,, \qquad i \in [n]\,, \qquad 1 \leq b \leq \ell\,,$$

so in particular $D_{i,b}^{(n,2)} = D_{a(i),b}^{(n)}$ where $a(i)$ is such that $i \in \mathsf{V}_n^a$. Of course, for $k = 0, 2, 4$,

$$m_{a,b}^{(n,k)} := \sum_{i \in \mathsf{V}_n^a} D_{i,b}^{(n,k)} = m_{b,a}^{(n,k)}\,, \quad m_{a,a}^{(n,k)} \text{ is even}, \qquad 1 \leq a, b \leq \ell\,. \tag{21}$$

Claim Conditional on a given sequence of sub-degrees $\{D_{i,b}^{(n,k)}\}$, the adjacency matrices $\mathbf{A}_{\mathsf{G}_n^{(k)}}$ for $k \in \{0, 2, 4\}$ all have the same conditional law.

Proof Observe that $\mathsf{G}_n = \mathsf{G}_n^{(0)}$ gives the same weight to each perfect matching of its half-edges, thus conditioning on $\{D_{i,b}^{(n,k)}\}$ amounts to specifying a subset of permissible matchings, on which the conditional distribution would be uniform. The same applies to the graphs $\mathsf{H}_{(a,b)}^{(n)}$ for all $1 \leq a \leq b \leq \ell$, each being an independently drawn uniform multigraph, and hence to their union $\mathsf{H}_n = \mathsf{G}_n^{(2)}$, thus establishing the claim for $k = 0, 2$. To treat $k = 4$, notice that the probability that the multigraph $\mathsf{H}_{(a,b)}^{(n)}$, $a \neq b$, given the sub-degrees $\{D_{i,b}^{(n,k)}\}$, features the adjacency matrix $\mathbf{a} := (a_{i,j})$ $(i \in \mathsf{V}_n^a, j \in \mathsf{V}_n^b)$, is

$$\frac{1}{m_{a,b}^{(n,k)}!}\left(\prod_{i \in \mathsf{V}_n^a} \frac{D_{i,b}^{(n,k)}!}{\prod_{j \in \mathsf{V}_n^b} a_{i,j}!}\right)\left(\prod_{j \in \mathsf{V}_n^b} D_{j,a}^{(n,k)}!\right) \propto \prod_{i \in \mathsf{V}_n^a} \prod_{j \in \mathsf{V}_n^b} \frac{1}{a_{i,j}!}$$

by the definition of the configuration model. As the distribution of a vector of t i.i.d. Poisson variables with mean λ, conditional on their sum being m, is multinomial with parameters $(m, \frac{1}{t}, \dots, \frac{1}{t})$, the analogous conditional probability under $\widetilde{\mathsf{H}}^{(n)}_{(a,b)}$ is

$$\prod_{i\in\mathsf{V}_n^a} \frac{D_{i,b}^{(n,k)}!}{\prod_{j\in\mathsf{V}_n^b} a_{i,j}!} |\mathsf{V}_n^b|^{-D_{i,b}^{(n,k)}} \propto \prod_{i\in V_a^n} \prod_{j\in\mathsf{V}_n^b} \frac{1}{a_{i,j}!} .$$

Lastly, the probability that $\mathsf{H}^{(n)}_{(a,a)}$, conditional on $\{D_{i,b}^{(n,k)}\}$, assigns to $\mathbf{a} = (a_{i,j})$ is

$$\prod_{i\in\mathsf{V}_n^a} \frac{D_i!}{2^{a_{i,i}}} \prod_{\substack{j\in\mathsf{V}_n^a\\ j>i}} \frac{1}{a_{i,j}!} \propto 2^{-\sum_i a_{i,i}} \prod_{\substack{i,j\in\mathsf{V}_n^a\\ j>i}} \frac{1}{a_{i,j}!} ,$$

whereas the analogous conditional probability under $\widetilde{\mathsf{H}}^{(n)}_{(a,b)}$ (now involving a vector that is multinomial with parameters $(D_{i,b}^{(n,k)}, \frac{1}{2t+1}, \frac{2}{2t+1}, \dots, \frac{2}{2t+1})$ for $t = |\{j \in \mathsf{V}_n^a : j \geq i\}|$, recalling the factor of 2 in the definition of the rate of loops under $\widetilde{H}^{(n)}_{(a,a)}$), is

$$\prod_{i\in\mathsf{V}_n^a} \frac{D_{i,b}^{(n,k)}!}{\prod_{\substack{j\in\mathsf{V}_n^a\\ j>i}} a_{i,j}!} 2^{-a_{i,i}} \left(\frac{2}{|\{j \in \mathsf{V}_n^a : j \geq i\}|} \right)^{-D_{i,b}^{(n,k)}} \propto 2^{-\sum_i a_{i,i}} \prod_{\substack{i,j\in\mathsf{V}_n^a\\ j>i}} \frac{1}{a_{i,j}!} .$$

This completes the proof of the claim.

We will introduce two auxiliary multigraphs $\mathsf{G}_n^{(1)}$ and $\mathsf{G}_n^{(3)}$ having the latter property, and further, the corresponding single-adjacency matrices (or single-edge sets $\mathsf{E}_n^{(k)}$), can be coupled in such a way that

$$\sum_{k=1}^{4} \mathbb{E}\Big[\big|\mathsf{E}_n^{(k)} \triangle \mathsf{E}_n^{(k-1)}\big|\Big] = o(n\omega_n) . \tag{22}$$

It follows that, under the resulting coupling, both $\mathbb{E}[\mathrm{tr}\,\big((\mathbf{A}_n - \mathbf{A}'_n)^2\big)] = o(n\omega_n)$ and $\mathbb{E}[\mathrm{tr}\,\big((\mathbf{A}'_n - \widetilde{\mathbf{A}}_n)^2\big)] = o(n\omega_n)$, yielding Proposition 4 via the Hoffman–Wielandt bound.

Proceeding to construct the multigraph $\mathsf{G}_n^{(1)}$, write, for all $i \in [n]$ and $1 \leq b \leq \ell$,

$$D_{i,b}^{(n,1)} = D_{i,b}^{(n,0)} \wedge D_{i,b}^{(n,2)} , \tag{23}$$

then further uniformly reduce the number of potential half-edges in $\mathsf{G}_n^{(1)}$ until achieving (21) for $k = 1$. That is, if (23) yields $m_{a,b}^{(n,1)} > m_{b,a}^{(n,1)}$ for some $a \neq b$, we uniformly choose and eliminate $m^{(n,1)}_{a,b} - m^{(n,1)}_{b,a}$ potential half-edges leading from V^a_n to V^b_n and accordingly adjust $\{D^{(n,1)}_{i,b}, i \in \mathsf{V}^a_n\}$, an operation which only affects the constraint (21) for that particular $a \neq b$. With Observation 1 in mind, construct two *bridge* copies of the random multigraph $\mathsf{G}^{(1)}_n$ with the adjusted sub-degrees $\{D^{(n,1)}_{i,b}\}$, as follows:

- For each i and b, mark as BLUE(b) a uniformly chosen subset of $D^{(n,1)}_{i,b}$ half-edges incident to vertex i, the other part of which is, according to $\mathsf{G}^{(0)}_n$, in V^b_n.
- Retain for $\mathsf{G}^{(1)}_n$ every edge of $\mathsf{G}^{(0)}_n$ where both parts are marked with BLUE.
- After removing all non-BLUE half-edges of $\mathsf{G}^{(0)}_n$, complete the construction of $\mathsf{G}^{(1)}_n$ by uniformly matching, for each $a \geq b$, all unmatched BLUE(b) half-edges of V^a_n to all unmatched BLUE(a) half-edges of V^b_n.
- A second copy of $\mathsf{G}^{(1)}_n$ is obtained by repeating the preceding construction, now with $\mathsf{G}^{(2)}_n$ taking the role of $\mathsf{G}^{(0)}_n$.

Replacing in the above procedure the multigraph $\mathsf{G}^{(0)}_n$ by the multigraph $\mathsf{G}^{(4)}_n$, the same construction produces a multigraph $\mathsf{G}^{(3)}_n$ having sub-degrees

$$D^{(n,3)}_{i,b} \leq D^{(n,2)}_{i,b} \wedge D^{(n,4)}_{i,b}, \tag{24}$$

and two *bridge* copies of $\mathsf{G}^{(3)}_n$ which are coupled (using such BLUE marking), to $\mathsf{G}^{(2)}_n$ and $\mathsf{G}^{(4)}_n$, respectively.

Next, as for (22), recall that $|\mathsf{E}^{(k)}_n \triangle \mathsf{E}^{(k-1)}_n| \leq |\mathsf{E}_{\mathsf{G}^{(k)}_n} \triangle \mathsf{E}_{\mathsf{G}^{(k-1)}_n}|$, which under our coupling is at most the number of edges of $\mathsf{G}^{(2[k/2])}_n$ that had at least one non-BLUE part. This in turn is at most

$$\Delta^{(n)} := \sum_{a,b=1}^{\ell} |m^{(n,k)}_{a,b} - m^{(n,k-1)}_{a,b}|.$$

Our construction is such that $m^{(n,0)}_{a,b} \wedge m^{(n,2)}_{a,b} \geq m^{(n,1)}_{a,b}$ and $m^{(n,4)}_{a,b} \wedge m^{(n,2)}_{a,b} \geq m^{(n,3)}_{a,b}$. Further, if the sub-degrees of bridge multigraphs were set by (23), then

$$m^{(n,0)}_{a,b} + m^{(n,2)}_{a,b} - 2m^{(n,1)}_{a,b} = \sum_{i \in \mathsf{V}^a_n} |D^{(n,0)}_{i,b} - D^{(n)}_{a,b}| := \Delta^{(n,1)}_{a,b},$$

for any $1 \leq a, b \leq \ell$, with analogous identities relating $m^{(n,3)}_{a,b}$ and $\Delta^{(n,3)}_{a,b}$. Since (21) holds for $k = 0, 2, 4$, while $m^{(n,1)}_{a,b} \wedge m^{(n,1)}_{b,a}$, $b < a$ are not changed by the $\mathsf{G}^{(1)}_n$ sub-degree adjustments (and similarly for the $\mathsf{G}^{(3)}_n$ sub-degree adjustments), we deduce that

$$\Delta^{(n)} \leq 2 \sum_{a,b=1}^{\ell} \Delta^{(n,1)}_{a,b} + 2 \sum_{a,b=1}^{\ell} \Delta^{(n,3)}_{a,b}.$$

Thus, we have (22) as soon as we show that for any $1 \le a, b \le \ell$,

$$\mathbb{E}\Delta_{a,b}^{(n,1)} + \mathbb{E}\Delta_{a,b}^{(n,3)} = o(n\omega_n)\,,$$

which by our choice of $\{D_{a,b}^{(n)}\}$ follows from having for any fixed $i \in \mathsf{V}_n^a$,

$$\mathbb{E}|\omega_n^{-1} D_{i,b}^{(n,0)} - q_{a,b}| + \mathbb{E}|\omega_n^{-1} D_{i,b}^{(n,4)} - q_{a,b}| = o(1)\,. \tag{25}$$

For $i \in \mathsf{V}_n^a$ the variable $D_{i,b}^{(n,4)}$ is Poisson with mean $(1 + o(1))\lambda_{a,b}^{(n)} m_b^{(n)} = \omega_n q_{a,b}(1 + o(1))$ (see (19)), hence $\mathbb{E}|\omega_n^{-1} D_{i,b}^{(n,4)} - q_{a,b}| \to 0$. Similarly, $D_{i,b}^{(n,0)}$ counts how many of the $d_a\omega_n$ half-edges emanating from such i, are paired by the uniform matching of the half-edges of G_n, with half-edges from the subset E_n^b of those incident to V_n^b. With $|\mathsf{E}_n^b| = d_b\omega_n m_b^{(n)}$, the probability of a specific half-edge paired with an element of E_n^b is $\mu_n = (|\mathsf{E}_n^b| - 1_{\{a=b\}})/(2|\mathsf{E}_{\mathsf{G}_n}| - 1) \to d_b\nu_b$, hence $\omega_n^{-1}\mathbb{E}D_{i,b}^{(n,0)} = d_a\mu_n \to q_{a,b}$. It is not hard to verify that two specific half-edges incident to $i \in \mathsf{V}_n^a$ are both paired with elements of E_n^b with probability $v_n = \mu_n^2(1 + o(1))$. Consequently,

$$\mathrm{Var}(\omega_n^{-1} D_{i,b}^{(n,0)}) \le d_a \frac{\mu_n}{\omega_n} + d_a^2(v_n - \mu_n^2) \to 0\,,$$

yielding the L^2-convergence of $\omega_n^{-1} D_{i,b}^{(n,0)}$ to $q_{a,b}$ and thereby establishing (25).

Step III We proceed to verify (20) for the single-adjacency matrices $\widetilde{\mathbf{A}}_n$ of $\widetilde{\mathsf{H}}_n$. To this end, as argued before, such weak convergence as in (20) is not affected by changing $o(n\omega_n)$ of the entries of $\widetilde{\mathbf{A}}_n$, so WLOG we modify the law of number of loops in $\widetilde{\mathsf{H}}_n$ incident to each $i \in \mathsf{V}_n^a$ to be a $\mathrm{Po}(\lambda_{a,a}^{(n)})$ variable, yielding the symmetric matrix $\widetilde{\mathbf{A}}_n$ of independent upper triangular Bernoulli$(p_{a,b}^{(n)})$ entries, where $p_{a,b}^{(n)} = 1 - \exp(-\lambda_{a,b}^{(n)})$ when $i \in \mathsf{V}_n^a$ and $j \in \mathsf{V}_n^b$. In particular, the rank of $\mathbb{E}\widetilde{\mathbf{A}}_n$ is at most ℓ, so by Lidskii's theorem we get (20) upon proving that $\mathcal{L}^{\hat{\mathbf{B}}_n} \Rightarrow \nu_{\hat{D}} \boxtimes \sigma_{\mathrm{SC}}$ in probability, for $\hat{\mathbf{B}}_n := \omega_n^{-1/2}(\widetilde{\mathbf{A}}_n - \mathbb{E}\widetilde{\mathbf{A}}_n)$, a symmetric matrix of uniformly (in n) bounded, independent upper-triangular entries $\{\hat{Z}_{ij}\}$, having zero mean and variance $v_{a,b}^{(n)} := \omega_n^{-1} p_{a,b}^{(n)}(1 - p_{a,b}^{(n)}) = \frac{1}{n} d_a d_b(1 + o(1))$ when $i \in \mathsf{V}_n^a$, $j \in \mathsf{V}_n^b$. As a special case of Remark 1 (corresponding to piecewise-constant diagonal matrices with values $\{d_a\}_{a=1}^{\ell}$), such convergence holds for the symmetric matrices $\mathbf{B}_n$, whose independent centered Gaussian entries Z_{ij} have variance $v_{a,b}^{(n)}$ when $i \in \mathsf{V}_n^a$ and $j \in \mathsf{V}_n^b$, subject to on-diagonal rescaling $\mathbb{E}Z_{ii}^2 = 2v_{a(i),a(i)}^{(n)}$. As in the classical proof of Wigner's theorem by the moment's method (cf. [1, Sec. 2.1.3]), it is easy to check that for any fixed $k = 1, 2, \ldots$,

$$\mathbb{E}\Big[\frac{1}{n}\,\mathrm{tr}(\hat{\mathbf{B}}_n^k)\Big] = \mathbb{E}\Big[\frac{1}{n}\,\mathrm{tr}(\mathbf{B}_n^k)\Big](1 + o(1))\,,$$

since both expressions are dominated by those cycles of length k that pass via each entry of the relevant matrix exactly twice (or not at all). Further, adapting the concentration argument of [1, Sec. 2.1.4] we deduce that as in the Wigner's case, $\langle x^k, \mathcal{L}^{\hat{\mathbf{B}}_n} - \mathbb{E}\mathcal{L}^{\hat{\mathbf{B}}_n}\rangle \to 0$ in probability, for each fixed k, thereby completing the Proof of Theorem 1(a).

To prove Theorem 1(b), recall that $|\mathsf{E}_n \triangle \mathscr{E}_n| \le \Delta_n$ for any coupling of the pair of matching which generate the graphs G_n and $\mathscr{G}_n$. Appealing to Proposition 1 and the bound (7) following it, we get that under the coupling μ provided by that proposition,

$$\mathbb{E}_\mu[|\mathsf{E}_n \triangle \mathscr{E}_n|] \le \mathbb{E}_\mu[\Delta_n] \le \sqrt{2\mathbb{E}_\mu[\Delta_n(\Delta_n - 1)]} \le 4b_n\,,$$

where (recalling from (1) that $\omega_n = (2 + o(1))|\mathsf{E}_n|/n$)

$$b_n^2 := \sqrt{2|\mathsf{E}_n|}\sum_{j=1}^n D_j^2 = (1 + o(1))n^{3/2}\sqrt{\omega_n}\mathbb{E}_{U_n}(D_{U_n}^{(n)})^2$$
$$= (1 + o(1))n^{3/2}\omega_n^{5/2}\mathbb{E}_{U_n}(\hat{D}_{U_n}^{(n)})^2 = o(n^2\omega_n^2)$$

via our assumption on the RHS of (3); thus, $\mathbb{E}_\mu[|\mathsf{E}_n \triangle \mathscr{E}_n|] = o(n\omega_n)$. We claim that Lemma 1 then concludes the proof. To see this, set $\hat{\mathbf{B}}'_n \equiv \omega_n^{-1/2}\mathbf{A}_{\mathscr{G}_n}$ and further let $\hat{\mathbf{A}}_n \equiv \omega_n^{-1/2}\mathbf{A}_n$ for the single-adjacency matrix $\mathbf{A}_n$ associated with $\mathbf{A}_{\mathsf{G}_n}$. Since the entries of $\mathbf{A}_n$ and $\mathbf{A}_{\mathscr{G}_n}$ may differ at most by one from each other, (2) implies that

$$\mathbb{E}_\mu\Big[\frac{1}{n}\,\mathrm{tr}\big((\hat{\mathbf{A}}_n - \hat{\mathbf{B}}'_n)^2\big)\Big] \le \frac{2}{n\omega_n}\mathbb{E}_\mu[|\mathsf{E}_n \triangle \mathscr{E}_n|] \to 0\,,$$

as required for Lemma 1.

Proof of Corollary 1 The assumed growth of ω_n yields (2) out of (1). In case of G_n, the latter amounts to

$$\frac{1}{n}\sum_{i=1}^n \hat{D}_i^{(n)} \to 1\,, \qquad \text{in probability,} \tag{26}$$

which we get by applying the L^2-WLLN for triangular arrays with uniformly bounded second moments. The same reasoning yields the required uniform integrability in (3), namely, that when $n \to \infty$ followed by $r \to \infty$

$$\frac{1}{n}\sum_{i=1}^n \hat{D}_i^{(n)} 1_{\{\hat{D}_i^{(n)} \ge r\}} \to 0\,, \qquad \text{in probability.} \tag{27}$$

Further, applying the weak law for non-negative triangular arrays $\{(\hat{D}_i^{(n)})^2\}$ of uniformly bounded mean, at truncation level $b_n \gg n$, it is not hard to deduce that

$$\frac{1}{b_n}\sum_{i=1}^{n}(\hat{D}_i^{(n)})^2 \to 0\,, \qquad \text{in probability,} \tag{28}$$

whereupon, considering $b_n = n/\sqrt{\omega_n/n}$ results with the RHS of (3). Next, recall that the empirical measures $\mathcal{L}^{\hat{\mathbf{A}}_n}$ of i.i.d. $\hat{D}_i^{(n)}$ converge in probability to the weak limit $\nu_{\hat{D}}$ of the laws of $\hat{D}_1^{(n)}$. Thus, Theorem 1(a) applies for G_n of degrees $[\omega_n \hat{D}_i^{(n)}]$, yielding Corollary 1 in this case.

Turning to the case of uniform simple graphs, thanks to (27), truncating the degrees $[\omega_n \hat{D}_i^{(n)}]$ at some $\bar{d}_n \gg \omega_n$ removes at most $o(n\omega_n)$ edges from E_n. Thus, such truncation neither affects (1), nor the preceding verification of (3). Further, such truncation alters only $o(n)$ degrees, yielding the same limit $\nu_{\hat{D}}$ for $\mathcal{L}^{\hat{\mathbf{A}}_n}$. In view of Theorem 1(b), the stated convergence of $\mathcal{L}^{\hat{\mathbf{A}}_{\mathscr{G}_n}}$ holds, provided that $\{[\omega_n \hat{D}_i^{(n)}] \wedge \bar{d}_n\}$ are graphical WHP as $n \to \infty$. To this end, inspired by the proof of [3, Theorem 1(d)], recall from the Erdős-Gallai theorem, that integers $d_1 \ge d_2 \ge \cdots \ge d_n \ge 0$ are graphical if

$$2\sum_{i=1}^{j} d_i \le j(j-1) + \sum_{i=1}^{n}\min(j, d_i)\,, \qquad \forall\, 1 \le j \le n\,. \tag{29}$$

Thanks to (2) we can fix $j_n = o(n)$ such that $j_n/\sqrt{n\omega_n} \to \infty$. The LHS of (29) is in our setting at most $2\omega_n \sum_i \hat{D}_i^{(n)}$, which in view of (26) is for $j > j_n$ negligible in comparison with the term $j(j-1)$ on the RHS of (29). Denoting by $o_p(1)$ the LHS of (28) at $b_n = n^2/j_n \gg n$, we further have here that the LHS of (29) is at most

$$2\min\Big(j\bar{d}_n, n\omega_n\, o_p(1)\Big)\,, \qquad \forall\, 1 \le j \le j_n\,. \tag{30}$$

The Paley–Zygmund inequality yields $\inf_n \mathbb{P}(\hat{D}_i^{(n)} \ge 2/3) \ge 2\delta$, for some $\delta > 0$. Hence,

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{D_i^{(n)} \ge \omega_n/3\}} > \delta\,, \qquad \text{in probability.}$$

This yields that the right-most term in (29) is for all large n and $j \in [n]$, at least

$$\delta \min(jn, n\omega_n/3)\,,$$

which in turn exceeds (30) (as $\bar{d}_n = o(n)$), thus completing the proof.

3 Coupling Simple Graphs and Multigraphs: Proof of Proposition 1

Fixing *graphical* degrees $D_1 \geq D_2 \geq \cdots \geq D_n$, let $\mathsf{m}_n := \sum_i D_i = 2|\mathsf{E}_n|$. Enumerate the m_n half-edges as follows: each half-edge e is identified with a vertex $v(e) \in [n]$; the first D_1 half-edges have $v(e) = 1$, the next D_2 have $v(e) = 2$ and so on. A matching of half-edges $m : [\mathsf{m}_n] \mapsto [\mathsf{m}_n]$ is an involution without fixed points (i.e., $m(e) = m^{-1}(e)$ and $m(e) \neq e$ for all $e \in [\mathsf{m}_n]$). A coupled pair of multigraphs $(\mathsf{G}_n, \mathscr{G}_n)$ is hereby represented by a pair of matching (X, Y), restricting $\mathsf{Y}(\cdot)$ to the non-empty collection of matching that correspond to a simple graph; namely, $v(e) \neq v(\mathsf{Y}(e))$ (no loops) and $\{v(e), v(\mathsf{Y}(e))\} \neq \{v(f), v(\mathsf{Y}(f))\}$ (no multiple edges) for any $f \neq \{e, \mathsf{Y}(e)\}$.

Starting from any such pair of matching $(\mathsf{X}_0, \mathsf{Y}_0)$, consider the switching Markov chain $(\mathsf{X}_k, \mathsf{Y}_k)$ that proceeds as following (see also Fig. 4):

- Uniformly choose $e \neq f \in [\mathsf{m}_n]$ and disconnect their matching in X_k and Y_k;
- Reconnect e with f, and $\mathsf{X}_k(e)$ with $\mathsf{X}_k(f)$, to get the match X_{k+1};
- If reconnecting e with f and $\mathsf{Y}_k(e)$ with $\mathsf{Y}_k(f)$ yields a simple graph, set this to be Y_{k+1}. Otherwise, leave $\mathsf{Y}_{k+1} = \mathsf{Y}_k$ unchanged.

We say that coupling succeeds in the k-th step if the proposed move to Y_{k+1} results in a simple graph, otherwise saying that the coupling failed (in the k-th step).

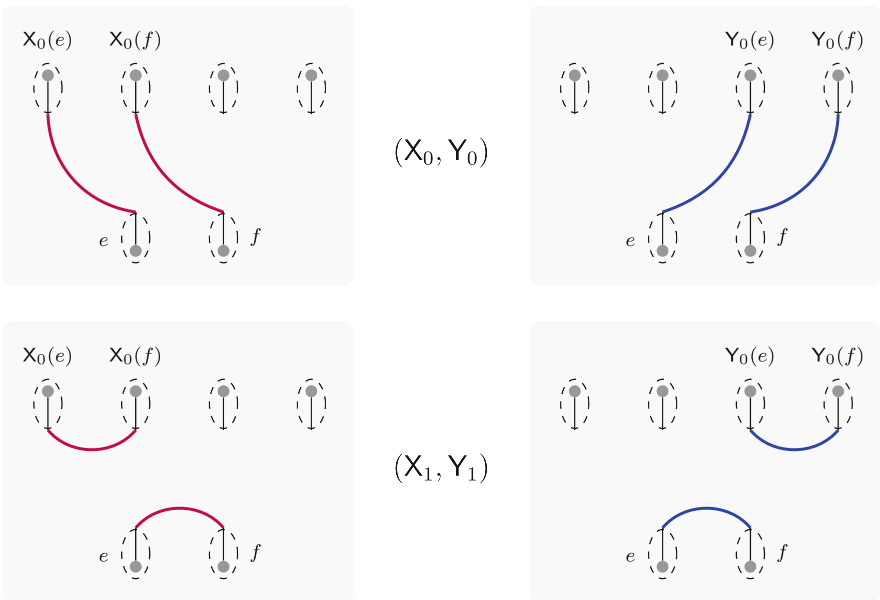


Fig. 4 Coupling of the chains $(\mathsf{X}_t, \mathsf{Y}_t)$ corresponding to $(\mathsf{G}_n, \mathscr{G}_n)$

The marginal (X_k) evolves as a Markov chain in the space of all matching, with the marginal (Y_k) likewise evolving as a Markov chain in the non-empty subset of all matching that correspond to simple graphs with the specified degrees. These switching chains are further reversible with respect to the corresponding uniform measures. Both marginal chains have been extensively studied as means of sampling uniform graphs subject to given degrees. In particular, it is well-known ([18]; cf. also the recent work [10]) that each of these marginals is an irreducible Markov chain. Having a non-empty finite state space, the Markov chain $(\mathsf{X}_k, \mathsf{Y}_k)$ admits an invariant probability measure μ, and by the preceding, any such μ is a coupling between the random multigraph G_n and the corresponding uniformly simple graph $\mathcal{G}_n$ of the specified degrees.

Denoting by

$$\mathsf{C}_k \equiv \{e \in [\mathsf{m}_n] : \mathsf{X}_k(e) = \mathsf{Y}_k(e)\},$$

the common part of the two matching $\mathsf{X}_k, \mathsf{Y}_k$, note that under an invariance measure $\mathbb{E}_\mu[|\mathsf{C}_k|]$ must be independent of k. We further have the following lower bound on the change between $|\mathsf{C}_{k+1}|$ and $|\mathsf{C}_k|$:

$$|\mathsf{C}_{k+1}| - |\mathsf{C}_k| \geq 2\mathbb{1}_{\{e,f \notin \mathsf{C}_k\}} - 4\mathbb{1}_{\{\text{coupling fails in step } k\}}\,. \tag{31}$$

Indeed, (31) is verified by enumerating over the seven possible cases for $e, f \in [\mathsf{m}_n]$:

I. $\mathsf{X}_0(e) = \mathsf{Y}_0(e) = f$;
II. $\mathsf{X}_0(e) = \mathsf{Y}_0(e) \neq f, \mathsf{X}_0(f) = \mathsf{Y}_0(f)$;
III. $\mathsf{X}_0(e) = \mathsf{Y}_0(e) \neq f, \mathsf{X}_0(f) \neq \mathsf{Y}_0(f)$ or $\mathsf{X}_0(f) = \mathsf{Y}_0(f) \neq e, \mathsf{X}_0(e) \neq \mathsf{Y}_0(e)$;
IV. $\mathsf{X}_0(e) = f \neq \mathsf{Y}_0(e)$ or $\mathsf{Y}_0(e) = f \neq \mathsf{X}_0(e)$;
V. $\mathsf{X}_0(e) = \mathsf{Y}_0(f), \mathsf{X}_0(f) = \mathsf{Y}_0(e)$;
VI. $\mathsf{X}_0(e) = \mathsf{Y}_0(f), \mathsf{X}_0(f) \neq \mathsf{Y}_0(e)$ or $\mathsf{X}_0(f) = \mathsf{Y}_0(e), \mathsf{X}_0(e) \neq \mathsf{Y}_0(f)$;
VII. $e, f, \mathsf{X}_0(e), \mathsf{X}_0(f), \mathsf{Y}_0(e), \mathsf{Y}_0(f)$ are six distinct half-edges.

The corresponding value of $|\mathsf{C}_1|-|\mathsf{C}_0|$ in each of these cases are given in Table 1, from which it follows that under an invariant measure μ,

$$0 = \mathbb{E}[|\mathsf{C}_1| - |\mathsf{C}_0|] \geq 2\,\mathbb{P}(e, f \notin \mathsf{C}_0) - 4\,\mathbb{P}(\text{coupling fails})\,. \tag{32}$$

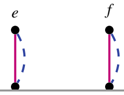
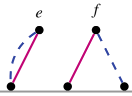
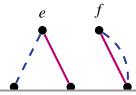
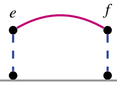
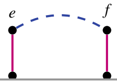
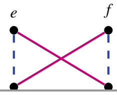
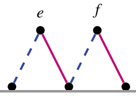
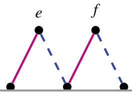
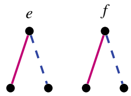
For the first term on the RHS of (32),

$$\mathbb{P}(e, f \notin \mathsf{C}_0 \mid \mathsf{C}_0) = \left(\frac{\mathsf{m}_n - |\mathsf{C}_0|}{\mathsf{m}_n}\right)\left(\frac{\mathsf{m}_n - |\mathsf{C}_0| - 1}{\mathsf{m}_n - 1}\right). \tag{33}$$

We consequently get that the LHS of (6) is at most $2\mathsf{m}_n(\mathsf{m}_n - 1)\,\mathbb{P}(\text{coupling fails})$.

For the latter, note that the coupling fails only under one of the following scenarios, where we introduce

Table 1 Analysis of the change in the size of the common part of the two matchings after one step of the coupling. In cases marked by $^{(\star)}$, the difference could be larger if $\mathsf{X}_0(\mathsf{Y}_0(e)) = \mathsf{Y}_0(f)$ or $\mathsf{Y}_0(\mathsf{X}_0(e)) = \mathsf{X}_0(f)$

		$\lvert \mathsf{C}_1\rvert - \lvert \mathsf{C}_0\rvert$	
Case	Criterion	Success	Failure
I	$\mathsf{X}_0(e) = f = \mathsf{Y}_0(e)$ 	0	—
II	$\mathsf{X}_0(e) = \mathsf{Y}_0(e) \neq f$ $\mathsf{X}_0(f) = \mathsf{Y}_0(f)$ 	0	−4
III	$\mathsf{X}_0(e) = \mathsf{Y}_0(e) \neq f$ $\mathsf{X}_0(f) \neq \mathsf{Y}_0(f)$ 	0	−2
	$\mathsf{X}_0(f) = \mathsf{Y}_0(f) \neq e$ $\mathsf{X}_0(e) \neq \mathsf{Y}_0(e)$ 		
IV	$\mathsf{X}_0(e) = f \neq \mathsf{Y}_0(e)$ 	$^{(\star)}$ ≥ 2	0
	$\mathsf{Y}_0(e) = f \neq \mathsf{X}_0(e)$ 		
V	$\mathsf{X}_0(e) = \mathsf{Y}_0(f)$ $\mathsf{X}_0(f) = \mathsf{Y}_0(e)$ 	4	0
VI	$\mathsf{X}_0(e) = \mathsf{Y}_0(f)$ $\mathsf{X}_0(f) \neq \mathsf{Y}_0(e)$ 	2	0
	$\mathsf{X}_0(f) = \mathsf{Y}_0(e)$ $\mathsf{X}_0(e) \neq \mathsf{Y}_0(f)$ 		
VII	$e, \mathsf{X}_0(e), \mathsf{Y}_0(e),$ $f, \mathsf{X}_0(f), \mathsf{Y}_0(f)$ are all distinct 	$^{(\star)}$ ≥ 2	$^{(\star)}$ ≥ 0

(a) *a loop*: $v(e) = v(f)$ or $v(\mathsf{Y}_0(e)) = v(\mathsf{Y}_0(f))$;
(b) *multiple edges*: $v(e)$ is connected to $v(f)$ in $\mathsf{Y}_0 \setminus \{(e, \mathsf{Y}_0(e)), (f, \mathsf{Y}_0(f))\}$, or $v(\mathsf{Y}_0(e))$ is connected to $v(\mathsf{Y}_0(f))$ in $\mathsf{Y}_0 \setminus \{(e, \mathsf{Y}_0(e)), (f, \mathsf{Y}_0(f))\}$.

As $(\mathsf{Y}_0(e), \mathsf{Y}_0(f))$ has the same (uniform) distribution as (e, f), we thus deduce that

$$\frac{1}{2}\mathbb{P}(\text{coupling fails}) \leq \mathbb{P}(v(e) = v(f)) + \mathbb{P}(v(e) \text{ connected to } v(f))\,.$$

With q_{ij} denoting the probability that $i \neq j$ are adjacent in Y_0, clearly

$$\mathbb{P}(v(e) \text{ connected to } v(f)) = \sum_{i \neq j} \frac{(D_i - 1)(D_j - 1)}{\mathsf{m}_n(\mathsf{m}_n - 1)} q_{ij} .$$

Similarly, recalling that $\sum_j q_{ij} = D_i$ for any $i \in [n]$, we have that

$$\mathbb{P}(v(e) = v(f)) = \sum_{i=1}^{n} \frac{D_i(D_i - 1)}{\mathsf{m}_n(\mathsf{m}_n - 1)} = \sum_{i \neq j} \frac{(D_i + D_j)/2 - 1}{\mathsf{m}_n(\mathsf{m}_n - 1)} q_{ij} .$$

Adding these expressions and reducing the sum by symmetry to $j > i$, we arrive at

$$\frac{1}{2}\mathbb{P}(\text{coupling fails}) \le \frac{1}{\mathsf{m}_n(\mathsf{m}_n - 1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (2D_i D_j - D_i - D_j) q_{ij} . \tag{34}$$

With $j \mapsto D_j$ non-decreasing and $\sum_{j>i} q_{ij} \le D_i$, by replacing q_{ij} with $\mathbb{1}_{\{j \le i + D_i\}}$ we upper bound the RHS of (34). Combining this with (32)–(33) establishes (6), thereby concluding the Proof of Proposition 1. □

4 Analysis of the Limiting Density

Remark 5 With $\nu^{(2)}$ denoting the pushforward of ν by the map $x \mapsto x^2$ (that is, the weak limit of $\mathcal{L}^{\boldsymbol{\Lambda}_n^2}$), we have similarly to Remark 1 that $\mu_{\text{MP}} = \nu^{(2)} \boxtimes \sigma_{\text{SC}}^{(2)}$, where the pushforward $\sigma_{\text{SC}}^{(2)}$ (of density $(2\pi)^{-1}\sqrt{4/x - 1}$ on $[0, 4]$), is the limiting empirical distribution of singular values of $n^{-1/2}\widetilde{\mathbf{X}}_n$.

Proof of Proposition 2 The matrix $\mathbf{M}_n := n^{-1}\widetilde{\mathbf{X}}_n \boldsymbol{\Lambda}_n^2 \widetilde{\mathbf{X}}_n^\star$ has the same ESD as $n^{-1}\boldsymbol{\Lambda}_n \widetilde{\mathbf{X}}_n \widetilde{\mathbf{X}}_n^\star \boldsymbol{\Lambda}_n$. Thus, μ_{MP} is also the limiting ESD for $\mathbf{M}_n$ (see [12, 15]). Taking $\mathcal{L}^{\boldsymbol{\Lambda}_n} \Rightarrow \nu$ with $d\nu/d\nu_{\hat{D}}(x) = x$ yields the Cauchy–Stieltjes transform $G_{\mu_{\text{MP}}}(z) = h(z)$ which is the unique decaying to zero as $|z| \to \infty$, $\mathbb{C}_+$-valued analytic on $\mathbb{C}_+$, solution of

$$h = \left(\mathbb{E}\left[\frac{\hat{D}^2}{1+h\hat{D}}\right] - z\right)^{-1} = -z^{-1}\mathbb{E}\left[\frac{\hat{D}}{1+h\hat{D}}\right] . \tag{35}$$

Indeed, the LHS of (35) merely re-writes the fact that $\xi(\cdot)$ of (11) is such that $\xi(h(z)) = z$ on $\mathbb{C}_+$, while having $\int x d\nu_{\hat{D}} = 1$, one thereby gets the RHS of (35) by elementary algebra. Recall [2, Prop. 5(a)] that the Cauchy–Stieltjes transform of the symmetric measure $\widetilde{\mu}$ having the pushforward $\widetilde{\mu}^{(2)} = \mu_{\text{MP}}$ under the map $x \mapsto x^2$, is given for $\Re(z) > 0$ by $g(z) = zh(z^2) : \mathbb{C}_+ \mapsto \mathbb{C}_+$, which by the RHS of (35) satisfies for $\Re(z) > 0$,

$$g = -\mathbb{E}\Big[\frac{\hat{D}}{z + g\hat{D}}\Big]. \tag{36}$$

By the symmetry of the measure $\widetilde{\mu}$ on $\mathbb{R}$ we know that $g(-\bar{z}) = -\bar{g}(z)$ thereby extending the validity of (36) to all $z \in \mathbb{C}_+$. Applying the implicit function theorem in a suitable neighborhood of $(-z^{-1}, g) = (0, 0)$ we further deduce that $g(z) = G_{\widetilde{\mu}}(z)$ is the unique $\mathbb{C}_+$-valued, analytic on $\mathbb{C}_+$ solution of (36) tending to zero as $\Im(z) \to \infty$. Recall the S-transform defined via (4)–(5) for $\varphi \neq \delta_0$ supported on $\mathbb{R}_+$ and similarly for symmetric measure ψ. In particular (see [2, Eqn. (20]),

$$S_{\sigma_{\mathrm{SC}}}(w) = w^{-1/2}.$$

Further, from (4) we see that (36) results with $m_{\nu_{\hat{D}}}(-z^{-1}g) = g^2$, yielding

$$S_{\nu_{\hat{D}}}(g^2) = -(1 + g^{-2})z^{-1}g.$$

Since $S_\mu(w) = S_{\nu_{\hat{D}}}(w)S_{\sigma_{\mathrm{SC}}}(w)$, we get $S_\mu(g^2) = -(1+g^{-2})z^{-1}$ and consequently $m_\mu(-z^{-1}) = g^2$. The latter amounts to

$$f(z) := -z^{-1}(1 + g^2) = \int \frac{1}{-t - z} d\mu(t), \tag{37}$$

which since μ is symmetric, matches the stated relation $f(z) = G_\mu(z)$ of (8).

Proof of Proposition 3 Recall from (37) that $f(z) = -zh(z^2)^2 - z^{-1}$ for $z \in \mathbb{C}_+$ and $\Re(z) > 0$. When $z \to x \in (0, \infty)$ we further have that $h(z^2) \to h(x^2)$ and hence

$$\frac{1}{\pi}\Im(f(z)) \to -\frac{1}{\pi}\Im(xh(x^2)^2) = -2\Re(h(x^2))\widetilde{\boldsymbol{\rho}}(x), \tag{38}$$

where the last identity is due to (9). Thus, for a.e. $x > 0$ the density $\boldsymbol{\rho}(x)$ exists and given by Plemelj formula, namely the RHS of (38). The continuity of $x \mapsto h(x)$ implies the same for the symmetric density $\boldsymbol{\rho}(x)$, thereby we deduce the validity of (10) at *every* $x \neq 0$. While proving [16, Thm. 1.1] it was shown that $h(z)$ extends analytically around each $x \in \mathbb{R} \setminus \{0\}$ where $\Im(h(x)) > 0$ (see also Remark 4). In particular, (10) implies that $\boldsymbol{\rho}(x)$ is real analytic at any $x \neq 0$ where it is positive. Further, in view of (10), the support identity $\mathrm{supp}(\mu) = \mathrm{supp}(\widetilde{\mu})$ is an immediate consequence of having $\Re(h(x)) < 0$ for all $x > 0$ (as shown in Lemma 2). Similarly, the stated relation with $\mathrm{supp}(\mu_{\mathrm{MP}})$ follows from the explicit relation $\widetilde{\boldsymbol{\rho}}(x) = |x|\boldsymbol{\rho}_{\mathrm{MP}}(x^2)$. Finally, Lemma 2 provides the stated bounds on $\widetilde{\boldsymbol{\rho}}$ and

$\boldsymbol{\rho}$ (see (39) and (40), respectively), while showing that if $\nu_{\hat{D}}(\{0\}) = 0$ then μ is absolutely continuous.

Our next lemma provides the estimates we deferred when proving Proposition 3.

Lemma 2 *The function $g(z) = G_{\widetilde{\mu}}(z)$ satisfies*

$$|g(z)| \leq 1 \wedge \frac{2}{|\Re(z)|}, \qquad \forall z \in \mathbb{C}_+ \cup \mathbb{R} \tag{39}$$

and (36) holds for $z \in \mathbb{C}_+ \cup \mathbb{R} \setminus \{0\}$, resulting with $\Re(h(x)) < 0$ for $x > 0$. In addition

$$\boldsymbol{\rho}(x) \leq \frac{1}{\pi}\big((\mathbb{E}\hat{D}^{-2})^{1/2} \wedge 4|x|^{-3}\big) \qquad \forall x \in \mathbb{R}, \tag{40}$$

and if $\nu_{\hat{D}}(\{0\}) = 0$, then $\mu(\{0\}) = 0$.

Proof As explained when proving Proposition 2, by the symmetry of $\widetilde{\mu}$, we only need to consider $\Re(z) \geq 0$. Starting with $z \in \mathbb{C}_+$, let

$$z = x + i\eta \qquad \text{for } x \geq 0 \text{ and } \eta > 0,$$
$$g(z) = -y + i\gamma \qquad \text{for } y \in \mathbb{R} \text{ and } \gamma > 0.$$

Then, separating the real and imaginary parts of (36) gives

$$y = \mathbb{E}\Big[\hat{D}(x - y\hat{D})\hat{W}^{-2}\Big], \quad \gamma = \mathbb{E}\Big[\hat{D}(\eta + \gamma\hat{D})\hat{W}^{-2}\Big], \tag{41}$$

where $\hat{W} := |z + g(z)\hat{D}|$ must be a.s. strictly positive (or else $\gamma = \infty$). Next, defining

$$A = A(z) := \mathbb{E}[\hat{D}\hat{W}^{-2}], \qquad B = B(z) := \mathbb{E}[\hat{D}^2\hat{W}^{-2}], \tag{42}$$

both of which are positive and finite (or else $\gamma = \infty$), translates (41) into

$$y = Ax - By, \qquad \gamma = A\eta + B\gamma.$$

Therefore,

$$y = \frac{Ax}{1+B}, \qquad \gamma = \frac{A\eta}{1-B}. \tag{43}$$

Since $\gamma > 0$, necessarily $0 < B < 1$ and $y \geq 0$ is strictly positive iff $x > 0$. Next, by (36), Jensen's inequality and (42),

$$|g| \leq \mathbb{E}\Big[\hat{D}\hat{W}^{-1}\Big] := V(z) \leq \sqrt{B} \leq 1. \tag{44}$$

Further, letting $D \sim \nu$ be the size-biasing of $\hat{D}$ and $W := |z + g(z)D|$, we have that

$$g(z) = -\mathbb{E}[(z + g(z)D)^{-1}], \quad V = \mathbb{E}[W^{-1}], \quad A = \mathbb{E}[W^{-2}]. \tag{45}$$

With $B < 1$ we thus have by (43), (45) and Jensen's inequality, that

$$\frac{|x|A}{2} \le \frac{|x|A}{1+B} = |y| \le |g| \le V \le \sqrt{A}.$$

Consequently, $|g(z)| \le \sqrt{A} \le 2/|x|$ as claimed. Next, recall [16, Theorem 1.1] that $h(z) \to h(x)$ whenever $z \to x \neq 0$, hence same applies to $g(\cdot)$ with (39) and the bound $B(z) \le 1$, also applicable throughout $\mathbb{R} \setminus \{0\}$. Further, having $z_n \to x \neq 0$ implies that $|\Re(z_n)|$ is bounded away from zero, hence $\{A(z_n)\}$ are uniformly bounded. In view of (45), this yields the uniform integrability of $(z_n + g(z_n)D)^{-1}$ and thereby its L_1-convergence to the absolutely-integrable $(x + g(x)D)^{-1}$. Appealing to the representation (45) of $g(z)$ we conclude that (36) extends to $\mathbb{R} \setminus \{0\}$. Utilizing (36) at $z = x > 0$ we see that $0 < |g(x)|^2 \le A(x)$ due to (45). Hence, from (41) we have as claimed,

$$\Re(h(x^2)) = x^{-1}\Re(g(x)) = \frac{-A(x)}{1 + B(x)} < 0.$$

From (43) we have that $g(z) = i\gamma$ when $z = i\eta$, where by (36), for any $\delta > 0$,

$$\gamma = \mathbb{E}\Big[\frac{\hat{D}}{\eta + \gamma\hat{D}}\Big] \ge \frac{\delta}{\eta + \gamma\delta}\nu_{\hat{D}}([\delta, \infty)).$$

Taking $\eta \downarrow 0$ followed by $\delta \downarrow 0$ we see that $\gamma(i\eta) \to \gamma(0) = 1$, provided $\nu_{\hat{D}}(\{0\}) = 0$. By definition of the Cauchy–Stieltjes transform and bounded convergence, we have then

$$\mu(\{0\}) = -\lim_{\eta\downarrow 0}\Re(i\eta f(i\eta)) = 1 - [\lim_{\eta\downarrow 0}\gamma(i\eta)]^2 = 0,$$

due to (37) (and having $\Re(g(i\eta)) = 0$). Finally, from (36) and the LHS of (37) we have that $f(z) = -\mathbb{E}[(z + g(z)\hat{D})^{-1}]$ throughout $\mathbb{C}_+$, hence by Cauchy–Schwarz

$$|f(z)| \le \mathbb{E}[\hat{W}^{-1}] \le \sqrt{B(z)\mathbb{E}[\hat{D}^{-2}]} \le \mathbb{E}[\hat{D}^{-2}]^{1/2}$$

is uniformly bounded when $\mathbb{E}\hat{D}^{-2}$ is finite. Up to factor π^{-1} this yields the stated uniform bound on $\boldsymbol{\rho}(x)$, namely the RHS of (38). At any $x > 0$ the latter is bounded above also by $\frac{1}{\pi x}|g(x)|^2$, with (40) thus a consequence of (39).

Proof of Corollary 2 Fixing $\alpha > \eta > 0$ we have that

$$\nu_{\hat{D}}(\{\alpha\}) = q_o\,, \quad \nu_{\hat{D}}(\{\eta\}) = 1 - q_o$$

and since $1 = \mathbb{E}\hat{D} = \alpha q_o + \eta(1-q_o)$, further $\alpha > 1 > \eta$. By Remark 4 we identify $\operatorname{supp}(\mu)$ upon examining the regions in which $\xi'(-v) > 0$ for $\mathbb{R}$-valued $v \notin \{0, \alpha^{-1}, \eta^{-1}\}$. Since $\Re(h(x)) < 0$ for $x > 0$ (see Lemma 2), for $\operatorname{supp}(\mu) \cap \mathbb{R}_+$ it suffices to consider the sign of

$$\xi'(-v) = \frac{1}{v^2} - \frac{q\alpha^2}{(1-v\alpha)^2} - \frac{(1-q)\eta^2}{(1-v\eta)^2}\,,$$

when $v \in (0,\infty) \setminus \{\alpha^{-1}, \eta^{-1}\}$ and $q := \alpha q_o$. Observe that $\xi'(-v) > 0$ for such v iff

$$\begin{aligned} P(v) &:= av^3 + bv^2 + cv + d \\ &= -2\alpha\eta(q\eta + (1-q)\alpha)v^3 + \left(q\eta^2 + 4\alpha\eta + (1-q)\alpha^2\right)v^2 - 2(\alpha+\eta)v + 1 > 0\,. \end{aligned}$$

Noting that $\lim_{v\to\infty} P(v) = -\infty$ and $\lim_{v\downarrow 0} P(v) = 1$, we infer from Remark 4 that $\operatorname{supp}(\mu)$ has holes iff $P(v)$ has three distinct positive roots. As Descartes' rule of signs is satisfied ($a, c < 0$ and $b, d > 0$), the latter occurs iff the discriminant $\mathsf{D}(P)$ is positive. Evaluating $\mathsf{D}(P)$ shows that

$$\mathsf{D}(P) = b^2c^2 - 4ac^3 - 4b^3d + 18abcd - 27a^2d^2 = 4q(1-q)(\alpha-\eta)^2\big(\alpha\phi - q\theta\big)\,,$$

where

$$\theta := (\alpha-\eta)(\alpha+\eta)^3\,, \qquad \phi := (\alpha - 2\eta)^3\,.$$

Having $q = \alpha q_o$ and $\theta > 0$ we conclude that $\mathsf{D}(P) > 0$ iff $\phi/\theta > q_o$. That is

$$\frac{\phi}{\theta} = \frac{(\alpha-2\eta)^3}{(\alpha-\eta)(\alpha+\eta)^3} > \frac{1-\eta}{\alpha-\eta} = q_o\,.$$

For $\varphi := 3\eta/(\alpha+\eta)$ and $\eta \in (0,1)$ this translates into $1 - \varphi > (1-\eta)^{1/3}$, or equivalently

$$\frac{\alpha}{\eta} + 1 = \frac{3}{\varphi} > \frac{3}{1-(1-\eta)^{1/3}}\,,$$

as stated in (12).

Acknowledgments The authors wish to thank Nick Cook, Alice Guionnet, Allan Sly and Ofer Zeitouni for many helpful discussions. A.D. was supported in part by NSF grant DMS-1613091. E.L. was supported in part by NSF grants DMS-1513403 and DMS-1812095.

References

1. Anderson, G.W., Guionnet, A., Zeitouni, O.: An Introduction to Random Matrices. Cambridge Studies in Advanced Mathematics, vol. 118. Cambridge University Press, Cambridge (2010)
2. Arizmendi E.O., Pérez-Abreu, V.: The S-transform of symmetric probability measures with unbounded supports. Proc. Amer. Math. Soc. **137**(9), 3057–3066 (2009)
3. Arratia, R., Liggett, T.M.: How likely is an i.i.d. degree sequence to be graphical? Ann. Appl. Probab. **15**(1B), 652–670 (2005)
4. Bercovici, H., Voiculescu, D.: Free convolution of measures with unbounded support. Indiana Univ. Math. J. **42**(3), 733–773 (1993)
5. Bordenave, C.: Spectrum of random graphs. In: Advanced Topics in Random Matrices. Panoramas et Synthèses, vol. 53, pp. 91–150. Société Mathématique de France, Paris (2017)
6. Bordenave, C., Lelarge, M.: Resolvent of large random graphs. Random Struct. Algorithms **37**(3), 332–352 (2010)
7. Coste, S., Salez, J.: Emergence of extended states at zero in the spectrum of sparse random graphs (2018). Preprint `arXiv:1809.07587`
8. Dumitriu, I., Pal, S.: Sparse regular random graphs: spectral density and eigenvectors. Ann. Probab. **40**(5), 2197–2235 (2012)
9. Erdős, P., Gallai, T.: Graphs with prescribed degrees of vertices. Mat. Lapok **11**, 264–274 (1960)
10. Greenhill, C., Sfragara, M.: The switch Markov chain for sampling irregular graphs and digraphs. Theoret. Comput. Sci. **719**, 1–20 (2018)
11. Hachem, W., Hardy, A., Najim, J.: Large complex correlated Wishart matrices: fluctuations and asymptotic independence at the edges. Ann. Probab. **44**(3), 2264–2348 (2016)
12. Marčenko, V.A., Pastur, L.A.: Distribution of eigenvalues for some sets of random matrices. Math. USSR-Sbornik **1**(4), 457–483 (1967)
13. Rao, N.R., Speicher, R.: Multiplication of free random variables and the S-transform: the case of vanishing mean. Electron. Comm. Probab. **12**, 248–258 (2007)
14. Salez, J.: Spectral atoms of unimodular random trees. J. Eur. Math. Soc. **22**(2), 345–363 (2020)
15. Silverstein, J.W., Bai, Z.D.: On the empirical distribution of eigenvalues of a class of large-dimensional random matrices. J. Multivariate Anal. **54**(2), 175–192 (1995)
16. Silverstein, J.W., Choi, S.-I.: Analysis of the limiting spectral distribution of large-dimensional random matrices. J. Multivariate Anal. **54**(2), 295–309 (1995)
17. Tao, T.: Topics in Random Matrix Theory. Graduate Studies in Mathematics, vol. 132. American Mathematical Society, Providence (2012)
18. Taylor, R.: Constrained switchings in graphs. In: Combinatorial Mathematics, VIII (Geelong, 1980). Lecture Notes in Mathematics, vol. 884, pp. 314–336. Springer, Berlin (1981)
19. Tran, L.V., Vu, V.H., Wang, K.: Sparse random graphs: eigenvalues and eigenvectors. Random Struct. Algorithms **42**(1), 110–134 (2013)
20. Voiculescu, D.: Multiplication of certain noncommuting random variables. J. Operator Theory **18**(2), 223–235 (1987)

Upper Bounds on the Percolation Correlation Length

Hugo Duminil-Copin, Gady Kozma, and Vincent Tassion

In memory of our dear friend Vladas Sidoravicius 1963–2019

Abstract We study the size of the near-critical window for Bernoulli percolation on $\mathbb{Z}^d$. More precisely, we use a quantitative Grimmett–Marstrand theorem to prove that the correlation length, both below and above criticality, is bounded from above by $\exp(C/|p - p_c|^2)$. Improving on this bound would be a further step towards the conjecture that there is no infinite cluster at criticality on $\mathbb{Z}^d$ for every $d \geq 2$.

Keywords Percolation · Correlation length

1 Introduction

1.1 Critical Percolation

The main open question in percolation theory is to understand the behaviour at criticality, i.e. when p is equal to p_c, and in particular to prove that there does not exist an infinite cluster at p_c (precise definitions will be given below, in Sect. 2).

Conjecture 1 For every $d \geq 2$, $\mathbb{P}_{p_c}[0 \leftrightarrow \infty] = 0$.

H. Duminil-Copin (✉)
Unige, Genève, Switzerland

I.H.E.S, Bures-Sur-Yvette, France
e-mail: hugo.duminil@unige.ch

G. Kozma
Weizmann Institute of Science, Rehovot, Israel
e-mail: gady.kozma@weizmann.ac.il

V. Tassion
ETH, Zurich, Switzerland
e-mail: vincent.tassion@math.ethz.ch

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_16

This conjecture has been solved for $d = 2$ [25, 30] and for $d \geq 11$ [21] based on ideas pioneered in [12, 26]. An important result related to the techniques of our paper is the fact that there is no percolation on a half space [8]. Further, the result is also known for graphs of the form $\mathbb{Z}^2 \times G$ with G finite; see [17, 19]. On transitive graphs with rapid growth, additional tools are available, and the following cases are known: non-amenable graphs [9], graphs with exponential growth [28], and recently some graphs with stretched-exponential growth [27].

A natural scheme to attack the conjecture on $\mathbb{Z}^d$ is to find a $\delta > 0$ and a sequence of events $\mathscr{E}_n$ depending on edges in the box $\Lambda_n := \{-n, \dots, n\}^d$ only, such that *for any p*,

$$\exists n > 0 \text{ s.t. } \mathbb{P}_p[\mathscr{E}_n] > 1 - \delta \quad \Longleftrightarrow \quad \mathbb{P}_p[0 \leftrightarrow \infty] > 0. \qquad (\star)$$

If such a sequence exists, the set of p such that $\mathbb{P}_p[0 \leftrightarrow \infty] > 0$ is an open set since it is the union of the open sets (indexed by n) $\{p : \mathbb{P}_p[\mathscr{E}_n] > 1 - \delta\}$ (this set is open since $p \mapsto \mathbb{P}_p[\mathscr{E}_n]$ is continuous).

Of course, this strategy is tempting, but the main difficulty is that the $\Longrightarrow$ and the $\Longleftarrow$ implications involved in ($\star$) are difficult to prove simultaneously. One may for instance easily check the $\Longrightarrow$ implication by asking a lot on $\mathscr{E}_n$, but then the $\Longleftarrow$ one becomes difficult, and vice-versa. To illustrate this trade-off phenomenon, let us give a few examples of possible sequences $(\mathscr{E}_n)$, going from the strongest criterion (meaning the one for which the $\Longrightarrow$ implication is the easiest to prove) to the weakest one (meaning the one for which $\Longrightarrow$ is the hardest).

Example 1 Let $\mathscr{E}_n$ be the event that $\Lambda_{n/10}$ is connected to $\partial\Lambda_n := \Lambda_n \setminus \Lambda_{n-1}$ and that the second largest cluster in Λ_n has radius smaller than $n/10$. In this case, a coarse-graining argument similar to [7] implies the $\Longrightarrow$ implication easily. Proving $\Longleftarrow$ is still open in particular because of the difficulty to exclude the existence of many large clusters avoiding each other.

Example 2 Let $\mathscr{E}_n$ be the intersection of the events that $(\pm n, 0) + \Lambda_{n/2}$ are connected in Λ_{2n} and that there exists at most one cluster in Λ_{2n} going from $(\pm n, 0) + \Lambda_{n/2}$ to $(\pm n, 0) + \partial\Lambda_n$. A coarse-graining argument may be used to prove $\Longrightarrow$ but $\Longleftarrow$ remains open due to the same reason as the previous condition.

In general, uniqueness of clusters going from one area to another one is a key difficulty in these problems. This might be related to the fact that in high dimensions Λ_n indeed hosts many disjoint clusters at p_c, see [2]. In order to circumvent this difficulty, one can make different choices for $\mathscr{E}_n$.

Example 3 Let $\mathscr{E}_n$ be the same event as in the second example, but with $(\pm n, 0) + \Lambda_{n/2}$ replaced by $(\pm n, 0) + \Lambda_{u_n}$, with u_n much smaller than $n/2$. In this case, the implication $\Longrightarrow$ is as before, and does not depend on u_n. As for the implication $\Longleftarrow$, as u_n becomes smaller the connectivity part becomes harder and the uniqueness part becomes easier.

Recently, a paper of Cerf [16], based on [6], provided a beautiful insight on how big u_n must be taken to have that with large probability, $\Lambda_n := \{-n, \dots, n\}^d$, the box of size n, contains at most one cluster going from Λ_{u_n} to $\partial\Lambda_n$. We will come back to this later in the introduction, but let us mention the result right now.

Given $1 \le m \le n$, consider the set of clusters in the configuration restricted to the box Λ_n, and define $A_2(m, n)$ to be the event that there exists at least two disjoint such clusters intersecting both Λ_m and $\partial\Lambda_n$.

Proposition 1 (Cerf) *Let $d \ge 2$. There exists $\alpha = \alpha(d) \in (0, 1)$ such that for any $p \in [0, 1]$ and n large enough,*

$$\mathbb{P}_p\left[A_2(n^\alpha, n)\right] \le \frac{1}{n^\alpha}.$$

We fill some details on this proposition in Sect. 7. Let us finish by a last example, which is very simple but interesting for the discussion that follows.

Example 4 Let $\mathscr{E}_n$ be the event that the box Λ_N is connected to $\{n\} \times \{-n, \ldots, n\}^{d-1}$, with $N = N(\delta) > 0$ independent of n. Here, $\Longleftarrow$ follows easily from the ergodicity of $\mathbb{P}_p$ and FKG but again the $\Longrightarrow$ implication seems difficult to obtain.

The search for a good sequence of events $\mathscr{E}_n$ has been at the heart of attempts to prove the conjecture. An important development was made in [22]. In this paper, the authors considered the sequence of events $\mathscr{E}_n$ defined in the fourth example. As mentioned above, the $\Longrightarrow$ seems extremely difficult to derive. Nevertheless, Grimmett and Marstrand introduced a clever renormalisation scheme allowing to prove the following weaker version of the implication: for any $\varepsilon > 0$, there exists $\delta > 0$ such that for every n,

$$\mathbb{P}_p[\mathscr{E}_n] > 1 - \delta \quad \Longrightarrow \quad \mathbb{P}_{p+\varepsilon}[0 \stackrel{\mathrm{Slab}_n^d}{\longleftrightarrow} \infty] > 0$$

where $\mathrm{Slab}_n^d := \mathbb{Z}^2 \times \{-n, \ldots, n\}^{d-2}$. In words, the implication can be proved if one allows some sprinkling. As suggested in [22], if one could get rid of the sprinkling by ε in the previous statement, then the conjecture would follow.

The goal of this paper is to prove a quantitative version of the Grimmett–Marstrand argument by bounding the critical point of Slab_n^d in terms of n. In the language of the Grimmett–Marstrand theorem, we will be interested in how small ε can be taken as a function of n. We believe that improving how small ε can be taken is a good intermediate problem for the conjecture. Getting bounds is non-trivial and requires some understanding of the critical phase. As a consequence, each improvement on the existing bounds should shed a new light on the critical behaviour.

There is a quantity which is intimately related to $p_c(\mathrm{Slab}_n^d)$, called the *correlation length*, which appears repeatedly in physics. In order to have a statement which is independent of the Grimmett–Marstrand theorem, we choose to first state our main result in terms of the correlation length.

1.2 An Upper Bound on the Correlation Length

For $p < p_c$, the probability $\mathbb{P}_p[0 \leftrightarrow \partial\Lambda_n]$ decays exponentially fast in n (see [3, 18, 35]). The rate at which this happens is known as the *correlation length* ξ_p, namely

$$\xi_p := \overline{\lim_{n\to\infty}} -\frac{n}{\log \mathbb{P}_p[0 \leftrightarrow \partial\Lambda_n]}.$$

For $p > p_c$, the correlation length is also defined, but the formula is slightly modified:

$$\xi_p := \overline{\lim_{n\to\infty}} -\frac{n}{\log \mathbb{P}_p[0 \leftrightarrow \partial\Lambda_n, 0 \not\leftrightarrow \infty]}.$$

Again, the probability decays exponentially fast so ξ_p is finite. This is due to the following. Grimmett and Marstrand [22] showed that for any $p > p_c$, there exists $n \geq 1$ such that

$$\mathbb{P}_p\big[0 \xleftrightarrow{\text{Slab}_n^d} \infty\big] > 0.$$

And the exponential decay follows from that by the results of [14]. Let us mention that in fact, both limits exist. We could not find a reference for this fact, but it follows using standard methods, see e.g. [23, Section 6.2] for the subcritical case and [15] for the supercritical case (both prove the existence of the limit with a different definition of the correlation length, but the proofs work also with our definition and the values are equal).

Our main result is the following.

Theorem 1 *Let $d \geq 3$. There exists $C = C(d) > 0$ such that for any $p \neq p_c$,*

$$\xi_p \leq \exp(C|p - p_c|^{-2}).$$

The results below and above p_c are different in nature (even though the same proof gives both), a point which will become clearer when we discuss the proof in the next section. In particular, the use of [22] to connect slabs and the correlation length mentioned above is used only for $p > p_c$.

Our bound on ξ_p is far from the truth. Conjecturally, one has $\xi_p = |p - p_c|^{-\nu+o(1)}$, where $o(1)$ tends to 0 as $p \to p_c$ ($p \neq p_c$) and ν is given by

$$\nu = \begin{cases} \frac{4}{3} & \text{if } d = 2, \\ 0.87\ldots & \text{if } d = 3, \\ 0.69\ldots & \text{if } d = 4, \\ 0.56\ldots & \text{if } d = 5, \\ \frac{1}{2} & \text{if } d \geq 6. \end{cases}$$

Some physics references are [1, 33, 43]. The predictions for $d = 3, 4, 5$ are numerical, while the prediction for $d = 2$ is based on conformal field theory, quantum gravity or Coulomb gas formalism, and the prediction for $d \geq 6$ on the fact that the model should have a mean-field behaviour. For site percolation on the triangular lattice, $\xi_p = |p - p_c|^{-4/3+o(1)}$ was proved in [41] using the conformal invariance of the model proved in [40], the theory of Schramm–Löwner evolution and scaling relations obtained by Kesten in [32] (such scaling relations were proved under the hyper-scaling hypothesis [10] which is expected to be valid for $d \leq 5$). In fact, Russo–Seymour–Welsh theory [37, 39] combined with [32] imply that there exists $C > 0$ such that $\xi_p \leq |p - p_c|^{-C}$ for Bernoulli bond percolation on $\mathbb{Z}^2$. For $d \geq 19$, $\xi_p = |p - p_c|^{-1/2+o(1)}$ was proved in [24, 26] for $p < p_c$. Let us remark that *lower* polynomial bounds may be achieved. We could not found a proof in the literature for this fact, so we include a proof sketch in Sect. 8.

1.3 A Quantitative Grimmett–Marstrand Theorem

The theory of static renormalisation, developed throughout the eighties [5, 8, 14, 15, 22], allows to relate the correlation length, percolation in slabs, and various events of the type discussed in Sect. 1.1. It is a deep theory and we will not attempt to survey it here. But, as already explained, it motivates us to state a version of our main theorem in terms of slabs.

Definition 1 Throughout the paper we denote by p_n the smallest $p < p_c$ such that $\xi_p = n$.

Theorem 2 *Fix $d \geq 3$. There exists a constant $C = C(d) > 0$ such that for every $n \geq 3$,*

$$\mathbb{P}_{p_n + \frac{C}{\sqrt{\log n}}}[0 \xleftrightarrow{\mathrm{Slab}_n^d} \infty] \geq \frac{1}{2\sqrt{\log n}},$$

In particular, we have that

$$p_c(\mathrm{Slab}_n^d) < p_c + \frac{C}{\sqrt{\log n}}.$$

Let us explain the main elements in the proof of Theorems 1 and 2 (they share 95 percent of the proof). The proof is composed of the following 3 steps, each of which requires to increase the probability somewhat.

Step 1. The result of Chayes and Chayes [13] stating that $\mathbb{P}_{p_c+\varepsilon}[0 \leftrightarrow \infty] \geq \varepsilon$ ([3] was the first unconditioned proof). We take the opportunity to give a new proof of this inequality, based on the ideas of [18].

Step 2. The result of Kahn, Kalai and Linial ([29], see also [11, 38, 42]) that any boolean function with small individual influences has at least logarithmic total influence. We apply this to the function $\mathbb{1}\{\Lambda_n \leftrightarrow \infty\}$ for some n. To show that all individual influences are small (as $n \to \infty$) we use a geometric argument connecting the probability that a certain edge is pivotal to the same probability for nearby edges. Thus from the information that $\mathbb{P}_{p_c+\varepsilon}[0 \leftrightarrow \infty] \geq \varepsilon$ we can get $\mathbb{P}_{p_c+2\varepsilon}[\Lambda_n \leftrightarrow \infty] \geq 1 - \delta$ (with appropriate connections between the parameters ε, n and δ).

Step 3. A "seedless" renormalisation scheme, based on ideas of [34]. In Grimmett–Marstrand the renormalisation follows by finding seeds, i.e. small boxes (say of size n) all whose edges are open, which are on the boundary of a much larger box, say of size N [22]. A first version of our argument which used the same scheme gave $p_c(\text{Slab}_n) < p_c + C/\log\log\log\log n$. Here the path that already exists inside the N-box is used in place of the seeds, each piece of it, if sufficiently separated, can be used independently. Proposition 1 plays a crucial role in the argument. Sprinkling is used as in [22], so eventually we get a renormalisation scheme at $p_c + 3\varepsilon$.

The value $\varepsilon = 1/\sqrt{\log n}$ comes from the interaction of Steps 2 and 3. In Step 3 we do an ε-sprinkling and the proof requires connections happening with probability at least $1 - \exp(1/\varepsilon)$. This forces the δ of Step 2 to be smaller than $\exp(-1/\varepsilon)$. But this forces n to be $\exp(1/\varepsilon^2)$ since our estimate of the total influence is only logarithmic in n. Thus, the use of [29] is the main constraining factor.

Finally, let us remark on the subcritical case in Theorem 1, i.e. on the bound of ξ_p for $p < p_c$. It is a corollary from the supercritical result. There are various ways to perform this conclusion, but here the simplest was simply not to start all the process (i.e. Steps 1–3) from p_c but rather from an appropriate $p < p_c$ where ξ_p is sufficiently large. Thus we prove both results in one fell swoop. Let us stress again, though, that it is the supercritical result which is central and the subcritical result is merely a corollary.

In Sects. 3–5 we detail these steps, one step per section. In the last sections we prove Theorems 1 and 2, as well as Proposition 1.

2 Preliminaries

Fix an integer $d \geq 2$. Two vertices x and y of $\mathbb{Z}^d$ are said to be *neighbours* (denoted $x \sim y$) if $\|x - y\|_2 = 1$. In such a case, $\{x, y\}$ is called an *edge* of $\mathbb{Z}^d$. The set of edges is denoted by $E(\mathbb{Z}^d)$. For $n \geq 1$, introduce the box $\Lambda_n := \{-n, \ldots, n\}^d$ and its (vertex) boundary $\partial\Lambda_n := \Lambda_n \setminus \Lambda_{n-1}$. Also, we define $\text{Slab}_n^d := \mathbb{Z}^2 \times \{-n, \ldots, n\}^{d-2}$.

A percolation configuration $\omega = (\omega(e) : e \in E(\mathbb{Z}^d))$ is an element of $\{0, 1\}^{E(\mathbb{Z}^d)}$. If $\omega(e) = 1$, the edge e is said to be *open*, otherwise it is said to be *closed*. Let $S \subset \mathbb{Z}^d$. Two vertices x and y are said to be *connected in* S (in ω) if there exists a path $x = v_0 \sim v_1 \sim v_2 \sim \cdots \sim v_k = y$ of vertices in S such that $\omega(\{v_i, v_{i+1}\}) = 1$ for every $0 \leq i < k$. Let A and B be two subsets of S, we write

$A \overset{S}{\leftrightarrow} B$ if some vertex of A is connected in S to some vertex of B, and $A \overset{S}{\leftrightarrow} \infty$ if $A \overset{S}{\leftrightarrow} \partial\Lambda_n$ holds for any $n \geq 1$. If $S = \mathbb{Z}^d$, we drop it from the notation and simply write $A \leftrightarrow B$ and $A \leftrightarrow \infty$. A *cluster* is a maximal set of vertices that are connected together in ω.

For $p \in [0, 1]$, consider the Bernoulli bond percolation measure $\mathbb{P}_p$ on $\{0, 1\}^{E(\mathbb{Z}^d)}$ under which the variables $\omega(e)$ with $e \in E(\mathbb{Z}^d)$ are i.i.d. Bernoulli random variables with parameter p. Define $p_c = p_c(d) \in (0, 1)$ such that $\mathbb{P}_p[0 \leftrightarrow \infty]$ is 0 when $p < p_c$ and strictly positive when $p > p_c$. See [23, Theorem 1.10] for the fact that indeed $0 < p_c < 1$.

An event A is *increasing* if it is stable to opening edges. The FKG inequality states that increasing events are positively correlated, see [23, Theorem 2.2]. The edge e is *pivotal for the event A* if the configurations ω^e and ω_e defined by

$$\omega^e(f) = \begin{cases} \omega(f) & \text{if } f \neq e \\ 1 & \text{if } f = e, \end{cases} \qquad \text{and } \omega_e(f) = \begin{cases} \omega(f) & \text{if } f \neq e \\ 0 & \text{if } f = e \end{cases}$$

satisfy $\omega^e \in A$ and $\omega_e \notin A$.

We will denote by c and C arbitrary constants which depend only on the dimension d (and occasionally other parameters, which will be noted). Their value may change from formula to formula, and even inside the same formula. Occasionally we will number them for clarity. We will use c for constants which are sufficiently small and C for constants which are sufficiently large.

3 The Result of Chayes and Chayes

In this section, we prove the following (recall that p_n is the smallest $p < p_c$ such that $\xi_p = n$).

Proposition 2 *For every n large enough,*

$$\mathbb{P}_{p_n + \frac{1}{\sqrt{\log n}}}[0 \leftrightarrow \partial\Lambda_n] \geq \frac{1}{\sqrt{\log n}}. \tag{1}$$

(For the purpose of the supercritical result it would have been enough to know this at $p_c + 1/\sqrt{\log n}$, which is exactly the original result of Chayes and Chayes, but for the subcritical result we want to know it with p_c replaced by p_n.)

Proof Given a finite set S containing 0, and a parameter $p \in [0, 1]$, define

$$\varphi_p(S) := \sum_{\substack{x \sim y \\ x \in S,\, y \notin S}} p\, \mathbb{P}_p[0 \overset{S}{\leftrightarrow} x].$$

Fix $n \geq 1$. Let us recall two relations between this quantity and the one-arm probability, established in [18]. First, for every $S \subset \Lambda_n$ containing 0, the last displayed equation of Section 2.1 of [18] gives the upper bound

$$\mathbb{P}_p[0 \leftrightarrow \partial\Lambda_{nk}] \leq \varphi_p(S)^{k-1}. \tag{2}$$

for every $k \geq 1$. Also, the quantity $\varphi_p(S)$ can be used to bound the derivative of the one-arm probability. Lemma 2.1 of [18] states that for every $p \in [0, 1]$,

$$\frac{d}{dp}\mathbb{P}_p[0 \leftrightarrow \partial\Lambda_n] \geq \frac{1}{p(1-p)} \cdot \Big[\inf_{0 \in S \subset \Lambda_n} \varphi_p(S)\Big] \cdot (1 - \mathbb{P}_p[0 \leftrightarrow \partial\Lambda_n]). \tag{3}$$

The proof of Proposition 2 can be easily derived from the two equations above. If for some $p \in [0, 1]$, there exists a subset S of Λ_n with $\varphi_p(S) < \frac{1}{e}$, then one deduces immediately from (2) that

$$\mathbb{P}_p[0 \leftrightarrow \partial\Lambda_k] \leq e^{-A\lfloor k/n\rfloor - 1}, \qquad A > 1,$$

which implies that $\xi_p < n$. As a consequence, $\varphi_{p_n}(S) \geq \frac{1}{e}$ for any set S included in Λ_n containing 0. Since $\varphi_p(S)$ is increasing in p, we have $\varphi_p(S) \geq \frac{1}{e}$ for any $p \geq p_n$, and the differential inequality (3) gives that for every $p \geq p_n$,

$$\frac{d}{dp}\mathbb{P}_p[0 \leftrightarrow \partial\Lambda_n] \geq \tfrac{4}{e}(1 - \mathbb{P}_p[0 \leftrightarrow \partial\Lambda_n]). \tag{4}$$

Now, set $p_n' := p_n + 1/\sqrt{\log n}$. Either $\mathbb{P}_{p_n'}[0 \leftrightarrow \partial\Lambda_n] > 1 - \frac{e}{4}$, or integrating (4) between p_n and p_n' gives (1). This concludes the proof.

4 Sharp Threshold

In this section we prove the following result.

Proposition 3 *For every* $0 < \beta < 1$*, there exists* $C = C(\beta, d) > 0$ *such that for every* n

$$\mathbb{P}_{p_n + C/\sqrt{\log n}}[\Lambda_{n^\beta} \leftrightarrow \partial\Lambda_n] \geq 1 - e^{-\sqrt{\log n}}.$$

We may assume without loss of generality that $\beta < \alpha$, where α is chosen such that the statement of Proposition 1 holds. Further, we may assume n to be sufficiently large, as for small n and large C we would have $p_n + C/\sqrt{\log n} > 1$, making the claim trivial. Set $m := \lfloor n^\beta \rfloor$ for brevity. The proposition will follow, using standard arguments, once we prove the following lemma.

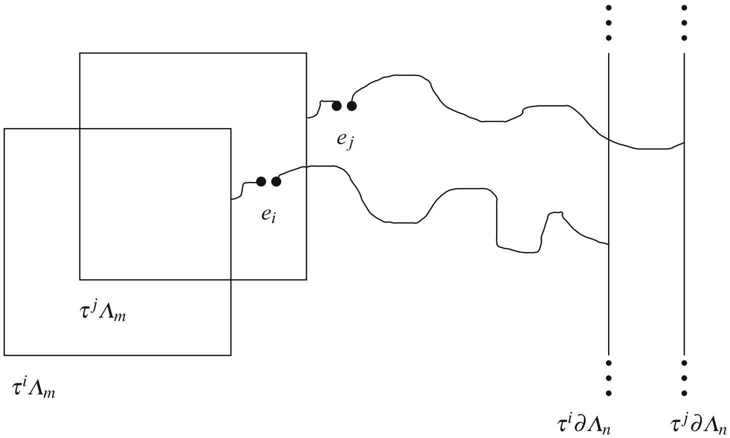


Fig. 1 Two edges e_i, e_j for $i < j$, and the corresponding translated boxes

Lemma 1 *With n, α, β and m as above, and for any p,*

$$\mathbb{P}_p[e \text{ is a closed pivotal for } \Lambda_m \leftrightarrow \partial \Lambda_n] \leq \frac{1}{m^{\alpha/4}}. \tag{5}$$

Proof Fix an edge $e \in E$ and distinguish between two cases, depending on whether the edge e is close to $\partial \Lambda_m \cup \partial \Lambda_n$ or not. Write ρ for the L^∞-distance between the edge e and $\partial \Lambda_m \cup \partial \Lambda_n$.

If $\rho \geq m^{1/4}$, then observe that a translated version of the event $A_2(m^{\alpha/4}, m^{1/4})$ must occur around the edge e when the edge is a closed pivotal. Therefore, Proposition 1 implies that (5) holds.

The more difficult case is when $\rho \leq m^{1/4}$. Let us first assume that e is at a distance smaller than $m^{1/4}$ of Λ_m. Then there exists a translation τ by a vector in $\Lambda_{m^{1/4}}$ such that e belongs to the translate $\tau \Lambda_m$ of Λ_m by τ, and further, such that $e \in \tau^i \Lambda_m$ (where τ^i denotes the i-th iterate of τ) for every $i \in \{1, \ldots, I\}$, where $I := \lfloor \frac{1}{2} m^{3/4} \rfloor$. For $0 \leq i < I$ define the edges $e_i = \tau^i e$. It follows that for every $i < j$, both endpoints of the edge e_i belong to $\tau^j \Lambda_m$, see Fig. 1. Define also the event

$$B_i := \{e_i \text{ is a closed pivotal for } \tau^i \Lambda_m \leftrightarrow \tau^i \partial \Lambda_n\}.$$

Writing M for the number of indices i for which B_i occurs, translation invariance and the Cauchy–Schwarz inequality imply

$$(I \cdot \mathbb{P}_p[B_0])^2 = \mathbb{E}_p[M]^2 \leq \mathbb{E}_p[M^2] = \mathbb{E}_p[M] + 2 \sum_{i<j} \mathbb{P}_p[B_i \cap B_j]. \tag{6}$$

Let us bound probabilities on the right-hand side. Fix $i < j$ and assume that $B_i \cap B_j$ occurs. Then, we claim that there must exist two disjoint clusters in $\Lambda_{n/2}$ crossing the annulus between Λ_{2m} and $\Lambda_{n/2}$. Indeed, one extremity x_i of e_i must be connected to the boundary of $\tau^i \Lambda_n$, and one extremity x_j of e_j must be connected to the boundary of $\tau^j \Lambda_n$. The fact that e_j is a *closed* pivotal implies in particular that $\tau^j \Lambda_m \nleftrightarrow \tau^j \partial \Lambda_n$ and hence, since x_i belongs to $\tau^j \Lambda_m$, it is not connected to the boundary of $\tau^j \Lambda_n$ so that the clusters of x_i and x_j in the box $\Lambda_{n/2}$ must be disjoint (see again Fig. 1). For n large enough, we have $2m \leq (n/2)^\alpha$ and Proposition 1 implies that

$$\mathbb{P}_p[B_i \cap B_j] \leq \frac{1}{2m}.$$

Plugging this estimate in (6) and using the trivial bound $M \leq I$, we obtain

$$\mathbb{P}_p[B_0]^2 \leq \frac{1}{I} + \frac{1}{m} \leq \frac{1}{\sqrt{m}},$$

provided n is large enough. This completes the proof in this case.

The exact same reasoning also works if one assumes that the edge e is within distance $m^{1/4}$ of the boundary of Λ_n. Consider a translation τ by a vector in $\Lambda_{m^{1/4}}$ such that e does *not* belong to $\tau \Lambda_n$. One can define the edges e_i and the events B_i as above. In this case, for $i < j$, the edge e_i does not belong to $\tau^j \Lambda_n$ and the same reasoning as above concludes the proof.

Proof of Proposition 3 We use the following standard sharp threshold result for Boolean functions (see e.g. [42, Corollary 1.2]): for any $\delta > 0$, there exists a constant $c' = c'(\delta) > 0$ such that for any increasing event A depending on a finite set E of edges, and any $p \in [\delta, 1-\delta]$,

$$\frac{d}{dp}\mathbb{P}_p[A] \geq c' \log\left(\frac{1}{\max\{\mathbb{P}_p\left[e \text{ pivotal for } A\right] : e \in E\}}\right) \cdot \mathbb{P}_p[A](1 - \mathbb{P}_p[A]). \tag{7}$$

We apply (7) to the event $A = \{\Lambda_m \leftrightarrow \partial \Lambda_n\}$, bounding the pivotality probability inside the log using Lemma 1. Note that we use here the fact that pivotality is independent of the status of the edge, hence the probability of being *closed* pivotal (which is what we get from Lemma 1) is $1-p$ times the probability of being pivotal, as needed in (7). We get that for any $\delta > 0$, there exists $c = c(\delta, \beta, d) > 0$ such that for every $p \in [\delta, 1-\delta]$ and every n large enough,

$$\frac{f'(p)}{f(p)(1-f(p))} \geq c \log n, \quad \text{where } f(p) = \mathbb{P}_p\left[\Lambda_{n^\beta} \overset{\partial}{\longleftrightarrow} \Lambda_n\right]. \tag{8}$$

Set $p_n' := p_n + \frac{1}{\sqrt{\log n}}$. By Proposition 2,

$$\mathbb{P}_{p_n'}\left[\Lambda_{n^\beta} \leftrightarrow \partial\Lambda_n\right] \geq \mathbb{P}_{p_n'}\left[0 \leftrightarrow \partial\Lambda_n\right] \geq \frac{1}{\sqrt{\log n}}.$$

We now integrate the differential inequality (8). Define p_n'' by $f(p_n'') = \frac{1}{2}$ and then throughout the interval $[p_n', p_n'']$ we can remove the factor $1 - f(p)$ from the denominator, and pay only by halving the constant on the right hand side of (8). This gives $\log(f)' \geq c \log n$ which we integrate and get that p_n'' must be no more than $p_n' + C \log\log n / \log n$. Similarly, in the interval $[p_n'', p_n'' + C/\sqrt{\log n}]$ we remove the factor $f(p)$ from the denominator, get $-\log(1 - f)' \geq c \log n$ and arrive at the conclusion that $f(p_n'' + C/\sqrt{\log n}) \geq 1 - \exp(-\sqrt{\log n})$, if $C = C(\beta, \delta)$ is sufficiently large. This is exactly the conclusion of the proposition (with a larger C to compensate for replacing p_n'' with p_n).

Remark 1 Proposition 3 can be directly obtained using Section 3 of [20] with the definition of the event A_k being, for $0 \leq k \leq \frac{1}{2}n^\beta$,

$$A_k = A_k(n) := \{\Lambda_{n^\beta/2+k} \longleftrightarrow \partial\Lambda_n\}.$$

Roughly speaking, since all the events A_k have a probability larger than $1/\sqrt{\log n}$ at p_n, the argument in [20] implies that at every $p \geq p_n$, one of the events A_k has a logarithmic derivative larger than $c \log n$ for some small constant c. A careful manipulation enables to prove that one of the events A_k (and therefore $A_{n^\beta/2}$) must have probability larger than $1 - e^{-\sqrt{\log n}}$ at $p_n + C/\sqrt{\log n}$. We believe that the present solution is simpler in the case of Bernoulli percolation and may have further applications, even though the other alternative does not use the Aizenman–Kesten–Newman estimate on the probability of the two-arm event.

5 The Seedless Renormalisation Scheme

The normalisation scheme we will work with uses four different scales, which we will denote by $k < K < n < N$. The most important is the scale between K and n, where we will insert $1/\varepsilon^2$ boxes of size K and use the independence between these boxes to get to an event with high probability. The scales between k and K; and between n and N will be used for gluing paths using Proposition 1 (the second one, between n and N, is used only for resolving a technical issue of connecting to a specific facet and is less important). Here is the exact formulation which we will use.

Theorem 3 *Fix $d \geq 3$. There exists a constant $C = C(d) > 0$ such that the following holds. Assume that for some $p \in [0, 1]$ and some $\varepsilon > 0$, there exist $1 \leq k \leq K \leq n \leq N < \infty$ such that $K \leq \varepsilon^2 n$ and*

(a) $\mathbb{P}_p[0 \leftrightarrow \partial\Lambda_N] \geq \varepsilon$,
(b) $\mathbb{P}_p[\Lambda_k \leftrightarrow \partial\Lambda_N] \geq 1 - \exp(-\frac{1}{\varepsilon})$,
(c) $\mathbb{P}_p[A_2(k, K)] \leq \exp(-\frac{1}{\varepsilon})$ *and* $\mathbb{P}_p[A_2(n, N)] \leq \exp(-\frac{1}{\varepsilon})$.

Then

$$\mathbb{P}_{p+C\varepsilon}[0 \xleftrightarrow{\mathrm{Slab}^d_{2N}} \infty] \geq \tfrac{\varepsilon}{2}.$$

Again, the reader who is interested only in the supercritical case may mentally replace "some $p \in [0, 1]$" with "$p_c + 2\varepsilon$" as in the proof sketch in the introduction. But for the subcritical result we will apply it at $p + 2\varepsilon$ for a slightly subcritical p, and we do not know, eventually, if $p + 2\varepsilon$ is sub- or supercritical.

The proof is divided into two parts. In the first one, we prove the following intermediate statement.

Lemma 2 *Assume that conditions (a), (b) and (c) hold. Then, there exists some $c > 0$ (depending on d only) such that for every connected set $S \ni 0$ with a diameter larger than n,*

$$\mathbb{P}_p\left[S \overset{\Lambda_N}{\longleftrightarrow} F(N)\right] \geq 1 - 2\exp[-c/\varepsilon],$$

where $F(N) := \{(x_1, \ldots, x_d) \in \partial\Lambda_N \ : \ x_1 = N, x_2 \geq 0, \ldots, x_d \geq 0\}$.

Here and below we call sets such as $F(N)$ "quarter-faces" even though this name is correct only in $d = 3$.

Remark 2 The introduction of quarter-faces is a purely technical step and should not worry the reader. Indeed, the probability of connecting to a quarter-face is easily compared to the probability of connecting to the boundary of the box. To this end, divide $\partial\Lambda_N$ into $d2^d$ quarter-faces $F_1, \ldots, F_{d2^d}$. Using the Harris-FKG inequality (sometimes called "the square root trick" when used in this way, see [23, equation (11.14)]) together with (b), we find that

$$\mathbb{P}_p[\Lambda_k \overset{\Lambda_N}{\longleftrightarrow} F(N)] \geq 1 - \exp[-1/(\varepsilon d 2^d)]. \tag{9}$$

The conclusion of Lemma 2 can be understood as a strengthening of the condition (b) where the box Λ_k is replaced by arbitrary sufficiently large sets, and the boundary of Λ_N is replaced by one of its quarter-faces. Using it, we will be able to construct an infinite cluster in Slab^d_{2N} by propagating it using local connections. Heuristically, if the cluster of the origin is connected to a large box Λ away from 0, then it must contain a large set, which is sufficient to propagate this cluster to other boxes neighbouring Λ. The condition on connectedness of arbitrary large sets was introduced in the work of Martineau and Tassion [34], where it was established using abstract measurability arguments. The main contribution here is to make it quantitative.

The proof of Theorem 3 is now organised as follows. In Sect. 5.1, we prove Lemma 2. The proof of the main theorem is then concluded in Sect. 5.2.

5.1 Connections to Arbitrary Sets

In this section we prove Lemma 2. Without loss of generality we may assume ε is sufficiently small (as otherwise by choosing c sufficiently small the claim is trivially true). Below, the constants c_i depend on d only.

Let $p \in [0, 1]$, $\varepsilon > 0$ and $k \le K \le n \le N$ be such that $K \le \varepsilon^2 n$ and the three conditions **(a)**, **(b)** and **(c)** hold. Fix a connected set S containing 0 with a diameter at least n. Without loss of generality, we may assume $S \subset \Lambda_n$.

Consider a family of points $x_1, \dots, x_\ell \in S$ such that the boxes $Q''_i := x_i + \Lambda_K$ are all disjoint and included in Λ_n. Also, introduce the smaller box $Q'_i := x_i + \Lambda_k$. Note that we may choose $\ell \ge c_1/\varepsilon^2$ such points, so let us fix $\ell = \lceil c_1/\varepsilon^2 \rceil$.

For every $i \in \{1, \dots, \ell\}$, define the two events

$$\begin{aligned} E_i &:= \{x_i \longleftrightarrow \partial Q''_i\} \cap \{\exists \text{ unique cluster in } Q''_i \text{ from } Q'_i \text{ to } \partial Q''_i\}, \\ B_i &:= \{Q'_i \nleftrightarrow \partial \Lambda_N\}. \end{aligned}$$

By translation invariance and conditions **(a)** and **(c)**,

$$\begin{aligned} \mathbb{P}_p\,[E_i] &\ge \mathbb{P}_p\left[x_i \longleftrightarrow \partial Q''_i\right] - \mathbb{P}_p\,[A_2(k, K)] \\ &\ge \varepsilon - \exp(-1/\varepsilon) \ge \varepsilon/2. \end{aligned}$$

Since the boxes Q''_i are disjoint, the events E_i are independent, and hence

$$\mathbb{P}_p\,[\cup E_i] \ge 1 - 2e^{-c_2\varepsilon\ell} \ge 1 - 2e^{-c_3/\varepsilon}.$$

where the second inequality is from our requirement that $\ell \ge c_1/\varepsilon^2$.

Now, **(b)** implies that for every i,

$$\mathbb{P}_p\,[B_i] \le \exp[-1/(d2^d\varepsilon)].$$

Indeed, find a quarter-face F of $x_i + \Lambda_N$ outside Λ_{N-1} and then apply (9) (with 0 shifted to x_i) to get

$$\mathbb{P}_p\,[B_i] \le \mathbb{P}_p\left[Q'_i \nleftrightarrow F\right] \overset{(9)}{\le} \exp[-1/(d2^d\varepsilon)].$$

By a union bound we have

$$\mathbb{P}_p[\cup B_i] \le \ell \exp[-1/(d2^d \varepsilon)] \le \exp(-c/\varepsilon)$$

where the last inequality is by our assumption that $\ell \le c_1/\varepsilon^2 + 1$ and that ε is sufficiently small.

Assume $E_i \setminus B_i$ occurred for some i. Then we know that $x_i \leftrightarrow \partial Q_i''$ (from the first part of E_i), that $Q_i' \leftrightarrow \partial \Lambda_N$ (from the negation of B_i) and that the two clusters performing these two connections are the same (from the second part of E_i). We get that there is a path from x_i to $\partial \Lambda_N$, and in particular from S to $\partial \Lambda_N$. This gives

$$\begin{aligned}\mathbb{P}_p[S \leftrightarrow \partial \Lambda_N] &\ge \mathbb{P}_p[\exists i \text{ s.t. } E_i \setminus B_i] \\ &\ge \mathbb{P}_p[\cup E_i] - \mathbb{P}_p[\cup B_i] \\ &\ge 1 - 2e^{-c_4/\varepsilon}.\end{aligned}$$

It remains to replace the boundary of $\partial \Lambda_N$ in the equation above by the quarter-face $F(N)$. Yet, because we assumed $S \subset \Lambda_n$,

$$\mathbb{P}_p[S \overset{\Lambda_N}{\longleftrightarrow} F(N)] \ge \mathbb{P}_p[\{\Lambda_n \leftrightarrow F(N)\} \cap \{S \leftrightarrow \partial \Lambda_N\} \cap A(n, N)^c] \ge 1 - Ce^{-c_5/\varepsilon}$$

thanks to **(b)** (again in the form (9)) and **(c)**. This concludes the proof. □

5.2 Renormalisation

To prove Theorem 3 we couple a growing exploration process on the slab with a growing exploration process on a rescaled version of the square lattice. One will need a simple condition for a growing exploration process on $\mathbb{Z}^2$ to contain an infinite cluster. Therefore, before moving to the proof, we describe a particular type of exploration process on $\mathbb{Z}^2$ and give a sufficient condition for the existence of an infinite connected component.

Fix an arbitrary ordering of the edges of $\mathbb{Z}^2$. Let $\{0\} = A_0 \subset A_1 \subset A_2 \ldots$ and $\emptyset = B_0 \subset B_1 \subset B_2 \ldots$ be two growing sequences of subsets of $\mathbb{Z}^2$. We say that the sequence $X_t = (A_t, B_t)$ is an *exploration sequence* if for every $t \ge 0$,

$$\begin{aligned}&X_{t+1} = X_t \text{ if there is no edge connecting } A_t \text{ to } (A_t \cup B_t)^c,\\ &X_{t+1} = (A_t \cup \{x_t\}, B_t) \text{ or } X_{t+1} = (A_t, B_t \cup \{x_t\}) \text{ otherwise,}\end{aligned}$$

where x_t is the endpoint in $(A_t \cup B_t)^c$ of the minimal edge connecting A_t to $(A_t \cup B_t)^c$ (here and below, when we write minimal we mean with respect to the chosen ordering of the edges of $\mathbb{Z}^2$). A typical example of a random exploration sequence results from the exploration of the cluster of the origin in a site percolation process on $\mathbb{Z}^2$. In this case, the set A_t corresponds to the open sites discovered after t steps of exploration and B_t is the discovered part of the (closed) boundary of the cluster.

We say that an exploration sequence *percolates* if the set $\cup_{t\geq 0} A_t$ is infinite. The following lemma, proved in [22, Lemma 1], gives a sufficient condition for a random exploration sequence to percolate.

Lemma 3 *Let p_c^{site} be the critical parameter of Bernoulli site percolation on $\mathbb{Z}^2$. Let $X_t = (A_t, B_t)$ be a random exploration sequence and assume that there exists some $q > p_c^{\text{site}}$ such that for every $t \geq 0$,*

$$\mathbb{P}\big(B_{t+1} = B_t \mid X_0, \dots, X_t\big) \geq q \text{ a.s.},$$

then the process X percolates with probability larger than a constant $c = c(q) > 0$ that can be taken arbitrarily close to 1 provided that q is close enough to 1.

We now return to the proof of the theorem. For every $x \in \mathbb{Z}^2$, set $\Lambda_x = Nx + \Lambda_N \subset \mathbb{Z}^d$ and $\widetilde{\Lambda}_x = Nx + \Lambda_{2N} \subset \mathbb{Z}^d$, where for both we identify $x = (x_1, x_2) \in \mathbb{Z}^2$ with $(x_1, x_2, 0, \dots, 0) \in \mathbb{Z}^d$. We will identify 0 with $(0, 0)$ so Λ_0 is the box of size N centered at 0 in $\mathbb{Z}^d$.

Let ω be a Bernoulli percolation of parameter p in Slab_{2N}^d and for every $x \in \mathbb{Z}^2$, let ω^x be a $\lambda\varepsilon$-percolation on $\widetilde{\Lambda}_x$, where λ is some constant to be fixed later. We assume that ω and the ω^x's are independent of each other. We will prove that the origin is connected to infinity in

$$\omega_{\text{total}} := \omega \vee \big(\vee_{x\in\mathbb{Z}^2}\, \omega^x\big)$$

with a probability which is larger than $\varepsilon/2$ (the notation $\vee$ stands for the maximum, or the union of the open edges if one prefers). This will conclude the proof since ω_{total} is stochastically dominated by a $(p + 25 \cdot \lambda\varepsilon)$-percolation—each edge of the slab appears in at most 25 boxes $\widetilde{\Lambda}_x$ (note that the number 25 does not depend on d because x is taken only in $\mathbb{Z}^2$).

To prove this claim, define an increasing sequence of percolation configurations $(\omega_t)_{t\geq 0}$ in the slab, coupled with a random exploration sequence $X_t = (A_t, B_t)$ in $\mathbb{Z}^2$. Given a percolation configuration ω in the slab, let $\mathscr{C}(\omega)$ be the set of vertices that are connected inside $\mathbb{Z}^2 \times \{-2N, \dots, 2N\}^{d-2}$ to 0 by a path of ω.

Definition 2 Set $X_0 = (A_0, B_0) := (\{0\}, \emptyset)$ and $\omega_0 = \omega$. For every $t \geq 0$, let ω_{t+1} and X_{t+1} be constructed from ω_t and X_t as follows. If there is no edge connecting A_t to $(A_t \cup B_t)^c$, define $X_{t+1} = X_t$. Otherwise, let $x = x_t$ be the extremity in $(A_t \cup B_t)^c$ of the minimal edge connecting A_t to $(A_t \cup B_t)^c$ and define

$$\omega_{t+1} := \omega_t \vee \omega^x,$$

$$X_{t+1} := \begin{cases} (A_t \cup \{x\}, B_t) & \text{if } 0 \leftrightarrow \Lambda_x \text{ in } \omega_{t+1}, \\ (A_t, B_t \cup \{x\}) & \text{otherwise.} \end{cases}$$

Remark 3 There is something unorthodox in the exploration process just defined, as we are not constraining the length of the paths that are created in each step. For example, it is possible that the path in $\mathbb{Z}^d$ that is responsible to connect 0 to $\Lambda_{(0,1)}$ goes much further than N.

Returning to the construction, we have the following two properties:

(i) $\omega_\infty := \vee_{t\geq 0}\, \omega_t \leq \omega_{\text{total}}$,
(ii) if (X_t) percolates, then 0 is connected to infinity in ω_∞.

We now wish to prove a third property which, when combined with the previous two and **(a)**, concludes the proof.

(iii) $\mathbb{P}[X \text{ percolates} \mid 0 \leftrightarrow \partial\Lambda_0 \text{ in } \mathscr{C}(\omega_0)] \geq 1/2$.

The proof relies on an application of Lemma 3. In order to apply this lemma, let us fix $q > p_c^{\text{site}}(\mathbb{Z}^2)$ in such a way that $c(q) \geq 1/2$ and try to prove

$$\mathbb{P}\big(B_{t+1} = B_t \mid X_0, \ldots, X_t\big) \geq q \ \text{ a.s.}$$

Since $B_{t+1} = B_t$ as soon as there is no edge connecting A_t to $(A_t \cup B_t)^c$, we can focus on the case where the minimal edge e connecting A_t to $(A_t \cup B_t)^c$ is well defined, and therefore its endpoint x in $(A_t \cup B_t)^c$ also is. In this case, we have $B_{t+1} = B_t$ if 0 is connected to Λ_x in ω_{t+1}. Since $X_0, \ldots, X_t$ and the event that x is well defined are measurable with respect to $\mathscr{C}(\omega_0), \ldots, \mathscr{C}(\omega_t)$, it suffices to show that for any admissible $C_0, C_1, \ldots, C_t$, we have

$$\mathbb{P}\big(\Lambda_0 \leftrightarrow \Lambda_x \text{ in } \omega_{t+1} \mid \mathscr{C}(\omega_0) = C_0, \ldots, \mathscr{C}(\omega_t) = C_t\big) \geq q \ \text{ a.s.},$$

which would follow if we showed that for every admissible C_t,

$$\mathbb{P}\big(C_t \leftrightarrow \Lambda_x \text{ in } \omega_t \vee \omega^x \mid \omega_t|_{\partial_E C_t} \equiv 0\big) \geq q. \tag{10}$$

Now, observe that any admissible C_t must intersect $\Lambda_{x'}$, where x' is the endpoint of e in A_t. Furthermore, the diameter of C_t must be at least N (here is where we use the conditioning over the event $0 \leftrightarrow \partial\Lambda_0$). Let y be a vertex of $C_t \cap \Lambda_{x'}$. Since at least one of the quarter-faces of $y + \Lambda_N$ is included in Λ_x (see an example in Fig. 2, where it is denoted by F; the reader is kindly requested to imagine the third dimension), Lemma 2 (applied after shifting 0 to y) implies

$$\mathbb{P}_p[C_t \xleftrightarrow{y+\Lambda_N} \Lambda_x \text{ in } \omega \vee \omega^x] \geq 1 - 2e^{-c/\varepsilon}.$$

Since the event $C_t \leftrightarrow \Lambda_x$ is increasing, we may replace $\omega \vee \omega^x$ with $\omega_t \vee \omega^x$. Since ω^x is independent of ω_t, Lemma 4 below shows that (10) holds, provided the constant λ is large enough. This concludes the proof of Item (iii) and therefore of Theorem 3. □

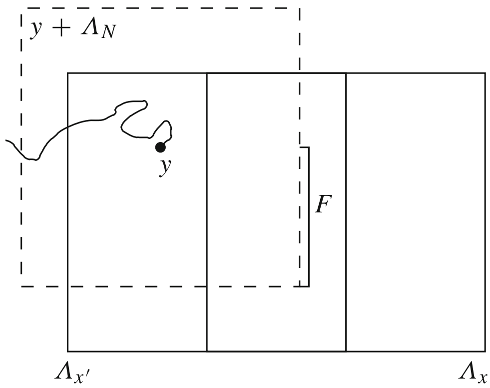


Fig. 2 An example of a quarter-face of $y + \Lambda_N$ that is included in Λ_x

For the next (and last) lemma, it will be convenient to have a notation for the edge boundary restricted to a fixed set. Fix therefore a set $R \subset \mathbb{Z}^d$, and define

$$\Delta A = \Big\{\{x, y\} \subset R \ :\ |x - y| = 1,\ x \in A,\ y \in R \setminus A\Big\}.$$

Lemma 4 *For any* $\delta, \eta > 0$*, there exists* $\lambda > 0$ *such that for any* $p \in [\delta, 1 - \delta]$ *and* $\varepsilon > 0$*, as well as any* $A, B \subset R$*,* $\mathbb{P}_p[A \overset{R}{\longleftrightarrow} B] \geq 1 - 2\exp(-\eta/\varepsilon)$ *implies that*

$$\mathbb{P}\big[A \overset{R}{\longleftrightarrow} B \text{ in } \omega \vee \tilde{\omega} \,\big|\, \omega(e) = 0, \forall e \in \Delta A\big] \geq 1 - \delta,$$

where ω *is a Bernoulli percolation configuration satisfying* $\mathbb{P}[\omega(e) = 1] \geq p$ *for every* e*, and* $\tilde{\omega}$ *a Bernoulli percolation of parameter* $\lambda\varepsilon$ *which is independent of* ω*.*

Proof If $A \cap B \neq \emptyset$, the result is obvious. We therefore assume $A \cap B = \emptyset$. Also, introduce the event E that $\omega(e) = 0$ for all $e \in \Delta A$ and the set W defined by

$$W = \Big\{\{x, y\} \in \Delta A \text{ and } y \overset{R\setminus A}{\longleftrightarrow} B \text{ in } \omega\Big\}.$$

Any path from A to B in R, open in ω, must use at least one edge of W. Consequently, for any $t \in \mathbb{N}$, we have

$$\mathbb{P}_p[A \overset{R}{\not\leftrightarrow} B] \geq (1 - p)^{t-1}\, \mathbb{P}_p[|W| < t].$$

Then, using that $|W| \geq t$ is independent of the event E, we deduce that, still for an arbitrary t,

$$\begin{aligned}\mathbb{P}[A \overset{R}{\longleftrightarrow} B \text{ in } \omega \vee \tilde{\omega} \mid E] &\geq \mathbb{P}[\exists e \in W : \tilde{\omega}(e) = 1, W \geq t \mid E]\\ &\geq (1 - (1 - \lambda\varepsilon)^t)\mathbb{P}[W \geq t]\end{aligned}$$

$$
\begin{aligned}
&\geq (1-(1-\lambda\varepsilon)^t)\Big(1-\frac{\mathbb{P}_p[A \overset{R}{\nleftrightarrow} B]}{(1-p)^{t-1}}\Big)\\
&\geq (1-(1-\lambda\varepsilon)^t)\Big(1-\frac{\exp(-\eta/\varepsilon)}{(1-p)^{t-1}}\Big).
\end{aligned}
$$

Choosing $\lambda = \lambda(\delta, \eta)$ large enough, the result follows by optimizing on t.

6 Proofs of Theorems 1 and 2

Theorem 2 follows immediately from what was already proved. Indeed, we use Theorem 3 with

$$
\begin{aligned}
p &= p_n + \lambda/\sqrt{\log n},\\
\varepsilon &= 1/\sqrt{\log n},\\
(k, K, n, N) &= (n^{\alpha^3}, n^{\alpha^2}, n^{\alpha}, n),
\end{aligned}
$$

where α is given by Proposition 1, and where λ is some sufficiently large constant. By the definition of α, condition **(c)** of Theorem 3 is satisfied when n is large enough. Condition **(a)** follows from Proposition 2, while Condition **(b)** follows from Proposition 3, if only λ is sufficiently large (we use Proposition 3 with $\beta = \alpha^3$).

The $p < p_c$ case of Theorem 1 is identical. Given $p < p_c$ we define $n = \lfloor \xi_p \rfloor$ and then $p_n \leq p$. Using Theorem 3 in the same way and with the same parameters as above gives that at $p_n + C/\sqrt{\log n}$ we already have percolation in a slab, and in particular it is above p_c. Hence

$$
p_c \leq p_n + \frac{C}{\sqrt{\log n}} \leq p + \frac{C}{\sqrt{\log \xi_p}}
$$

which is identical to $\xi_p \leq \exp(C(p_c - p)^{-2})$, as claimed.

For the case $p > p_c$ of Theorem 1 we need to estimate the probability to percolate in a slab *starting from the boundary of the slab*. It will be slightly more convenient to work in the "other slab", $\mathrm{OSlab}_n = \{-n, \ldots, n\} \times \mathbb{Z}^{d-1}$. Define

$$
\theta(p, n) := \mathbb{P}_p[(-n, 0, \ldots, 0) \xleftrightarrow{\mathrm{OSlab}_n} \infty].
$$

To estimate $\theta(p, n)$ we use the fact that at p_c

$$
\sum_{x \in \partial\Lambda_n} \mathbb{P}_{p_c}[0 \overset{\Lambda_n}{\longleftrightarrow} x] \geq c.
$$

(this is well-known, and in fact we already gave a proof of that while proving Proposition 2). By symmetry the same holds for the bottom face, i.e.

$$\sum_{x\in\{-n\}\times\{-n,\dots,n\}^{d-1}} \mathbb{P}_{p_c}[0 \overset{\Lambda_n}{\longleftrightarrow} x] \geq c.$$

Hence for some x on this face we have $\mathbb{P}_{p_c}[0 \overset{\Lambda_n}{\longleftrightarrow} x] \geq cn^{1-d}$. By translation invariance we get for some $y \in \{0\} \times \{-n, \dots, n\}^{d-1}$ that

$$\mathbb{P}_{p_c}[(-n, 0, \dots, 0) \overset{\mathrm{OSlab}_n}{\longleftrightarrow} y] \geq cn^{1-d}.$$

Let now $p > p_c$ and define n such that $\mathbb{P}_p[0 \overset{\mathrm{Slab}_n}{\longleftrightarrow} \infty] \geq 1/2\sqrt{\log n}$. By Theorem 2 we may take $n \leq \exp(C(p - p_c)^{-2})$. We may certainly replace Slab_n with OSlab_n, as it is larger. We get

$$\begin{aligned}\theta(p,n) &\geq \mathbb{P}_p[(-n, 0, \dots, 0) \overset{\mathrm{OSlab}_n}{\longleftrightarrow} y, y \overset{\mathrm{OSlab}_n}{\longleftrightarrow} \infty] \\ &\geq \mathbb{P}_p[(-n, 0, \dots, 0) \overset{\mathrm{OSlab}_n}{\longleftrightarrow} y]\mathbb{P}_p[y \overset{\mathrm{OSlab}_n}{\longleftrightarrow} \infty] \geq cn^{1-d} \cdot \frac{1}{2\sqrt{\log n}}\end{aligned}$$

where we used FKG, translation invariance and the fact that the event $(-n, 0, \dots, 0) \leftrightarrow y$ is monotone. By Chayes et al. [14, Theorem 5], $\xi_p \leq n/\theta(p,n)$ and Theorem 1 is established. □

7 On Proposition 1

Proposition 1 was not stated in this generality in the paper of Cerf [16] (in that paper, the polynomial upper bound is stated for $p = p_c$). Here we explain how the arguments of [16] can be adapted to get a bound which is uniform in $0 \leq p \leq 1$.

Proof of Proposition 1 It suffices to prove the estimate above for $p \in (\delta, 1-\delta)$, for some fixed $\delta > 0$ small enough. Indeed if p is close to 0 or 1, one can easily prove the bounds of the proposition using standard perturbative arguments. Now, using for example the inequality above Proposition 5.3 of [16], we see that there exists a constant $\kappa > 0$ such that for every $\delta \leq p \leq 1 - \delta$, and every $n \geq 1$

$$\mathbb{P}_p[A_2(0,n)] \leq \frac{\kappa \log n}{\sqrt{n}}.$$

We will prove that there exists $C \geq 1$ large enough such that, uniformly in $\delta \leq p \leq 1-\delta$, we have for every $n \geq 2$

$$\mathbb{P}_p[A_2(n, n^C)] \leq \frac{1}{n},$$

which concludes the proof. In the proof of [16], the fact that $p = p_c$ was used in order to obtain a lower bound on the two-point function (see Lemma 6.1 in [16]). One can replace the input coming from the hypothesis that $p = p_c$ by the following simple argument. Fix $n \geq 2$. Since $\mathbb{P}_p[A_2(n, n^C)] \leq \mathbb{P}_p[\Lambda_n \leftrightarrow \partial\Lambda_{2n}]$, we can assume that the probability that there exists an open path from Λ_n to $\partial\Lambda_{2n}$ is larger than $1/n$. Therefore, by the union bound, there must exist a point x at the boundary of Λ_n that is connected to $x + \partial\Lambda_n$ with probability larger than $\frac{1}{n|\partial\Lambda_n|}$. Hence, by translation invariance, 0 is connected to $\partial\Lambda_n$ with probability larger than $\frac{1}{n|\partial\Lambda_n|}$, and the union bound again implies that for every $m \leq n$,

$$\sum_{y\in\partial\Lambda_m} \mathbb{P}_p[0 \overset{\Lambda_m}{\longleftrightarrow} y] \geq \frac{1}{n|\partial\Lambda_n|}.$$

Using this estimate, one can repeat the argument of Lemma 6.1 in [16] to show that there exists a constant $C > 0$ (independent of n and p) such that

$$\forall x, y \in \Lambda_n \quad \mathbb{P}_p[x \overset{\Lambda_{2n}}{\longleftrightarrow} y] \geq \frac{1}{n^C}.$$

Then one can conclude the proof using the estimate above together with Corollary 7.3 in [16].

8 A Lower Bound

In this section, we make a few remarks on lower bounds for the correlation length. We first note that [4] shows that at $p < p_c$ the expected size of the cluster ("the susceptibility") is at least $1/(p_c - p)$, and this shows that $\xi_p \geq (p_c - p)^{-1/d}$. Newman [36] shows a lower bound also on the truncated susceptibility for $p > p_c$ but he makes assumptions on the behaviour of critical percolation which are still unknown: we could not complete Newman's argument without assuming Conjecture 1.

Here we will sketch a proof that $\xi_p \geq (p_c - p)^{-2/d+o(1)}$ in the case that $p < p_c$, leaving the more complicated case of $p > p_c$ for the future. Let $N > 0$, and let E be the event that there exists an easy-way crossing of the box $3N \times \cdots \times 3N \times N$. By Kesten [31, §5.1] there is a constant $c_1(d)$ such that if $\mathbb{P}_q[E] < c_1$ then $q < p_c$.

Let now $p < p_c$. Standard arguments using supermultiplicativity (see [14]) show that for every x with $|x| > N$ we have $\mathbb{P}_p(0 \leftrightarrow x) \leq \exp(-N/\xi_p)$. Hence there exists $N \leq C\xi_p \log \xi_p$ such that $\mathbb{P}_p[E] < \frac{1}{2}c_1(d)$. Now, it is well-known that for any boolean function f the total influence $I(f)$ satisfies $I(f) \leq \sqrt{n \mathrm{var} f / p(1-p)}$. Defining $F(p) = \mathbb{E}_p[f]$ this gives, for p bounded away from 0 and 1, $\sqrt{F}' \leq C\sqrt{n}$. We apply this for f being the indicator of the event E and get

$$\frac{d}{dp}\sqrt{\mathbb{P}_p[E]} \leq C' N^{d/2} \leq C''(\xi_p \log \xi_p)^{d/2}.$$

Hence at $q := p + c_2(\xi_p \log \xi_p)^{-d/2}$ for some c_2 sufficiently small, we would have $\mathbb{P}_q[E] < c_1(d)$ and hence $q < p_c$. The claim follows.

Acknowledgments This work is based on earlier unpublished results achieved with Vladas Sidoravicius, we wish to take this opportunity to thank Vladas for his contributions. Many readers of the first version made important suggestions and corrections, including Rob van den Berg, Raphael Cerf, Roger Silva. We thank Ariel Yadin for asking what is known about lower bounds.

HDC was supported by the ERC CriBLaM, a chair IDEX from Paris-Saclay, a grant from the Swiss NSF and the NCCR SwissMap also funded by the Swiss NSF. GK was supported by the Israel Science Foundation, by Paul and Tina Gardner and by the Jesselson Foundation. VT was supported by the NCCR SwissMap, funded by the Swiss NSF.

References

1. Adler, J., Meir, Y., Aharony, A., Harris, A.B.: Series sutdy of percolation moments in general dimension. Phys. Rev. B **41**(13), 9183–9206 (1990)
2. Aizenman, M.: On the number of incipient spanning clusters. Nucl. Phys. B **485**(3), 551–582 (1997). https://doi.org/10.1016/S0550-3213(96)00626-8
3. Aizenman, M., Barsky, D.J.: Sharpness of the phase transition in percolation models. Commun. Math. Phys. **108**(3), 489–526 (1987). https://projecteuclid.org/euclid.cmp/1104116538
4. Aizenman, M., Newman, C.M.: Tree graph inequalities and critical behavior in percolation models. J. Stat. Phys. **36**(1–2), 107–143 (1984). https://doi.org/10.1007/BF01015729
5. Aizenman, M., Delyon, F., Souillard, B.: Lower bounds on the cluster size distribution. J. Stat. Phys. **23**(3), 267–280 (1980). https://doi.org/10.1007/BF01011369
6. Aizenman, M., Kesten, H., Newman, C.M.: Uniqueness of the infinite cluster and continuity of connectivity functions for short and long range percolation. Commun. Math. Phys. **111**(4), 505–531 (1987). https://projecteuclid.org/euclid.cmp/1104159720
7. Antal, P., Pisztora, A.: On the chemical distance for supercritical Bernoulli percolation. Ann. Probab. **24**(2), 1036–1048 (1996). https://projecteuclid.org/euclid.aop/1039639377
8. Barsky, D., Grimmett, G.R., Newman, C.M.: Percolation in half-spaces: equality of critical densities and continuity of the percolation probability. Probab. Theory Relat. Fields **90**(1), 111–148 (1991). https://doi.org/10.1007/BF01321136
9. Benjamini, I., Lyons, R., Peres, Y., Schramm, O.: Critical percolation on any nonamenable group has no infinite clusters. Ann. Probab. **27**(3), 1347–1356 (1999). https://projecteuclid.org/euclid.aop/1022677450
10. Borgs, C., Chayes, J.T., Kesten, H., Spencer, J.: Uniform boundedness of critical crossing probabilities implies hyperscaling. Random Struct. Algorithms **15**(3–4), 368–413 (1999). Statistical physics methods in discrete probability, combinatorics, and theoretical computer science (Princeton, NJ, 1997). https://doi.org/10.1002/(SICI)1098-2418(199910/12)15:3/4%3C368::AID-RSA9%3E3.0.CO;2-B

11. Bourgain, J., Kahn, J., Kalai, G., Katznelson, Y., Linial, N.: The influence of variables in product spaces. Isr. J. Math. **77**(1–2), 55–64 (1992). https://link.springer.com/article/10.1007%2FBF02808010
12. Brydges, D., Spencer, T.: Self-avoiding walk in 5 or more dimensions. Commun. Math. Phys. **97**(1–2), 125–148 (1985). https://projecteuclid.org/euclid.cmp/1103941982
13. Chayes, J.T., Chayes, L.: Inequality for the infinite-cluster density in Bernoulli percolation. Phys. Rev. Lett. **56**(16), 1619–1622 (1986). https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.56.1619
14. Chayes, J.T., Chayes, L., Newman, C.M.: Bernoulli percolation above threshold: an invasion percolation analysis. Ann. Probab. **15**(4), 1272–1287 (1987). https://www.jstor.org/stable/2244002?seq=1#metadata_info_tab_contents
15. Chayes, J.T., Chayes, L., Grimmett, G.R., Kesten, H., Schonmann, R.H.: The correlation length for the high-density phase of Bernoulli percolation. Ann. Probab. **17**(4), 1277–1302 (1989). https://www.jstor.org/stable/2244436?seq=1#metadata_info_tab_contents
16. Cerf, R.: A lower bound on the two-arms exponent for critical percolation on the lattice. Ann. Probab. **43**(5), 2458–2480 (2015). https://projecteuclid.org/euclid.aop/1441792290
17. Damron, M., Newman, C.M., Sidoravicius, V.: Absence of site percolation at criticality in $\mathbb{Z}^2 \times \{0, 1\}$. Random Struct. Algorithms **47**(2), 328–340 (2015). https://doi.org/10.1002/rsa.20544
18. Duminil-Copin, H., Tassion, V.: A new proof of the sharpness of the phase transition for Bernoulli percolation on $\mathbb{Z}^d$. L'Enseignement Mathématique **62**(1/2), 199–206 (2016). https://www.ems-ph.org/journals/show_abstract.php?issn=0013-8584&vol=62&iss=1&rank=12
19. Duminil-Copin, H., Sidoravicius, V., Tassion, V.: Absence of infinite cluster for critical bernoulli percolation on slabs. Commun. Pure Appl. Math. **69**(7), 1397–1411 (2016). https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.21641
20. Duminil-Copin, H., Raoufi, A., Tassion, V.: A new computation of the critical point for the planar random-cluster model with $q \geq 1$. Ann. Inst. Henri Poincaré **1**(54), 422–436 (2018). https://projecteuclid.org/euclid.aihp/1519030834
21. Fitzner, R., van der Hofstad, R.: Mean-field behavior for nearest-neighbor percolation in $d > 10$. Electron. J. Probab. **22**(43), 1–65 (2017). https://projecteuclid.org/euclid.ejp/1493777019
22. Grimmett, G.R., Marstrand, J.M.: The supercritical phase of percolation is well behaved. Proc. Roy. Soc. Lond. Ser. A **430**(1879), 439–457 (1990). https://doi.org/10.1098/rspa.1990.0100
23. Grimmett, G.R.: Percolation. In: Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 321, 2nd edn. Springer, Berlin (1999)
24. Hara, T.: Mean-field critical behaviour for correlation length for percolation in high dimensions. Probab. Theory Relat. Fields **86**(3), 337–385 (1990). https://link.springer.com/article/10.1007%2FBF01208256
25. Harris, T.E.: A lower bound for the critical probability in a certain percolation process. Proc. Camb. Philos. Soc. **56**(1), 13–20 (1960). https://doi.org/10.1017/S0305004100034241
26. Hara, T., Slade, G.: Mean-field critical behaviour for percolation in high dimensions. Commun. Math. Phys. **128**(2), 333–391 (1990). https://projecteuclid.org/euclid.cmp/1104180434
27. Hermon, J., Hutchcroft, T.: No percolation at criticality on certain groups of intermediate growth (preprint, 2018). https://arxiv.org/abs/1809.11112
28. Hutchcroft, T.: Critical percolation on any quasi-transitive graph of exponential growth has no infinite clusters. Comptes Rendus Math. **354**(9), 944–947 (2016). https://www.sciencedirect.com/science/article/pii/S1631073X16301352?via%3Dihub
29. Kahn, J., Kalai, G., Linial, N.: The influence of variables on boolean functions. In: 29th Symposium on the Foundations of Computer Science, pp. 68–90. White Planes, IEEE (1988). https://doi.ieeecomputersociety.org/10.1109/SFCS.1988.21923
30. Kesten, H.: The critical probability of bond percolation on the square lattice equals 1/2. Commun. Math. Phys. **74**(1), 41–69 (1980). https://projecteuclid.org/euclid.cmp/1103907931
31. Kesten, H.: Percolation Theory for Mathematicians. Progress in Probability and Statistics, vol. 2. Birkhäuser, Boston (1982). http://pi.math.cornell.edu/~kesten/kesten-book.html

32. Kesten, H.: A scaling relation at criticality for 2D-percolation. In: Percolation Theory and Ergodic Theory of Infinite Particle Systems (Minneapolis, 1984–1985). The IMA Volumes in Mathematics and its Applications, vol. 8, pp. 203–212. Springer, New York (1987). https://link.springer.com/chapter/10.1007%2F978-1-4613-8734-3_12
33. LeClair, A., Squires, J.: Conformal bootstrap for percolation and polymers. J. Stat. Mech. Theory Exp. **2018**(12), 1–19 (2018). https://doi.org/10.1088/1742-5468/aaf10a
34. Martineau, S., Tassion, V.: Locality of percolation for Abelian Cayley graphs. Ann. Probab. **45**(2), 1247–1277 (2017). https://projecteuclid.org/euclid.aop/1490947319
35. Menshikov, M.V.: Совпадение критических точек в задачах перколяции. Dokl. Akad. Nauk SSSR **288**(6), 1308–1311 (1986). http://mi.mathnet.ru/eng/dan8543. English translation: Coincidence of critical points in percolation problems. Soviet Math. Dokl. **33**, 856–859 (1986)
36. Newman, C.M.: Some critical exponent inequalities for percolation. J. Stat. Phys. **45**(3–4), 359–368 (1986). https://doi.org/10.1007/BF01021076
37. Russo, L.: A note on percolation. Z. Wahrsch. Verw. Gebiete **43**(1), 39–48 (1978). https://link.springer.com/article/10.1007%2FBF00535274
38. Russo, L.: An approximate zero-one law. Z. Wahrsch. Verw. Gebiete **61**(1), 129–139 (1982). https://doi.org/10.1007/BF00537230
39. Seymour, P.D., Welsh, D.J.A.: Percolation probabilities on the square lattice. Ann. Discrete Math. **3**, 227–245 (1978). Advances in graph theory (Cambridge Combinatorial Conference, Trinity College, Cambridge, 1977). https://ac.els-cdn.com/S0167506008705090/1-s2.0-S0167506008705090-main.pdf?_tid=4ebacdef-40d0-4243-8f78-6500a710c2ef&acdnat=1548694546_f81db87832a5d4daa7d1505c1acf48fa
40. Smirnov, S.: Critical percolation in the plane: conformal invariance, Cardy's formula, scaling limits. C. R. Acad. Sci. Paris Sér. I Math. **333**(3), 239–244 (2001). https://doi.org/10.1016/S0764-4442(01)01991-7
41. Smirnov, S., Werner, W.: Critical exponents for two-dimensional percolation. Math. Res. Lett. **8**(5–6), 729–744 (2001). http://dx.doi.org/10.4310/MRL.2001.v8.n6.a4
42. Talagrand, M., On Russo's approximate zero-one law. Ann. Probab. **22**(3), 1576–1587 (1994). https://www.jstor.org/stable/2245033
43. Wikipedia, Percolation critical exponents. https://en.wikipedia.org/wiki/Percolation_critical_exponents. Accessed Jan. 2020

The Roles of Random Boundary Conditions in Spin Systems

Eric O. Endo, Aernout C. D. van Enter, and Arnaud Le Ny

Abstract Random boundary conditions are one of the simplest realizations of quenched disorder. They have been used as an illustration of various conceptual issues in the theory of disordered spin systems. Here we review some of these results.

Keywords Random boundary conditions · Quenched disordered systems · Chaotic size dependence · Weak versus strong uniqueness · Metastates

1 Introduction

In the theory of disordered systems, and in particular in the theory of spin glasses, for which the existence of phase transitions on the lattice is so far unproven and the nature of the conjectured transition even among theoretical physicists is a topic of controversy, the traditional approach of selecting different Gibbs states by imposing boundary conditions geared towards a preferred state is ineffective, as we don't know which (Gibbs or ground) preferred states there could be to select from.

Physically, it then makes more sense considering boundary conditions which are independent of the interactions. In the case of spin glasses those could be fixed, periodic or free, for example. Indeed, choosing boundary conditions which depend

E. O. Endo
NYU-ECNU Institute of Mathematical Sciences at NYU Shanghai, Shanghai, China
e-mail: eoe2@nyu.edu

A. C. D. van Enter (✉)
Bernoulli Institute, University of Groningen, Groningen, Netherlands
e-mail: aenter@phys.rug.nl

A. Le Ny
LAMA UMR CNRS 8050, UPEC, Université Paris-Est, Créteil, France
e-mail: arnaud.le-ny@u-pec.fr

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_17

on a realization of a set of random interactions is not a physically feasible procedure. For an early discussion of this point we refer to [49].

The mathematical theory of disordered spin systems has been described for example in [8, 39, 44]. See also [43] for a reader-friendly introduction to the issues which show up in the spin-glass problem.

Having boundary conditions independent of the interactions has played a role in the proper definition of spin-glass Edwards-Anderson order parameters [12, 50] and also in the issue of weak versus strong uniqueness of Gibbs measures [12, 17, 27].

A possibility which can naturally occur in disordered models is non-convergence of finite-volume Gibbs measures in the thermodynamic limit ("Chaotic Size Dependence") [40, 48]. If the finite-volume states don't converge, it still might be the case that distributional limits exist. Such limiting objects then are called "metastates". Tractable examples of them mean-field models, see e.g. [9, 29, 35], but for short-range lattice models one usually needs to consider somewhat simplified models, as e.g. in [52, 53].

Considering deterministic models with random boundary conditions can provide suitable illustrations of various conceptual issues. For some descriptions of the analogy between spin glasses with fixed boundary conditions and deterministic models with random boundary conditions, see [3, 17, 40, 48]. Short-range Ising models have been studied in [52, 53], and more recently one-dimensional long-range models have been considered [21]. Here we review some of the results which were found in those examples.

2 Background and Notation on Disordered Spin Models

2.1 Spin Models and Disorder

We will consider spin models in which we denote spin configurations (respectively spins at site i) by σ (respectively σ_i), on a connected, infinite and locally-finite graph G. The state space is Ω_0, the spin configuration space is Ω_0^G.

For a subset of vertices $G' \subset G$, we denote by $\sigma_{G'} = (\sigma_i)_{i \in G'}$ the configuration restricted on G'. Define $\Omega_0^{G'}$ to be the set of configurations on G'.

Let $\Phi = (\Phi_X)_{X \subset G,\ X \text{ finite}}$ be a family of $\mathscr{F}_X$-measurable functions $\Phi_X : \Omega_0^G \to \mathbb{R}$, where $\mathscr{F}_X$ is the local σ-algebra generated by the cylinders σ_X; we will call such a family an interaction.

Given a finite-volume $\Lambda \subset G$, finite-volume Hamiltonians are expressed in terms of interactions

$$H_\Lambda(\sigma_\Lambda b_{\Lambda^c}) = \sum_{X \cap \Lambda \neq \emptyset} \Phi_X(\sigma_\Lambda b_{\Lambda^c}).$$

Here b is an arbitrary boundary condition, an element of Ω_0^G, identified with its projections $\Omega_0^{\Lambda^c}$, and

$$\sigma_\Lambda b_{\Lambda^c} = \begin{cases} \sigma_i, & \text{if } i \in \Lambda, \\ b_i, & \text{if } i \in \Lambda^c. \end{cases}$$

From such Hamiltonians one constructs finite-volume Gibbs measures on volume Λ, with boundary condition b on Λ^c and inverse temperature $\beta > 0$,

$$\mu^b_{\Lambda,\beta}(\sigma_\Lambda) = \frac{1}{Z^b_{\Lambda,\beta}} e^{-\beta H_\Lambda(\sigma_\Lambda b_{\Lambda^c})}.$$

The normalization

$$Z^b_{\Lambda,\beta} = \sum_{\sigma \in \Omega_0^\Lambda} e^{-\beta H_\Lambda(\sigma_\Lambda b_{\Lambda^c})}$$

is called partition function. Under appropriate summability conditions on the interaction, in the thermodynamic limit (infinite-volume) Gibbs measures exist, also known as DLR measures. For the theory of infinite-volume Gibbs measures we refer to [23, 28, 45, 51].

Although the theory applies in wider generality, we will tend to restrict ourselves to Ising spins $\sigma_i \in \Omega_0 = \{-1, 1\}$.

Disordered systems depend on another random parameter, the disorder parameter η. This disorder parameter can describe either bond randomness or site randomness in the interactions, which then become random $\mathscr{F}_X$-measurable functions $\Phi^\eta_X(\cdot)$ for each X finite. Usually the variables η are independent random variables with a distribution which is translation invariant and which depends on the shapes of the subsets of the lattice X.

There exists an extensive literature, both in (rigorous and nonrigorous) theoretical physics and mathematical physics, on disordered systems. Here we refer to [8, 39, 43, 44] for some further mathematical and conceptual background and theory on them.

We warn the reader, moreover, that the well-known random-bond equivalent-neighbour Sherrington-Kirkpatrick model of a spin-glass, although it has been rigorously solved by Guerra and Talagrand, following the ideas of Parisi, in many aspects is exceptional and many statements which apply to it have no equivalent statement in the context we discuss. For some of the arguments on these issues, see [8, 39–43, 46, 47].

2.2 Examples: From Spin Glass to Mattis Disorder to Random Boundary Conditions

Popular examples of disordered Ising systems include:

1. **Edwards-Anderson spin-glasses:** on $G = \mathbb{Z}^d$ with Hamiltonian

$$H_\Lambda(\sigma_\Lambda b_{\Lambda^c}) = -\sum_{\substack{i,j\in\Lambda\\ i\neq j}} \eta_{i,j} J(i-j)\sigma_i\sigma_j - \sum_{\substack{i\in\Lambda\\ j\in\Lambda^c}} \eta_{i,j} J(i-j)\sigma_i b_j,$$

   where the distribution of the (bond-random) $\eta_{i,j}$ is symmetric and depends only on $|i-j|$. They are usually taken as centered Gaussian or symmetric ± 1.
2. **Random-field Ising models:** on $G = \mathbb{Z}^d$ with Hamiltonian

$$H_\Lambda(\sigma_\Lambda b_{\Lambda^c}) = -\sum_{\substack{i,j\in\Lambda\\ i\neq j}} J(i-j)\sigma_i\sigma_j - \sum_{\substack{i\in\Lambda\\ j\in\Lambda^c}} J(i-j)\sigma_i b_j - \lambda\sum_{i\in\Lambda}\eta_i\sigma_i,$$

   where the η_i are (site-random) i.i.d. and symmetrically distributed random variables. Just as with Edwards-Anderson models, the most considered distributions are centered Gaussian and Bernoulli distributions.
3. **Mattis spin glasses:** on $G = \mathbb{Z}^d$ with Hamiltonian

$$H_\Lambda(\sigma_\Lambda b_{\Lambda^c}) = -\sum_{\substack{i,j\in\Lambda\\ i\neq j}} J(i-j)\eta_i\eta_j\sigma_i\sigma_j - \sum_{\substack{i\in\Lambda\\ j\in\Lambda^c}} J(i-j)\eta_i\eta_j\sigma_i b_j,$$

   where again the (site-random) η_i are i.i.d. and symmetrically distributed, typically ± 1.

So far the theory of Edwards-Anderson spin glasses lacks examples in which it is clear that phase transitions occur.[1]

The random-field Ising models have phase transitions in case of nearest-neighbor models in dimension at least 3, and also in dimension 1 if we consider the long-range interaction with sufficiently slow decay. These results all agree with heuristic predictions, based on some form of an Imry-Ma argument [32]. In such an argument one compares the (free-)energy cost of an excitation due to the spin interactions with the energy gain due to the magnetic field term.

This argument has been rigorized in a number of cases, sometimes requiring a serious mathematical analysis [1, 7, 11, 14, 36].

[1] C.M. Newman and D.L. Stein, among others, have raised the question of proving phase transitions for Edwards-Anderson models at various occasions. See for example http://web.math.princeton.edu/~aizenman/OpenProblems_MathPhys/9803.SpinGlass.html. No progress seems to have been made since then.

The Mattis spin-glasses on the other hand, by the random gauge transformation $\sigma_i' = \eta_i \sigma_i$, are equivalent to ferromagnetic models. Thus the existence or not of a phase transition typically only requires understanding the ferromagnet. We note that for a finite-volume Gibbs measure a fixed boundary condition by this random gauge transformation is mapped to a random boundary condition [40, 41].

We also notice that Mattis disorder is the same as single-pattern Hopfield disorder, which has been considered in particular in the mean-field version; for more on Hopfield models see e.g. [8, 9, 35].

3 Earlier Results and New Heuristics on Random Boundary Conditions

In the results which we review below, we always impose boundary conditions which are drawn from a symmetric i.i.d. product (Bernoulli) measure. This does not preselect the phase, and such boundary conditions are sometimes called "incoherent" (as introduced in [42], see also [53]).

3.1 Weak Versus Strong Uniqueness

We say that a model displays weak uniqueness of the Gibbs measure if for each choice of boundary condition almost surely (for almost all choices of the random interaction) the same infinite-volume Gibbs measure is approached. Strong uniqueness holds if there exists a unique Gibbs measure for the model for almost all choices of the interaction.

It is known that one-dimensional high-temperature long-range spin-glass models display weak uniqueness without strong uniqueness [25, 27]. Other examples where this occurs are the nearest-neighbour Ising spin-glass models on a tree, between the critical temperature and the spin-glass temperature [17].

Similarly to what happens in Mattis models, one can transform the disorder to the boundary, and in the temperature interval between the ferromagnetic transition T_c (below which plus and minus boundaries produce different states in the thermodynamic limit) and the free-boundary-purity (or spin-glass) transition temperature T_{SG}, below which the limiting Gibbs measure obtained with free boundary conditions becomes non-extremal, there is weak uniqueness of the Gibbs measure without strong uniqueness.

Similar behaviour (weak but not strong uniqueness) has also been derived for a Potts-Mattis model on $\mathbb{Z}^d$ for high q, at the transition temperature, with $d \geq 2$ [12].

3.2 Nearest-Neighbour Ising Models at Low Temperatures, Metastates

We first summarize here the results derived and described in [52, 53] for the nearest-neighbour Ising model on $\mathbb{Z}^d$. If we consider an Ising model on a box of size N^d with random boundary conditions, the ground state energy and also the low-temperature free energy satisfy a weak version of the local central limit theorem. One obtains estimates for the probability of the boundary term of a boundary of size N to lie, not in finite intervals (as in the proper local limit theorem) or in intervals of size $\sqrt{N}$ (as in the ordinary central limit theorem), but in intervals of size N^{δ}, with some δ between 0 and $\frac{1}{2}$. This still suffices to show that the probability that the boundary free energy is close to zero goes to zero at a fast enough rate.

This can be used to show that the boundary (free) energy in a reasonably precise way scales like $N^{\frac{d-1}{2}}$. From this it follows in particular that the (free) energy difference between plus and minus phase diverges, with large enough probability, and thus randomly one of the two tends to be preferred.

The distributional limit behaviour can be described in terms of "metastates", objects which were introduced by Aizenman-Wehr [1] and Newman-Stein [39, 41, 42] via different constructions, which then were shown to be equivalent, see also [19].

A metastate is a measure on Gibbs measures. In its support either extremal or non-extremal Gibbs measures, or both, can occur. If the support of a metastate contains more than one measure, it is called "dispersed".

In case the distribution is η-dependent, the translation covariant metastate in fact becomes a measure on measures (distributions) on Gibbs measures. Translation covariance here means that shifting the η induces a shift of the corresponding random Gibbs measures in the metastate.

The weight of a Gibbs measure in a metastate indicates the probability of finding that particular Gibbs measure for a randomly chosen volume for a particular realization of the interaction, or else, the probability of finding that Gibbs measure for a random realization of the interaction in a given large volume. If the Gibbs measures are random, this metastate necessarily is also a random object.

Although the notion of metastate has been developed for spin glasses, these have turned out to be so intractable that most examples which could be handled are either mean-field models with site-random variables (see for example [9, 29]), or other heavily simplified models.

In [52, 53] it was proven, for example, that the metastates obtained with random boundary conditions live on the (extremal) plus and minus measures of the nearest-neighbour Ising model. Whereas the simple case of ground states with weak (finite-energy) boundary conditions—that is, the bonds inside the volume are infinite, but boundary bonds are finite—is mathematically fairly straightforward, the low-temperature case required a careful analysis, making use of the technique of cluster expansions, including estimates on boundary contours, leading to a weak version of the local central limit theorem. But the analysis ended up providing essentially the same result as holds at $T = 0$. As the weights in the metastate are obtained by exponentiating and then normalizing the boundary (free) energies, divergence of those boundary terms leads to weights which are either zero or one.

When the sequence of volumes is sufficiently sparse, the plus and minus measures are in fact the only two limit points. To prove this, one has to exclude null-recurrent behaviour when taking sequences of increasing volumes. By taking sparse sequences, this allows one to apply a Borel-Cantelli argument.

3.3 Long-Range Ising Models, Metastates

In [21] we have started to extend the analysis of the metastate description to one-dimensional long-range Ising systems. There has been a substantial progress in the study of such low-dimensional long-range Ising models, which are known to display phase transitions [2, 10, 20, 24, 26, 30, 31, 33, 34]. For a number of more recent works on these models see [4–7, 13–16, 18, 22, 37, 38, 54, 55].

As our canonical example, on $G = \mathbb{Z}$ and $\alpha \in (1, 2]$, we consider the Dyson models with Hamiltonian

$$H_\Lambda(\sigma_\Lambda b_{\Lambda^c}) = -\sum_{\substack{i,j\in\Lambda\\ i\neq j}} |i-j|^{-\alpha}\sigma_i\sigma_j - \sum_{\substack{i\in\Lambda\\ j\in\Lambda^c}} |i-j|^{-\alpha}\sigma_i\eta_j,$$

where the site-random boundary η_j are i.i.d., symmetrically distributed random variables on $\{-1, 1\}$.

If we impose weak (finite-energy) random boundary conditions, on an interval of size N (so boundary bonds are finite, but bonds inside the volume are infinite) the ground states, which are the plus and minus configurations, have a difference in energy which scales as $N^{\frac{3}{2}-\alpha}$ when $\alpha < \frac{3}{2}$, and is almost surely bounded otherwise. This implies that for $\alpha > \frac{3}{2}$ the metastate lives on mixed ground state measures, whereas for $\alpha < \frac{3}{2}$, similarly to what occurs in higher-dimensional short-range models, the metastate has only the plus and minus states in its support. For positive temperatures analogous results for metastates on Gibbs measures are expected and partially proven ([21] in progress).

To prove that null-recurrence of the set of mixed states does not occur, in case the decay is slow enough, we again need to consider sufficiently sparse sequences of increasing volumes. Next to being needed for a Borel-Cantelli argument, this also allows us to treat the boundary energies of different volumes as (approximately) independent.

To obtain an almost sure statement, a local-limit-type argument, along the lines of the one discussed in Appendix B of [52] could then be invoked.

Denote by $\sigma_j^+ = +1$ for all $j \in \mathbb{Z}$. The main object we study is the formal expression

$$W^+ = \sum_{i<0}\sum_{j\geq 0} |i-j|^{-\alpha}\sigma_j^+\eta_i = \sum_{i<0} W_i.$$

This expression describes the interaction of a random boundary condition on the negative half-line with the plus ground state configuration on the right half-line.

As the W_i are independent, the expectation of each W_i is zero, and the variance is $\mathrm{Var}(W_i) = O(|i|^{2-2\alpha})$.

Alternatively, we can write $W = \sum_{j\geq 0} W'_j$, with the random variables W'_j having zero expectation, being strongly correlated and satisfying $\mathbb{E}((W'_j)^2) = O(|j|^{1-2\alpha})$, so $(\mathrm{Var}(W'_j))^{1/2} = O(|j|^{\frac{1}{2}-\alpha})$. Now instead of the sum of the variances, we need to consider a sum of the –non-independent– W_j themselves.

Therefore it follows that, whether one considers either a plus interval of size N with a random boundary, or, alternatively, a random interval of size N with a plus boundary, both scale like $N^{\frac{3}{2}-\alpha}$.

We remark that the sum of left and right boundary energy terms on both sides of a large enough interval, again can be written as a sum of similar form and for that reason satisfies the same scaling. This provides the scaling of the boundary energies mentioned above.

The boundary terms of a finite interval consist of a left boundary term and a right boundary term, when the interval is large those can be treated as more or less independent.

Let μ^+_β and μ^-_β be the thermodynamic limit of the plus b.c. Gibbs measures $\mu^+_{\Lambda,\beta}$ and minus b.c. $\mu^-_{\Lambda,\beta}$, respectively.

We notice that if two boundary energies are only differing by a finite amount, the limiting Gibbs measures (or ground state measures in the zero-temperature framework) are absolutely continuous with respect to each other. This happens almost surely when $\alpha > \frac{3}{2}$. In that case, W is a well-defined, almost surely finite random variable with some non-trivial distribution. The weight distribution $\lambda = \lambda(\mu) \in [0, 1]$ on mixed Gibbs measures μ given by

$$\mu = \lambda\mu^+_\beta + (1-\lambda)\mu^-_\beta,$$

obtained by exponentiating the boundary energies and normalizing them, also has a non-trivial distribution.

This means in particular that the measures in the support of this distribution are different mixtures of the plus and minus states. Therefore the metastate lives on different mixtures, rather than on pure states.

So far we have derived the results described above for low enough temperatures when $\alpha > \frac{3}{2}$, and for $T = 0$ with finite boundary terms when $\alpha < \frac{3}{2}$. Just as in the nearest-neighbour case, the extension to positive temperatures in the second case requires a sophisticated low-temperature (contour expansion) analysis, which is in progress.

We remark moreover that it follows from [19] that a metastate supported on pure states also exists; however, its construction will have to be different than just imposing independent random boundary conditions (possibly by making use of a maximizing procedure, or by considering highly correlated boundary conditions). For the ferromagnet this is immediate, for the Mattis version of our models less so.

We also notice here that an earlier mean-field example of metastates living on mixed states appeared already in [35]. In that case, bulk disorder was present. Mean-field models lacking boundary terms, the origin of mixed states occurring in the metastate therefore seems rather different from our random-boundary Dyson model example.

4 Conclusion, Final Remarks

Random boundary conditions for ferromagnets play a similar role as fixed boundary for spin glasses. In the case of spin glasses with Mattis disorder, there is in fact direct map between those two cases.

Random boundary conditions can be used to illustrate the concepts of weak and strong uniqueness, as well as describing various metastate scenarios. In particular, in one-dimensional Dyson models with a decay power between $\frac{3}{2}$ and 2, they lead in a natural way to examples where the phenomenon of dispersed metastates living on mixed Gibbs measures appear, a phenomenon which apparently is new for lattice systems.

Acknowledgments We dedicate this contribution to the memory of Vladas Sidoravicius. His enthusiasm as well as his broad interests, which included both long-range models and the new phenomena occurring in random versions of a wide class of statistical mechanics models, have often been a stimulus and inspiration for all of us. We thank our collaborators and discussion partners on this and related projects, Rodrigo Bissacot, Loren Coquille, Roberto Fernández, Bruno Kimura, Christof Külske, Pierre Picco, Wioletta Ruszel, Cristian Spitoni and Evgeny Verbitskiy.

References

1. Aizenman, M., Wehr, J.: Rounding effects of quenched randomness on first-order phase transitions. Comm. Math. Phys. **130**, 489–528 (1990)
2. Aizenman, M., Chayes, J., Chayes, L., Newman, C.: Discontinuity of the magnetization in the one-dimensional $1/\mid x-y\mid^2$ percolation, Ising and Potts models. J. Stat. Phys. **50**(1/2), 1–40 (1988)
3. Banavar, J.R., Cieplak, M., Cieplak, M.Z.: Influence of boundary conditions on random unfrustrated systems. Phys. Rev. B **26**, 2482–2489 (1982)
4. Berger, N., Hoffman, C., Sidoravicius, V.: Nonuniqueness for specifications in $l^{2+\epsilon}$. Erg. Th. Dyn. Syst. **38**, 1342–1352 (2018)
5. Berghout, S., Fernández, R., Verbitskiy, E.: On the relation between Gibbs and g-measures. Ergodic Theory Dyn. Syst. **39**, 3224–3249 (2019)
6. Bissacot, R., Endo, E.O., van Enter, A.C.D., Le Ny, A.: Entropic repulsion and lack of the g-measure property for Dyson models. Comm. Math. Phys. **363**, 767–788 (2018)
7. Bissacot, R., Endo, E.O., van Enter, A.C.D., Kimura, B., Ruszel, W.M.: Contour methods for long-range Ising models: weakening nearest-neighbor interactions and adding decaying fields. Ann. Henri Poincaré **19**, 2557–2574 (2018)
8. Bovier, A.: Statistical Mechanics of Disordered Systems. Cambridge University Press, Cambridge (2006)

9. Bovier, A., Gayrard, V.: Metastates in the Hopfield model in the replica symmetric regime. Math. Phys. Anal. Geom. **1**, 107–144 (1998)
10. Bramson, M., Kalikow, S.: Non-uniqueness in g-functions. Israel J. Math. **84**, 153–160 (1993)
11. Bricmont, J., Kupiainen, A.: Phase transition in the 3d random field Ising model. Comm. Math. Phys. **116**, 539–572 (1988)
12. Campanino, M., van Enter, A.C.D.: Weak versus strong uniqueness of Gibbs measures: a regular short-range example. J. Phys. A **28**, L45–L47 (1995)
13. Cassandro, M., Ferrari, P.A., Merola, I., Presutti, E.: Geometry of contours and Peierls estimates in $d = 1$ Ising models with long range interactions. J. Math. Phys. **46**(5), 0533305 (2005)
14. Cassandro, M., Orlandi, E., Picco, P.: Phase transition in the 1D random field Ising model with long range interaction. Comm. Math. Phys. **288**, 731–744 (2009)
15. Cassandro, M., Merola, I., Picco, P., Rozikov, U.: One-dimensional Ising models with long range interactions: cluster expansion, phase-separating point. Comm. Math. Phys. **327**, 951–991 (2014)
16. Cassandro, M., Merola, I., Picco, P.: Phase separation for the long range one-dimensional Ising model. J. Stat. Phys. **167**(2), 351–382 (2017)
17. Chayes, J.T., Chayes, L., Sethna, J., Thouless, D.: A mean field spin glass with short range interactions. Comm. Math. Phys. **106**, 41–89 (1986)
18. Coquille, L., van Enter, A.C.D., Le Ny, A., Ruszel, W.M.: Absence of dobrushin states for $2d$ long-range Ising models. J. Stat. Phys. **172**, 1210–1222 (2018)
19. Cotar, C., Jahnel, B., Külske, C.: Extremal decomposition for random Gibbs measures: from general metastates to metastates on extremal random Gibbs measures. Electron. Commun. Probab. **18**, paper 95 (2018)
20. Dyson, F.J.: Existence of a phase transition in a one-dimensional Ising ferromagnet. Comm. Math. Phys. **12**, 91–107 (1969)
21. Endo, E.O., van Enter, A.C.D., Le Ny, A.: Paper in preparation
22. Fernández, R., Gallo, S., Maillard, G.: Regular g-measures are not always Gibbsian. El. Comm. Prob. **16**, 732–740 (2011)
23. Friedli, S., Velenik, Y.: Statistical Mechanics of Lattice Systems: A Concrete Mathematical Introduction. Cambridge University Press, Cambridge (2017)
24. Fröhlich, J., Spencer, T.: The phase transition in the one-dimensional Ising model with $1/r^2$ interaction energy. Comm. Math. Phys. **84**, 87–101 (1982)
25. Fröhlich, J., Zegarlinski, B.: The high-temperature phase of long-range spin glasses. Comm. Math. Phys. **110**, 121–155 (1987)
26. Fröhlich, J., Israel, R.B., Lieb, E.H., Simon, B.: Phase transitions and reflection positivity. I. General theory and long range lattice models. Comm. Math. Phys. **62**, 1–34 (1978)
27. Gandolfi, A., Newman, C.M., Stein, D.L.: Exotic states in long-range spin glasses. Comm. Math. Phys. **157**, 371–387 (1993)
28. Georgii, H.O.: Gibbs Measures and Phase Transitions. De Gruyter Studies in Mathematics, vol. 9, 2nd edn. De Gruyter, Berlin, 1988 (2011)
29. Iacobelli, G., Külske, C.: Metastates in finite-type mean-field models: visibility, invisibility, and random restoration of symmetry. J. Stat. Phys. **140**, 27–55 (2010)
30. Imbrie, J.: Decay of correlations in one-dimensional Ising model with $J_{ij} =| i - j |^{-2}$. Comm. Math. Phys. **85**, 491–515 (1982)
31. Imbrie, J., Newman, C.M.: An intermediate phase with slow decay of correlations in one-dimensional $1/ \mid x - y \mid^2$ percolation, Ising and Potts models. Comm. Math. Phys. **118**, 303–336 (1988)
32. Imry, Y., Ma, S.-K.: Random-field instability of the ordered state of continuous symmetry. Phys. Rev. Lett. **35**, 1399–1401 (1975)
33. Johansson, K.: Condensation of a one-dimensional lattice gas. Comm. Math. Phys. **141**, 41–61 (1991)
34. Kac, M., Thompson, C.J.: Critical behaviour of several lattice models with long-range interaction. J. Math. Phys. **10**, 1373–1386 (1968)

35. Külske, C.: Metastates in disordered mean-field models: random field and Hopfield models. J. Stat. Phys. **88**, 1257–1293 (1997)
36. Littin, J.: Work in progress
37. Littin, J., Picco, P.: Quasiadditive estimates on the Hamiltonian for the one-dimensional long-range Ising model. J. Math. Phys. **58**, 073301 (2017)
38. Marchetti, D., Sidoravicius, V., Vares, M.E.: Oriented percolation in one-dimensional $\frac{1}{|x-y|^2}$ percolation models. J. Stat. Phys. **139**, 941–959 (2010)
39. Newman, C.M.: Topics in Disordered Systems. Lecture Notes in Mathematics. Birkhäuser, Basel (1997)
40. Newman, C.M., Stein, D.L.: Multiple states and thermodynamic limits in short-range Ising spin-glass models. Phys. Rev. B **46**, 973–982 (1992)
41. Newman, C.M., Stein, D.L.: Metastate approach to thermodynamic chaos. Phys. Rev. B **55**, 5194–5211 (1997)
42. Newman, C.M., Stein, D.L.: The state(s) of replica symmetry breaking: mean field theories vs. short-ranged spin glasses. J. Stat. Phys. **106**, 213–244 (2002)
43. Newman, C.M., Stein, D.L.: Spin Glasses and Complexity. Princeton University Press, Princeton (2013)
44. Pétritis, D.: Equilibrium statistical mechanics of frustrated spin glasses: a survey of mathematical results. Ann. Inst. H. Poincaré, Phys. Théorique. **64**, 255–288 (1996)
45. Ruelle, D.: Thermodynamic Formalism, 2nd edn. Cambridge University Press, Cambridge (2004)
46. Talagrand, M.: Mean Field Models for Spin Glasses, vol. I. Basic Examples. Springer, Berlin (2011)
47. Talagrand, M.: Mean Field Models for Spin Glasses, vol. II. Advanced Replica Symmetry and Low Temperature. Springer, Berlin (2011)
48. van Enter, A.C.D.: Stiffness exponent, number of pure states, and Almeida-Thouless line in spin-glasses. J. Stat. Phys. **60**, 275–279 (1990)
49. van Enter, A.C.D., Fröhlich, J.: Absence of symmetry breaking for N-vector spin glass models in two dimensions. Comm. Math. Phys. **98**, 425–432 (1985)
50. van Enter, A.C.D., Griffiths, R.B.: The order parameter in a spin glass. Comm. Math. Phys. **90**, 319–327 (1983)
51. van Enter, A.C.D., Fernández, R., Sokal, A.D.: Regularity properties and pathologies of position-space R.G. transformations: scope and limitations of Gibbsian theory. J. Stat. Phys. **72**, 879–1167 (1993)
52. van Enter, A.C.D., Netocný, K., Schaap, H.: On the Ising model with random boundary condition. J. Stat. Phys. **118**, 997–1056 (2005)
53. van Enter, A.C.D., Netocný, K., Schaap, H.G.: Incoherent boundary conditions and metastates. IMS Lecture Notes Monogr. Ser. Dyn. Stoch. **48**, 144–153 (2006)
54. van Enter, A.C.D., Le Ny, A.: Decimation of the Dyson-Ising ferromagnet. Stoch. Proc. Appl. **127**, 3776–3791 (2017)
55. van Enter, A.C.D., Kimura, B., Ruszel, W.M., Spitoni, C.: Nucleation for one-dimensional long-range Ising models. J. Stat. Phys. **174**, 1327–1345 (2019)

Central Limit Theorems for a Driven Particle in a Random Medium with Mass Aggregation

Luiz Renato Fontes, Pablo Almeida Gomes, and Rémy Sanchis

Abstract We establish central limit theorems for the position and velocity of the charged particle in the mechanical particle model introduced by Fontes, Jordão Neves and Sidoravicius (2000).

Keywords Mass aggregation · Markovian approximation · Central limit theorem

AMS 2010 Mathematics Subject Classification 60K35, 60J27

1 Introduction

We revisit the $1d$ mechanical particle model introduced in [4], where we have a charged particle initially standing at the origin, subjected to an electric field, in an environment of initially standing neutral particles of unit mass. Each neutral particle has randomly either an elastic nature or an inelastic nature. With the first kind of neutral particle, the charged particle collides in a totally elastic fashion. And the collisions of the charged particle with the second kind of neutral particle is totally inelastic. The neutral particles do not interact amongst themselves. Both kinds of neutral particles are initially randomly placed in space.

One dimensional mechanical models have been treated quite often in the literature. We refer to [1, 7–9] for models with purely elastic interactions; the latter two papers, in the first of which the neutral particles have independent lifetimes, establish the existence of a limit velocity for the charged particle, as well as an invariance principle for its position. For models with purely inelastic interactions,

L. R. Fontes
Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, SP, Brasil
e-mail: lrfontes@usp.br

P. A. Gomes · R. Sanchis (✉)
Instituto de Ciências Exatas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil
e-mail: pabloag@ufmg.br; rsanchis@mat.ufmg.br

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_18

we cite [6]. Other references may be found in [4]. See also [2, 5] for somewhat more recent studies of related models of each kind.

In [4], a law of large numbers was proved for the instantaneous velocity of the charged particle. In this article, we derive central limit theorems for both the position and the instantaneous velocity of that particle, in a sense completing the result of [4]; see Final Remarks of [4].

Our approach is similar to that of [4], namely, we first prove CLT's for the corresponding objects of a modified process, where there are no recollisions. The results for the original process are established by showing that the differences between the actual and modified quantities are negligible in the relevant scales.

2 The Model and Results

We consider a system of infinitely many point like particles in the non-negative real semi-axis $[0, \infty)$. At time 0 the system is static, every particle has velocity 0. There is a distinguished particle of mass 2 initially at the origin; we will call it the *tracer particle (t.p.)* (referred to before as the charged particle). The remaining particles (referred to before as neutral particles) have mass 1.[1] Let $\{\xi_i\}_{i\in\mathbb{N}}$ denote a family of i.i.d. positive random variables, with an absolutely continuous distribution, and finite mean $\mathbb{E}\xi_1 = \mu < \infty$, representing the initial interparticle distances. In this way, $S_i = \xi_1 + \cdots + \xi_i$ denotes the position of the i-th particle initially in front of the t.p. at time 0. Moreover, given a parameter $p \in (0, 1]$, and a family $\{\eta_i\}_{i\in\mathbb{N}}$ of i.i.d. Bernoulli random variables with success probability p, we say that the i-th particle is *sticky* if $\eta_i = 1$ and is *elastic* if $\eta_i = 0$. We assume $\{\xi_i\}_{i\in\mathbb{N}}$ and $\{\eta_i\}_{i\in\mathbb{N}}$ to be independent of one another.

A constant positive force F is turned on at time 0, and kept on. It acts solely on the tracer particle, producing in it an accelerated motion to the right. Collisions will thus take place in the system; we assume they occur only when involving the t.p., and suppose that all other particles do not interact among themselves. If at an instant $t > 0$, the t.p. collides with a sticky particle, then this is a perfectly inelastic collision, meaning that, upon collision, momentum is conserved and the energy of the two particle system is minimum, which in turn means that the t.p. incorporates the sticky particle, along with its mass, and the new velocity of the t.p. becomes (immediately after time t)

$$V(t^+) = \frac{M_t}{M_t + 1} V_t, \tag{1}$$

[1]The distinction of the initial mass of the t.p. with respect to the other particles, absent in [4], is for convenience only; any positive initial mass for the t.p. would not change our results, but values 1 or below would require unimportant complications in our arguments.

where V_t and M_t are respectively the velocity and mass of t.p. at time t. However, if the t.p. collides with an elastic particle which is moving at velocity v at the time of the collision, say t, then we have a perfectly elastic collision, where energy and momentum are preserved, and in this case, immediately after time t, the t.p. and the elastic particle velocities become, respectively,

$$V(t^+) = \frac{M_t - 1}{M_t + 1} V_t + \frac{2}{M_t + 1} v \text{ and}$$
$$v' = \frac{2M_t}{M_t + 1} V_t - \frac{M_t - 1}{M_t + 1} v, \tag{2}$$

where V_t and M_t are as above.

For $t \geq 0$, let V_t and Q_t denote the velocity and position of the t.p. at time t, respectively. As argued in [4], the stochastic process $(V_t, Q_t)_{t \geq 0}$ is well defined—see the discussion at the end of Section 2 of [4]; in particular there a.s. are no multiple collisions or infinitely many recollisions in finite time intervals—, and is determined by $\{\xi_i, \eta_i \ ; \ i \in \mathbb{N}\}$. Therefore we consider the product sample space $\Omega = \{(0, \infty) \times \{0, 1\}\}^{\mathbb{N}}$, and the usual product Borel σ-algebra, and the product probability measure $\mathbb{P} := \prod_{i \geq 1} [\mathbb{P}_{\xi_i} \otimes \mathbb{P}_{\eta_i}]$, where for $i \geq 1$, $\mathbb{P}_{\xi_i}$ and $\mathbb{P}_{\eta_i}$ denote the probability measures of ξ_i and η_i. We will make repeatedly make use of the notation

$$\bar{\xi}_i = \xi_i - \mu, \ \bar{\eta}_i = \eta_i - p.$$

From [4], we know that $\mathbb{P}$-almost surely, the velocity of the t.p. converges to a(n explicit) limit. More precisely, we have the following result.

Theorem 1 *The stochastic process* $(V_t, Q_t)_{t \geq 0}$ *is such that*

$$\lim_{t \to \infty} V_t = \sqrt{\frac{F\mu}{2 - p}} \ \mathbb{P} - a.s.$$

From now on we denote the limit velocity $\sqrt{F\mu/(2-p)}$ by V_L. The purpose of this paper is to show that the velocity V_t and position Q_t of the tracer particle satisfy central limit theorems. Our main results are as follows (where "$\Longrightarrow$" denotes convergence in distribution).

Theorem 2 *Let* $Var(\xi_1) = \sigma^2 < \infty$. *Then, as* $t \to \infty$,

$$\frac{Q_t - tV_L}{\sqrt{t}} \Longrightarrow \mathcal{N}(0, \sigma_q^2),$$

where $\sigma_q > 0$.

Theorem 3 *Let* $Var(\xi_1) = \sigma^2 < \infty$. *Then, as* $t \to \infty$,

$$\sqrt{t}(V_t - V_L) \Longrightarrow \mathcal{N}(0, \sigma_v^2),$$

where $\sigma_v > 0$.

3 Central Limit Theorems in a Modified Process

As mentioned in the Introduction, we first prove central limit theorem analogues of Theorems 2 and 3 for a modified process in which, when an elastic particle collides with the t.p., the elastic particle is annihilated and disappears from the system, and the velocity of the t.p. changes according to the formula (2), while collisions between the t.p. and sticky particles remain as in the original model. We denote the modified stochastic process by $(\bar{V}(t), \bar{Q}(t))_{t\geq 0}$, where $\bar{V}(t)$ and $\bar{Q}(t)$ are respectively the velocity and position of the t.p. in the modified system at time t.

In the modified model, for $i \geq 1$, the t.p. collides with the i-th particle only in the initial position of the latter particle, given by S_i; let us denote the instant when that collision occurs by $\bar{t}_i$, i.e., $\bar{Q}(\bar{t}_i) = S_i$. In this way, we can compute the $i-$th collision incoming and outgoing velocities $\bar{V}(\bar{t}_i)$ and $\bar{V}(\bar{t}_i^+)$, respectively, as follows. First note that, according the formulas (1) and (2), we have the following relations

$$\text{(a) } \bar{V}^2(\bar{t}_i \ ) = \bar{V}^2(\bar{t}_{i-1}^+) + \frac{2F\xi_i}{M(\bar{t}_i)};$$

$$\text{(b) } \bar{V}^2(\bar{t}_i^+) = \bar{V}^2(\bar{t}_i)\left[\frac{M(\bar{t}_i) + (\eta_i - 1)}{M(\bar{t}_i) + 1}\right]^2,$$

where $M(\bar{t}_i) = 2 + \sum_{l=1}^{i-1} \eta_l$.

Iterating this relations, we get for $i = 1, 2, \ldots$, that

$$\bar{V}^2(\bar{t}_i^+) = \sum_{j=1}^{i}\left[\frac{2F\xi_j}{M(\bar{t}_j)}\prod_{k=j}^{i}\left(\frac{M(\bar{t}_k) + (\eta_k - 1)}{M(\bar{t}_k) + 1}\right)^2\right]. \tag{3}$$

In [4], it is proved that, almost surely,

$$\lim_{t\to\infty} \bar{V}(t) = V_L.$$

Let us at this point set some notation. Given two random sequences $\{X_n\}_{n\in\mathbb{N}}$ and $\{Y_n\}_{n\in\mathbb{N}}$, we write $X_n = O(Y_n)$ if there almost surely exists $C > 0$, which may be a (proper) random variable, but does not depend on n, such that $|X_n| \leq CY_n$ for every $n \in \mathbb{N}$. And we say $X_n = o(Y_n)$ if X_n/Y_n almost surely converges to 0 as

$n \to \infty$. For simplicity, along the rest of the paper we denote $M(\bar{t}_i)$ by M_i. Notice that $M_1 = 2$ and $M_i = 2 + \sum_{k=1}^{i-1} \eta_k$, $i \geq 2$.

To obtain the central limit theorems for the modified process, we start with an estimate for the random term

$$X_{i,j} := \frac{1}{M_j} \prod_{k=j}^{i} \left(\frac{M_k + (\eta_k - 1)}{M_k + 1} \right)^2, \ 1 \leq j \leq i \text{ and } i \in \mathbb{N}. \tag{4}$$

Given $\varepsilon > 0$, for each $m \in \mathbb{N}$ we define the event

$$A_{m,\varepsilon} = \left\{ X_{i,j} \in \left((1-\varepsilon) \frac{j^{\zeta-1}}{p i^{\zeta}}, (1+\varepsilon) \frac{j^{\zeta-1}}{p i^{\zeta}} \right), \ \forall (i,j) \text{ such that } m \leq j \leq i \right\}, \tag{5}$$

where $\zeta := 2(2-p)/p$.

Lemma 1 *Let $X_{i,j}$ be as in (4), and $A_{m,\varepsilon}$ as in (5), where $\varepsilon > 0$ is otherwise arbitrary. Then we have that*

$$\lim_{m \to \infty} \mathbb{P}\left(A_{m,\varepsilon}\right) = 1.$$

Proof We first Taylor-expand the logarithm to write

$$\begin{aligned} \prod_{k=j}^{i} \left(\frac{M_k + (\eta_k - 1)}{M_k + 1} \right)^2 &= \exp \left\{ 2 \sum_{k=j}^{i} \log \left(1 - \frac{2 - \eta_k}{M_k + 1} \right) \right\} \\ &= \exp \left\{ -2 \sum_{k=j}^{i} \left[\frac{2-p}{M_k + 1} - \frac{\bar{\eta}_k}{M_k + 1} \right] + O\left(\sum_{k=j}^{i} \left(\frac{2 - \eta_k}{M_k + 1} \right)^2 \right) \right\}. \end{aligned} \tag{6}$$

Given $\delta > 0$, $m \in \mathbb{N}$, let $B_m^\delta = \left\{ M_j \in ((1-\delta)pj, (1+\delta)pj), \ \forall j \geq m \right\}$. It follows from the Law of Large Numbers that $P(B_m^\delta) \to 1$ a.s. as $m \to \infty$. In B_m^δ, we have

$$\sum_{k=1}^{\infty} \left(\frac{2 - \eta_k}{M_k + 1} \right)^2 \leq \sum_{k=1}^{m-1} \left(\frac{2 - \eta_k}{M_k + 1} \right)^2 + \frac{4}{p^2 (1-\delta)^2} \sum_{k=m}^{\infty} \frac{1}{k^2} < \infty. \tag{7}$$

Note also that

$$\sum_{k=j}^{i} \frac{1}{M_k + 1} = \sum_{k=j}^{i} \left(\frac{1}{M_k + 1} - \frac{1}{pk} \right) + \frac{1}{p} \left[\sum_{k=j}^{i} \frac{1}{k} - \int_j^i \frac{1}{x} dx \right] + \frac{1}{p} \int_j^i \frac{1}{x} dx. \tag{8}$$

Clearly the second term at the right-hand side of (8) goes to 0 as j and i goes to infinity. Let now $C_m = \left\{|M_j + 1 - jp| \le j^{2/3},\ \forall j \ge m\right\}$. It follows from Law of the Iterated Logarithm that $\lim_{m\to\infty} \mathbb{P}(C_m) = 1$. In $B_m^\delta \cap C_m$ we have

$$\left|\sum_{k=1}^{\infty}\left(\frac{1}{M_k+1} - \frac{1}{pk}\right)\right| \le \left|\sum_{k=1}^{m-1}\left(\frac{1}{M_k+1} - \frac{1}{pk}\right)\right| + \frac{1}{p(1-\delta)}\sum_{k=m}^{\infty}\frac{|M_k+1-kp|}{k^2}$$
$$\le \left|\sum_{k=1}^{m-1}\left(\frac{1}{M_k+1} - \frac{1}{pk}\right)\right| + \sum_{k=m}^{\infty}\frac{1}{k^{4/3}} < \infty. \quad (9)$$

We also write

$$\sum_{k=1}^{\infty}\frac{\bar{\eta}_k}{M_k+1} = \sum_{k=1}^{\infty}\left[\bar{\eta}_k\left(\frac{1}{M_k+1} - \frac{1}{pk}\right)\right] + \sum_{k=1}^{\infty}\frac{\bar{\eta}_k}{pk}. \quad (10)$$

We may apply Kolmogorov's Two-series Theorem to obtain that $\sum_{k=1}^{\infty}\bar{\eta}_k/k$ converges a.s., and proceeding as in the estimation leading to (9), we may conclude that the first term in the right-hand side of (10) is also convergent in the event $B_m^\delta \cap C_m$.

To conclude, due to (6)–(10), taking $\delta > 0$ sufficient small and m sufficient large, we have that, in the event $B_m^\delta \cap C_m$,

$$\prod_{k=j}^{i}\left(\frac{M_k + (\eta_k - 1)}{M_k+1}\right)^2 \in (1 \pm \varepsilon)\exp\left\{-\zeta\int_j^i \frac{1}{x}dx\right\}. \quad (11)$$

Recalling now the definition of $X_{i,j}$ and $A_{m,\varepsilon}$ in (4) and (5), respectively, we have that (11) implies that $B_m^\delta \cap C_m \subset A_{m,\varepsilon}$, and the result follows. □

We now turn our attention to $S_n - \bar{t}_n V_L$, for which we will prove a central limit theorem, as a step to establish Theorem 2, as follows.

Proposition 1 *Let $Var(\xi_1) = \sigma^2 < \infty$. Then, as $n \to \infty$,*

$$\frac{S_n - \bar{t}_n V_L}{\sqrt{n}} \Longrightarrow \mathcal{N}(0, \hat{\sigma}_q^2), \quad (12)$$

where $\hat{\sigma}_q > 0$.

The proof of this result consists of a number of steps which take most of this section.

From elementary physics relations, the time taken for the t.p. to go from S_{i-1} to S_i is given by

$$\bar{t}_i - \bar{t}_{i-1} = \frac{\bar{V}(\bar{t}_i) - \bar{V}(\bar{t}_{i-1}^+)}{F/M_i} = \frac{2\xi_i\left(\bar{V}(\bar{t}_i) - \bar{V}(\bar{t}_{i-1}^+)\right)}{2\xi_i F/M_i} = \frac{2\xi_i}{\bar{V}(\bar{t}_i) + \bar{V}(\bar{t}_{i-1}^+)}.$$

Thus, we may write

$$
\begin{aligned}
S_n - \bar{t}_n V_L &= \sum_{i=1}^{n} \left[\xi_i \left(1 - \frac{2V_L}{\bar{V}(\bar{t}_i) + \bar{V}(\bar{t}_{i-1}^{+})} \right) \right] \\
&= \sum_{i=1}^{n} \left[\xi_i \left(\frac{\bar{V}(\bar{t}_i) + \bar{V}(\bar{t}_{i-1}^{+}) - 2V_L}{\bar{V}(\bar{t}_i) + \bar{V}(\bar{t}_{i-1}^{+})} \right) \right] \\
&= \sum_{i=1}^{n} \left[\frac{2\xi_i \left(\bar{V}(\bar{t}_{i-1}^{+}) - V_L \right)}{\bar{V}(\bar{t}_i) + \bar{V}(\bar{t}_{i-1}^{+})} \right] + \sum_{i=1}^{n} \left[\xi_i \left(\frac{\bar{V}(\bar{t}_i) - \bar{V}(\bar{t}_{i-1}^{+})}{\bar{V}(\bar{t}_i) + \bar{V}(\bar{t}_{i-1}^{+})} \right) \right]. \qquad (13)
\end{aligned}
$$

Note that

$$
\frac{\bar{V}(\bar{t}_i) - \bar{V}(\bar{t}_{i-1}^{+})}{\bar{V}(\bar{t}_i) + \bar{V}(\bar{t}_{i-1}^{+})} = \frac{2F\xi_i}{M_i \left(\bar{V}(\bar{t}_i) + \bar{V}(\bar{t}_{i-1}^{+}) \right)^2}. \qquad (14)
$$

Since $\bar{V}(\bar{t}_i) + \bar{V}(\bar{t}_{i-1}^{+})$ converges to the constant $2V_L$, the Law of Large Numbers and (14) imply that

$$
\sum_{i=1}^{n} \left[\xi_i \left(\frac{\bar{V}(\bar{t}_i) - \bar{V}(\bar{t}_{i-1}^{+})}{\bar{V}(\bar{t}_i) + \bar{V}(\bar{t}_{i-1}^{+})} \right) \right] = O\left(\sum_{i=1}^{n} \frac{\xi_i^2}{i} \right). \qquad (15)
$$

Let $\tilde{S}_0 = 0$ and $\tilde{S}_k = \sum_{i=1}^{k} \xi_i^2$, $k \geq 1$. Assuming $\mathbb{E}\xi_1^2 < \infty$, we have that

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\xi_i^2}{i} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\tilde{S}_i - \tilde{S}_{i-1}}{i} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n-1} \frac{\tilde{S}_i}{i(i+1)} + \frac{\tilde{S}_n}{n^{3/2}} = o(1). \qquad (16)
$$

Noticing that $\bar{V}(\bar{t}_i) = \bar{V}(\bar{t}_{i-1}^{+}) + 2F\xi_i \left(\bar{V}(\bar{t}_i) + \bar{V}(\bar{t}_{i-1}^{+}) \right) / M_i$, we find that

$$
\begin{aligned}
\frac{\bar{V}(\bar{t}_{i-1}^{+}) - V_L}{\bar{V}(\bar{t}_i) + \bar{V}(\bar{t}_{i-1}^{+})} &= \frac{\bar{V}(\bar{t}_{i-1}^{+}) - V_L}{2V_L} + \left[\frac{\bar{V}(\bar{t}_{i-1}^{+}) - V_L}{\bar{V}(\bar{t}_i) + \bar{V}(\bar{t}_{i-1}^{+})} - \frac{\bar{V}(\bar{t}_{i-1}^{+}) - V_L}{2V_L} \right] \\
&= \frac{\bar{V}(\bar{t}_{i-1}^{+}) - V_L}{2V_L} + \frac{\left(\bar{V}(\bar{t}_{i-1}^{+}) - V_L \right) \left(2V_L - 2\bar{V}(\bar{t}_{i-1}^{+}) \right)}{2V_L \left(\bar{V}(\bar{t}_i) + \bar{V}(\bar{t}_{i-1}^{+}) \right)} - \frac{2F\xi_i}{2V_L M_i}. \qquad (17)
\end{aligned}
$$

In particular,

$$\sum_{i=1}^{n}\left[\frac{2\xi_i\left(\bar V(\bar t_{i-1}^{+})-V_L\right)}{\bar V(\bar t_i)+\bar V(\bar t_{i-1}^{+})}\right]=$$
$$\sum_{i=1}^{n}\left[\frac{\xi_i\left(\bar V(\bar t_{i-1}^{+})-V_L\right)}{V_L}\right]+O\left(\sum_{i=1}^{n}\left[\xi_i\left(\bar V(\bar t_{i-1}^{+})-V_L\right)^2\right]\right)+O\left(\sum_{i=1}^{n}\frac{\xi_i^2}{i}\right). \tag{18}$$

Proceeding in an analogous way, we obtain that

$$\sum_{i=1}^{n}\left[\frac{\xi_i\left(\bar V(\bar t_{i-1}^{+})-V_L\right)}{V_L}\right]=$$
$$\sum_{i=1}^{n}\left[\frac{\xi_i\left(\bar V(\bar t_{i-1}^{+})^2-V_L^2\right)}{2V_L^2}\right]+O\left(\sum_{i=1}^{n}\left[\xi_i\left(\bar V(\bar t_{i-1}^{+})-V_L\right)^2\right]\right). \tag{19}$$

To simplify notation, for each $i\in\mathbb{N}$, we henceforth denote $\bar V(\bar t_i^{+})$ simply by $\bar V_i$. The following lemma will be useful now; we postpone its proof till the end of this section.

Lemma 2 *Let* $Var(\xi_1)=\sigma^2<\infty$ *and let* $\epsilon>0$. *The velocities* $\{\bar V_i\}_{i\in\mathbb{N}}$ *are such that* $\bar V_i-V_L=o(1/i^{1/2-\epsilon})$. *In particular,*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\xi_i\left(\bar V_{i-1}-V_L\right)^2\right]=o(1).$$

By (13)–(19) and Lemma 2, in order to establish Proposition 1 it is enough to show that as $n\to\infty$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i(\bar V_{i-1}^2-V_L^2)\Longrightarrow\mathcal{N}(0,\tilde\sigma_q^2), \tag{20}$$

for some $\tilde\sigma_q>0$; we then of course have $\hat\sigma_q=\tilde\sigma_q/(2V_L^2)$. For that, the strategy we will follow is to expand the expression on the left of (20) into several terms, one of which depends only on the interparticle distances $\{\xi_i\}_{i\in\mathbb{N}}$, another one depending only on the stickiness indicator random variables $\{\eta_k\}_{k\in\mathbb{N}}$; for each of those terms we can apply Lindeberg-Feller's Central Limit Theorem; upon showing that the remaining terms are negligible, the result follows.

Recalling that $\zeta = 2(2-p)/p$, (3) and (4), we start with

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\xi_{i+1}(\bar{V}_i^2 - V_L^2)\right] =$$
$$\frac{2F}{\sqrt{n}}\sum_{i=1}^{n}\left[\xi_{i+1}\sum_{j=1}^{i}\left(\xi_j X_{i,j} - \mu\frac{j^{\zeta-1}}{pi^{\zeta}}\right)\right] +$$
$$\frac{2F\mu}{p\sqrt{n}}\sum_{i=1}^{n}\left[\xi_{i+1}\left(\frac{1}{i}\sum_{j=1}^{i}\left(\frac{j}{i}\right)^{\zeta-1} - \int_0^1 x^{\zeta-1}dx\right)\right]. \quad (21)$$

The term on the left of expression within parentheses in the second term on the right hand side of (21) is a Riemann sum for the term to its right; we conclude that the full expression within parenthesis on the right hand side of (21) is an $O(1/i)$, and we may thus conclude that the second term on the right-hand side of (21) is an $o(1)$, and proceed by dropping that term and focusing on the first term, which we write as follows.

$$\frac{2F}{\sqrt{n}}\sum_{i=1}^{n}\left[\xi_{i+1}\sum_{j=1}^{i}\left(\xi_j X_{i,j} - \mu\frac{j^{\zeta-1}}{pi^{\zeta}}\right)\right] =$$
$$\frac{2F}{\sqrt{n}}\sum_{i=1}^{n}\left[\bar{\xi}_{i+1}\sum_{j=1}^{i}\left(\xi_j X_{i,j} - \mu\frac{j^{\zeta-1}}{pi^{\zeta}}\right)\right] + \frac{2F\mu}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{i}\left(\xi_j X_{i,j} - \mu\frac{j^{\zeta-1}}{pi^{\zeta}}\right)$$
$$:= V_n + W_n. \quad (22)$$

Now writing

$$\sum_{j=1}^{i}\left(\xi_j X_{i,j} - \mu\frac{j^{\zeta-1}}{pi^{\zeta}}\right) = \sum_{j=1}^{i}\bar{\xi}_j X_{i,j} + \mu\sum_{j=1}^{i}\left(X_{i,j} - \frac{j^{\zeta-1}}{pi^{\zeta}}\right),$$

V_n given in (22) becomes

$$V_n = \frac{2F}{\sqrt{n}}\sum_{i=1}^{n}\left[\bar{\xi}_{i+1}\sum_{j=1}^{i}\left(\xi_j X_{i,j} - \mu\frac{j^{\zeta-1}}{pi^{\zeta}}\right)\right] =$$
$$\frac{2F}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{i}\bar{\xi}_{i+1}\bar{\xi}_j X_{i,j} + \frac{2F\mu}{\sqrt{n}}\sum_{i=1}^{n}\left[\bar{\xi}_{i+1}\sum_{j=1}^{i}\left(X_{i,j}(\omega) - \frac{j^{\zeta-1}}{pi^{\zeta}}\right)\right]$$
$$=: V_{1,n} + V_{2,n}. \quad (23)$$

We will show in Lemmas 6 and 7 below that $V_{1,n}$ and $V_{2,n}$ are negligible.

Analogously, W_n given in (22) becomes

$$W_n = \frac{2F\mu}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{i} \left(\xi_j X_{i,j} - \mu \frac{j^{\zeta-1}}{pi^{\zeta}} \right) =$$

$$\frac{2F\mu}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{i} \bar{\xi}_j X_{i,j} + \frac{2F\mu^2}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{i} \left(X_{i,j} - \frac{j^{\zeta-1}}{pi^{\zeta}} \right)$$

$$=: W_{1,n} + W_{2,n}, \tag{24}$$

and $W_{1,n}$ is further broken down into

$$\begin{aligned} W_{1,n} &= \frac{2F\mu}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{i} \bar{\xi}_j X_{i,j} \\ &= \frac{2F\mu}{p\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{i} \frac{j^{\zeta-1}}{i^{\zeta}} \bar{\xi}_j + \frac{2F\mu}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{i} \bar{\xi}_j \left(X_{i,j} - \frac{j^{\zeta-1}}{pi^{\zeta}} \right) \\ &=: W_{3,n} + W_{4,n}. \end{aligned} \tag{25}$$

One may readily verify the conditions of Lindeberg-Feller's CLT to obtain

Lemma 3 *Let* $Var(\xi_1) = \sigma^2 < \infty$. *For* $1 \le j \le n$, *set* $a_{j,n} = j^{\zeta-1} \sum_{i=j}^{n} \frac{1}{i^{\zeta}}$. *Then, as* $n \to \infty$,

$$W_{3,n} = \frac{2F\mu}{p\sqrt{n}} \sum_{j=1}^{n} a_{j,n} \bar{\xi}_j \Longrightarrow \mathcal{N}(0, \sigma_w^2),$$

where $\sigma_w = \frac{2F\mu}{p\sqrt{\zeta}} \sigma$.

In Lemma 8 below we show that $W_{4,n}$ is negligible.

Let us now focus on $W_{2,n}$. To alleviate notation, for each $1 \le j \le i$, set

$$Y_{i,j} = \log \left[\prod_{k=j}^{i} \left(\frac{M_k + (\eta_k - 1)}{M_k + 1} \right)^2 \right] = 2 \sum_{k=j}^{i} \log \left(1 - \frac{2 - \eta_k}{M_k + 1} \right), \tag{26}$$

thus $X_{i,j} = e^{Y_{i,j}}/M_j$, and therefore,

$$W_{2,n} = \frac{2F\mu^2}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{i} \left(X_{i,j} - \frac{j^{\zeta-1}}{pi^{\zeta}} \right) =$$

$$\frac{2F\mu^2}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{i} \left[\left(\frac{1}{M_j} - \frac{1}{pj} \right) \frac{j^{\zeta}}{i^{\zeta}} \right] + \frac{2F\mu^2}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{i} \left[\left(\frac{1}{M_j} - \frac{1}{pj} \right) \left(e^{Y_{i,j}} - \frac{j^{\zeta}}{i^{\zeta}} \right) \right] +$$

$$\frac{2F\mu^2}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{i} \left[\frac{1}{pj} \left(e^{Y_{i,j}} - \frac{j^{\zeta}}{i^{\zeta}} \right) \right] =: Z_{1,n} + Z_{2,n} + Z_{3,n}. \tag{27}$$

Lemma 4 *$Z_{2,n}$, as defined in (27), is an $o(1)$.*

Proof Note that, as defined in (26) and (27),

$$\begin{aligned} |Z_{2,n}| &= \left| \frac{2F\mu^2}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{i} \left[\left(\frac{1}{M_j} - \frac{1}{pj} \right) \left(e^{Y_{i,j}} - \frac{j^{\zeta}}{i^{\zeta}} \right) \right] \right| \\ &= \left| \frac{2F\mu^2}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{i} \left[\frac{j^{\zeta}}{i^{\zeta}} \left(\frac{1}{M_j} - \frac{1}{pj} \right) \left(\exp\left\{ Y_{i,j} + \zeta \int_j^i \frac{1}{x} dx \right\} - 1 \right) \right] \right| \\ &\leq \frac{2F\mu^2}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{i} \left[\frac{j^{\zeta}}{i^{\zeta}} \left| \frac{1}{M_j} - \frac{1}{pj} \right| \left| Y_{i,j} + \zeta \int_j^i \frac{1}{x} dx \right| \right]. \end{aligned} \tag{28}$$

For each $i \geq j \geq 1$, we define

$$R_{i,j} = Y_{i,j} + \zeta \int_j^i x^{-1} dx. \tag{29}$$

It follows from (28) that

$$|Z_{2,n}| = O\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{i} \left[\frac{j^{\zeta}}{i^{\zeta}} \left| \frac{1}{M_j} - \frac{1}{pj} \right| \left| R_{i,j} \right| \right] \right). \tag{30}$$

As we see in (6) and (26), $R_{i,j}$ can be written as

$$R_{i,j} = \zeta \int_j^i x^{-1}dx - 2\sum_{k=j}^{i} \frac{2-p}{M_k+1} + 2\sum_{k=j}^{i} \frac{\bar{\eta}_k}{M_k+1} + O\left(\sum_{k=j}^{i}\left(\frac{2-\eta_k}{M_k+1}\right)^2\right) =$$

$$\zeta\left[\int_j^i \frac{1}{x}dx - \sum_{k=j}^{i}\frac{1}{k}\right] + \sum_{k=j}^{i}\left(\frac{\zeta}{k} - \frac{2(2-p)}{M_k+1}\right) + 2\sum_{k=j}^{i}\left[\frac{\bar{\eta}_k}{M_k+1} - \frac{\bar{\eta}_k}{p(k-1)+3}\right] +$$

$$2\sum_{k=j}^{i}\frac{\bar{\eta}_k}{p(k-1)+3} + O\left(\sum_{k=j}^{i}\left(\frac{2-\eta_k}{M_k+1}\right)^2\right) := R_{i,j}^{(1)} + \cdots + R_{i,j}^{(5)}. \tag{31}$$

One readily checks by elementary deterministic estimation that for all $i \geq j \geq 1$, $|R_{i,j}^{(1)}|$ can be bounded above by $1/j$.

Let now $0 < \delta < 1/4$ be fixed. The Law of Large Numbers and the Law of the Iterated Logarithm, there a.s. exists $j_0 \in \mathbb{N}$ such that $|R_{i,j}^{(2)}|$, $|R_{i,j}^{(3)}|$ and $|R_{i,j}^{(5)}|$ are bounded above by $1/j^{1/2-\delta}$, for every $i \geq j \geq j_0$.

To study $|R_{i,j}^{(4)}|$, we apply Hoeffding's Inequality to obtain, for every $i \geq j \geq 1$,

$$\mathbb{P}\left(\left|\sum_{k=j}^{i}\frac{\bar{\eta}_k}{p(k-1)+3}\right| \geq \frac{1}{j^{1/2-\delta}}\right) \leq \exp\left\{-2\Big/\left(j^{1-2\delta}\sum_{k=j}^{i}\frac{1}{(p(k-1)+3)^2}\right)\right\}. \tag{32}$$

We next apply a variation of Lévy's Maximal Inequality, namely Proposition 1.1.2 in [3], combined with (32), to get that

$$\begin{aligned}
\mathbb{P}&\left(\max_{i\geq j}\left|\sum_{k=j}^{i}\frac{\bar{\eta}_k}{p(k-1)+3}\right| \geq \frac{3}{j^{1/2-\delta}}\right) \\
&= \lim_{n\to\infty}\mathbb{P}\left(\max_{j\leq i\leq n}\left|\sum_{k=j}^{i}\frac{\bar{\eta}_k}{p(k-1)+3}\right| \geq \frac{3}{j^{1/2-\delta}}\right) \\
&\leq 3\lim_{n\to\infty}\max_{j\leq i\leq n}\mathbb{P}\left(\left|\sum_{k=j}^{i}\frac{\bar{\eta}_k}{p(k-1)+3}\right| \geq \frac{1}{j^{1/2-\delta}}\right) \\
&\leq 3\exp\left\{-2\Big/\left(j^{1-2\delta}\sum_{k=j}^{\infty}\frac{1}{(p(k-1)+3)^2}\right)\right\}.
\end{aligned} \tag{33}$$

Since the latter term is summable, we conclude that almost surely exists $j_0 \in \mathbb{N}$ such that $|R_{i,j}^{(4)}| \leq 3/j^{1/2-\delta}$, for every $i \geq j \geq j_0$. Collecting all the bounds, we find that a.s.

$$|R_{i,j}| \leq |R_{i,j}^{(1)}| + \cdots + |R_{i,j}^{(5)}| < 3/j^{1/2-\delta} \tag{34}$$

for every $i \geq j$ sufficiently large. Recalling that $M_j = 2 + \sum_{l=1}^{j-1} \eta_l$, we have, as consequence of the Law of the Iterated Logarithm and the Law of Large Numbers, that $|1/M_j - 1/(pj)| = o(1/j^{3/2-\delta})$, and the result follows from (30). □

It follows from (33) that $R_{i,j}$ is uniformly bounded in i, j by a proper random variable. We may thus write

$$\begin{aligned} Z_{3,n} &= \frac{2F\mu^2}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{i} \left[\frac{1}{pj} \left(e^{Y_{i,j}} - \frac{j^\zeta}{i^\zeta} \right) \right] = \frac{2F\mu^2}{p\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{i} \left[\frac{j^{\zeta-1}}{i^\zeta} \left(e^{R_{i,j}} - 1 \right) \right] \\ &= \frac{2F\mu^2}{p\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{i} \frac{j^{\zeta-1}}{i^\zeta} R_{i,j} + O\left(\frac{2F\mu^2}{p\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{i} \frac{j^{\zeta-1}}{i^\zeta} R_{i,j}^2 \right) =: Z'_{3,n} + \tilde{Z}_{3,n}. \end{aligned} \tag{35}$$

Since, almost surely, for every $i \geq j$ sufficiently large, we have the bound $|R_{i,j}^{(1)}| + |R_{i,j}^{(5)}| \leq 1/j^{2/3}$, it follows that

$$\frac{2F\mu^2}{p\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{i} \left[\frac{j^{\zeta-1}}{i^\zeta} \left(R_{i,j}^{(1)} + R_{i,j}^{(5)} \right) \right] = o(1).$$

Considering only the term $R_{i,j}^{(2)}$ of $R_{i,j}$ in (31), its contribution to $Z'_{3,n}$ in (35) is

$$\begin{aligned} &2(2-p)\frac{2F\mu^2}{p\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{i} \left[\frac{j^{\zeta-1}}{i^\zeta} \sum_{k=j}^{i} \left(\frac{1}{pk} - \frac{1}{M_k+1} \right) \right] = \\ &\zeta \frac{2F\mu^2}{\sqrt{n}} \sum_{i=1}^{n} \sum_{k=1}^{i} \left[\left(\frac{1}{pk} - \frac{1}{M_k+1} \right) \sum_{j=1}^{k} \frac{j^{\zeta-1}}{i^\zeta} \right] = \\ &\frac{2F\mu^2}{\sqrt{n}} \sum_{i=1}^{n} \sum_{k=1}^{i} \left[\left(\frac{1}{pk} - \frac{1}{M_k} \right) \frac{k^\zeta}{i^\zeta} \right] + o(1) = -Z_{1,n} + o(1), \end{aligned} \tag{36}$$

where $Z_{1,n}$ is defined in (27). We may remark at this point that combining (36) and (27) drops $Z_{1,n}$ out of the overall computation.

Let us now estimate the contribution of $R^{(3)}_{i,j}$ to $Z'_{3,n}$ in (35), recalling that $M_k = 2 + \sum_{l=1}^{k-1} \eta_l$ and setting $\bar{M}_k = -\sum_{l=1}^{k} \bar{\eta}_l$:

$$\frac{4F\mu^2}{p\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{i} \left[\frac{j^{\zeta-1}}{i^{\zeta}} \sum_{k=j}^{i} \left(\frac{\bar{\eta}_k}{M_k+1} - \frac{\bar{\eta}_k}{p(k-1)+3} \right) \right] =$$

$$\frac{4F\mu^2}{p\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{i} \left[\frac{j^{\zeta-1}}{i^{\zeta}} \sum_{k=j}^{i} \frac{\bar{\eta}_k \bar{M}_{k-1}}{(M_k+1)(p(k-1)+3)} \right] =$$

$$\frac{4F\mu^2}{p\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{i} \left[\frac{j^{\zeta-1}}{i^{\zeta}} \sum_{k=j}^{i} \frac{\bar{\eta}_k \bar{M}_{k-1}}{(p(k-1)+3)^2} \right] +$$

$$\frac{4F\mu^2}{p\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{i} \left[\frac{j^{\zeta-1}}{i^{\zeta}} \sum_{k=j}^{i} \left(\frac{\bar{\eta}_k \bar{M}_{k-1}}{p(k-1)+3} \left(\frac{1}{M_k+1} - \frac{1}{p(k-1)+3} \right) \right) \right]$$

$$=: Z_{5,n} + Z_{6,n}. \qquad (37)$$

Let us fix $0 < \alpha < 1/2$; the Law of the Iterated Logarithm and the Law of Large Numbers give us that

$$\left| \frac{\bar{\eta}_k \bar{M}_{k-1}}{p(k-1)+3} \left(\frac{1}{M_k+1} - \frac{1}{p(k-1)+3} \right) \right| = o\left(\frac{1}{k^{2-\alpha}} \right).$$

Since $0 < \alpha < 1/2$, it follows that $Z_{6,n} = o(1)$.

We will study the asymptotic behavior of $Z_{5,n}$ in Lemma 9.

We now estimate the contribution of $R^{(3)}_{i,j}$ to $Z'_{3,n}$ in (35):

$$\frac{4F\mu^2}{p\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{i-1} \left[\frac{j^{\zeta-1}}{i^{\zeta}} \sum_{k=j}^{i-1} \frac{\bar{\eta}_k}{k} \right] = \frac{4F\mu^2}{p\sqrt{n}} \sum_{i=1}^{n} \sum_{k=1}^{i-1} \left[\frac{\bar{\eta}_k}{k} \sum_{j=1}^{k} \frac{j^{\zeta-1}}{i^{\zeta}} \right]$$

$$= \frac{4F\mu^2}{\zeta p\sqrt{n}} \sum_{i=1}^{n} \sum_{k=1}^{i-1} \frac{k^{\zeta-1}}{i^{\zeta}} \bar{\eta}_k + o(1) =: Z_{4,n} + o(1). \qquad (38)$$

By a routine verification of the conditions of the Lindeberg-Feller CLT we get the following result.

Lemma 5 *As $n \to \infty$*

$$Z_{4,n} = \frac{4F\mu^2}{\zeta p\sqrt{n}} \sum_{i=1}^{n} \sum_{k=1}^{i-1} \frac{k^{\zeta-1}}{i^\zeta} \bar{\eta}_k \Longrightarrow \mathcal{N}(0, \sigma_z^2),$$

where $\sigma_z = 4F\mu^2\sqrt{\frac{1-p}{p\zeta^3}}$.

Let us now estimate $\tilde{Z}_{3,n}$ in (35). From (34) it readily follows that

$$\frac{2F\mu^2}{p\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{i} \frac{j^{\zeta-1}}{i^\zeta} R_{i,j}^2 = o(1),$$

and thus $\tilde{Z}_{3,n} = o(1)$.

So far we have argued that

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big[\xi_{i+1}(\bar{V}_i^2 - V_L^2)\Big] &= \big(W_{3,n} + Z_{4,n}\big) + \big(V_{1,n} + V_{2,n} + W_{4,n} + Z_{5,n}\big) + o(1) \\ &=: G_n + H_n + o(1), \end{aligned} \tag{39}$$

where $W_{3,n}$, $W_{4,n}$, $Z_{4,n}$, $V_{1,n}$, $V_{2,n}$ and $Z_{5,n}$ are defined, respectively, in (25), (38), (23) and (37). By the independence of $W_{3,n}$ and $Z_{4,n}$, we have by Lemmas 3 and 5 that $G_n \Longrightarrow \mathcal{N}(0, \tilde{\sigma}_q^2)$, where $\tilde{\sigma}_q^2 = \sigma_w^2 + \sigma_z^2$. To establish (20), it is then enough to show that $H_n = o(1)$, which we do in the following lemmas, one for each of the constituents of H_n.

Lemma 6 *Assume $Var(\xi_1) = \sigma^2 < \infty$. Then $V_{1,n} = o(1)$.*

Proof First fix $\delta > 0$. Given $\varepsilon > 0$, Lemma 1 states that exists $m \in \mathbb{N}$ such that $\mathbb{P}(A_{m,\varepsilon}^c) < \varepsilon/2$. Recall the definition of $X_{i,j}$ in (4), and that $\{X_{i,j}, i \geq j \geq 1\}$ and $\{\xi_n\}_{n\in\mathbb{N}}$ are independent.

$$\begin{aligned} &\mathbb{P}\left(\left|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{i} \bar{\xi}_{i+1}\bar{\xi}_j X_{i,j}\right| > \delta\right) \\ &\qquad \leq \mathbb{P}\left(\left|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{i} \bar{\xi}_{i+1}\bar{\xi}_j X_{i,j} 1_{A_{m,\varepsilon}}\right| > \delta\right) + \frac{\varepsilon}{2}. \end{aligned} \tag{40}$$

It follows from definition of $A_{m,\varepsilon}$ in (5) that $X_{i,j}1_{A_{m,\varepsilon}} \leq (1+\varepsilon)[j^{\zeta-1}/(pi^\zeta)]$ for all $i \geq j \geq m$. Using this and by Markov's Inequality, we get that the first term on the right of (40) is bounded above by

$$\frac{1}{\delta^2 n}\sum_{i=1}^{n}\sum_{j=1}^{i}\mathbb{E}(\bar{\xi}_{i+1})^2\mathbb{E}(\bar{\xi}_j)^2\,\mathbb{E}(X_{i,j}^2 1_{A_{m,\varepsilon}})$$

$$=\frac{(1+\varepsilon)^2\sigma^4}{p^2\delta^2 n}\sum_{i=1}^{n}\sum_{j=m}^{i}\frac{j^{2\zeta-2}}{i^{2\zeta}}+o(1)=o(1).$$

Since $\delta > 0$ and $\varepsilon > 0$ are arbitrary, the combination of this inequality and (40) yields the result. □

Lemma 7 *Assume $Var(\xi_1) = \sigma^2 < \infty$. Then $V_{2,n} = o(1)$.*

Proof Arguing similarly as in the Proof of Lemma 6, given $\delta > 0$ and $\varepsilon > 0$, we have that m large enough

$$\mathbb{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\bar{\xi}_{i+1}\sum_{j=1}^{i}\left(X_{i,j}-\frac{j^{\zeta-1}}{pi^{\zeta}}\right)\right]\right|>\delta\right)\le$$

$$\mathbb{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\bar{\xi}_{i+1}\sum_{j=1}^{i}\left(X_{i,j}-\frac{j^{\zeta-1}}{pi^{\zeta}}\right)1_{A_{m,\varepsilon}}\right]\right|>\delta\right)+\frac{\varepsilon}{2} \tag{41}$$

and since $|X_{i,j}(\omega) - j^{\zeta-1}/(pi^{\zeta})|1_{A_{m,\varepsilon}} \le (\varepsilon j^{\zeta-1})/(pi^{\zeta})$ for all $i \ge j \ge m$, we get that the first term on the right of (41) is bounded above by

$$\frac{1}{\delta^2 n}\sum_{i=1}^{n}\left[\mathbb{E}(\bar{\xi}_{i+1})^2\,\mathbb{E}\left(\sum_{j=1}^{i}\left(X_{i,j}(\omega)-\frac{j^{\zeta-1}}{pi^{\zeta}}\right)1_{A_{m,\varepsilon}}\right)^2\right]$$

$$\le\frac{\sigma^2}{\delta^2 n}\sum_{i=1}^{n}\mathbb{E}\left[\sum_{j=1}^{i}\left|X_{i,j}(\omega)-\frac{j^{\zeta-1}}{pi^{\zeta}}\right|1_{A_{m,\varepsilon}}\right]^2$$

$$\le\varepsilon^2\frac{\sigma^2}{\delta^2 n}\sum_{i=1}^{n}\left(\sum_{j=m}^{i}\frac{j^{\zeta-1}}{i^{\zeta}}\right)^2+o(1)\le\frac{\varepsilon}{2},$$

as soon as n is large enough, and the result follows upon substitution in (41), since δ and ε are arbitrary. □

Lemma 8 *Assume $Var(\xi_1) = \sigma^2 < \infty$. Then $W_{4,n} = o(1)$.*

Proof Similar to the Proof of Lemma 7. □

Lemma 9 *Assume* $Var(\xi_1) = \sigma^2 < \infty$. *Then* $Z_{5,n} = o(1)$.

Proof Changing the order of summation, we find that $Z_{5,n}$ equals constant times

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} L_{k,n} \bar{\eta}_k \bar{M}_{k-1},$$

where $L_{k,n} = \frac{1}{k^2} \left(\sum_{j=1}^{k} j^{\zeta-1} \right) \left(\sum_{i=k}^{n} \frac{1}{i^{\zeta}} \right)$, which is bounded above by constant times $1/k$ uniformly in j and n. Now by Markov:

$$\mathbb{P}\left(\left| Z_{5,n} \right| \geq \delta \right) \leq \frac{\text{const}}{\delta^2 n} \sum_{k=1}^{n} \frac{1}{k^2} \mathbb{E}(\bar{M}_{k-1}^2) \leq \frac{\text{const}}{\delta^2} \frac{1}{n} \sum_{k=1}^{n} \frac{1}{k} = o(1),$$

and we are done. □

We still owe a proof for Lemma 2.

Proof of Lemma 2 Since $\bar{V}_i^2 - V_L^2 = (\bar{V}_i - V_L)(\bar{V}_i + V_L)$ and almost surely $\bar{V}_i$ converges to V_L, to prove the first claim is enough to show that $(\bar{V}_i^2 - V_L^2) = o(1/i^{1/2-\epsilon})$. We write

$$\bar{V}_i^2 - V_L^2 = 2F \sum_{j=1}^{i} \left[\xi_j X_{i,j} \right] - \frac{2F\mu}{p} \int_0^1 x^{\zeta-1} dx.$$

$$= 2F \sum_{j=1}^{i} \left[\xi_j X_{i,j} - \mu \frac{j^{\zeta-1}}{p i^{\zeta}} \right] + \frac{2F\mu}{p} \left[\frac{1}{i} \sum_{j=1}^{i} \left(\frac{j}{i} \right)^{\zeta-1} - \int_0^1 x^{\zeta-1} dx \right].$$

The second term on the right-hand side of this equation is an $O(1/i)$. We break down the first term as follows

$$2F \sum_{j=1}^{i} \bar{\xi}_j \frac{j^{\zeta-1}}{p i^{\zeta}} + 2F \sum_{j=1}^{i} \bar{\xi}_j \left(X_{i,j} - \frac{j^{\zeta-1}}{p i^{\zeta}} \right) + 2F\mu \sum_{j=1}^{i} \left(X_{i,j} - \frac{j^{\zeta}}{p i^{\zeta}} \right). \tag{42}$$

Setting $\bar{S}_0 = 0$ and $\bar{S}_k := \sum_{l=1}^{k} \bar{\xi}_k$, $k \in \mathbb{N}$, we write the first term on the right of (42) as

$$\sum_{j=1}^{i} \left[(\bar{S}_j - \bar{S}_{j-1}) \frac{j^{\zeta-1}}{p i^{\zeta}} \right] = \sum_{j=1}^{i-1} \left[\bar{S}_j \left(\frac{j^{\zeta-1}}{p i^{\zeta}} - \frac{(j+1)^{\zeta-1}}{p i^{\zeta}} \right) \right] + \frac{\bar{S}_i}{p i} = o(1/i^{1/2-\epsilon}),$$

where the last equality follows by the Law of the Iterated Logarithm.

Analogously, we write the second term on the right of (42) as

$$\sum_{j=1}^{i-1}\left[\bar{S}_j\left(X_{i,j}-X_{i,j+1}\right)\right]+\sum_{j=1}^{i-1}\left[\bar{S}_j\left(\frac{(j+1)^{\zeta-1}}{pi^{\zeta}}-\frac{j^{\zeta-1}}{pi^{\zeta}}\right)\right]+\bar{S}_i\left(X_{i,i}-\frac{1}{ip}\right). \tag{43}$$

Recalling (4), one readily checks that that $|X_{i,j}-X_{i,j+1}|=O\left(|X_{i,j+1}|/(M_j+1)\right)$. Given $\varepsilon>0$, by Lemma 1 we a.s. find an $m\in\mathbb{N}$ such that $|X_{i,j+1}|\leq(1+c)(j+1)^{\zeta-1}/(pi^{\zeta})$ for every $i\geq j\geq m$. Therefore, again by the Law of Large Numbers and the Law of the Iterated Logarithm, the three terms on (43) are $o(1/i^{1/2-\epsilon})$.

To deal with the third and last term on the right of (42), we may proceed similarly as in the analysis of $W_{2,n}$ above—recall (24), (27), (29) and (35). We write

$$\begin{aligned}\sum_{j=1}^{i}\left[X_{i,j}-\frac{j^{\zeta}}{pi^{\zeta}}\right]&=\sum_{j=1}^{i}\left[\frac{j^{\zeta}}{i^{\zeta}}\left(\frac{1}{M_j}-\frac{1}{pj}\right)\right]\\&+\sum_{j=1}^{i}\left[\frac{j^{\zeta}}{i^{\zeta}}\left(\frac{1}{M_j}-\frac{1}{pj}\right)\left(R_{i,j}+O(R_{i,j}^2)\right)\right]+\sum_{j=1}^{i}\left[\frac{j^{\zeta-1}}{pi^{\zeta}}\left(R_{i,j}+O(R_{i,j}^2)\right)\right].\end{aligned} \tag{44}$$

In the Proof of Lemma 4, we have shown that almost surely, for $i\geq j$ sufficiently large, $|R_{i,j}|\leq 1/j^{1/2-\epsilon}$, and we also argued that $\left(1/M_j-1/(pj)\right)=o(1/j^{3/2-\epsilon})$. Using this estimates, we readly get that each of the terms on the right hand side of (44) is an $o(1/i^{1/2-\epsilon})$, for $0<\epsilon<1/4$, and thus, so is the left hand side of (44), and we are done with the first claim of the lemma.

To argue the last claim of the lemma, note that

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\xi_i\left(\bar{V}_{i-1}-V_L\right)^2\right]&=o\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\xi_i}{i^{1-2\epsilon}}\right)\\&=o\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\mu}{i^{1-2\epsilon}}\right)+o\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\bar{\xi}_i}{i^{1-2\epsilon}}\right)\\&=o(1/n^{1/2-2\epsilon}),\end{aligned}$$

where the last equality holds by the hypothesis that ξ_1 has finite second moment and the Two Series Theorem, and we are done. □

Proceeding analogously as in the Proof of Proposition 1, similarly breaking down the relevant quantities, we may also obtain a central limit theorem for the velocity of the t.p. on the modified process (at collision times), namely

Proposition 2 *Let $Var(\xi_1) = \sigma^2 < \infty$. Then, as $n \to \infty$,*

$$\sqrt{n}\left(\bar{V}_n - V_L\right) \Longrightarrow \mathcal{N}(0, \hat{\sigma}_v^2),$$

where $\hat{\sigma}_v > 0$.

4 Central Limit Theorem for the Original Process

In this section, we prove our main results.

4.1 Proof of Theorem 2

For each, $i \in \mathbb{N}$, let t_i be the instant when the t.p. collides for the first time with the initial $i-$th particle in the line; more precisely, t_i is such that $Q(t_i) = S_i$. It is enough to show a CLT along (t_i), and for that it suffices to establish a version of Proposition 1 with barred quantities replaced by respective unbarred quantities, which amounts to replacing $\bar{t}_n$ by t_n in (12), namely showing that $(S_n - t_n V_L)/\sqrt{n} \Longrightarrow \mathcal{N}(0, \hat{\sigma}_q^2)$. Theorem 2 readily follows with $\sigma_q^2 = \frac{V_L}{\mu}\hat{\sigma}_q^2$.

We use Proposition 1 and a comparison between $\bar{t}_i$ and t_i to conclude our proof. Due to Proposition 1, it is enough to argue that

$$\frac{t_n - \bar{t}_n}{\sqrt{n}} = o(1). \tag{45}$$

Let $s_1, s_2, \ldots$ be the instants when the t.p. recollides with a moving elastic particle, whose velocities will be, respectively, denoted by $v_1, v_2, \ldots$. As follows from the remarks in the Introduction on the fact that the dynamics is a.s. well defined—see paragraph right below (2)—these sequences are well defined, and $s_1, s_2, \ldots$ has no limit points. We also recall that, for each $l \in \mathbb{N}$, $V(s_l)$ and $V(s_l^+)$ denote the velocities of the t.p. immediately before and at the l-th recollision, respectively.

For each $j \in \mathbb{N}$ we define

$$\Delta(j) := \sum_{s_l \in [t_{j-1}, t_j]} \left[V^2(s_l) - V^2(s_l^+)\right] \text{ and } \delta(j) := \sum_{s_l \in [t_{j-1}, t_j]} \left[V(s_l) - v_l\right]. \tag{46}$$

As follows from what has been pointed out in the above paragraph, these sums are a.s. well defined and consist of finitely many terms.

Let $v : [0, \infty) \longrightarrow \mathbb{R}$ denote the function that associates the position x to the velocity of the t.p. at x, that is, $v(x) = V(Q^{-1}(x))$. We analogously define $\bar{v} : [0, \infty) \longrightarrow \mathbb{R}$ for the modified process. We have that

$$t_n = \int_0^{S_n} \frac{1}{v(x)}\, dx \text{ and } \bar{t}_n = \int_0^{S_n} \frac{1}{\bar{v}(x)}\, dx.$$

In this way, (45) becomes

$$\int_0^{S_n} \left(\frac{1}{v(x)} - \frac{1}{\bar{v}(x)} \right) dx = o(n^{1/2}),$$

and due to convergence of $v(x)$ and $\bar{v}(x)$, it is enough to argue that

$$\int_0^{S_n} \left(\bar{v}^2(x) - v^2(x) \right) dx = o(n^{1/2}).$$

Torricelli's equation, (1), (2) and (46), give us that, for each $i \in \mathbb{N}$, at position $x \in [S_{i-1}, S_i)$,

$$\bar{v}^2(x) - v^2(x) = \sum_{j=1}^{i-1} \left[\Delta(j) \prod_{k=j}^{i-1} \left(\frac{M_k + (\eta_k - 1)}{M_k + 1} \right)^2 \right] + \sum_{s_l \in [S_{i-1}, x)} \left(V^2(s_l) - V^2(s_l^+) \right).$$

Therefore, we have the following upper bound

$$\int_0^{S_n} \left(\bar{v}^2(x) - v^2(x) \right) dx \le \sum_{i=1}^{n} \left[\xi_i \sum_{j=1}^{i-1} \left(\Delta(j) \prod_{k=j}^{i-1} \left(\frac{M_k + (\eta_k - 1)}{M_k + 1} \right)^2 \right) \right] + \sum_{i=1}^{n} \xi_i \Delta(i).$$

Turning back to (2), we have that,

$$V(s_j) - V(s_j^+) = V(s_j) - \left(\frac{M(s_j) - 1}{M(s_j) + 1} V(s_j) + \frac{2}{M(s_j) + 1} v_j \right) = \frac{2(V(s_j) - v_j)}{M(s_j) + 1}.$$

And therefore, again by the fact that $V(\cdot)$ is convergent, recalling (46), we have that

$$\Delta(j) = O\left(\frac{\delta(j)}{M_j + 1} \right);$$

moreover, recalling (4), we have that

$$\sum_{i=1}^{n}\left[\xi_{i+1}\sum_{j=1}^{i}\left(\Delta(j)\prod_{k=j}^{i}\left(\frac{M_k+(\eta_k-1)}{M_k+1}\right)^2\right)\right]+\sum_{i=1}^{n}\xi_{i+1}\Delta(i+1)=$$
$$O\left(\sum_{i=1}^{n}\left[\xi_{i+1}\sum_{j=1}^{i}\delta(j)X_{i,j}\right]+\sum_{i=1}^{n}\frac{\xi_{i+1}\delta(i+1)}{i+1}\right).$$

By Lemma 1,

$$\sum_{i=1}^{n}\left[\xi_{i+1}\sum_{j=1}^{i}\delta(j)X_{i,j}\right]=\sum_{j=1}^{n}\left[\delta(j)\sum_{i=j}^{n}\xi_{i+1}X_{i,j}\right]=O\left(\sum_{j=1}^{n}\left[\delta(j)j^{\zeta-1}\sum_{i=j}^{n}\frac{\xi_{i+1}}{i^{\zeta}}\right]\right).$$

Since $\mathbb{E}\xi^2<\infty$, Borel-Cantelli lemma readily implies that for every $\epsilon>0$, $\mathbb{P}(\xi_{n+1}>\epsilon\sqrt{n}\text{ i.o.})=0$. Thus,

$$\sum_{j=1}^{n}\left[\delta(j)j^{\zeta-1}\sum_{i=j}^{n}\frac{\xi_{i+1}}{i^{\zeta}}\right]=O\left(\sum_{j=1}^{n}\left[\delta(j)j^{\zeta-1}\sum_{i=j}^{n}\frac{\epsilon\sqrt{i}}{i^{\zeta}}\right]\right)=$$
$$\epsilon\sqrt{n}O\left(\sum_{j=1}^{n}\left[\delta(j)j^{\zeta-1}\sum_{i=j}^{n}\frac{1}{i^{\zeta}}\right]\right)=\epsilon\sqrt{n}O\left(\sum_{j=1}^{n}\delta(j)\right),$$

and also

$$\sum_{i=1}^{n}\frac{\xi_{i+1}\delta(i+1)}{i+1}=O\left(\sum_{i=1}^{n}\delta(i+1)\right).$$

By Lemma 10, we are done, since $\epsilon>0$ is arbitrary.

Lemma 10 *Let $\delta(j)$ as defined in (46). Almost surely,*

$$\sum_{j=1}^{\infty}\delta(j)<\infty. \tag{47}$$

Proof This result is already contained more or less explicitly in [4], in the argument to prove Theorem 1—see discussion on page 803 of [4]. For completeness and simplicity, circularity notwithstanding, we present an argument relying on Theorem 1 directly.

There a.s. exists a time T_0 such that there are no recollisions with standing particles met by the t.p. after T_0. This is because at large times, the velocity of the t.p. is close enough to V_L and its mass close enough to infinity, so that new collisions with standing elastic particles will give them velocity roughly $2V_L$, and thus they will be thence unreachable by the t.p. This means that we have only finitely many particles that recollide with the t.p.

We may also conclude by an elementary reasoning using Theorem 1 that if a particle collides infinitely often with the t.p., then its velocity may never exceed V_L. Let $u_1, u_2, \dots$ denote the recollision times with such a particle, and $v(u_1), v(u_2), \dots$, its velocity at such times, respectively. As we can deduce from (2), $v(u_{i+1}) > V(u_i)$; thus,

$$\sum_{i=1}^{\infty} [V(u_i) - v(u_i)] < \sum_{i=1}^{\infty} [v(u_{i+1}) - v(u_i)] \le V_L,$$

and (47) follows. □

4.2 *Proof of Theorem 3*

By Proposition 2, and the convergences of both V_n and $\bar{V}_n$, and after similar considerations as at the beginning of Sect. 4.1, we find that it is enough to prove that

$$\sqrt{n}\left(\bar{V}_n^2 - V_n^2\right) = o(1) \tag{48}$$

(so that in the end we get that Theorem 3 holds with $\sigma_v^2 = \frac{\mu}{V_L}\hat{\sigma}_v^2$).

Recalling (4.1), we have that

$$\bar{V}_n^2 - V_n^2 = \bar{v}^2(S_n) - v^2(S_n) = \sum_{j=1}^{n-1}\left[\Delta(j)\prod_{k=j}^{i-1}\left(\frac{M_k + (\eta_k - 1)}{M_k + 1}\right)^2\right] + \Delta(n).$$

Proceeding similarly as in the Proof of Theorem 2, we find that

$$\sqrt{n}\left(\bar{V}_n^2 - V_n^2\right) = O\left(\sqrt{n}\sum_{j=1}^{n}\left[\delta(j)\frac{j^{\zeta-1}}{n^{\zeta}}\right] + \frac{\delta(n+1)}{\sqrt{n}}\right).$$

By Lemma 10, given $\epsilon > 0$, there almost surely exists $j_0 \in \mathbb{N}$ such that $\sum_{j \ge j_0} \delta(j) \le \epsilon/2$. Thus,

$$\sqrt{n}\sum_{j=1}^{n}\left[\delta(j)\frac{j^{\zeta-1}}{n^{\zeta}}\right] \le \frac{1}{n^{\zeta-1/2}}\sum_{j=1}^{j_0}\delta(j)j^{\zeta-1} + \sum_{j>j_0}\delta(j) \le \epsilon,$$

for n sufficiently large. Lemma 10 implies that $\delta(n) = o(1)$. Since $\epsilon > 0$ is arbitrary, (48) follows.

Acknowledgments L.R. Fontes acknowledges support of CNPq (grant 311257/2014-3), and FAPESP (grant 2017/10555-0). P.A. Gomes acknowledges support of CAPES and FAPESP (grant 2020/02636-3). R. Sanchis acknowledges support of CAPES, CNPq and FAPEMIG (PPM 0600-16). The authors would like to thank Marco Aymone for useful discussions.

References

1. Boldrighini, C., Pellegrinotti, A., Presutti, E., Sinai, Ya., Soloveitchik, M.: Ergodic properties of a semi-infinite one-dimensional system of statistical mechanics. Comm. Math. Phys. **101**, 363–382 (1985)
2. Buttà, P., Caglioti, E., Marchioro, C.: On the violation of Ohm's law for bounded interactions: a one dimensional system. Comm. Math. Phys. **249**, 353–382 (2004)
3. De La Peña, V.H., Giné, E.: Decoupling: From Dependence to Independence. Probability and Its Applications. Springer, Berlin (1999)
4. Fontes, L.R.G., Jordão Neves, E., Sidoravicius, V.: Limit velocity for a driven particle in a random medium with mass aggregation. Ann. Inst. H. Poincaré Probab. Statist. **3**6, 787–805 (2000)
5. Lifshits, M., Shi, Z.: Aggregation rates in one-dimensional stochastic systems with adhesion and gravitation. Ann. Probab. **33**, 53–81 (2005)
6. Martin, Ph.A., Piasecki, J.: Aggregation dynamics in a self-gravitating one-dimensional gas. J. Statist. Phys. **84**, 837–857 (1996)
7. Pellegrinotti, A., Sidoravicius, V., Vares, M.E.: Stationary state and diffusion for a charged particle in a one-dimensional medium with lifetimes. Teor. Veroyatnost. i Primenen. **44**, 796–825 (2000); Reprinted in Theory Probab. Appl. **44**, 697–721 (1999)
8. Sidoravicius, V., Triolo, L., Vares, M.E.: On the forced motion of a heavy particle in a random medium. I. Existence of dynamics. I Brazilian School in Probability (Rio de Janeiro, 1997). Markov Process. Relat. Fields **4**, 629–647 (1998)
9. Sidoravicius, V., Triolo, L., Vares, M.E.: Mixing properties for mechanical motion of a charged particle in a random medium. Comm. Math. Phys. **219**, 323–355 (2001)

Structural Properties of Conditioned Random Walks on Integer Lattices with Random Local Constraints

Sergey Foss and Alexander Sakhanenko

To the memory of Vladas

Abstract We consider a random walk on a multidimensional integer lattice with random bounds on local times, conditioned on the event that it hits a high level before its death. We introduce an auxiliary "core" process that has a regenerative structure and plays a key role in our analysis. We obtain a number of representations for the distribution of the random walk in terms of the similar distribution of the "core" process. Based on that, we prove a number of limiting results by letting the high level to tend to infinity. In particular, we generalise results for a simple symmetric one-dimensional random walk obtained earlier in the paper by Benjamini and Berestycki (J Eur Math Soc 12(4):819–854, 2010).

Keywords Conditioned random walk · Bounded local times · Regenerative sequence · Potential regeneration · Separating levels · Skip-free distributions

MSC Classification 60K15, 60K37, 60F99, 60G99

Research is supported by RSF research grant No. 17-11-01173-Ext.

S. Foss (✉)
School of MACS, Heriot-Watt University, Riccarton, Edinburgh, UK

Novosibirsk State University, Novosibirsk, Russia
e-mail: sergueiorfoss25@gmail.com

A. Sakhanenko
Novosibirsk State University, Novosibirsk, Russia
e-mail: aisakh@mail.ru

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_19

1 Introduction

Consider a d-dimensional random walk

$$S_t = (S_t[1], \dots, S_t[d]) = S_0 + \sum_{j=1}^{t} \xi_j, \quad t = 0, 1, 2, \dots, \tag{1}$$

on the integer lattice $\mathbb{Z}^d$, where $\xi_j = (\xi_j[1], \dots, \xi_j[d])$, $j = 1, 2, \dots$ are i.i.d. random vectors that do not depend on the initial value S_0. The random variable

$$L_t(x) = \sum_{j=0}^{t} \mathbf{1}\{S_j = x\}, \quad x \in \mathbb{Z}^d, \tag{2}$$

counts the number of visits to (or local time at) state x by time $t = 0, 1, 2, \dots$. We assume that, for each $x \in \mathbb{Z}^d$, the number of possible/allowed visits to state x is limited above by a counting number $H(x) \geq 0$. Let

$$T_* = \inf\{t \geq 0 : L_t(S_t) > H(S_t)\} \leq \infty \tag{3}$$

be the first time when the number of visits to any state exceeds its upper limit. If T_* is finite, we assume that the random walk is "killed" at the time instant T_* (or it "dies", or "freezes" at time T_*).

Thus, we consider a multidimensional integer-valued random walk in a changing random environment, where initially each point x is characterized by a random number $H(x)$ of allowed visits to it. At any time t, the random walk jumps from $x = S_t$ to S_{t+1} and changes the environment at point x by decreasing the number of remaining allowed visits by 1. As a natural example, consider a model of a random walk on atoms of a "harmonic crystal" (see, e.g., [8] and [3]). An electron jumps from one atom to another, taking from a visited atom for the next jump a fixed unit of energy, which cannot be recovered. Thus, if S_t is a position of the electron at time t, then a unit of energy is sufficient for it to have a next jump to position $S_{t+1} = S_t + \xi_{t+1}$, which may be in any direction from S_t. We interpret the first coordinate $S_t[1]$ of S_t as its *height* and assume further that the height cannot increase by more than one unit:

$$\xi_t[1] \leq 1 \quad a.s., \qquad t = 1, 2, \dots. \tag{4}$$

When the electron arrives at an atom with insufficient energy level, it "freezes" there. We may formulate two natural tasks. Firstly, to find the asymptotics, as $n \to \infty$, of the probability of the event B_n that the electron reaches the level n before it "freezes", i.e.

$$B_n := \{\alpha(n) < T_*\} \tag{5}$$

with $\alpha(n)$ being the hitting time of the level n:

$$\alpha(n) := \inf\{t \geq 0 : S_t[1] \geq n\} = \inf\{t \geq 0 : S_t[1] = n\}, \tag{6}$$

where the latter equality follows from the skip-free property (4).

Secondly, given that the electron is still active by the time of hitting level n, a question of interest is the asymptotic, as n increases, of the conditional distribution of the electron's sample path.

To clarify the presentation, we will use the low-case "star" in the probability $\mathbf{P}_*(\cdot)$ in order to underline the influence of the random environment. We omit the "star" in $\mathbf{P}(\cdot)$ if the environment is not involved.

In [3], a simple symmetric one-dimensional random walk on the integers ("one-dimensional atoms") has been considered under the assumptions that

$$\mathbf{P}(\xi_1 = 1) = 1/2 = \mathbf{P}(\xi_1 = -1), \quad S_0 = 0 \quad \text{and} \quad H(x) = L_0 = const \geq 2$$

for all $x \in \mathbb{Z}$. The latter means that initially each atom has a fixed (the same for all) amount of energy L_0. The authors showed that

$$\mathbf{P}_*(B_n) \sim const \cdot q^n \quad \text{as} \quad n \to \infty, \quad \text{where} \quad 0 < q < 1.$$

Based on that, they proved (see Theorem 5 in [3]) convergence of the conditional distributions:

$$\mathbf{P}_*((S_0, \ldots, S_k) \in A \mid B_n) \to \mathbf{P}((\overline{S}_0, \ldots, \overline{S}_k) \in A), \tag{7}$$

for any $k = 0, 1, 2, \ldots$ and all $A \subset \mathbb{Z}^{(k+1)\times d}$, where $\mathbb{Z}^{K\times d}$ denotes the space of vectors $\mathbf{x} = (x_1, \ldots, \mathbf{x}_K)$ having d-dimensional vectors as their components. Further, it was shown in [3] that the limiting sequence $\{\overline{S}_k = \overline{S}_k[1]\}$ in (7) has a regenerative structure (see Definition 3 below for details) and increases to infinity with a linear speed, i.e.

$$\overline{S}_n[1]/n \to a_1 \quad \text{a.s. as} \quad n \to \infty, \tag{8}$$

where $0 < a_1 < 1/L_0$.

In our paper, we consider a multivariate random walk on the integer lattice with random local constraints. We generalise the model of [3] in three directions: we consider more general distributions of jumps, many dimensions, and random local constraints. We develop the approach introduced in [3], with a number of essential differences. The main difference is that we first focus on the analysis of the structure of the initial random walk $\{S_t\}$. In particular, we introduce a notion of n-separating levels which often exist in our model. The analysis of properties of such random levels allows us to introduce a sequence of random vectors $\{\overline{S}_t\}$ with specially chosen joint distribution. We call $\{\overline{S}_t\}$ the *core* random sequence, or the *core* random process.

There are several advantages of studying the core process. We show that its structure (a) does not involve any counting constraints, (b) does not involve an environment, (c) operates with proper distributions only, and (d) the core process has a (strongly) regenerative structure with an infinite sequence of random regenerative levels $\{\overline{\nu}_i\}$ (see Definition 3 for details).

We obtain a number of interesting representations for the conditional distribution of the random walk $\{S_t\}$ in random environment $\{H(x)\}$, linked to the distribution of the core sequence $\{\overline{S}_t\}$. These representations allow us to obtain a number of novel results. For example, we show that

$$\mathbf{P}_*(B_n) = \psi_0 q^n \mathbf{P}(\overline{B}_n), \quad \text{where} \quad \overline{B}_n := \cup_{m=0}^{n} \{\overline{\nu}_m = n\}, \tag{9}$$

for well-defined positive constants ψ_0 and $0 < q \le 1$, and that

$$\mathbf{P}_*((S_0, \ldots, S_k) \in A \mid B_n) = \mathbf{P}((\overline{S}_0, \ldots, \overline{S}_k) \in A \mid \overline{B}_n), \tag{10}$$

for any $n \ge k = 0, 1, 2, \ldots$ and all $A \in \mathbb{Z}^{(k+1)\times d}$. Here event $\overline{B}_n$ occurs iff n coincides with one of the regenerative levels of the core random walk.

Finally, we obtain the desired limiting result (7) as a simple corollary of (10), which is a generalisation of Theorem 5 in [3].

Our analysis is based on renewal theory and other direct probabilistic methods that are proven to be productive for studies of various types of random walks and, in particular, for random walks in random environment, see e.g. E. Bolthauzen and A.-S. Sznitman [7] and many references therein, including the pioneering papers [11] and [13].

We have to mention that a number of known results for conditioned random walks that do not have local-time constraints (see, e.g. [6] and [1]) may be represented, in some particular cases, as corollaries of our results, see Remark 2 in Sect. 7 for details.

There is a number of publications on random walks with constraints on local times. We have already mentioned papers [3] and [8]. The paper [3] was, in fact, the initial point of our studies, and we have made a number of preliminary observations in [16] where we considered a reasonable one-dimensional generalisation of the discrete-time model in [3] with non-random boundary constraints. Papers [4] and [12] deal with a different problem: they consider a random walk on the line (see also [2] for a generalisation onto a class of Markov processes), assuming that the initial energy level $H(x)$ of a point $x > 0$ is a deterministic function of x that increases to infinity with x. These papers analyse recurrence/transience properties of the random walk that depend on the shape of the function $H(x)$. A generalisation of the model onto random trees may be found in [5]. Papers [10, 15] and [14] are more distant, they discuss unconditioned regenerative phenomena that depend on an infinite future, in a number of situations.

To conclude, in the present paper we provide a unified treatment of the conditional regenerative phenomenon in a class of multivariate random walks on the integers with changing random constraints on the numbers of visits.

The paper is organized as follows. In Sect. 2, we introduce the main assumptions on the model and the notions of separating and regenerative levels. In Sect. 3, we first introduce and discuss the structure of the model connected with the existence of random n-separating levels. After that we describe the core random sequence and its structure, and, finally, formulate the Representation Theorem and limiting results as its Corollaries. Then Sects. 4–6 are devoted to the proofs. We have to note that, in the proof of the main auxiliary result, the Key Theorem, we follow the approach developed in [3]. We conclude with Sect. 7 containing a few remarks.

2 Main Assumptions and Definitions

In this section, we present our main assumptions (A1)–(A4) and further technical assumptions, and introduce the so-called separating and regenerative levels that play the key role in our studies.

2.1 Basic Assumptions

For $n \in \mathbb{Z}$, introduce a half-space of $\mathbb{Z}^d$

$$\mathbb{Z}^d_{n+} := \{x = (x[1], x[2], \dots, x[d]) \in \mathbb{Z}^d : x[1] \geq n\}. \tag{11}$$

The following assumptions $(A1)$–$(A3)$, are supposed to hold throughout the paper.

(**A1**) The increments $\{\xi_t : t \geq 1\}$ of the random walk $\{S_t\}$ from (1) are i.i.d. random vectors taking values in $\mathbb{Z}^d$, and their first components have a skip-free distribution:

$$\sum_{k=-\infty}^{1} \mathbf{P}(\xi_1[1] = k) = 1 \quad \text{and} \quad \mathbf{P}(\xi_1[1] = 1) > 0.$$

(**A2**) The random constraints $\{H(x),\ x \in \mathbb{Z}^d\}$ are non-negative integer-valued random variables which may take the infinite value: for any $x \in \mathbb{Z}^d$,

$$\sum_{l=0}^{\infty} \mathbf{P}(H(x) = l) + \mathbf{P}(H(x) = \infty) = 1. \tag{12}$$

Moreover, the next three families of random variables

$$\{S_0;\ H(x),\ x \notin \mathbb{Z}^d_{0+}\}, \quad \{\xi_i,\ i \geq 1\} \quad \text{and} \quad \{H(x),\ x \in \mathbb{Z}^d_{0+}\}$$

are mutually independent, $S_0[1] \le 0$ a.s. and $\mathbf{P}_*(B_0) > 0$.

(**A3**) The family $\{H(x) : x \in \mathbb{Z}^d_{0+}\}$ consists of i.i.d. random variables with

$$\mathbf{P}(1 \le H(0) \le \infty) = 1.$$

We may interpret Assumption (A3) as follows: at time $t = 0$ the environment in $\mathbb{Z}^d_{0+}$ is stochastically homogeneous, so is "virgin" (see, also, Remark 1 in Sect. 7). Then condition $\mathbf{P}_*(B_0) > 0$ in Assumption (A2) may be read as "the random walk S_t arrives at the virgin domain of the random environment with a positive probability."

Assumptions $(A1)$–$(A3)$ yield that, for any $n \ge 0$,

$$\begin{aligned}\mathbf{P}_*(B_n) &\ge \mathbf{P}_*(\alpha(0) < T_*,\ \xi_{\alpha(0)+j}[1] = 1,\ H(S_{\alpha(0)+j}) > 0,\ j = 1, \dots, n)\\ &= \mathbf{P}_*(B_0)\mathbf{P}^n(\xi_1[1] = 1) > 0, \end{aligned} \tag{13}$$

where the events B_n were introduced in (5). Thus, for all $n \ge 0$ the event B_n occurs with positive probability and hence, as we can see later, all conditional probabilities in all our main assertions are well defined.

2.2 *Technical Assumption and Comments*

We have certain flexibility in the initial value S_0 and in the random environment $\{H(x)\}$ outside the set $\mathbb{Z}^d_{0+}$. Recall that we use notation $\mathbf{P}_*(\cdot)$ for probabilities of events where the environment is involved. We will also use special notation, $\mathbf{P}_0$ and $\mathbf{P}_+$, for two particular environments when $S_0 = 0$. For any event B, let

$$\mathbf{P}_0(B) := \mathbf{P}_*(B \mid S_0 = 0,\ H(y) = 0\ \ \forall y \notin \mathbb{Z}^d_{0+}\}, \tag{14}$$

$$\mathbf{P}_+(B) := \mathbf{P}_*(B \mid S_0 = 0,\ H(y) = \infty\ \ \forall y \notin \mathbb{Z}^d_{0+}\}. \tag{15}$$

In (14), *it is prohibited for the random walk to visit any states* $y \notin \mathbb{Z}^d_{0+}$, and (15) corresponds to the case where there is *no restrictions on the number of visits to any of the states* $y \notin \mathbb{Z}^d_{0+}$. Clearly,

$$\mathbf{P}_+(B_0) = \mathbf{P}_0(B_0) = 1 \quad \text{and} \quad \mathbf{P}_+(B_n) \ge \mathbf{P}_0(B_n) \ge \mathbf{P}^n(\xi_1[1] = 1) > 0\ \ \forall n \ge 0. \tag{16}$$

For the classical random walk (no environment), introduce two stopping times:

$$\beta_0 := \inf\{t > 0 : \xi_1[1] + \dots + \xi_t[1] = 0\} \le \beta_{0,0} := \inf\{t > 0 : \xi_1 + \dots + \xi_t = 0\} \le \infty.$$

We will need the following assumption:

(**A4**) If $\mathbf{P}(\beta_0 < \infty) = 1$ then $\mathbf{P}(\beta_{0,0} < \infty) > 0$. And

$$\text{if } \mathbf{P}(\beta_0 = \beta_{0,0} < \infty) = 1, \text{ then } \mathbf{E}H(0) < \infty.$$

It is clear that assumption (A4) is fulfilled in the following cases:

(a) $\mathbf{E}\xi_1[1] \neq 0$ (including the cases $\mathbf{E}\xi_1[1] > 0$ and $-\infty \leq \mathbf{E}\xi_1[1] < 0$);
(b) $\mathbf{E}\xi_1[1] = 0$ and $0 < \mathbf{P}(\beta_{0,0} < \infty) < 1$;
(c) $\mathbf{E}\xi_1[1] = 0$, $\mathbf{P}(\beta_{0,0} < \infty) = 1$ and $\mathbf{E}H(0) < \infty$.

Thus, our results do not work only in the next two cases:

(d) $\mathbf{E}\xi_1[1] = 0$ and $\mathbf{P}(\beta_{0,0} < \infty) = 0$;
(e) $\mathbf{E}\xi_1[1] = 0$, $\mathbf{P}(\beta_{0,0} < \infty) = 1$, and $\mathbf{E}H(0) = \infty$.

Note that the case (d) is degenerate in the spirit of our paper, since it corresponds to the situation where the random walk visits each state at most once.

Note also that the cases (c) and (e) relate to essentially one- or two-dimensional random walks only.

2.3 Separating and Regenerative Levels

For a finite or infinite sequence $\mathbf{y} = (y_0, y_1, y_2, \ldots)$ of $\mathbb{Z}^d$-valued vectors and for any $n \geq 0$, we let

$$\alpha(n|\mathbf{y}) := \inf\{t \geq 0 : y_t[1] \geq n\} \leq \infty, \tag{17}$$

where $y_t[1]$ is the first coordinate of y_t, for $t = 0, 1, \ldots$. Here and throughout the paper, we follow the standard conventions that

$$\inf \emptyset = \infty, \quad \sup \emptyset = -\infty \quad \text{and} \quad \sum_{k \in \emptyset} a_k = 0. \tag{18}$$

Definition 1 A number $k \geq 0$ is a "separating level" of the sequence $\mathbf{y}$ if

$$\alpha(k|\mathbf{y}) < \infty \quad \text{and} \quad \max_{0 \leq t < \alpha(k|\mathbf{y})} y_t[1] < k = y_{\alpha(k|\mathbf{y})}[1] \leq \inf_{t > \alpha(k|\mathbf{y})} y_t[1].$$

Definition 2 A number $k \in \{0, 1, \ldots, n\}$ is an "n-separating level" of the sequence $\mathbf{y}$ if

$$\sup_{0 \leq t < \alpha(k|\mathbf{y})} y_t[1] < k = y_{\alpha(k|\mathbf{y})}[1] \leq \inf_{\alpha(k|\mathbf{y}) < t < \alpha(n|\mathbf{y})} y_t[1] \quad \text{and} \quad \alpha(n|\mathbf{y}) < \infty.$$

For $n \ge 0$, let $\eta(n|\mathbf{y})+1$ counts the number of n-separating levels; and let $\varkappa(n|\mathbf{y})$ be the supremum of all $k < n$ such that k is an n-separating level.

These levels play an important role in our analysis. One can see that if k is an n-separating level, then it may not be an N-separating level for $N > n$ and, hence, it may be not a separating level. For example, $k = n$ is always the last n-separating level if $\alpha(n|\mathbf{y})$ is finite, but it is not an $(n+1)$-separating level if $y_{\alpha(n)+1}[1] < 0$.

In what follows, a "block" is any collection of random variables that may contain a random number of these variables.

Definition 3 A random sequence $\overline{S} = (\overline{S}_0, \overline{S}_1, \dots)$ is *strongly regenerative with regenerative levels* $\overline{\nu}_0 < \overline{\nu}_1 < \dots < \overline{\nu}_n < \dots$, if $\{\overline{\nu}_i\}$ is an infinite sequence of proper integer-valued random variables such that, firstly, the following "blocks" of random variables

$$\{\overline{\nu}_i-\overline{\nu}_{i-1},\ \overline{\alpha}(\overline{\nu}_i)-\overline{\alpha}(\overline{\nu}_{i-1}),\ (\overline{S}_{\overline{\alpha}(\overline{\nu}_{i-1})+t}-\overline{S}_{\overline{\alpha}(\overline{\nu}_{i-1})},\ t=1,2,\dots,\overline{\alpha}(\overline{\nu}_i)-\overline{\alpha}(\overline{\nu}_{i-1}))\},$$

for $i \ge 1$, are i.i.d. and do not depend on the initial "block" $\{\overline{\nu}_0,\ \overline{\alpha}(\overline{\nu}_0),\ (\overline{S}_t : t \le \overline{\alpha}(\overline{\nu}_0))\}$, and, secondly,

$$\inf_{t\ge\overline{\alpha}(\overline{\nu}_i)} \overline{S}_t[1] = \overline{S}_{\overline{\alpha}(\overline{\nu}_i)}[1] = \overline{\nu}_i > \sup_{0\le t<\overline{\alpha}(\overline{\nu}_i)} \overline{S}_t[1], \quad i = 0, 1, 2, \dots.$$

We then say that $\overline{\alpha}(\overline{\nu}_i)$ is the *regenerative time* that corresponds to regenerative level ν_i. One can view n-separating levels as "potential candidates" for regenerative levels and talk about "potential regeneration".

3 Main Results

In Sect. 3.1 we introduce a renewal equation for the random walk with local constraints and introduce its splitting into random blocks. In Sect. 3.2 we present the Key Theorem and introduce a sequence of independent blocks related to the core sequence. Based on that, we provide a formal definition of the core process in Sect. 3.3 . After that we present our main results in Sects. 3.4 and 3.5.

3.1 On the Structure of the Random Walk

Note that earlier notation (6) matches (17) as follows: $\alpha(n) = \alpha(n|S)$, for $S = (S_0, S_1, \dots)$. For each $n \ge 0$, we let

$$\eta_*(n) := \begin{cases} \eta(n|S) \text{ if } \alpha(n) < T_*(n), \\ -1, \qquad \text{otherwise,} \end{cases} \quad \text{and } \varkappa_*(n) := \begin{cases} \varkappa(n|S), \text{ if } \eta_*(n) \ge 1, \\ -\infty, \quad \text{if } \eta_*(n) < 1. \end{cases} \tag{19}$$

So, $\eta_*(n) + 1$ counts the number of n-separating levels in the case where the event $B_n = \{\eta_*(n) \geq 0\}$ occurs. Note that if the event B_n occurs, then $k = n$ is the largest n-separating level, and $k = \varkappa_*(n)$ is the second largest n-separating level, if it exists, i.e. when $\eta_*(n) \geq 1$. Clearly,

$$\{\varkappa_*(n) > -\infty\} = \{0 \leq \varkappa_*(n) \leq n-1\} = \{\eta_*(n) \geq 1\} \subset \{\eta_*(n) \geq 0\} = B_n. \tag{20}$$

Further, $\mathbf{P}_0(\eta_*(n) = 0) = 1$ because, under the "0-environment", level 0 is n-separating for any n such that $\alpha(n) < T_*$.

The random walk under consideration has the following renewal-type Property.

Property 1 Under the assumptions $(A1)$–$(A3)$, for any $n > k \geq 0$,

$$\mathbf{P}_*(B_n, \varkappa_*(n) = k) = \mathbf{P}_*(B_k) \cdot \mathbf{P}_0(\varkappa_*(n-k) = 0),$$

and then the following renewal equation holds:

$$\mathbf{P}_*(B_n) = \mathbf{P}_*(\eta_*(n) = 0) + \sum_{k=0}^{n-1} \mathbf{P}_*(B_k) \cdot \mathbf{P}_0(\varkappa_*(n-k) = 0), \quad n = 1, 2, \ldots. \tag{21}$$

In particular,

$$\mathbf{P}_0(B_n) = \sum_{k=0}^{n-1} \mathbf{P}_0(B_k) \cdot \mathbf{P}_0(\varkappa_*(n-k) = 0), \quad n = 1, 2, \ldots. \tag{22}$$

For $n > 0$ with $\eta_*(n) \geq 0$, let

$$0 \leq \nu_0(n) < \ldots < \nu_{\eta_*(n)}(n) = n$$

be the sequence of all n-separating levels (where $\nu_0(n) = \nu_{\eta_*(n)}(n) = n$ if $\eta_*(n) = 0$). In the case $\eta_*(n) \geq 1$, we may find all n-separating levels by the backward recursion:

$$\varkappa_*(\nu_i(n)) = \nu_{i-1}(n), \quad i = \eta_*(n), \eta_*(n) - 1, \ldots, 1.$$

For $n > 0$ with $\eta_*(n) \geq i \geq 1$, we let

$$\lambda_i(n) := \nu_i(n) - \nu_{i-1}(n), \quad T_i(n) := \alpha(\nu_i(n)), \quad \tau_i(n) := T_i(n) - T_{i-1}(n).$$

We need more notation. Introduce the random vectors

$$\mathbf{S}_K = (S_0, \ldots, S_K), \quad \mathbf{S}_{K,N} = (S_{K,K+1}, \ldots, S_{K,N}), \quad N > K \geq 0, \tag{23}$$

where

$$S_{K,K+t} := S_{K+t} - S_K = \sum_{j=1}^{t} \xi_{K+j}, \quad t = 0, 1, \dots .$$

On the event $B_n = \{\eta_*(n) \geq 0\}$, introduce a random block

$$(\nu_0(n),\ T_0(n),\ \mathbf{S}_{T_0(n)}). \tag{24}$$

This is the *initial block* of our random walk. Further, if $\eta_*(n) \geq 1$, then we may introduce consecutive blocks of random variables:

$$(\lambda_i(n), \tau_i(n),\ \mathbf{S}_{T_{i-1}(n),T_i(n)}), \quad i = 1, 2, \dots, \eta_*(n), \tag{25}$$

where $\lambda_i(n)$ is the *height* of the i-th block and $\tau_i(n)$ its *duration*. Property 1 shows that there is a certain conditional independence of each block in (25) from the previous blocks. We present these properties in full in Theorem 2 below. After that a representation for the joint distributions of random blocks from (24) and (25) will be given in Corollary 2.

3.2 Key Theorem (the Main Auxiliary Result)

The following technical result plays a central role in our studies. It will be proved in Sect. 5.

Theorem 1 *Under the assumptions* $(A1)$–$(A4)$, *there exists a number* $q \in (0, 1]$ *such that,*

$$\sum_{k=1}^{\infty} \mathbf{P}_0(\varkappa_*(k) = 0)/q^k = 1, \tag{26}$$

$$1 \leq \mu := \sum_{k=1}^{\infty} k\mathbf{P}_0(\varkappa_*(k) = 0)/q^k < \infty, \tag{27}$$

$$0 < \psi_0 := \sum_{m=0}^{\infty} \mathbf{P}_*(\eta_*(m) = 0)/q^m < \infty. \tag{28}$$

Properties (26)–(28) allow us to introduce an infinite sequence

$$(\overline{\nu}_0, \overline{T}_0, \widetilde{S}_{\overline{T}_0}) \quad \text{and} \quad (\overline{\lambda}_i, \overline{\tau}_i, \widetilde{Y}_{i,\overline{\tau}_i}), \quad i = 1, 2, \dots, \tag{29}$$

of mutually independent random blocks with special distributions, where

$$\widetilde{S}_{\overline{T}_0} = (\overline{S}_0, \ldots, \overline{S}_{\overline{T}_0}) \quad \text{and} \quad \widetilde{Y}_{i,\overline{\tau}_i} = (\overline{Y}_{i,1}, \ldots, \overline{Y}_{i,\overline{\tau}_i}) \tag{30}$$

are random vectors of random lengths. We determine their distributions step by step. First, we let

$$\mathbf{P}(\overline{\nu}_0 = k) = \mathbf{P}_*(\eta_*(k) = 0)/(\psi_0 q^k), \quad k = 0, 1, \ldots, \tag{31}$$

$$\mathbf{P}(\overline{\lambda}_i = l) = \mathbf{P}_0(\varkappa_*(l) = 0)/q^l, \quad l = 1, 2, \ldots. \tag{32}$$

Thus, we have determined the distributions of random vectors $\overline{\nu}_0$ and $\overline{\lambda}_i$ as Cramér-type transforms of the characteristics of the initial random walk $\{S_t\}$. By Theorem 1, the random vectors $\overline{\nu}_0$ and $\overline{\lambda}_i$ have proper distributions and

$$1 \le \mu = \mathbf{E}\overline{\lambda}_1 < \infty, \quad \mathbf{P}(\overline{\lambda}_1 = 1) = \mathbf{P}_0(\varkappa_*(1) = 0)/q \ge \mathbf{P}(\xi_1[1] = 1)/q > 0. \tag{33}$$

We determine next the distributions of other components of the vectors in (29). We let

$$\mathbf{P}(\overline{T}_0 = K, \widetilde{S}_K = \mathbf{y}_K | \overline{\nu}_0 = k) := \mathbf{P}_*(\alpha(k) = K < T_*, \mathbf{S}_K = \mathbf{y}_K | \eta_*(k) = 0), \tag{34}$$

for any $K \ge k + 1 \ge 1$ and $\mathbf{y}_K \in \mathbb{Z}^{(K+1)\times d}$; and then

$$\mathbf{P}(\overline{\tau}_i = L, \widetilde{Y}_{i,L} = \mathbf{x}_L | \overline{\lambda}_i = l) := \mathbf{P}_0(\alpha(l) = L < T_*, \mathbf{S}_{0,L} = \mathbf{x}_L | \varkappa_*(l) = 0), \tag{35}$$

for any $L \ge l \ge 1$ and $\mathbf{x}_L \in \mathbb{Z}^{L\times d}$.

Thus, we have introduced all joint distributions of random elements from (29). All these distributions are proper, since they are determined by proper distributions from (31), (32), (34) and (35). By the construction, with probability 1

$$\overline{\nu}_0 \ge 0, \ \ \overline{T}_0 \ge 0, \ \text{and}\ \overline{\tau}_i \ge \overline{\lambda}_i = \widetilde{Y}_{i,\overline{\tau}_i}[1] \ge 1, \quad \text{for all } i \ge 1. \tag{36}$$

Moreover, the random vectors $\{(\overline{\lambda}_i, \overline{\tau}_i, \widetilde{Y}_{i,\overline{\tau}_i}), \ \ i = 1, 2, \ldots\}$ are i.i.d.

3.3 Sample-Path Construction of the Core Random Sequence

Using mutually independent random blocks introduced in (29), we may define random variables

$$\overline{\nu}_m = \overline{\nu}_0 + \sum_{i=1}^{m} \overline{\lambda}_i > \overline{\nu}_{m-1}, \quad \overline{T}_m = \overline{T}_0 + \sum_{i=1}^{m} \overline{\tau}_i > \overline{T}_{m-1}, \quad m = 1, 2, \ldots. \tag{37}$$

Now we introduce random vectors $\overline{S}_j$ for all $j \ge 0$ using an induction argument. For $j \le \overline{T}_0$ they are given in (30). Suppose we have defined $\overline{S}_j$ for all $j \le \overline{T}_{i-1}$. Then we let

$$\overline{S}_{\overline{T}_{i-1}+j} := \overline{S}_{\overline{T}_{i-1}} + \overline{Y}_{i,j}, \quad j = 1, \ldots, \overline{\tau}_i = \overline{T}_i - \overline{T}_{i-1}. \tag{38}$$

Thus, we have defined $\overline{S}_j$ for all $j \le \overline{T}_i$. Repeating this procedure for all $i = 1, 2, \ldots$ we define random vectors $\overline{S}_j$ for all $j \ge 0$.

Similar to (23), we introduce vectors with multivariate components:

$$\widetilde{S}_N = (\overline{S}_0, \ldots, \overline{S}_N), \quad \widetilde{S}_{K,N} = (\overline{S}_{K+1} - \overline{S}_K, \ldots, \overline{S}_N - \overline{S}_K), \quad N > K \ge 0. \tag{39}$$

Consider now the random blocks

$$(\overline{\nu}_0, \overline{T}_0, \widetilde{S}_{\overline{T}_0}) \quad \text{and} \quad (\overline{\lambda}_i, \overline{\tau}_i, \widetilde{S}_{\overline{T}_{i-1}, \overline{T}_i}), \quad i = 1, 2, \ldots, \tag{40}$$

and note that, by (38) the i-th block in (40) coincides with the i-th block in (29). Thus, all blocks in (40) are mutually independent and all of them, but the initial, are i.i.d.

3.4 Representation Theorem

We are now ready to present our main results. The following statement summarises the main structural properties of the core random sequence and provides an inverse formulae for the distributions of the random walk in terms of the core process.

Let $\mathbb{Z}_*^d := \cup_{n=1}^{\infty} \mathbb{Z}^{n \times d}$. We consider $\mathbb{Z}_*^d$ as the state space for random sequences of random lengths.

Theorem 2 *Under the assumptions* $(A1)$–$(A4)$*, for any set* $\mathscr{A} \subset \mathbb{Z}_*^d$ *and for each* $n \ge m \ge 0$*,*

$$\begin{aligned} &\mathbf{P}_*(\alpha(n) < T_*,\ \eta_*(n) = m,\ (S_0, \ldots, S_{\alpha(n)}) \in \mathscr{A}) \\ &\quad = \psi_0 q^n \mathbf{P}(\overline{\nu}(m) = n,\ (\overline{S}_0, \ldots, \overline{S}_{\overline{\alpha}(n)}) \in \mathscr{A}). \end{aligned} \tag{41}$$

Thus, the distribution of the trajectory of the core random sequence has the same support with the distribution of the trajectory of the initial random walk (any finite sample path has positive probabilities to occur simultaneously for the core sequence and for the random walk, however these probabilities may differ). In particular, for all $j = 1, 2, \ldots$ the following inequalities hold with probability 1:

$$\overline{\xi}_j[1] \le 1 \quad \text{and} \quad \overline{S}_j[1] \le j, \quad \text{where} \quad \overline{\xi}_j = \overline{S}_j - \overline{S}_{j-1}. \tag{42}$$

Since $B_n = \{\alpha(n) < T_*\} = \{0 \le \eta_*(n) \le n\}$, we have from (41) that, for any set $A \subset \mathbb{Z}^{(k+1)\times d}$,

$$\mathbf{P}_*((S_0,\dots,S_k)\in A,\ B_n) = \sum_{m=0}^{n} \mathbf{P}_*((S_0,\dots,S_k)\in A,\ \eta_*(n)=m) \tag{43}$$

$$= \sum_{m=0}^{n} \psi_0 q^n \mathbf{P}((\overline{S}_0,\dots,\overline{S}_k)\in A,\ \overline{\nu}(m)=n) = \psi_0 q^n \mathbf{P}((\overline{S}_0,\dots,\overline{S}_k)\in A,\ \overline{B}_n).$$

Now (9) follows from (43) with $A = \mathbb{Z}^{(k+1)\times d}$. Equating the ratio of the left-hand sides of (43) and (9) to the ratio of the right-hand sides leads to (10). Repeating these arguments with $\alpha(n)$ in place of k, we obtain that, for any set $\mathscr{A} \subset \mathbb{Z}_*^d$ and for each $n \ge 0$,

$$\mathbf{P}_*((S_0,\dots,S_{\alpha(n)})\in \mathscr{A} \,|\, B_n) \tag{44}$$

$$= \mathbf{P}((\overline{S}_0,\dots,\overline{S}_{\overline{\alpha}(n)})\in \mathscr{A} \,|\, \overline{B}_n) \le \frac{\mathbf{P}((\overline{S}_0,\dots,\overline{S}_{\overline{\alpha}(n)})\in \mathscr{A})}{\mathbf{P}(\overline{B}_n)}.$$

Thus, we have proved

Corollary 1 *Under the assumptions* (*A*1)–(*A*4)*, relations* (9)*,* (10)*,* (43) *and* (44) *hold.*

Note that $\overline{c} := \inf_{n\ge 0} \mathbf{P}(\overline{B}_n) > 0$ as it follows from the first convergence in (47) below.

3.5 Limiting Results

Representation (43) allows us to obtain a number of limiting theorems using the standard renewal arguments. First of all, we can see from (43) that

$$\mathbf{P}(\overline{B}(n)\,|\,\overline{\nu}_0=0) = \mathbf{P}(\overline{\nu}_m = n \quad \text{for some} \quad m \ge 0 \,|\, \overline{\nu}_0 = 0)$$

$$= \sum_{m=0}^{n} \mathbf{P}(\overline{\nu}_i = n\,|\,\overline{\nu}_0=0) = V_n := \mathbf{I}\{n=0\} + \sum_{m=1}^{n} \mathbf{P}\left(\sum\nolimits_{i=1}^{m} \overline{\lambda}_i = n\right)$$

is the renewal function of the undelayed renewal process with i.i.d. increments $\{\overline{\lambda}_i,\ i=1,2,\dots\}$ satisfying (33).

Now consider the probabilities

$$U_n := \mathbf{P}(\overline{A}_k \cap \overline{B}_n), \quad \text{where} \quad \overline{A}_k := \{(\overline{S}_0,\dots,\overline{S}_k)\in A\},\ A \in \mathbb{Z}^{(k+1)\times d}.$$

Note that $S_j[1] \le k \le \overline{\nu}_k < \overline{\nu}_i$ for all $0 \le j \le k < i$ by (42) . Hence, the event $\overline{A}_k$ does not depend on the random variables $\{\overline{\lambda}_i = \overline{\nu}_i - \overline{\nu}_{i-1} : i > k\}$. Then

$$\begin{aligned} U_{n,l} &:= \mathbf{P}\big(\overline{B}_n \mid \overline{A}_k,\ \overline{\nu}_k = l \le n\big) \\ &= \mathbf{P}\Big(n = \overline{\nu}_m = \sum\nolimits_{i=1}^m \overline{\lambda}_i \ \text{ for some } m \ge 0 \,\Big|\, \overline{A}_k,\ l = \overline{\nu}_k = \sum\nolimits_{i=1}^k \overline{\lambda}_i\Big) \\ &= \mathbf{P}\Big(n - l = \sum\nolimits_{i=k+1}^m \overline{\lambda}_i \quad \text{for some} \quad m \ge k\Big) = V_{n-l}. \end{aligned}$$

Hence, by the total probability formula,

$$U_n - \mathbf{P}(\overline{A}_k \cap \overline{B}_n, \overline{\nu}_k > n) = \sum_{l=k}^{n} \mathbf{P}(\overline{A}_k,\ \overline{\nu}_k = l) \cdot U_{n,l} = \sum_{l=k}^{n} \mathbf{P}(\overline{A}_k,\ \overline{\nu}_k = l) V_{n-l}.$$

Thus, the differences $U_n - \mathbf{P}(\overline{A}_k,\ \overline{B}_n, \overline{\nu}_k > n)$ satisfy the renewal equation, where $\mathbf{P}(\overline{\nu}_k > n) \to 0$ as $n \to \infty$. So, by (33) and the local renewal theorem, as $n \to \infty$,

$$V_n \to 1/\mu \ \text{ and } \ U_n \to \sum_{l=k}^{\infty} \mathbf{P}(\overline{A}_k,\ \overline{\nu}_k = l)/\mu = \mathbf{P}(\overline{A}_k)/\mu. \tag{45}$$

Substituting (45) into (43) and (9) leads to the following statement

$$\mathbf{P}_*((S_0, \dots, S_k) \in A,\ B_n)/q^n \to \psi_0 \mathbf{P}((\overline{S}_0, \dots, \overline{S}_k) \in A)/\mu, \tag{46}$$

$$\mathbf{P}(\overline{B}_n) \to 1/\mu \quad \text{and} \quad \mathbf{P}_*(B_n)/q^n \to \psi_0/\mu. \tag{47}$$

In particular, (7) takes place. Next, using (36) and (37) we obtain from the Strong Law of Large Numbers that

$$\begin{aligned} \frac{\overline{S}_{\overline{T}_n}[1]}{n} = \frac{\overline{\nu}_n}{n} &= \frac{\overline{\nu}_0 + \overline{\lambda}_1 + \dots \overline{\lambda}_n}{n} \to \mathbf{E}\overline{\lambda}_1 = \mu \in (0, \infty) \quad a.s., \\ \frac{\overline{T}_n}{n} &= \frac{\overline{T}_0 + \overline{\tau}_1 + \dots \overline{\tau}_n}{n} \to \theta := \mathbf{E}\overline{\tau}_1 \in [\mu, \infty] \quad a.s. \end{aligned} \tag{48}$$

(In what follows we use the standard convention that $c/\theta = 0$ when $\theta = \infty$.) Thus, we have proved the following result.

Theorem 3 *Under the assumptions* $(A1)$–$(A4)$, *for all* $k \ge 0$ *and any* $A \subset \mathbb{Z}^{(k+1)\times d}$ *convergences* (46), (47) *and* (7) *hold. In addition, by the Strong Law of Large Numbers, convergence* (8) *takes place with* $a_1 := \mu/\theta \in [0, 1]$.

We would like to say that Theorem 3 was the initial aim of our studies. A simple proof of Theorem 3 (given above) shows the power of Theorem 2. In [3], direct analytical arguments have been used to establish for a simple symmetric random walk a limiting result similar to Theorem 3.

Further, Theorem 2 allows us to obtain an estimate (51) for the rate of convergence of the conditioned original process in the multidimensional generalization (49) of convergence (8):

Theorem 4 *Suppose that* $\mathbf{E}\overline{Y}_{1,\overline{\tau}_1}[j]$ *is finite for some* $j \geq 1$*. Then, under the assumptions* $(A1)$–$(A4)$*, for the* j*'th component of limiting process we have:*

$$\frac{\overline{S}_n[j]}{n} \to a_j := \frac{\mathbf{E}\overline{Y}_{1,\overline{\tau}_1}[j]}{\theta} = \frac{\mathbf{E}\big[\overline{S}_{\overline{T}_1}[j] - \overline{S}_{\overline{T}_0}[j]\big]}{\theta} \quad a.s., \tag{49}$$

which is equivalent to the convergence:

$$\forall\, \varepsilon > 0 \quad \mathbf{P}(\sup_{t \geq m} |\overline{S}_t[j]/t - a_j| > \varepsilon) \to 0 \quad as \quad m \to \infty. \tag{50}$$

In addition, for any $\varepsilon > 0$ *and each* $n \geq m > 0$

$$\mathbf{P}(\sup_{t \in [m,\alpha(n)]} |S_t[j]/t - a_j| > \varepsilon \mid B_n) \leq \overline{C}\mathbf{P}(\sup_{t \geq m} |\overline{S}_t[j]/t - a_j| > \varepsilon), \tag{51}$$

where $\overline{C} := 1/\overline{c} = 1/\inf_{n \geq 0} \mathbf{P}(\overline{B}_n) < \infty$.

In particular, (49)*,* (50) *and* (51) *hold for* $j = 1$ *with* $a_1 := \mu/\theta \in [0, 1]$.

Indeed, (49) follows from (48) and the following simple corollary from the Strong Law of Large Numbers:

$$\frac{\overline{S}_{\overline{T}_n}[j]}{n} = \frac{\overline{S}_{\overline{T}_0}[j] + \overline{Y}_{1,\overline{\tau}_1}[j] + \ldots \overline{Y}_{n,\overline{\tau}_n}[j]}{n} \to \mathbf{E}\overline{Y}_{1,\overline{\tau}_1}[j] \in (-\infty, \infty) \quad \text{a.s.}$$

Finally, the inequality in (51) follows from (44).

4 Proofs of Property 1 and Auxiliary Lemmas

In this section, we introduce a number of auxiliary notation, define shifts of "virgin" environment, formulate and prove a number of lemmas, and, finally, complete with the proof of Property 1.

4.1 Additional Notation

In the proofs we will frequently use notation

$$H_t(x) := H(x) - L_t(x) = H_{t-1}(x) - \mathbf{1}\{S_t = x\}, \quad t = 0, 1, 2, \ldots,$$

with $H_{-1}(x) := H(x)$. Thus, $H_t(x)$ is the number of allowed visits to state x after time $t + 0$.

We need a number of further notation. Let

$$h(n) := \min_{0 \le t \le n} H_t(S_t) = \min\{h(n-1),\ H_n(S_n) - 1\}, \quad n = 0, 1, 2, \dots .$$

It follows from (3) that, for all $n, N \ge 0$,

$$\{T_* > N\} = \{h(N) \ge 0\}, \tag{52}$$
$$B_n = \{\alpha(n) < T_*\} = \{h(\alpha(n)) \ge 0\} = \{h(\alpha(n) - 1) \ge 0\}.$$

The latter equality follows from condition $H(x) \ge 1$ for $x[1] \ge 0$.

In what follows, we consider a random walk that starts at time $t \ge 0$ from a state x, rather that at time $t = 0$ from the state S_0. The following notation will be helpful:

$$\alpha_t(l) = \inf\{j \ge 0 : S_{t,t+j}[1] = l\}, \quad h_t(l, x) := \inf_{0 \le j < \alpha_t(l)} H_{t+j}(x + S_{t,t+j}), \tag{53}$$
$$s(t, L) := \inf_{0 \le j < L} S_{t,t+j}[1], \qquad s_t(l) := s(t, t + \alpha_t(l)), \tag{54}$$

for $t, l, L \ge 0$, where notation $S_{t,j} := S_j - S_t$ for $t \ge j$ was introduced earlier. Note that $\alpha_0(l) = \alpha(l)$ for all $l \ge 0$.

Later on we will use the following properties of notation from (53) and (54):

$$\alpha(l + m) = \alpha(l) + \alpha_{\alpha(l)}(m), \{s(0, T + \alpha_T(l)) \ge 0\} = \{s(0, T) \ge 0, S_T + s_T(l) \ge 0\},$$
$$\{0 < T \le T + \alpha_T(l) < T_*\} = \{T > 0,\ h(T - 1) \ge 0,\ h_T(l, S_T) \ge 0\}, \tag{55}$$

for any random or non-random $T \ge 0$ and each $l \ge 0$ and $m > 0$.

Note that, given $\alpha(n) < \infty$,

$$S_{\alpha(n)} \in \mathbb{Z}^d_n := \{x = (x[1], x[2], \dots, x[d]) \in \mathbb{Z}^d : x[1] = n\}.$$

4.2 Shifts of Virgin Environment

For any $j \ge t \ge 0$, introduce random variable

$$L_{t,t+j}(x) = \sum_{k=0}^{j} \mathbf{1}\{S_{t,t+k} = x\}, \quad x \in \mathbb{Z}^d,$$

which, similarly to (2), counts the number of visits to state x within time interval $(t, t+j]$. For each $k \geq 0$, introduce the following (possibly, improper) random variables:

$$H^{(k)}(y) = \begin{cases} H(y), & y \in \mathbb{Z}^d_{k+}, \\ \infty, & y \notin \mathbb{Z}^d_{x[1]+}, \end{cases} \quad \text{so that } H^{(x[1])}(x+y) = \begin{cases} H(x+y), & y \in \mathbb{Z}^d_{0+}, \\ \infty, & y \notin \mathbb{Z}^d_{0+}, \end{cases} \tag{56}$$

for all $x \in \mathbb{Z}^d_{0+}$. For $t, k \geq 0$ and $x \in \mathbb{Z}^d_{0+}$, let

$$h_t^{(k)}(l, x) := \inf_{0 \leq j < \alpha_t(l)} [H^{(k)}(x + S_{t,t+j}) - L_{t,t+j}(x + S_{t,t+j})]. \tag{57}$$

The function $H^{(k)}(y)$ describes the environment which is virgin for all $y \in \mathbb{Z}^d_{k+}$ and which has no restrictions on the number of visits to all states $y \notin \mathbb{Z}^d_{k+}$. The function $h_t^{(k)}(l, x)$ describes the behaviour in this environment of a random walk that starts at time $t \geq 0$ from the state x. Inequality (73) below shows that this environment has characteristics that dominate the corresponding characteristics of any of our initial environments.

Note that

$$\{h_t^{(k)}(l, x) = h_t(l, x), \ s_t(l) \geq 0\} \subset \{\sup_{0 \leq j < t} S_j[1] < k\}. \tag{58}$$

We use symbol ∞ in place of 0 in (56) because we like to use in Sect. 5 the following result (with $\mathbf{P}(\cdot) = \mathbf{P}_+(\cdot)$):.

Lemma 1 *Under the assumptions* (A1)–(A3) *and for each fixed* $l \geq 0$, *given the event* $\{S_t = x\}$ *occurs, the joint conditional distribution of the random variables from the following family*

$$\alpha_t(l), \ s_t(l), \ \mathbf{S}_{t,t+\alpha_t(l)}, \ h_t^{(x[1])}(l, x), \ ; \xi_{t+j}, \ j \geq 1$$

does not depend on $t \geq 0$ *and on* $x \in \mathbb{Z}^d_{0+}$. *In particular, for all* $\mathscr{C} \subset \mathbb{Z}^d_*$

$$\begin{aligned} &\mathbf{P}(\mathbf{S}_{t,t+\alpha_t(l)} \in \mathscr{C}, \ h_t^{(x[1])}(l, x) \geq 0, \ s_t(l) \geq 0 | S_t = x) \\ &= \mathbf{P}(\mathbf{S}_{0,\alpha_0(l)} \in \mathscr{C}, \ h_0(l, 0) \geq 0, \ s_0(l) \geq 0) \\ &= \mathbf{P}_0(\alpha(l) < T_*, \ \mathbf{S}_{0,\alpha(l)} \in \mathscr{C}). \end{aligned} \tag{59}$$

Proof The first assertion follows directly from assumptions (A1)–(A3) and, in particular, from the time/space homogeneity of the random walk and from the homogeneity of the random environment in the positive half-space $\mathbb{Z}^d_{0+}$. To get (59) we use (58) too. □

4.3 Auxiliary Lemmas

Suppose that a random variable $T \geq 0$ is such that

$$\{T \geq 0\} = \cup_{t=0}^{\infty}\{T = t,\ S_T \in \mathbb{X}(t)\} \quad \text{for some} \quad \mathbb{X}(t) \subset \mathbb{Z}^d. \tag{60}$$

For a fixed $l > 0$ and arbitrary sets $\mathscr{A}, \mathscr{C} \subset \mathbb{Z}_*^d$, consider the event

$$\tilde{D} := \{T + \alpha_T(l) < T_*,\ \mathbf{S}_T \in \mathscr{A},\ \mathbf{S}_{T,T+\alpha_T(l)} \in \mathscr{C}\}. \tag{61}$$

Using (52) and (55), we may represent (61) in the form

$$\tilde{D} = \{T < \infty,\ h(T-1) \geq 0,\ h_T(l, S_T) \geq 0,\ \mathbf{S}_T \in \mathscr{A},\ \mathbf{S}_{T,T+\alpha_T(l)} \in \mathscr{C}\}.$$

For fixed $t \geq 0$ and $x \in \mathbb{Z}^d$, introduce events

$$\tilde{A}_{t,x} :=\{T{=}t,\ h(t{-}1) \geq 0,\ \mathbf{S}_t \in \mathscr{A},\ S_t{=}x\},\ \ \tilde{C}_{t,x} :=\{h_t(l,x) \geq 0,\ \mathbf{S}_{t,\alpha(n)} \in \mathscr{C}\}.$$

Clearly,

$$\mathbf{P}_*(\tilde{D}) = \sum_{t=0}^{\infty} \sum_{x \in \mathbb{X}(t)} \mathbf{P}_*(\tilde{A}_{t,x} \cdot \tilde{C}_{t,x}). \tag{62}$$

Thus, we have the following elementary

Lemma 2 *Suppose that a random variable $T \geq 0$ satisfies condition (60). Then for all $\mathscr{A}, \mathscr{C} \subset \mathbb{Z}_*^d$ and each $l > 0$ equality (62) takes place. In addition, if for all $t \geq 0$ and $x \in \mathbb{X}(t)$ events $\tilde{A}_{t,x}$ and $\tilde{C}_{t,x}$ are pairwise independent and $\mathbf{P}_*(\tilde{C}_{t,x})$ does not dependent on $t \geq 0$ and $x \in \mathbb{X}(t)$, then we have*

$$\mathbf{P}_*(\tilde{D}) = \mathbf{P}_*(T < T_*,\ \mathbf{S}_T \in \mathscr{A}) \cdot \mathbf{P}_*(\alpha(l) < T_*,\ \mathbf{S}_{0,\alpha(l)} \in \mathscr{C}). \tag{63}$$

One can observe that the sequence $\{S_t,\ H_t(x) : x \in \mathbb{Z}^d\}$, $t = 0, 1, 2, \ldots,$ of infinite-dimensional random variables forms an infinite-dimensional Markov chain. In the proofs below we apply Lemma 2 four times for stopping times $T \geq 0$ of this Markov chain.

Lemma 3 *Under the assumptions* $(A1)$–$(A3)$,

$$\begin{aligned} &\mathbf{P}_*(\alpha(n) < T_*,\ \mathbf{S}_{\alpha(k)} \in \mathscr{A},\ s(\alpha(k), \alpha(n)) \geq 0,\ \mathbf{S}_{\alpha(k),\alpha(n)} \in \mathscr{C}) \\ &\quad = \mathbf{P}_*(\alpha(k) < T_*,\ \mathbf{S}_{\alpha(k)} \in \mathscr{A}) \cdot \mathbf{P}_0(\alpha(l) < T_*,\ \mathbf{S}_{0,\alpha(l)} \in \mathscr{C}) \end{aligned} \tag{64}$$

for all $\mathscr{A}, \mathscr{C} \subset \mathbb{Z}_*^d$ and each $n > k \geq 0$ (where $l := n - k > 0$).

Proof We will apply Lemma 2 with $T = \alpha(l)$ and $\mathbb{X}(t) = \mathbb{Z}_k^d$. We have from Lemma 1 that probability $\mathbf{P}_*(C_{t,x})$ does not depends on $t \geq 0$ and $x \in \mathbb{Z}_k^d$. Hence, by (58)

$$\mathbf{P}_*(\tilde{C}_{t,x}) = \mathbf{P}_*(\tilde{C}_{0,0}) = \mathbf{P}_0(\alpha(l) < T_*,\ \mathbf{S}_{0,\alpha(l)} \in \mathscr{C}). \tag{65}$$

For any fixed $t \geq 0$ and $x \in \mathbb{Z}_k^d$, random variables $\alpha(k)$, $\mathbf{S}_t$ and $h(t-1)$ that define the event $\tilde{A}_{t,x}$ are functions only of the variables from the following two families:

$$\{\xi_j : j \leq t\} \quad \text{and} \quad \{H(y) : y \notin \mathbb{Z}_{k+}^d\}. \tag{66}$$

On the other hand, all random variables that determine the event $\tilde{C}_{t,x}$, are functions only of random variables from the following two families:

$$\{\xi_j : j > t\} \quad \text{and} \quad \{H(y) : y \in \mathbb{Z}_{k+}^d\}. \tag{67}$$

Since the families in (67) and (66) do not overlap, they are independent. Hence, events $\tilde{A}_{t,x}$ and $\tilde{C}_{t,x}$ are independent too. This fact, together with (65), allows us to apply Lemma 2 to get (64). □

Lemma 4 *Under the assumptions* $(A1)$–$(A3)$*,*

$$\begin{aligned} &\mathbf{P}_*(\alpha(n) < T_*,\ \mathbf{S}_{\alpha(k)} \in \mathscr{A},\ \varkappa_*(n) = k,\ \mathbf{S}_{\alpha(k),\alpha(n)} \in \mathscr{C}) \\ &= \mathbf{P}_*(\alpha(k) < T_*,\ \mathbf{S}_{\alpha(k)} \in \mathscr{A}) \cdot \mathbf{P}_0(\varkappa_*(n-k) = 0,\ \mathbf{S}_{0,\alpha_0(n-k)} \in \mathscr{C}) \end{aligned} \tag{68}$$

for any $n > k \geq 0$ *and all* $\mathscr{A}, \mathscr{C} \subset \mathbb{Z}_*^d$.

Proof For each $n \geq 1$ introduce the following subset of $\mathbb{Z}_*^d$:

$$\mathscr{C}_n^+ := \{(y_1, y_2, \ldots) \in B_n^+ : \varkappa(n|\mathbf{y}) = 0 \quad \text{for} \quad \mathbf{y} = (0, y_1, y_2, \ldots)\}. \tag{69}$$

We assume in (69) that $y_0 = 0$ to avoid problems with the definition of the value $\alpha(n|\mathbf{y})$. It follows from (69) that

$$\{\alpha(n) < T_*, \varkappa_*(n) = k\} = \{\alpha(n) < T_*,\ s(\alpha(k), \alpha(n)) \geq 0,\ \mathbf{S}_{\alpha(k),\alpha(n)} \in \mathscr{C}_{n-k}^+\}, \tag{70}$$

where in (70) we used also that $\{\varkappa_*(n) = k\} \subset \{s(\alpha(k), \alpha(n)) \geq 0\}$.

If we compare now (68) and (70) with (64), we can observe that (68) is a particular case of (64), given that we replace in (64) $\mathscr{C}$ by $\mathscr{C} \cap \mathscr{C}_{n-k}^+$. □

4.4 Proof of Property 1

The first assertion of Property 1 immediately follows from Lemma 4 with $\mathscr{A} = \mathscr{C} = \mathbb{Z}_*^d$ since, in this case, we have from (68) that

$$\mathbf{P}_*(B_n, \varkappa_*(n) = k) = \mathbf{P}_*(\alpha(n) < T_*, \ \varkappa_*(n) = k) \tag{71}$$
$$= \mathbf{P}_*(\alpha(k) < T_*) \cdot \mathbf{P}_0(\varkappa_*(n-k) = 0) = \mathbf{P}_*(B_k) \cdot \mathbf{P}_0(\varkappa_*(n-k) = 0).$$

Since $B_n = \{\eta_*(n) \geq 0\}$ and $\{\eta_*(n) \geq 1\} = \{\varkappa_*(n) \geq 0\}$ by (20), we have, for $n = 1, 2, \ldots$,

$$\mathbf{P}_*(B_n) = \mathbf{P}_*(\eta_*(n) = 0) + \mathbf{P}_*(\eta_*(n) \geq 1)$$
$$= \mathbf{P}_*(\eta_*(n) = 0) + \sum_{k=0}^{n-1} \mathbf{P}_*(\alpha(n) < T_*, \ \varkappa_*(n) = k). \tag{72}$$

Thus, Property 1 follows from (71) and (72).

5 Proof of Theorem 1

We will use functions $H^{(k)}(y)$ and $h_t^{(k)}(l, x)$ introduced in (56) and (57), that have been already applied in Lemma 1. These functions have the following useful properties:

$$H^{(k)}(y) \geq H_t(y) \quad \text{and} \quad h_t^{(k)}(l, x) \geq h_t(l, x) \quad \forall y \in \mathbb{Z}^d, \ \forall x \in \mathbb{Z}_{0+}^d, \forall \, t, l \geq 0. \tag{73}$$

5.1 Main Lemma

We are going to prove

Property 2 Under the assumptions $(A1)$–$(A4)$, there exists a constant $C < \infty$ such that

$$\forall \, n \geq 0 \quad \mathbf{P}_+(B_n) \leq C\mathbf{P}_0(B_n). \tag{74}$$

The proof is based on several lemmas. Introduce the following stopping time:

$$\rho := \inf\{t > 0 : S_t[1] = 0 \quad \text{but} \quad S_t \neq 0\} \leq \infty.$$

So ρ is the time of the first return to level 0 by the first component of our random walk, given that at least one of other coordinates differs from 0.

Lemma 5 *For any $n > 0$*

$$P_\rho := \mathbf{P}_+(\rho < \alpha(n) < T_*) \leq \mathbf{P}(\rho < \infty) \cdot \mathbf{P}_+(B_n). \tag{75}$$

Proof It follows from (55) that

$$\{\rho < \alpha(n) < T_*\} = \{\rho < \infty,\ \alpha_\rho(n) < \infty,\ h(\rho - 1) \geq 0,\ h_\rho(n, S_\rho) \geq 0\}.$$

Since $h_\rho(n, S_\rho) \leq h_\rho^{(0)}(n, S_\rho)$ by (73), we have

$$\{\rho < \alpha(n) < T_*\} \subset D := \{\rho < \infty,\ \alpha_\rho(n) < \infty,\ h_\rho^{(0)}(n, S_\rho) \geq 0\}. \tag{76}$$

Introduce the events

$$A_{t,x} := \{\rho = t,\ S_t = x\}, \quad C_{t,x} := \{\alpha_t(n) < \infty,\ h_t^{(0)}(n, x) \geq 0\}.$$

By Lemma 1, probability $\mathbf{P}(C_{t,x})$ does not depends on $t \geq 0$ and $x \in \mathbb{Z}_0^d$. Hence, by (58)

$$\mathbf{P}_*(C_{t,x}) = \mathbf{P}_*(\alpha_0(n) < \infty,\ h_0^{(0)}(\alpha_0(n), 0) \geq 0) = \mathbf{P}_+(\alpha_0(n) < T_*) = \mathbf{P}_+(B_n), \tag{77}$$

since $\alpha_0(n) = \alpha(n)$.

Now we apply Lemma 2 with $T = \rho$ and $\mathbb{X}(t) = \mathbb{X}_0 := \mathbb{Z}_0^d \setminus \{0\}$, and with $h_k^{(k)}(l, x)$ in place of $h_k(l, S_k)$. For fixed values $t > 0$ and $x \in \mathbb{Z}_0^d$, random variables $\alpha_t(n)$ and $h_t(n, x)$ are functions only of random variables from (67) with $k = 0$, since $H(y) = \infty$ for all $y \notin \mathbb{Z}_{k+}^d$.

On the other hand, event $A_{t,x}$ does not depend on the environment and is determined by the variables $\{\xi_j : j \leq t\}$. Hence, events $\tilde{A}_{t,x}$ and $\tilde{C}_{t,x}$ do not depend on each other, and we may apply Lemma 2. Using also (77) and (76), we obtain

$$\mathbf{P}_*(\rho < \alpha(n) < T_*) \leq \mathbf{P}_*(D) = \sum_{t=1}^{\infty} \sum_{x \in \mathbb{Z}_0^d \setminus \{0\}} \mathbf{P}(A_{t,x}) = \sum_{t=1}^{\infty} \mathbf{P}(\rho = t, S_t \neq 0) = \mathbf{P}(\rho < \infty).$$

Thus (75) is proved. ◻

According to (18) introduce the following stopping times:

$$\rho_0 = 0 \quad \text{and} \quad \rho_i := \inf\{t > \rho_{i-1} : S_t = 0\} \leq \infty, = \quad i = 1, 2, \ldots.$$

So ρ_i is the time of the i-th return to 0 of our random walk. It is easy to see that, for any $n > 0$,

$$\mathbf{P}_+(B_n) \le \mathbf{P}_+(\rho < \alpha(n) < T_*) + \sum_{i=0}^{\infty} \mathbf{P}(D_i), \tag{78}$$

where

$$D_i = D_i(n) := \{\rho_i < \alpha(n) < \min(\rho_{i+1}, \rho) \le \infty,\ \alpha(n) < T_*\}.$$

Lemma 6 *For any $n > 0$*

$$\mathbf{P}_+(D_i) \le \mathbf{P}(\rho_i < \infty) \cdot \mathbf{P}(H(0) > i) \cdot \mathbf{P}_0(B_n). \tag{79}$$

Proof Underline that, on the event $\{\rho_i < \alpha(n) < \min(\rho_{i+1}, \rho)\}$, we have that $s(\rho_i, \alpha(n)) > 0$, due to the skip-free property of the random walk. Thus

$$D_i \subset \hat{D}_i := \{\rho_i < \alpha(n) < T_*,\ s(\rho_i, \alpha(n)) > 0\}. \tag{80}$$

Since $S_{\rho_i} = 0$, we have from (55) that

$$\hat{D}_i = \{\rho_i < \infty,\ \rho_i < \rho,\ \alpha_{\rho_i}(n) < \infty,\ h(\rho_i - 1) \ge 0,\ h_{\rho_i}(n, 0) \ge 0\}. \tag{81}$$

Since $H_{t+j}(y) \le H_j(y)$ for all $y \in \mathbb{Z}^d$ and $t, j \ge 0$, we have from (53) that

$$h_t(n, 0) = \inf_{0 \le j < \alpha_t(n)} H_{t+j}(S_{t,t+j}) \le \tilde{h}_t(n, 0) := \inf_{0 \le j < \alpha_t(n)} H_j(S_{t,t+j}) \tag{82}$$

for all possible $t \ge 0$. Note also that $h(\rho_i) \le H_{\rho_i}(0)$. This fact and (80)–(82) with $t = \rho_i$ yield

$$\hat{D}_i \subset \tilde{D}_i := \{\rho_i < \infty,\ \rho_i < \rho,\ \alpha_{\rho_i}(n) < \infty,\ H_{\rho_i}(0) \ge 0,\ \tilde{h}_{\rho_i}(n, 0) \ge 0,\ s(\rho_i, \alpha(n)) > 0\}.$$

Introduce the events

$$A_{i,t} := \{\rho > \rho_i = t,\ H_t(0) \ge 0\},$$
$$C_t := \{\alpha_t(n) < \infty,\ \tilde{h}_t(n, 0) \ge 0,\ s(t, \alpha(n)) > 0\}. \tag{83}$$

Comparing definition (82) with that in (53) and (54), we can see that $\tilde{h}_t(n, 0) = h_t(n, 0)$ and that probability $\mathbf{P}_*(C_t)$ does not depend on $t \ge 0$. Hence,

$$\mathbf{P}_*(C_t) = \mathbf{P}_*(C_0) = \mathbf{P}_*(\alpha_0(n) < \infty,\ \tilde{h}_0(n, 0) \ge 0,\ s(0, \alpha(n)) > 0) \tag{84}$$
$$\le \mathbf{P}_*(\alpha_0(n) < \infty,\ h_0(n, 0) \ge 0,\ s(0, \alpha(n)) \ge 0) = \mathbf{P}_0(\alpha_0(n) < T_*) = \mathbf{P}_0(B_n),$$

because $\alpha_0(n) = \alpha(n)$. We have also from (83) that

$$\mathbf{P}_*(A_{i,t}) = \mathbf{P}(\rho > \rho_i = t)\mathbf{P}(H_t(0) \geq 0) = \mathbf{P}(\rho > \rho_i = t)\mathbf{P}(H(0) > i), \quad (85)$$

because $H_{\rho_i}(0) = H_0(0) - i = H(0) - i - 1$.

Now we apply Lemma 2 with $T = \rho_i$ and $\mathbb{X}(t) = \{0\}$, and with $\tilde{h}_t(n, 0)$ in place of $h_t(n, 0)$. Now note that, for each value $t > 0$, under the condition $s(t, \alpha(n)) > 0$, the random variables $\alpha_t(n)$, $s(t, \alpha(n))$ and $h_t(n, x)$ (which determine event C_t) are functions only of random variables from (67), with $k = 1$. On the other hand, event $A_{i,t}$ is defined by the variable $H_t(0)$ and by the family $\{\xi_j : j \leq t\}$. Hence, the events $A_{i,t}$ and C_t are independent and we may apply Lemma 2 again. Using also (84) and (85), we get:

$$\mathbf{P}_+(D_i) \leq \mathbf{P}_+(\hat{D}_i) \leq \mathbf{P}_+(D_i) = \sum_{t=1}^{\infty} \mathbf{P}(A_{i,t})\mathbf{P}(C_t) \leq \sum_{t=1}^{\infty} \mathbf{P}(\rho > \rho_i = t)\mathbf{P}(H(0) > i)\mathbf{P}_0(B_n).$$

So, inequality (79) follows. □

Introduce the notation

$$p^* := \mathbf{P}(\rho < \infty), \quad p_1 := \mathbf{P}(\rho_1 < \infty,\ \rho_1 < \rho),$$
$$C^* := \sum_{i=0}^{\infty} p_1^i \mathbf{P}(H(0) > i) \leq 1 + \mathbf{E}H(0).$$

Substituting the results of Lemmas 5 and 6 into (78), we obtain

$$1 - \mathbf{P}_+(B_n) \leq C^* \mathbf{P}_0(B_n) + p^* \mathbf{P}_+(B_n).$$

Thus, under the assumptions $(A1)$–$(A3)$,

$$(1 - p^*)\mathbf{P}_+(B_n) \leq C^* \mathbf{P}_0(B_n). \quad (86)$$

One can easily conclude that, under any of assumptions (a)–(c) in $(A4)$, the following inequalities hold:

$$p_* < 1 \quad \text{and} \quad C_* < \infty. \quad (87)$$

Here is the only place in the paper where the assumption $(A4)$ is used.

From (86) and (87) we obtain the assertion of Property 2 with $C = C^*/(1 - p^*)$.

Note that for $n = 0$ inequality (74) follows from (16) since $C \geq 1$. □

5.2 Using Submultiplicativity

In this subsection we prove first that

$$\forall k, l \geq 0 \quad \mathbf{P}_*(B_k)\mathbf{P}_0(B_l) \leq \mathbf{P}(B_{k+l}) \leq \mathbf{P}(B_k)\mathbf{P}_+(B_l). \tag{88}$$

Using this form of sub/supermultiplicativity we show that

$$1 \geq q_+ := \inf_{n\geq 1} \sqrt[n]{\mathbf{P}_+(B_n)} = q := \sup_{n\geq 1} \sqrt[n]{\mathbf{P}_-(B_n)} \geq \mathbf{P}(\xi_1[1] = 1) > 0. \tag{89}$$

After that, we prove the following

Property 3 Under the assumptions $(A1)$–$(A4)$, relations (88) and (89) take place. Moreover

$$0 < \mathbf{P}_*(B_0)/C \leq \mathbf{P}_*(B_n)/q^n \leq C\mathbf{P}_*(B_0) \leq C < \infty \quad \forall\, n \geq 0. \tag{90}$$

Note that for $l = 0$ inequality (88) immediately follows from (16). We prove now the following lemma.

Lemma 7 *Under the assumptions* $(A1)$–$(A3)$*, inequality* (88) *takes place for all* $k \geq 0$ *and* $l > 0$.

Proof Applying Lemma 3 with $n = k + l$, we get

$$\begin{aligned}\mathbf{P}_*(B_{k+l}) &= \mathbf{P}_*(\alpha(k+l) < T_*) \geq \mathbf{P}_*(\alpha(l+k) < T_*,\ s(\alpha(k), \alpha(k+l)) \geq 0)\\ &= \mathbf{P}_*(\alpha(k) < T_*) \cdot \mathbf{P}_0(\alpha(l) < T_*) = \mathbf{P}_*(B_k)\mathbf{P}_0(B_l).\end{aligned}$$

This is the first inequality in (88).

Next, it follows from (55) that

$$B_{k+l} = \{\alpha(k+l) < T_*\} = \{\alpha(k) < \infty,\ \alpha_k(l) < \infty,\ h(k-1) \geq 0,\ h_k(l, S_k) \geq 0\}.$$

Since $h_k(l, S_k) \leq h_k^{(k)}(l, S_k)$ by (73), we have

$$B_{k+l} \subset \tilde{D} := \{\alpha(k) < \infty,\ \alpha_k(l) < \infty,\ h(k-1) \geq 0,\ h_k^{(k)}(l, S_k) \geq 0\}.$$

Now we apply Lemma 2 with the same $T = \alpha(l)$ and $\mathbb{X}(t) = \mathbb{Z}_k^d$ as in the proof of Lemma 3, but with $h_k^{(k)}(l, x)$ in place of $h_k(l, S_k)$. Introduce the events

$$\tilde{A}_{t,x} := \{\alpha(k) = t, ,\ h(t-1) \geq 0,\ S_t = x\}, \quad \tilde{C}_{t,x} := \{\alpha_k(l) < \infty,\ h_k^{(k)}(l, x) \geq 0\}.$$

By Lemma 1, probability $\mathbf{P}(\tilde{C}_{t,x})$ does not depends on $t \geq 0$ and $x \in \mathbb{Z}_k^d$,

$$\mathbf{P}_*(\tilde{C}_{t,x}) = \mathbf{P}_*(\tilde{C}_{0,0}) = \mathbf{P}_*(\alpha_0(l) < \infty,\ h_0^{(0)}(\alpha_0(l)) \geq 0) \\ = \mathbf{P}_+(\alpha_0(l) < T_*) = \mathbf{P}_+(B_l), \tag{91}$$

since $\alpha_0(l) = \alpha(l)$ for $l > 0$.

Now, for fixed values $t > 0$ and $x \in \mathbb{Z}_k^d$, random variables $\alpha_t(n)$ and $h_t(n, x)$, which define event $\tilde{C}_{t,x}$, are functions only of random variables from (67), since $H(y) = \infty$ for all $y \notin \mathbb{Z}_{k+}^d$. On the other hand, event $\tilde{A}_{t,x}$ is defined by random variables $\alpha(k)$, $h(t-1)$ and S_t which are functions of the variables from (67). Hence, events $\tilde{A}_{t,x}$ and $\tilde{C}_{t,x}$ are independent and we can apply Lemma 2. Using also (91), we obtain $\mathbf{P}_*(B_{k+l}) \leq \mathbf{P}_*(\tilde{D}) = \mathbf{P}_*(B_k)\mathbf{P}_+(B_l)$ as a result.

Thus, second inequality in (88) is proved. □

Proof (of Property 3) Using probabilities $\mathbf{P}_0(\cdot)$ and $\mathbf{P}_+(\cdot)$ instead of $\mathbf{P}_*(\cdot)$, we have from (88) and (16) that, for all $k, l \geq 1$,

$$\mathbf{P}_0(B_k)\mathbf{P}_0(B_{kl-k}) \leq \mathbf{P}_0(B_{kl}) \leq \mathbf{P}_+(B_{kl}) \leq \mathbf{P}_+(B_{kl-l})\mathbf{P}_+(B_l).$$

Then the induction argument leads to

$$(\mathbf{P}_0)^l(B_k) \leq \mathbf{P}_0(B_{kl}) \leq \mathbf{P}_+(B_{kl}) \leq (\mathbf{P}_+)^k(B_l). \tag{92}$$

Taking the kth root of the both sides of inequality (92), we arrive to

$$\forall\, k, l \geq 1 \quad \sqrt[k]{\mathbf{P}_0(B_k)} \leq \sqrt[l]{\mathbf{P}_+(B_l)}. \tag{93}$$

Taking in (93) supremum in $k \geq 1$ and infimum in $l \geq 1$, we obtain $q \leq q_+$.

On another hand, from (74) and the definition of q_+ in (89), we have

$$q_+^n \leq \mathbf{P}_+(B_n) \leq C\mathbf{P}_0(B_n) \leq Cq^n. \tag{94}$$

Hence, $q_+ \leq \sqrt[n]{C}q \to q$. So, we proved that $q_+ \leq q$ and hence (89) follows from (13) with $\mathbf{P}_0(B_n) \geq \mathbf{P}^n(\xi_1[1] = 1)$.

Next, it follows from (88) and (74) with $k = 0$ and $l = n$ that

$$\mathbf{P}_*(B_n) \leq \mathbf{P}_*(B_0)\mathbf{P}_+(B_n) \leq C\mathbf{P}_*(B_0)\mathbf{P}_0(B_n) \leq C\mathbf{P}_*(B_0)q^n \leq Cq^n. \tag{95}$$

Here we also used (94). On the other hand, using again (88), (74) and (94), we get

$$\mathbf{P}_*(B_n) \geq \mathbf{P}_*(B_0)\mathbf{P}_0(B_n) \geq \mathbf{P}_*(B_0)\mathbf{P}_+(B_n)/C \geq \mathbf{P}_*(B_0)q_+^n/C. \tag{96}$$

Now, all inequalities in (90) follow from (95) and (96). □

5.3 *Proof of Theorem 1*

With q from (89), introduce the following notation:

$$a_n := \frac{\mathbf{P}_0(\varkappa_*(n) = 0)}{q^n}, \; b_n := \frac{\mathbf{P}_*(\eta_*(n) = 0)}{q^n}, \; u_n := \frac{\mathbf{P}_*(B_n)}{q^n}, \; v_n := \frac{\mathbf{P}_0(B_n)}{q^n}. \tag{97}$$

Multiplying equalities (21) and (22) by q^{-n}, we obtain for all $n \geq 1$ that

$$u_n = b_n + \sum_{k=0}^{n-1} u_k a_{n-k} = b_n + \sum_{l=1}^{n} a_l u_{n-l}, \tag{98}$$

$$v_n = \sum_{l=1}^{n} a_l v_{n-l}, \quad \text{where} \quad v_0 = 1 \quad \text{and} \quad a_1 > 0. \tag{99}$$

The last property in (99) follows from (33).

We have from (97), (89) and (90) that

$$u_0 = b_0 = \mathbf{P}_*(B_0) > 0, \;\; 0 < u_0/C \leq u_n \leq C < \infty, \;\; 0 < 1/C \leq v_n \leq 1 \;\; \forall n \geq 1. \tag{100}$$

In addition, we have from (19), (33) and (100) that

$$v_1 = a_1 > 0, \quad 0 \leq a_n \leq v_n \quad \text{and} \;\; 0 < v_1^n \leq v_n \leq 1 \;\; \forall n \geq 1. \tag{101}$$

There are two possible scenarios, either $a_n < 1$ for all n or $a_M = 1$ for some $M \geq 1$. We start with the latter case which is, in fact, degenerative.

Lemma 8 *If $a_M = 1$ for some $M \geq 1$, then $M = 1$ and the assertions of Theorem 1 do hold with $q = \mathbf{P}_0(\varkappa_*(1) = 0)$.*

Proof Since $1 = a_M \leq v_M \leq 1$ by (101), we have $v_M = 1$. Then, by (99),

$$v_M - a_M = 0 = \sum_{l=1}^{M-1} a_l v_{n-l} \geq a_1 v_{M-1} > 0 \quad \text{if} \quad M > 1.$$

So we must have $M = 1$. Then $v_1 = a_1 = 1 = v_1^n \leq v_n \leq 1$ for all $n \geq 1$ by (101). Hence, $v_n = 1$ for all $n \geq 1$ and, by (99),

$$v_n - a_1 v_{n-1} = 1 - 1 = 0 = \sum_{l=2}^{n} a_l v_{n-l} = \sum_{l=2}^{n} a_l \quad \text{when} \quad n \geq 2.$$

Thus, $a_l = 0$ for all $l \geq 2$ and Eq. (26) reduces to $\mathbf{P}_0(\varkappa_*(1) = 0)/q = 1$. Hence all assertions of Theorem 1 hold with $q = \mathbf{P}_0(\varkappa_*(1) = 0)$. □

Consider now the main case where $0 \le a_k < 1$ for all $k \ge 1$. It is known (see, for example, Section 13.4 in the 1st Volume of the Feller's book [9]) that there are only four possibilities for the solutions to equation (99):

(a) $0 < \alpha := \sum_{k\ge1} a_k < 1$ and $v_n \to 0$;
(b) $\alpha = 1$, $\mu = \sum_{k\ge1} k a_k = \infty$ and $v_n \to 0$;
(c) $\alpha = 1$, $1 \le \mu < \infty$ and $v_n \to 1/\mu > 0$ since $a_1 > 0$;
(d) $\alpha \in (1,\infty]$ and $v_n \to \infty$.

It is easy to see that (c) is the only possibility which does not contradict to inequalities (100). Hence, $\alpha = 1$, $\mu < \infty$, and (26) with (27) follow.

Now, we again use [9] to evaluate $\psi_0 = \sum_{k\ge1} b_k$. From (100) and (98) with $v_n \to 1/\mu > 0$ we obtain

$$C \ge u_n = \sum_{k=0}^{n} b_k u_{n-k} \to \sum_{k\ge0} b_k/\mu = \psi_0/\mu \ge b_0/\mu = \mathbf{P}_*(B_0)/\mu > 0$$

by assumption $(A2)$. So, we obtain inequality (28) with $C\mu \ge \psi_0 \ge \mathbf{P}_*(B_0) > 0$.

Thus, Theorem 1 is proved.

6 Proof of Theorem 2

We suppose that assumptions $(A1)$–$(A4)$ continue to hold.

We start with a few preliminary comments. If follows directly from (31) and (34) that, for any integers $K \ge k \ge 0$ and all vectors $\mathbf{y}_K = (y_0,\dots,y_K) \in \mathbb{Z}^{(K+1)\times d}$,

$$\begin{aligned} &\mathbf{P}_*(\alpha(k) = K < T_*, \eta_*(k) = 0, \mathbf{S}_K = \mathbf{y}_K) \\ &= \psi_0 q^k \mathbf{P}(\overline{v}_0 = k,\ \overline{T}_0 = K,\ \widetilde{S}_K = y_K)). \end{aligned} \tag{102}$$

Similarly, it follows from (32) and (35) that, for any integers $L \ge l \ge 1$ and all $\mathbf{x}_L = (x_1,\dots,x_L) \in \mathbb{Z}^{L\times d}$

$$\begin{aligned} \mathbf{P}_0(\alpha(l) = L < T_*, \varkappa_*(l) = 0, \mathbf{S}_{0,L} = \mathbf{x}_L) = q^l \mathbf{P}(\overline{\lambda}_0 = l, \overline{\tau}_0 = L, \widetilde{S}_{0,L} = \mathbf{x}_L) \\ = q^l \mathbf{P}(\overline{\lambda}_m \equiv \overline{v}_m - \overline{v}_{m-1} = l, \overline{\tau}_m \equiv \overline{T}_m - \overline{T}_{m-1} = L, \widetilde{S}_{\overline{T}_{m-1},\overline{T}_m} = \mathbf{x}_L). \end{aligned} \tag{103}$$

In the proof of the following lemma we repeat in more detail the description of the core random sequence, introduced in Sect. 3.3.

Lemma 9 *Suppose that numbers $N \ge n \ge m \ge 0$ and vector $\mathbf{y}_N = (y_0,\dots,y_N) \in \mathbb{Z}^{(N+1)\times d}$ are such that*

$$\alpha(n|\mathbf{y}_N) = N \ge 0 \quad \textit{and} \quad \eta(n|\mathbf{y}_N) = m \ge 0. \tag{104}$$

Then

$$\mathbf{P}_*(\alpha(n) = N < T_*, \eta(n) = m, \mathbf{S}_N = \mathbf{y}_N) \tag{105}$$
$$= \psi_0 q^n \mathbf{P}(\overline{\nu}(m) = n, \overline{\alpha}(n) = N, \widetilde{S}_N = \mathbf{y}_N).$$

Moreover, all random variables in (105) *are deterministic functions only of random variables from the initial block and from the first m blocks in* (40).

We will prove the lemma by induction in m. For $m = 0$, (105) follows from (102) (with k in place of n and K in place of N) that has been verified already.

Let m be a strictly positive number and suppose that (105) holds for all possible N and $\mathbf{y}_N$ in the case $\eta_*(n) = m - 1 \geq 0$. Now take the numbers and a vector satisfying (104). Then, for some integers k and K,

$$\varkappa(n|\mathbf{y}_N) = k \in [0, n-1] \quad \text{and} \quad \alpha(k|\mathbf{y}_N) = K \in [0, N-1]. \tag{106}$$

Let

$$\mathbf{y}_K = (y_0, \ldots, y_K), \quad \mathbf{y}_{K,N} = (y_{K+1} - y_K, \ldots, y_N - y_K), \quad N > K \geq 0. \tag{107}$$

We have from (106) that

$$\{\alpha(n) < T_*, \eta(n) = m, \varkappa(n) = k, \} = \{\alpha(n) < T_*, \eta(k) = m-1, \varkappa(n) = k\}. \tag{108}$$

Hence, by (107) and (108),

$$\mathbf{P}_*(\alpha(n) = N < T_*, \eta(n) = m, \mathbf{S}_N = \mathbf{y}_N) \tag{109}$$
$$= \mathbf{P}_*(\alpha(n) = N < T_*, \eta(n) = m, \varkappa(n) = k, \alpha(k) = K, \mathbf{S}_K = \mathbf{y}_K, \mathbf{S}_{K,N} = \mathbf{y}_{K,N}).$$

Now we apply Lemma 4 with special sets $\mathscr{A} = \{\mathbf{y}_K\}$ and $\mathscr{C} = \{\mathbf{y}_{K,N}\}$ containing only one trajectory each. Then

$$\mathbf{P}_*(\alpha(n) = N < T_*, \eta(n) = m, \mathbf{S}_N = \mathbf{y}_N) \tag{110}$$
$$= \mathbf{P}_*(\alpha(k) = K < T_*, \eta(k) = m-1, \mathbf{S}_K = \mathbf{y}_K)$$
$$\times \mathbf{P}_0(\alpha(l) = L < T_*, \varkappa(l) = 0, \mathbf{S}_{0,L} = \mathbf{y}_{K,N}).$$

Clearly, $\eta_*(k) = m - 1$ by (108). Hence, by the induction base, we have that

$$\mathbf{P}_*(\alpha(k) = K < T_*, \eta_*(k) = m-1, \mathbf{S}_K = \mathbf{y}_K) \tag{111}$$
$$= \psi_0 q^k \mathbf{P}(\widetilde{S}_K = \mathbf{y}_K, \overline{\alpha}(k) = K, \overline{\nu}_{m-1} = k).$$

Now use (103) with

$$\overline{\nu}_m = n, \quad \overline{\nu}_{m-1} = \varkappa_*(\overline{\nu}_m) = k, \qquad (112)$$
$$\overline{T}_m = \overline{\alpha}(\overline{\nu}_m) = N, \quad \overline{T}_{m-1} = \overline{\alpha}(\overline{\nu}_{m-1}) = K.$$

Let $l = n - k$, $L = N_K$ and $\mathbf{x}_L = \mathbf{y}_{K,N}$. Substituting (111) and (103) into (110), we obtain from (110) and (112) that

$$\begin{aligned} &\mathbf{P}_*(\alpha(n) = N < T_*, \eta_*(n) = m, \mathbf{S}_N = \mathbf{y}_N) \\ &= \psi_0 q^k \mathbf{P}(\widetilde{S}_K = \mathbf{y}_K, \overline{\alpha}(k) = K, \overline{\nu}_{m-1} = k) \\ &\times q^l \mathbf{P}(\overline{\nu}_m - \overline{\nu}_{m-1} = l, \overline{T}_m - \overline{T}_{m-1} = L, \widetilde{S}_{\overline{T}_{m-1},\overline{T}_m} = \mathbf{x}_L). \end{aligned} \qquad (113)$$

Notice that the m-th block in (40) is independent of the previous ones. Hence, (113) may be represented as

$$\begin{aligned} &\mathbf{P}_*(\alpha(n) = N < T_*, \eta_*(n) = m, \mathbf{S}_N = \mathbf{y}_N) \\ &= \psi_0 q^{k+l} \mathbf{P}(\overline{\nu}_m = n, \overline{T}_m = \overline{\alpha}(\overline{\nu}_m) = N, \widetilde{S}_{\overline{T}_{m-1}} = \mathbf{y}_K, \widetilde{S}_{\overline{T}_{m-1},\overline{T}_m} = \mathbf{x}_L = \mathbf{y}_{K,N}) \\ &= \psi_0 q^n \mathbf{P}(\overline{\nu}(m) = n, \overline{\alpha}(n) = N, \widetilde{S}_N = \mathbf{y}_N). \end{aligned}$$

So, we have completed the induction step. This ends the proof of Lemma 9. □

To prove Theorem 2, note that any set $\mathscr{A} \in \mathbb{Z}_*^d$ may be represented as

$$\mathscr{A} = \cup_{N=0}^{\infty} A_N, \quad \text{where} \quad A_N \subset \mathbb{Z}^{(N+1)\times d}, \quad N = 0, 1, 2, \ldots.$$

So, all vectors $\mathbf{y}_N = (y_0, \ldots, y_N)$ from A_N satisfy (104).

Then summing up the LHS's and RHS's of (105) over N and $\mathbf{y}_N \in A_N$ leads to (41).

Thus, we have finished with the proofs of all our results.

7 Remarks

Remark 1 In our Assumptions (A1)–(A3), we assume that the environment is "virgin" only in a half-space and that the random walk starts either from the other half-space or from a boundary point. Here is a scenario that may lead to such situation.

Assume that at some time instant $-\infty \le -N < 0$ in the past the whole environment in Z^d was "virgin", i.e. all the random variables $\{H_{-N}(x),\ x \in \mathbb{Z}^d\}$ were i.i.d. Assume that our random walk had started at time $t > -N$. This assumption implies that

$$H(x) = H_{-1}(x) \le H_{-N}(x), \quad \forall x \in \mathbb{Z}^d. \qquad (114)$$

We then assume that the trajectory of our random walk on the time interval $-N < t < 0$ is unobservable (it is the "dark history"), and that we start to observe the trajectory only at time $t = 0$ when we realize that the environment is still virgin in the half-space $\mathbb{Z}^d_{0+}$ (see (11) for definition), so that

$$H(x) = H_{-1}(x) = H_{-N}(x), \quad \forall x \in \mathbb{Z}^d_{0+}.$$

Thus we arrive to our model with $S_{-1} \notin \mathbb{Z}^d_{0+}$ (and, hence, with $S_0[1] \le 0$).

Note that our condition (12) is more general than (114).

Remark 2 Here is a link to random walks conditioned not to leave a certain subspace. We may consider the trajectory $S_0, S_1, \dots, S_{\alpha(n)}$ conditioned on the event that the first coordinate stays positive by time $\alpha(n)$, i.e. $B_n = \{\min_{0 \le t < \alpha(n)} S_t[1] \ge 0\}$. Then, in our notation, the event B_n may be represented as $B_n = \{\alpha(n) < T_*\}$ if we consider that "extreme" environment of the form: for $x = (x[1], \dots, x[d])$,

$$H(x) = 0 \text{ when } x[1] < 0, \quad \text{and} \quad H(x) = \infty \text{ when } x[1] \ge 0.$$

Thus, there is no restrictions on the upper half-space with $x[1] \ge 0$, and it is prohibited to visit the lower half-space with $x[1] < 0$.

Note that the case $\mathbf{E}\xi_1[1] > 0$ is simple, since here the initial sequence itself has a regenerative structure and (10), (9) and (7) take place with $q = 1$. In the case $\mathbf{E}\xi_1[1] < 0$, there is only one $q \in (0, 1)$ that solves the equation

$$\sum_{k=-1}^{\infty} q^k \mathbf{P}(\xi_1[1] = -k) = 1. \tag{115}$$

Applying the corresponding exponential change of measure (the Cramér transform) to the distribution of ξ_1, we obtain (10), (7) and (9) with $q < 1$ from (115).

Remark 3 We may present a more detailed version of Theorem 2, containing a formula that relates joint distributions of blocks from (24) and (25) with independent blocks of the core process. Consider arbitrary numbers such that

$$0 \le L_0 < \dots < L_m = n, \quad 0 \le K_i < \dots < K_m, \quad \mathbf{y} \in \mathbb{Z}^{(K_0+1)\times d},$$
$$1 \le l_i := L_i - L_{i-1} \le k_i := K_i - K_{i-1}, \quad \mathbf{x}_i \in \mathbb{Z}^{(K_i - K_{i-1})\times d}, \quad \forall i = 1, \dots, m. \tag{116}$$

Below we use the notation for vectors introduced in (23) and (39).

Corollary 2 *For any $n = L_m \ge m \ge 1$ and any numbers from* (116)

$$\mathbf{P}_*(\{\eta_*(n) = m, \nu_0(n) = L_0, \alpha(\nu_0(n)) = K_0, \mathbf{S}_{K_0} = \mathbf{y}\} \cap \cap_{i=1}^m \{\lambda_i(n) = l_i, \tau_i(n) = k_i, \mathbf{S}_{\overline{K_{i-1}, K_i}} = \mathbf{x}_i\})$$
$$= \psi_0 q^n \mathbf{P}(\overline{\nu}_0 = L_0, \alpha(\nu_0) = K_0, \tilde{S}_{K_0} = \mathbf{y}) \cdot \prod_{i=1}^m \mathbf{P}(\overline{\lambda}_i = l_i, \overline{\tau}_i = k_i, \widetilde{S}_{K_{i-1}K_i} = \mathbf{x}_i\}).$$

Comment that random vectors $\{\overline{\xi}_j,\ \ j = 1, 2 \ldots\}$ that were introduced in (42) may be dependent, notwithstanding that $\{\xi_j,\ \ j = 1, 2 \ldots\}$ were i.i.d. However, the random blocks

$$\left(\overline{\lambda}_i, \overline{\tau}_i, (\overline{\xi}_{K_{i-1}+1}, \ldots, \overline{\xi}_{K_i})\right),\ i = 1, 2 \ldots,$$

are i.i.d. and do not depend on the initial block $\left(\overline{S}_0, \overline{\nu}_0, \overline{T}_0,\ (\overline{\xi}_1, \ldots, \overline{\xi}_{\overline{T}_0})\right)$. This type of the phenomenon is typical for conditioning that involves infinite future: an i.i.d. sequence is transformed into a regenerative sequence. It appears even in the simplest scenario, for a one-dimensional random walk with positive drift, conditioned to stay positive (see, e.g., [10, 14, 15] for similar observations in "unconditioned" models).

Acknowledgments We are thankful to Vladas Sidoravicus who has discussed the topic with Sergey Foss. The authors thank also the anonymous referee for useful and supportive comments and suggestions.

References

1. Asmussen, S.: Conditioned limit theorems relating a random walk to its associate, with applications to risk reserve processes and the $GI/G/1$ queue. Adv. Appl. Probab. **14**(1), 143–170 (1982)
2. Barker, A.: Transience and recurrence of Markov processes with constrained local time. arXiv:1806.05965v3 [math.PR]
3. Benjamini, I., Berestycki, N.: Random paths with bounded local time. J. Eur. Math. Soc. **12**(4), 819–854 (2010)
4. Benjamini, I., Berestycki, N.: An integral test for the transience of a Brownian path with limited local time. Ann. I H P Probab. Stat. **47**(2), 539–558 (2011)
5. Berestycki, N., Gantert, N., Moerters, P., Sidorova, N.: Galton-Watson trees with vanishing martingale limits. J. Stat. Phys. **155**, 737–762 (2014)
6. Bertoin, J., Doney, R.A.: On conditioning a random walk to stay positive. Ann. Probab. **22**(4), 2152–2167 (1994)
7. Bolthausen, E., Sznitman, A.-S.: Ten Lectures on Random Media. DMV Seminar. Band, vol. 52. Springer, Basel (2002)
8. Bolthausen, E., Deuschel, J.-D., Giacomin, G.: Entropic repulsion and the maximum of the two dimensional harmonic crystal. Ann. Probab. **29**, 1670–1692 (2001)
9. Feller, W.: An Introduction to Probability Theory and Its Applications, vol. 1, 3rd edn. Wiley, New York (2008)
10. Foss, S., Zachary, S.: Stochastic sequences with a regenerative structure that may depend both on the future and on the past. Adv. Appl. Probab. **45**(4), 1083–1110 (2013)
11. Kesten, H., Kozlov, M.V., Spitzer, F.: A limit law for random walk in a random environment. Comp. Math. **30**(2), 145–168 (1975)
12. Kolb, M., Savov, M.: Transience and recurrence of a Brownian path with limited local time and its repulsion envelope. Ann. Probab. **44**(6), 4083–4132 (2016)
13. Kozlov, M.V.: A random walk on the line with stochastic structure. Theory Probab. Appl. **18**(2), 406–408 (1973)

14. Kuczek, T.: The central limit theorem for the right edge of supercritical oriented percolation. Ann. Probab. **17**, 1322–1332 (1989)
15. Mountford, T., Sweet, , T.: An extension of Kuczek's argument to nonnearest neighbor contact processes. J. Theor. Probab. **13**, 1061–1081 (2000)
16. Sakhanenko, A., Foss, S.: On the structure of a conditioned random walk on the integers with bounded local times. Sib. Electron. Math. Rep. **14**, 1265–1278 (2017)

Random Memory Walk

Alexander Fribergh, Daniel Kious, Vladas Sidoravicius, and Alexandre Stauffer

To the memory of Vladas Sidoravicius

Abstract We present a simple model of a random walk with partial memory, which we call the *random memory walk*. We introduce this model motivated by the belief that it mimics the behavior of the once-reinforced random walk in high dimensions and with small reinforcement. We establish the transience of the random memory walk in dimensions three and higher, and show that its scaling limit is a Brownian motion.

Keywords Self-interacting random walk · Once-reinforced random walk

MSC 2010 Subject Classifications 60K35

The author "Alexandre Stauffer" was supported by EPSRC Fellowship EP/N004566/1.

A. Fribergh
Department of Mathematics and Statistics, Université de Montréal, Montreal, QC, Canada
e-mail: fribergh@dms.umontreal.ca

D. Kious
Department of Mathematical Sciences, University of Bath, Bath, UK
e-mail: d.kious@bath.ac.uk

V. Sidoravicius
NYU-ECNU Institute of Mathematical Sciences at NYU Shanghai, Shanghai, China

Courant Institute of Mathematical Sciences, New York, USA

A. Stauffer (✉)
Dipartimento di Matematica e Fisica, Università Roma Tre, Rome, Italy

Department of Mathematical Sciences, University of Bath, Bath, UK
e-mail: astauffer@mat.uniroma3.it

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_20

1 Introduction

This short paper started in discussions between the authors during a visit to NYU Shanghai. The model we study here, which we call the *random memory walk*, was suggested by Vladas as a way to interpolate between the more well understood case of a random walk with bounded memory (similar to the so-called *senile random walk* [9, 10]) and the challenging model of once-reinforced random walk, which Vladas was fascinated about. In this paper we will discuss the behavior of the random memory walk. It turned out that the analysis of this model is quite simple once one looks at it from the right point of view.

We start this paper by explaining the once-reinforced random walk, some related models, and the main questions in this area, which motivated us (and, in particular, Vladas) to look at this model. Then we explain the link between the once reinforced random walk and the random memory walk, and proceed to the analysis of the latter model.

1.1 Once-Reinforced Random Walk (ORRW)

This is one of such models whose definition is very simple but whose analysis is far from trivial. In fact, despite being introduced about three decades ago, the behavior of the ORRW on $\mathbb{Z}^d$ is still not well understood, even at an intuitive level, and there are essentially no rigorous result about it.

We start defining the ORRW. Consider an infinite, locally finite graph $G = (V, E)$ with vertex set V and (non-oriented) edge set E, and with a distinguished vertex, called the *origin*, that we denote 0. Given a reinforcement parameter $\delta > 0$, the ORRW $(X_n)_{n\geq 0}$ is defined by the following dynamics. Start at time 0 by placing the random walk at the origin (i.e., $X_0 = 0$) and by assigning weight 1 to every edge of E. Then, at time $n \geq 1$, the random walk jumps to one of its neighbors with a probability proportional to the weight of each edge between them. Note that the first jump of the random walk is to a neighbor of 0 chosen uniformly at random. Whenever the walk jumps across an edge e for the first time, the weight of e is updated from 1 to $1 + \delta$, and then the weight of e is never updated again from that time onwards.

More formally, let E_n be the collection of edges crossed by the random walk up to time n, that is

$$E_n := \{e \in E : \exists k \in \{1, \dots, n\} \text{ s.t. } \{X_{k-1}, X_k\} = e\}. \tag{1}$$

At time $n \in \mathbb{N}$ and on the event $\{X_n = x\}$ with $x \in V$, the walk jumps to a neighbor $y \sim x$ with conditional probability

$$\mathbb{P}(X_{n+1} = y | \mathscr{F}_n) = \frac{1 + \delta \mathbf{1}\{\{x, y\} \in E_n\}}{\sum_{z:z\sim x} 1 + \delta \mathbf{1}\{\{x, y\} \in E_n\}},$$

where $(\mathscr{F}_n)$ is the filtration generated by the history of (X_n), i.e. $\mathscr{F}_n = \sigma(X_k, 0 \leq k \leq n)$ for any integer $n \geq 0$.

This random walk thus favors edges that it has already crossed in the past (which, as usual, we call the *range* of the walk), and δ regulates the strength with which the random walk favors its range. Intuitively, one could say that, as the random walk grows its range, it interacts with it by experiencing a *drift inwards* whenever it tries to move out of its range. In other words, the random walk is attracted to traverse edges that it has already traversed in the past, creating some sort of a small trap for the walk.

1.2 Expected Behavior of ORRW on $\mathbb{Z}^d$

It is particularly interesting to study the ORRW on $\mathbb{Z}^d$, $d \geq 2$, where interesting conjectures have been made. The ORRW was introduced by Davis [6] in 1990 as a simplification of the linearly edge-reinforced random walk, which was defined by Coppersmith and Diaconis in the late eighties. Coppersmith and Diaconis conjectured that the linearly edge-reinforced random walk undergoes a phase transition between recurrence and transience, but this was only established about 25 years later in a sequence of papers [1, 7, 15, 16].

When defining the ORRW, Davis expected that its analysis should be easier than for the linearly edge-reinforced random walk, but curiously the question regarding recurrence and transience remains completely open for the ORRW. Davis noticed in his paper that ORRW has a trivial behavior in dimension one, and conjectured that it is recurrent in dimension two.

It turns out that the ORRW is quite challenging to analyze due to the nature of its interaction and to the lack of monotonicity. Indeed, the *drift inwards* that we mentioned above means that, when the random walk is on the boundary of its range, it is slightly more likely that it goes back inside its range, a fact that could trigger us to think about recurrence. However, the range of the random walk at that place could be of a form such that the *drift inwards* translates to a drift *away from 0*.

Extremely interesting conjectures have been made about the behavior of the ORRW on $\mathbb{Z}^d$, $d \geq 3$, which are usually attributed to Vladas Sidoravicius and Vincent Beffara, and independently to Mike Keane. They conjectured that on $\mathbb{Z}^d$, $d \geq 3$, there exists a phase transition on the strength of the reinforcement parameter δ. That is, there should exist a critical parameter δ_c, a priori depending on the dimension, such that if $\delta < \delta_c$ then ORRW with parameter δ is transient, and if $\delta > \delta_c$ then ORRW is recurrent.

One can then ask finer questions about the model, for instance, regarding the scaling limit of the random walk in the transient regime, or the size and the shape of the range of the random walk in the recurrent regime. All these questions are, of course, still very much open.

It seems particularly interesting to try to study the asymptotic shape of the range in the recurrent case. Simulation suggests that there is a certain *shape theorem*: the range E_n of the walk at time n, when properly scaled by some polynomial in n, seem to converge to a deterministic shape. Nothing has been proved in this direction, and we refer the reader to the nice survey by Gady Kozma [14] where some pictures from simulations are presented.

1.3 Other Models Related to ORRW

A very nice explanation for why the aforementioned shape theorem result is true was usually given by Vladas by referring to what he called the *Glassy sphere model*. In this model, consider spheres of radius $n \geq 1$ simply put inside each other, like Matryoshka dolls. Then, start a random walk on $\mathbb{Z}^d$ from the origin which is reflected upon touching the smallest sphere. Once the random walk has touched the smallest sphere a number of times that is proportional to its size (i.e. n^{d-1} for the n-th sphere), the sphere is destroyed so that the random walk now gets reflected on the next sphere. It is straightforward to prove that the random walk in the glassy sphere model is recurrent in any dimensions.

One could believe that the ORRW for large δ follows the same behaviour as the glassy sphere model. In fact, if one believes that the range of ORRW for large δ grows like a ball, then once the ORRW has visited all vertices in a ball of radius n, it will roughly visit all the edges on the boundary of this ball before going too far away; hence it will "bump" on the boundary of this ball a number of times that is comparable to the size of the boundary. It is not at all clear to us whether this picture really corresponds to the actual behavior of ORRW. Though simulations suggest that this is indeed the case, one cannot disregard that simulations may not be very conclusive for model with such strong self interactions.

Other caricature models have been considered in order to try to understand the ORRW. Here is another model which Vladas recurrently mentioned and which seems very interesting but very challenging to analyze (we are not sure who this model should be attributed to). Consider a semi-infinite cylinder $(\mathbb{Z}/N\mathbb{Z}) \times \{n : n \leq N\}$. On every vertex at non-negative height, i.e. on $(\mathbb{Z}/N\mathbb{Z}) \times \{n : 0 \leq n \leq N\}$, put a so-called cookie. Then, start a random walk coming from $-\infty$. This random walk evolves like a simple random walk with the exception that, when it jumps on a vertex (z, h) where there is a cookie, then it instantaneously jumps to the vertex $(z, h-1)$ just below it and the cookie disappears. It is clear that this random walk is recurrent as it is essentially one dimensional, but interesting questions can be asked about the shape created by the remaining cookies. Indeed, one can consider the interface between the area without cookies and the area with cookies. This interface is intended to provide a simplistic picture of the microscopic behavior of the ORRW close to the boundary of its range for very large δ.

Note that the interface looks like a function; if we clear the cookie at a given vertex, then all the cookies from vertices below it will be cleared as well by the definition of the dynamics. Several questions arise from this model. For instance, stop the random walk when it reaches for the first time the height N. Then, how many cookies are left? How does the interface look like at that time? What is the height of the lowest remaining cookie? It is believed that, when the random walk first reaches height N, almost all the cookies have been eaten, with only $o(N)$ cookies remaining. It is also believed that the fluctuations of the interface should be of order $N^{2/3}$. A more daring guess would be that the interface, when the random walk first reaches height N, is related to KPZ.

Such questions also inspired Vladas to look at random walk on growing domains. In this case, there is a growing sequence of subsets of $\mathbb{Z}^d$ called *domains* and denoted by $D_0 \subset D_1 \subset D_2 \subset \cdots$, and a deterministic sequence of times $t_1 < t_2 < \cdots$ such that at a time $t \in [t_i, t_{i+1})$ the random walk jumps according to a simple random walk that is confined to be inside D_i (that is, the random walk is reflected at the boundary of D_i). So the sequence $t_1, t_2, \ldots$ gives the times at which the domain of the random walk grows. This model was studied by Vladas and others in [2, 3], and we refer the reader to [11] for more recent results.

1.4 ORRW in Other Graphs

We conclude this section by mentioning interesting results that have been proved about ORRW in graphs that are not $\mathbb{Z}^d$. Indeed, it is interesting to ask whether the phase transition between recurrence and transience can be observed on *some* graph.

The ORRW on ladders has been studied, i.e. on $\mathbb{Z} \times \{1, \ldots, k\}$ with $k \geq 2$. In this case, the ORRW should clearly be recurrent for all values of the parameter δ. First, Sellke [17] proved that ORRW is recurrent for $k = 2$, and showed that ORRW is recurrent for any $k \geq 2$ as long as δ is small enough. Then, Vervoort [21] wrote a draft paper giving an incomplete proof of recurrence for large reinforcement parameter, which despite having some gaps and mistakes, contained a very good core idea. This argument was later on cleaned and completed in [13].

The ORRW has also been analyzed on trees. The first result in that direction is the proof of transience on the binary tree in [8] for any value of the parameter δ, which shows that there is no phase transition on binary trees unlike the conjectured behavior on $\mathbb{Z}^d$. The lack of a phase transition has also been established on Galton-Watson trees by Collevecchio [4], who found a very elegant proof through a comparison to a branching process. In the hope of observing a phase transition, Kious and Sidoravicius [12] considered the ORRW on a particular family of trees, which grows only polynomially fast, and were able to prove the existence of a phase transition on such trees. Later, it was proved in [5] that the critical parameter δ_c of the ORRW of any tree is equal to the a quantity that was called the *branching-ruin number* of the tree. This quantity characterizes the size of the tree at the polynomial scale.

2 Random Memory Walk

Our motivation to study the random memory walk is to compare it to ORRW in high dimensions and with small reinforcement parameter. The rough idea is to say that, if the ORRW is transient and if the dimension is large enough, then the loops produced by the range of the ORRW should not be too large, and thus the random walk should not get to revisit its range a large number of times. Consequently, the ORRW would behave as if it had a finite random memory (given by the size of the local loops it produces).

We have no intention to argue that the random memory walk has the same behavior as the ORRW in high dimension; in particular, as we will see in the definition below, the random memory walk has a memory that is independent of the range of the walk, which is certainly not the case for the ORRW. Nonetheless, one may ask the question of whether the ORRW in high dimensions and for small reinforcement parameter shows a similar regenerative structure as the random memory walk studied in the present paper.

Now we define the random memory walk. As before, we denote the random walk by $(X_n)_{n\geq 0}$ starting from $X_0 = 0$. Let us denote by $R_{n,m}$ the last m edges visited by the random walk at time n; that is,

$$R_{n,m} = \{\{x, y\} \colon \exists i \in \{n - m + 1, \dots, n\} \text{ s.t. } \{X_{i-1}, X_i\} = \{x, y\}\},$$

with the convention that $R_{n,0} = \emptyset$. In order to decide its position at time $n + 1$, the random memory walk will have access to a memory of random length regarding its past. The length of this memory is given by the random variable K_n, where $K_0, K_1, \dots$ will form an i.i.d. sequence of nonnegative random variables. Then the distribution of the location of the random memory walk at time $n + 1$ will depend only on the current location of the walk (X_n) and on the information (the memory) regarding its K_n last positions which is given by R_{n,K_n}.

More precisely, define the filtration $\mathscr{F}_n = \sigma((X_i, K_i), i \leq n)$, for any $n \geq 0$. Assuming $X_n = x \in \mathbb{Z}^d$ and y is a neighbor of x, i.e. $|x - y| = 1$, the next step is distributed according to the following conditional probability:

$$\mathbb{P}[X_{n+1} = y \,|\, \mathscr{F}_n] = \frac{1 + \delta \mathbf{1}\{\{x, y\} \in R_{n,K_n}\}}{\sum_{z:z\sim x}(1 + \delta \mathbf{1}\{\{x, z\} \in R_{n,K_n}\})}, \tag{2}$$

where $\delta > 0$ is the reinforcement parameter. In other words, the random memory walk defined above jumps like the ORRW but reinforcing only the last edges in the range, where the number of edges chosen to be reinforced is a random variable that changes at each step and is given by the sequence $(K_n)_n$.

For the moment we will assume that

$$\mathbb{P}[K_0 = 0] = p_0 > 0.$$

The above assumption is not at all essential for the proof and is made here just to simplify the exposition. Later in Sect. 6, we explain how our proof can be adapted to remove the above assumption. In that section we also discuss a more general version of this model, where the probability of jump of the random walk is not given by (2) but is a more general function of R_{n,K_n}.

3 Our Results

We are now ready to state our two main theorems. Our first theorem established transience in dimensions at least 3.

Theorem 1 *Assume that* $\mathbb{E}(K_0) < \infty$*. Then, the random memory walk* $(X_n)_n$ *on* $\mathbb{Z}^d$*,* $d \geq 3$*, is transient almost surely.*

In our second result, we establish the scaling limit of the random memory walk under stronger assumptions.

Theorem 2 *If* $\mathbb{E}(K_0^3) < \infty$*, then* $(X_n)_n$ *satisfies a functional central limit theorem, that is, for any* $T > 0$*,*

$$\left(\frac{X_{\lfloor nt \rfloor}}{\sqrt{n}}\right)_{t\in[0,T]} \Rightarrow (B_t)_{t\in[0,T]} \,,$$

where the convergence holds in law, and where $(B_t)_t$ *is a non-degenerate d-dimensional Brownian motion.*

It may seem surprising that we require a finite third moment for the memory in the above result, instead of only a finite second moment. However, as we explain later in the paper, it seems that this is the best we can do with the techniques we use.

Remark 1 Following the arguments of this paper, one can easily prove that, if K_0 has a third moment, then the random memory walk is recurrent in dimension 2. It is indeed straightforward to prove that the sub-walk defined in Sect. 4 is recurrent for $d = 2$, using Proposition 1 below and Chung-Fuchs theorem.

4 Regeneration Structure Induced by the Memory and Transience of a Sub-walk

The main idea is to focus on the sequence $(K_n)_n$. We will define regeneration times, that is, times at which the random walk forgets its past and starts afresh. Once we are able to prove that such times happen infinitely often, we will be able to use classical arguments in order to prove Theorems 1 and 2.

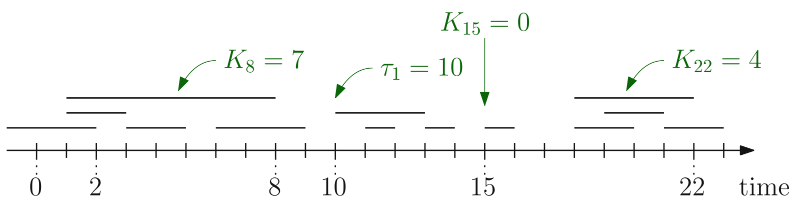


Fig. 1 Illustration of the regeneration time structure. The length of the horizontal line segment ending at coordinate i represents the variable K_i (which shows how far in the past the random walk needs to look at to decide where to be at time $i + 1$). Line segments are drawn at different heights for illustration purpose

Define $\tau_0 := 0$ and

$$\tau_1 = \inf\{n > 0 : K_{n+i} \le i, \ \forall i \ge 0\}.$$

Intuitively, if we consider time as the non-negative reals $\mathbb{R}_+$ and, for each integer $i \ge 0$, we draw a line segment between i and $i - K_i$, then τ_1 is the first position such that there is no line segment covering the edge $\{\tau_1 - 1, \tau_1\}$; see Fig. 1. Note that when the random walk decides to jump from its location at τ_1 to $\tau_1 + 1$, it does so as a step of simple random walk (that is, it just chooses a neighbor of X_{τ_1} uniformly at random and jumps there), and from that time onwards it will not take into account anymore the edges it traversed before time τ_1. Note also that τ_1 necessarily happens at a time for which $K_{\tau_1} = 0$; that is the reason why we consider the assumption that $\mathbb{P}[K_0 = 0] > 0$.

Now we show that τ_1 is finite almost surely.

Proposition 1 *We have that* $\mathbb{P}[\tau_1 < \infty] = 1$ *if and only if* $\mathbb{E}[K_0] < \infty$. *Moreover, for any integer* $m \ge 1$, *we have that* $\mathbb{E}(\tau_1^m) < \infty$ *if and only if* $\mathbb{E}(K_0^{m+1}) < \infty$.

Proof First, we note that if $\mathbb{P}[\tau_1 = 1] = 0$, then $\mathbb{P}[\tau_1 < \infty] = 0$ as, for any $n > 1$, we have

$$\begin{aligned}\mathbb{P}[\tau_1 = n] &= \mathbb{P}\left[\forall 1 \le k \le n-1, \exists j \ge 0 \text{ s.t. } K_{k+j} > j, K_{n+i} \le i, \forall i \ge 0\right] \\ &\le \mathbb{P}[K_{n+i} \le i, \forall i \ge 0] \\ &= \mathbb{P}[K_{1+i} \le i, \forall i \ge 0] \\ &= \mathbb{P}[\tau_1 = 1] = 0.\end{aligned}$$

Now, assume $\mathbb{P}[\tau_1 = 1] > 0$ and let us study the event $\{\tau_1 < \infty\}$. This event can be seen as successive trials of realizing the events $\{K_{n+i} \le i, \forall i \ge 0\}$, and these trials are independent and have probability $\mathbb{P}[\tau_1 = 1]$. To prove that properly, let us define recursively

$$T_1 = 1, \quad S_k = \inf\{i \ge 0 : K_{T_k + i} > i\}, \text{ and}$$
$$T_{k+1} = T_k + S_k + 1, \text{ for all } k \ge 1.$$

In words, $T_k + S_k$ is the first position after $T_k - 1$ for which the memory of the walk at that time (equivalently, the line segment that ends there) goes back all the way to $T_k - 1$. For example, in Fig. 1, we have that $T_1 + S_1 = 2$, and subsequently we get $T_2 = 3$ and $T_2 + S_2 = 3$.

The idea behind this definition is that if $\tau_1 > 1$, then we look for the value of S_1. This translates to checking the random variables $K_{T_1}, K_{T_1+1}, \ldots$ until finding the value of S_1. If we obtain that $S_1 < \infty$, then position $T_1 + S_1 + 1 = T_2$ is a possible candidate for τ_1. If it turns out that $\tau_1 > T_2$, then we look for S_2 and T_3. At each step of this procedure, say step $k \geq 1$, we will show that, regardless of the values of $T_1, T_2, \ldots, T_k$ and regardless of the values of $S_1, S_2, \ldots, S_{k-1}$, with positive probability we have that $S_k = \infty$, which in turn gives that $\tau_1 = T_{k+1} = T_k + S_k + 1$.

More formally, define $N = \inf\{k \geq 1 : S_k = \infty\}$. Using these random variables, we have that $\{\tau_1 < \infty\} = \{N < \infty\}$. Also, note that, $\mathbb{P}[S_1 = \infty] = \mathbb{P}[\tau_1 = 1]$ and, for any $k > 0$, conditional on T_k, S_k is distributed like S_1. Hence, one can write

$$\begin{aligned}\mathbb{P}[\tau_1 < \infty] &= \sum_{k=1}^{\infty} \mathbb{P}[N = k] \\ &= \sum_{k=1}^{\infty} \mathbb{P}\left[\bigcap_{i=1}^{k-1} \{S_i < \infty\}, S_k = \infty\right] \\ &= \sum_{k=1}^{\infty} \mathbb{P}\left[\tau_1 = 1\right]\left(1 - \mathbb{P}\left[\tau_1 = 1\right]\right)^{k-1} = 1.\end{aligned}$$

Hence, we have proved that if $\mathbb{P}\left[\tau_1 = 1\right] > 0$, then $\mathbb{P}\left[\tau_1 < \infty\right] = 1$. Finally, we can conclude the first statement of the proposition by noting that

$$\mathbb{P}\left[\tau_1 = 1\right] = \prod_{i=0}^{\infty} \mathbb{P}[K_0 \leq i] = \prod_{i=0}^{\infty} (1 - \mathbb{P}[K_0 > i]) \sim c e^{-\mathbb{E}(K_0)},$$

and therefore $\mathbb{P}\left[\tau_1 = 1\right] > 0$ if and only if $\mathbb{E}(K_0) < \infty$.

Now we turn to the second part of the proposition. For this purpose, note that, from the definition of $(T_k)_k$, $(S_k)_k$ and N,

$$\tau_1 = T_N = N + \sum_{k=1}^{N-1} S_k,$$

where we recall that N is distributed as a geometric random variable with parameter $\mathbb{P}\left[\tau_1 = 1\right]$, and where the random variables S_k appearing in the sum above are conditioned to be finite. Hence, τ_1 is essentially equal to a sum of a geometric number of independent random variables distributed like S_1 conditioned on $\{S_1 < \infty\}$. Therefore, for any $m \geq 1$, τ_1 has an m-th moment if and only if S_1 conditioned on $\{S_1 < \infty\}$ has an m-th moment. Now, for any $k \geq 0$, one can write

$$\mathbb{P}[S_1 = k | S_1 < \infty] = \frac{\mathbb{P}[K_{1+i} \leq i, \text{ for } 0 \leq i \leq k-1, \text{ and } K_{1+k} > k]}{\mathbb{P}[S_1 < \infty]}$$
$$= \frac{\prod_{i=0}^{k-1}(1 - \mathbb{P}[K_0 > i]) \times \mathbb{P}[K_0 > k]}{\mathbb{P}[S_1 < \infty]}.$$

Now, assume that $\mathbb{E}(K_0) < \infty$. In that case, as shown above, $\prod_{i=0}^{k-1}(1 - \mathbb{P}[K_0 > i])$ converges to a positive constant. Besides, we have $\mathbb{P}[S_1 < \infty] = \mathbb{P}[\tau_1 = 1] > 0$. Thus, there exist constants c_0 and c_1 such that

$$c_0 \mathbb{P}[K_0 > k] \leq \mathbb{P}[S_1 = k | S_1 < \infty] \leq c_1 \mathbb{P}[K_0 > k].$$

Thus, we have that, for any $m \geq 1$,

$$c_0 \sum_{k=1}^{\infty} k^m \mathbb{P}[K_0 > k] \leq \mathbb{E}\left(S_1^m \middle| S_1 < \infty\right) \leq c_1 \sum_{k=1}^{\infty} k^m \mathbb{P}[K_0 > k].$$

From there, it is clear that $\mathbb{E}\left(S_1^m \middle| S_1 < \infty\right)$ is finite if and only if K_0 has an $(m+1)$-th moment, which concludes the proof. □

The time $\tau_1 > 0$ is referred to as the first *regeneration time*. Let us denote

$$\mathscr{D}_n := \{K_{n+i} \leq i, \ \forall i \geq 0\} \tag{3}$$

the event on which $n \geq 0$ is a regeneration time.

By Proposition 1, we have that if $\mathbb{E}(K_0) < \infty$ then $\mathbb{P}[\mathscr{D}_1] > 0$, which easily implies that $\mathbb{P}[\mathscr{D}_0] > 0$. Therefore, we can safely define the conditional probability $\overline{\mathbb{P}}[\cdot] := \mathbb{P}[\cdot | \mathscr{D}_0]$ and we have that if $\mathbb{E}(K_0) < \infty$ then

$$\overline{\mathbb{P}}[\tau_1 < \infty] = \frac{\mathbb{P}[\tau_1 < \infty, \mathscr{D}_0]}{\mathbb{P}[\mathscr{D}_0]} = 1.$$

Also, we have that

$$\overline{\mathbb{P}}[\mathscr{D}_n] = \frac{\mathbb{P}[K_i \leq i, \ \forall 0 \leq i \leq n-1] \times \mathbb{P}[K_{n+i} \leq i, \ \forall i \geq 0]}{\mathbb{P}[K_i \leq i, \ \forall i \geq 0]}$$
$$= \mathbb{P}[K_i \leq i, \ \forall 0 \leq i \leq n-1]$$
$$\geq \mathbb{P}[\mathscr{D}_0] > 0. \tag{4}$$

We inductively define the sequence of regeneration times $\tau_n = \tau_{n-1} + \tau_1 \circ \theta_{\tau_{n-1}}$, where θ is the canonical shift. The following proposition is a classical result on regeneration times.

Proposition 2 *Assume that $\mathbb{E}(K_0) < \infty$. The random variables $(X_{\tau_n} - X_{\tau_{n-1}}, \tau_n - \tau_{n-1})_{n\geq 1}$ are independent and, except for $n = 1$, are distributed like (X_{τ_1}, τ_1) under $\overline{\mathbb{P}}$. In particular, all the regeneration times τ_n, $n \geq 1$, are finite $\mathbb{P}$-almost surely.*

Proof This easily follows from general and classical arguments. For instance, one can replicate the proof of Corollary 1.5 in [20], which comes from Proposition 1.3 and Theorem 1.4 therein. □

Note that, from the above, we have that if X is almost surely transient under $\overline{\mathbb{P}}$ then it is almost surely transient under $\mathbb{P}$ as $\liminf ||X_n|| \geq -||X_{\tau_1}|| + \liminf ||X_n - X_{\tau_1}||$. Nevertheless, it is not obvious that X satisfies a 0-1 law for transience, even under $\overline{\mathbb{P}}$.

In this section, we want first to prove transience and CLT for the walk (X_n) considered at regeneration times. For this purpose, define the walk $(Y_k)_{k\geq 0}$ on $\mathbb{Z}^d$ such that $Y_k = X_{\tau_k}$ for any $k \geq 0$.

Proposition 3 *If $\mathbb{E}(K_0) < \infty$, then the random walk (Y_k) is transient under $\overline{\mathbb{P}}$, and under $\mathbb{P}$.*

Proof Assume $\mathbb{E}(K_0) < \infty$. From Proposition 2, we have that, under $\overline{\mathbb{P}}$, $(Y_{k+1} - Y_k)_{k\geq 0}$ is a sequence of i.i.d. random variables. As the definition of the walk (X_n) is symmetric with respect to every direction of $\mathbb{Z}^d$, we have that, under $\overline{\mathbb{P}}$, the process $(Y_k)_k$ is a symmetric, genuinely d-dimensional random walk. We can then directly conclude the first statement by using Theorem T1, p.83 of Spitzer's book [18], that is Chung-Fuchs theorem. □

5 Transience and CLT for the Random Memory Walk

The proof of the CLT (Theorem 2) will come easily from classical arguments. On the other hand, the proof for transience (Theorem 1) requires some work as we want to derive it under minimal assumptions. The idea is that, once we know that the walk $(Y_k) = (X_{\tau_k})_k$ is transient, we need to prove that the random walk (X_n) cannot come back to zero between two regeneration times infinitely often.

Proof (Theorem 1) We will show the transience of (X_n). Note that X is transient, i.e. $||X_n|| \to \infty$, if and only if it visits 0 finitely often. We already know that the random walk $(Y_k)_k$ visits 0 only finitely often, which is equivalent to saying that there is only a finite number of indices i such that $X_{\tau_i} = 0$. We need to prove that X cannot come back to 0 between two regeneration times infinitely often.

Let us define the sequence of successive return times to 0 by $R_0 = 0$ and $R_i = \inf\{n > R_{i-1} : X_n = 0\}$, for $i \geq 1$. In the following computation, we use the fact that, every time X is back at 0 and this time is a regeneration time, it implies that Y is back at 0, thus this cannot happen infinitely often. Recall that $\mathscr{D}_n$ is the event that

n is a regeneration time; cf. (3). We have that

$$\begin{aligned}
&\overline{\mathbb{P}}[\cap_{i\geq 1}\{R_i < \infty\}] \\
&= \overline{\mathbb{P}}[\cap_{i\geq 1} \cup_{k\geq i} \{R_k < \infty, \mathscr{D}_{R_k}\}] + \overline{\mathbb{P}}[\cup_{i\geq 1} \cap_{k\geq i} \{R_k < \infty, \mathscr{D}^c_{R_k}\}] \\
&\leq \overline{\mathbb{P}}[Y \text{ visits } 0 \text{ i.o.}] + \overline{\mathbb{P}}[\cup_{i\geq 1} \cap_{k\geq i} \{R_k < \infty, \mathscr{D}^c_{R_k}\}] \\
&= \overline{\mathbb{P}}[\cup_{i\geq 1} \cap_{k\geq i} \{R_k < \infty, \mathscr{D}^c_{R_k}\}].
\end{aligned}$$

Hence, we obtain the bound

$$\overline{\mathbb{P}}[\cap_{i\geq 1}\{R_i < \infty\}] \leq \sum_{i\geq 1} \overline{\mathbb{P}}[\cap_{k\geq i}\{R_k < \infty, \mathscr{D}^c_{R_k}\}].$$

Let us fix an index $i \geq 1$ and prove that $\overline{\mathbb{P}}[\cap_{k\geq i}\{R_k < \infty, \mathscr{D}^c_{R_k}\}] = 0$.

We need to define inductively a sequence of stopping times that are all finite on $\cap_{k\geq i}\{R_k < \infty, \mathscr{D}^c_{R_k}\}$. First, define

$$\begin{aligned}
\tilde{R}_1 &= R_i \geq i \geq 1, \\
\tilde{S}_1 &= \tilde{R}_1 + \inf\{j \geq 0 : K_{\tilde{R}_1 + j} > j\}.
\end{aligned}$$

Note that $\tilde{S}_1$ is the first position after R_1 whose memory reaches back to before $\tilde{R}_1$; in other words, $\tilde{S}_1$ is the first position that shows that $\tilde{R}_1$ is not a regeneration time. Then, define inductively, for any $k \geq 1$,

$$\begin{aligned}
\tilde{R}_{k+1} &= \inf\{j > \tilde{S}_k : X_j = 0\} \geq k+1, \\
\tilde{S}_{k+1} &= \tilde{R}_{k+1} + \inf\{j \geq 0 : K_{\tilde{R}_{k+1}+j} > j\}.
\end{aligned}$$

The times $(\tilde{R}_n)_n$ are stopping times with respect to the filtration $\mathscr{F}_n := \sigma(X_k, K_{k-1}, 0 \leq k \leq n)$ and the times $(\tilde{S}_n)_n$ are stopping times with respect to $\sigma(X_k, K_k, 0 \leq k \leq n)$. Moreover, we have that

$$\begin{aligned}
\overline{\mathbb{P}}[\cap_{k\geq i}\{R_k < \infty, \mathscr{D}^c_{R_k}\}] &\leq \overline{\mathbb{P}}[\cap_{k\geq 1}\{\tilde{R}_k < \infty, \tilde{S}_k < \infty\}] \\
&= \lim_{N\to\infty} \overline{\mathbb{P}}[\cap_{k=1}^N\{\tilde{R}_k < \infty, \tilde{S}_k < \infty\}]. \qquad (5)
\end{aligned}$$

Now, note that, on the event $\{\tilde{R}_k < \infty\}$,

$$\begin{aligned}
\overline{\mathbb{P}}\left[\tilde{S}_k < \infty \middle| \mathscr{F}_{\tilde{R}_k}\right] &= 1 - \overline{\mathbb{P}}\left[\mathscr{D}_{\tilde{R}_k} \middle| \mathscr{F}_{\tilde{R}_k}\right] \\
&= 1 - \sum_{n\geq k} \mathbf{1}\{\tilde{R}_k = n\}\overline{\mathbb{P}}\left[\mathscr{D}_n | \mathscr{F}_n\right]
\end{aligned}$$

$$= 1 - \sum_{n \geq k} \mathbf{1}\{\tilde{R}_k = n\} \overline{\mathbb{P}}[\mathscr{D}_n]$$
$$\leq 1 - \mathbb{P}[\mathscr{D}_0],$$

where we used that $\mathscr{D}_n$ is independent of $\mathscr{F}_n$ and (4). Together with (5), we obtain that

$$\overline{\mathbb{P}}[\cap_{k \geq i}\{R_k < \infty, \mathscr{D}^c_{R_k}\}] \leq \lim_{N \to \infty} (1 - \mathbb{P}[\mathscr{D}_0])^N = 0.$$

This finally implies that

$$\overline{\mathbb{P}}[X \text{ is recurrent}] = \overline{\mathbb{P}}[\cap_{i \geq 1}\{R_i < \infty\}] = 0.$$

□

Proof (Theorem 2) We now establish the functional central limit theorem, assuming $\mathbb{E}(K_0^3) < \infty$. We will simply explain why it holds, as this can be proved by following classical results, for instance the proof of Theorem 4.1 in [19] (the only difference is that Brownian motion being non-degenerate comes much more easily in our case, as the process is fully symmetric). The idea of the proof is simply that a functional CLT holds for the random walk $(Y_k)_k$ as, for each $k \geq 1$, Y_k is a sum of i.i.d. random variables which are centered and square integrable (using our assumptions). This comes from Donsker's invariance principle. From there, one only needs an inversion argument for $k \mapsto \tau_k$, which comes from the fact that τ_k is also a sum of i.i.d. random variables (satisfying a law of large numbers), and for the first n regeneration times, the distances between successive regeneration times are small compared to $\sqrt{n}$ (in probability). This latter step is guaranteed by the fact that τ_1 has a finite second moment under $\overline{\mathbb{P}}$. □

6 Extensions

There are two main ways in which our results can be extended. The first one is that the assumption $\mathbb{P}(K_0 = 0) > 0$ is not necessary. The second one is that the jump distribution of the random memory walk does not need to have the form of a once-reinforced random walk, as stated in (2).

We start explaining how we can get over the assumption $\mathbb{P}(K_0 = 0) > 0$. This assumption might seem arbitrary at first, but this is actually equivalent to saying that, regardless of the past history of the random walk, the walker jumps to any given neighbor with a probability bounded below by a universal constant. Note that this is indeed the case when the jump distribution is as given by (2) since the probability that the walker jumps to any given neighbor is at least $\frac{1}{1+(2d-1)(1+\delta)}$, regardless of everything else. So even if we had $\mathbb{P}(K_0 = 0) = 0$, we could redefine the jump distribution and the distribution of K_0 to have $\mathbb{P}(K_0 = 0) > 0$.

This then leads us to look at different jump distributions for the walker. Consider the following more general version of the random memory walk. Define the filtration $\mathscr{F}_n = \sigma((X_k, K_k), k \leq n)$, for any $n \geq 0$. Assuming $X_n = x \in \mathbb{Z}^d$ and y is a neighbor of x, i.e. $|x - y| = 1$, the next step is distributed according to the following conditional probability:

$$\mathbb{P}[X_{n+1} = y | \mathscr{F}_n] = f\left(x, y, R_{n,K_n}\right), \tag{6}$$

where $f : \mathbb{Z}^d \times \mathbb{Z}^d \times \mathscr{E} \to (0, 1)$ is some predetermined function, and $\mathscr{E}$ denotes the set of all finite subsets of edges of $\mathbb{Z}^d$.

Then our proofs work provided f satisfies some symmetry assumption. Namely, it is enough to require that f is invariant under graph isomorphism. We also want to impose that either $\mathbb{P}(K_0 = 0) > 0$ or there exists a positive constant c so that for any neighboring vertices x and y, and any $R \in \mathscr{E}$ we have

$$f(x, y, R) \geq c.$$

References

1. Angel, O., Crawford, N., Kozma, G.: Localization for linearly edge reinforced random walks. Duke Math. J. **163**(5), 889–921 (2014)
2. Dembo, A., Huang, R., Sidoravicius, V.: Walking within growing domains: recurrence versus transience. Electron. J. Probab. **19**, 20 pp. (2014)
3. Dembo, A., Huang, R., Sidoravicius, V.: Monotone interaction of walk and graph: recurrence versus transience. Electron. Commun. Probab. **19**, 12 pp. (2014)
4. Collevecchio, A.: One the transience of processes defined on Galton-Watson trees. Ann. Probab. **34**(3), 870–878 (2006)
5. Collevecchio, A., Kious, D., Sidoravicius, V.: The Branching–Ruin number and the critical parameter of once? Reinforced random walk on trees. Commun. Pure Appl. Math. **73**(1), 210–236 (2020)
6. Davis, B.: Reinforced random walk. Probab. Theory Relat. Fields **84**(2), 203–229 (1990)
7. Disertori, M., Sabot, C., Tarrès, P.: Transience of edge-reinforced random walk. Commun. Math. Phys. **339**(1), 121–148 (2015)
8. Durrett, R., Kesten, H., Limic, V.: Once edge-reinforced random walk on a tree. Probab. Theory Relat. Fields **122**(4), 567–592 (2002)
9. Holmes, M.P.: The scaling limit of senile reinforced random walk. Electron. Commun. Probab. **14**, 104–115 (2009)
10. Holmes, M.P., Sakai, A.: Senile reinforced random walks. Stoch. Process. Appl. **117**, 1519–1539 (2007)
11. Huang, R.: On random walk on growing graphs. Ann. Inst. H. Poincaré Probab. Stat. **55**, 1149–1162 (2019)
12. Kious, D., Sidoravicius, V.: Phase transition for the once-reinforced random walk on Zd-like trees. Ann. Probab. **46**(4), 2121–2133 (2018)
13. Kious, D., Schapira, B., Singh, A.: Once reinforced random walk on Z x Gamma (2018). Preprint, arXiv:1807.07167
14. Kozma, G.: Reinforced random walks (2012). Preprint, arXiv:1208.0364

15. Sabot, C., Tarrès, P.: Edge-reinforced random walk, vertex-reinforced jump process and the supersymmetric hyperbolic sigma model. J. Eur. Math. Soc. **17**(9), 2353–2378 (2015)
16. Sabot, C., Zeng, X.: A random Schrödinger operator associated with the Vertex Reinforced Jump Process on infinite graphs. J. Am. Math. Soc. **32**, 311–349 (2019)
17. Sellke, T.: Recurrence of reinforced random walk on a ladder. Electron. J. Probab. **11**, 301–310 (2006)
18. Spitzer, F.: Principles of Random Walk, 2nd edn. Springer, Berlin (1976)
19. Sznitman, A.-S.: Slowdown estimates and central limit theorem for random walks in random environment. J. Eur. Math. Soc. **2**, 93–143 (2000)
20. Sznitman, A.-S., Zerner, M.: A law of large numbers for random walks in random environment. Ann. Probab. **27**(4), 1851–1869 (1999)
21. Vervoort, M.: Reinforced random walks (2002)

Exponential Decay in the Loop $O(n)$ Model on the Hexagonal Lattice for $n > 1$ and $x < \frac{1}{\sqrt{3}} + \varepsilon(n)$

Alexander Glazman and Ioan Manolescu

Abstract We show that the loop $O(n)$ model on the hexagonal lattice exhibits exponential decay of loop sizes whenever $n > 1$ and $x < \frac{1}{\sqrt{3}} + \varepsilon(n)$, for some suitable choice of $\varepsilon(n) > 0$.

It is expected that, for $n \leq 2$, the model exhibits a phase transition in terms of x, that separates regimes of polynomial and exponential decay of loop sizes. In this paradigm, our result implies that the phase transition for $n \in (1, 2]$ occurs at some critical parameter $x_c(n)$ strictly greater than that $x_c(1) = 1/\sqrt{3}$. The value of the latter is known since the loop $O(1)$ model on the hexagonal lattice represents the contours of the spin-clusters of the Ising model on the triangular lattice.

The proof is based on developing n as $1 + (n - 1)$ and exploiting the fact that, when $x < \frac{1}{\sqrt{3}}$, the Ising model exhibits exponential decay on any (possibly non simply-connected) domain. The latter follows from the positive association of the FK-Ising representation.

Keywords Loop models · $O(n)$ model · Phase diagram · Lattice models · Statistical mechanics · Ising model · Enhancement percolation

1 Introduction

The loop $O(n)$ model was introduced in [10] as a graphical model expected to be in the same universality class as the spin $O(n)$ model. The latter is a generalisation of the seminal Ising model [19] that incorporates spins contained on the n-dimensional

A. Glazman (✉)
Faculty of Mathematics, University of Vienna, Wien, Austrial
e-mail: alexander.glazman@univie.ac.at

I. Manolescu
Département de Mathématiques, Université de Fribourg, Fribourg, Switzerland
e-mail: ioan.manolescu@unifr.ch

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_21

sphere. See [22] for a survey of both $O(n)$ models. For integers $n > 1$, the connection between the loop and the spin $O(n)$ models remains purely heuristic. Nevertheless, the loop $O(n)$ model became an object of study in its own right; it is predicted to have a rich phase diagram [4] in the two real parameters $n, x > 0$. For $n = 0, 1, 2$ the loop $O(n)$ model is closely related to self-avoiding walk, the Ising model, and a certain random height model, respectively.

Let $\mathbb{H}$ denote the hexagonal lattice. A *domain* is a subgraph $\mathscr{D} = (V_{\mathscr{D}}, E_{\mathscr{D}})$ of $\mathbb{H}$ formed of the edges contained inside or along some simple cycle $\partial\mathscr{D} \subset E(\mathbb{H})$ (hereafter called a *loop*), and all endpoints of such edges. Write $F_{\mathscr{D}}$ for the set of faces of $\mathbb{H}$ delimited by edges of $\mathscr{D}$ only.

Configurations $\omega \in \{0, 1\}^{E(\mathscr{D})}$ will be identified to the subset of edges $e \in E_{\mathscr{D}}$ with $\omega(e) = 1$ (also called open edges) as well as to the spanning subgraph of $\mathscr{D}$ containing exactly these edges. A *loop configuration* is any element $\omega \in \{0, 1\}^{E_{\mathscr{D}}}$ that is even, which is to say that the degree of any vertex is 0 or 2 when ω is seen as a subgraph of $\mathscr{D}$. As such ω is the disjoint union of a set of loops of $\mathscr{D}$. Loops are allowed to run along the boundary edges, but may not terminate at boundary points.

For real parameters $n, x > 0$, let $\mathsf{Loop}_{\mathscr{D},n,x}$ be the measure on loop configurations given by

$$\mathsf{Loop}_{\mathscr{D},n,x}(\omega) = \frac{1}{Z_{\text{loop}}(\mathscr{D}, n, x)} \cdot x^{|\omega|} n^{\ell(\omega)},$$

where $|\omega|$ is the number of edges in ω, $\ell(\omega)$ is the number of loops in ω and $Z_{\text{loop}}(\mathscr{D}, n, x)$ is a constant called the partition function, chosen so that $\mathsf{Loop}_{\mathscr{D},n,x}$ is a probability measure.

We will consider that the origin 0 is a vertex of the hexagonal lattice and will always consider domains $\mathscr{D}$ containing 0. We say that the loop $O(n)$ model with edge-weight x exhibits exponential decay of loop lengths if there exists $c > 0$ such that for any $k \geq 1$ and any domain $\mathscr{D}$,

$$\mathsf{Loop}_{\mathscr{D},n,x}[\mathsf{R} \geq k] \leq \exp(-ck), \tag{1}$$

where R stands for the length of the biggest loop surrounding 0.

According to physics predictions [4, 20], the loop $O(n)$ model exhibits macroscopic loops when $n \in [0, 2]$ and $x \geq x_c(n) = \frac{1}{\sqrt{2+\sqrt{2-n}}}$; that is, the largest loop surrounding 0 has a diameter comparable to that of the largest ball centred at 0 and contained in $\mathscr{D}$. For all other values of n and x, the model is expected to exhibit exponential decay. Moreover, it was conjectured (see e.g. [18, Section 5.6]) that in the macroscopic-loops phase, the model has a conformally invariant scaling limit given by the Conformal Loop Ensemble (CLE) of parameter κ, where

$$\kappa = \begin{cases} \frac{4\pi}{2\pi - \arccos(-n/2)} \in [\frac{8}{3}, 4] & \text{if } x = x_c(n), \\ \frac{4\pi}{\arccos(-n/2)} \in [4, 8] & \text{if } x > x_c(n). \end{cases}$$

Our main result below is in agreement with the predicted phase diagram.

Theorem 1 *For any $n > 1$, there exists $\varepsilon(n) > 0$ such that the loop $O(n)$ model exhibits exponential decay* (1) *for all $x < \frac{1}{\sqrt{3}} + \varepsilon(n)$.*

Furthermore, ε may be chosen so that $\varepsilon(n) \sim C(n-1)^2$ as $n \searrow 1$, where $C = \frac{(1+\sqrt{3})^5}{9 \cdot 2^{13}}$.

Prior to our work, the best known bound on the regime of exponential decay for $n > 1$ was $x < \frac{1}{\sqrt{2+\sqrt{2}}} + \varepsilon(n)$ [25], where $\sqrt{2+\sqrt{2}}$ is the connective constant of the hexagonal lattice computed in [11]. Also, in [13] it was shown that when n is large enough the model exhibits exponential decay for any value of $x > 0$. Apart from the improved result, our paper provides a method of relating (some forms of) monotonicity in x and n; see Sect. 5 for more details.

Existence of macroscopic loops was shown for $n \in [1, 2]$ and $x = x_c(n) = \frac{1}{\sqrt{2+\sqrt{2-n}}}$ in [12], for $n = 2$ and $x = 1$ in [15], and for $n \in [1, 1+\varepsilon]$ and $x \in [1-\varepsilon, \frac{1}{\sqrt{n}}]$ in [9]. Additionally, for $n = 1$ and $x \in [1, \sqrt{3}]$ (which corresponds to the antiferromagnetic Ising model) as well as for $n \in [1, 2]$ and $x = 1$, a partial result in the same direction was shown in [9]. Indeed, it was proved that in this range of parameters, at least one loop of length comparable to the size of the domain exists with positive probability (thus excluding the exponential decay). All results appear on the phase diagram of Fig. 1.

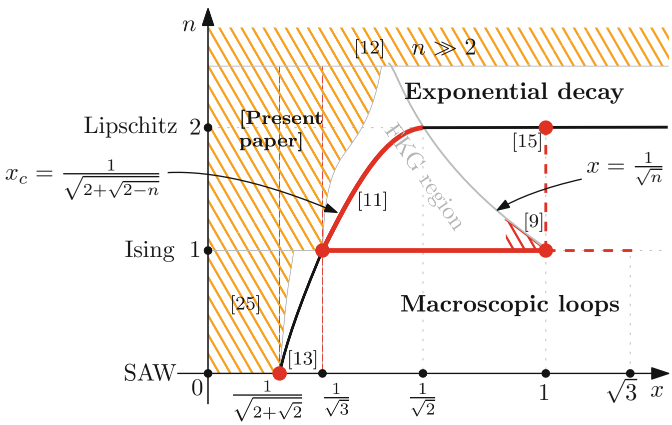


Fig. 1 The phase diagram of the loop $O(n)$ model. It is expected that above and to the left of the curve $x_c(n) = \frac{1}{\sqrt{2+\sqrt{2-n}}}$ (in black) the model exhibits exponential decay of loop lengths; below and on the curve, it is expected to have macroscopic loops and converge to CLE(κ) in the scaling limit. The convergence was established only at $n = 1, x = \frac{1}{\sqrt{3}}$ (critical Ising model [3, 8, 24]) and $n = x = 1$ (site percolation on $\mathbb{T}$ at $p_c = \frac{1}{2}$ [5, 23]). Regions where the behaviour was confirmed by recent results are marked in orange (for exponential decay) and red (for macroscopic loops). The relevant references are also marked

Let us mention that loop models were also considered on $\mathbb{Z}^d$. In particular, for n large enough, Chayes, Pryadko, and Shtengel [7] showed that the loop lengths exhibit exponential decay, for all values of x.

We finish the introduction by providing a sketch of our proof. There are three main steps in it. Fix $n > 1$. First, inspired by Chayes–Machta [6], we develop the partition function in $n = (n-1)+1$, so that it takes the form of the loop $O(1)$ model sampled on the vacant space of a weighted loop $O(n-1)$ model. Second, we use that the loop $O(1)$ model is the representation of the Ising model on the faces of $\mathbb{H}$; the latter exhibiting exponential decay of correlations for all $x < 1/\sqrt{3}$. Via the FK-Ising representation, this statement may be extended when the Ising model is sampled in the random domain given by a loop $O(n-1)$ configuration. At this stage we will have shown that the loop $O(n)$ model exhibits exponential decay when $x < 1/\sqrt{3}$. Finally, using enhancement techniques, we show that the presence of the loop $O(n-1)$ configuration strictly increases the critical parameter of the Ising model, thus allowing to extend our result to all $x < 1/\sqrt{3} + \varepsilon(n)$.

2 The Ising Connection

In this section we formalise a well-known connection between the Ising model (and its FK-representation) and the loop $O(1)$ model (see for instance [14, Sec. 3.10.1]). It will be useful to work with inhomogeneous measures in both models.

Fix a domain $\mathscr{D} = (V, E)$; we will omit it from notation when not necessary. Let $\mathbf{x} = (x_e)_{e\in E} \in [0,1]^E$ be a family of parameters. The loop $O(1)$ measure with parameters $\mathbf{x}$ is given by

$$\mathsf{Loop}_{\mathscr{D},1,\mathbf{x}}(\omega) = \mathsf{Loop}_{\mathbf{x}}(\omega) = \frac{1}{Z_{\mathrm{loop}}(\mathscr{D},1,\mathbf{x})}\Big(\prod_{e\in\omega} x_e\Big)\cdot \mathbf{1}_{\{\omega \text{ loop config.}\}} \quad \forall \omega \in \{0,1\}^E.$$

The percolation measure $\mathsf{Perco}_{\mathbf{x}}$ of parameters $\mathbf{x}$ consists of choosing the state of every edge independently, open with probability x_e for each edge $e \in E$,

$$\mathsf{Perco}_{\mathbf{x}}(\omega) = \Big(\prod_{e\in\omega} x_e\Big)\Big(\prod_{e\in E\setminus\omega}(1-x_e)\Big), \qquad \text{for all } \omega \in \{0,1\}^E.$$

Finally, associate to the parameters $\mathbf{x}$ the parameters $\mathbf{p} = (p_e)_E \in [0,1]^E$ defined by

$$p_e = p(x_e) = \frac{2x_e}{1+x_e}, \qquad \text{for all } e \in E.$$

Define the FK-Ising measure on $\mathscr{D}$ by

$$\Phi_{\mathbf{x}}(\omega) = \frac{1}{Z_{FK}(\mathbf{x})} \Big(\prod_{e\in\omega} p_e\Big)\Big(\prod_{e\in E\setminus\omega}(1-p_e)\Big) 2^{k(\omega)}, \qquad \text{for all } \omega \in \{0,1\}^E.$$

where $k(\omega)$ is the number of connected components of ω and $Z_{FK}(\mathbf{x})$ is a constant chosen so that $\Phi_{\mathbf{x}}$ is a probability measure.

When $\mathbf{x}$ is constant equal to some $x \in [0, 1]$, write x instead of $\mathbf{x}$. For $D \subset E$, write $\Phi_{D,x}$ and $\mathsf{Loop}_{D,1,x}$ for the FK-Ising and loop $O(1)$ measures, respectively, on $\mathscr{D}$ with inhomogeneous weights $(x\mathbf{1}_{\{e\in D\}})_{e\in E}$ (where $\mathbf{1}$ stands for the indicator function). These are simply the measures $\Phi_{\mathscr{D},x}$ and $\mathsf{Loop}_{\mathscr{D},1,x}$ conditioned on $\omega \cap D^c = \emptyset$.

Define a partial order on $\{0,1\}^E$ by saying that $\omega \le \omega'$ if $\omega_e \le \omega'_e$ for all $e \in E$. An event $A \subset \{0,1\}^E$ is called *increasing* if its indicator function is increasing with respect to this partial order.

Proposition 1 *Fix* $\mathbf{x} = (x_e)_{e\in E} \in [0,1]^E$ *and let* $\omega, \pi \in \{0,1\}^E$ *be two independent configurations chosen according to* $\mathsf{Loop}_{\mathbf{x}}$ *and* $\mathsf{Perco}_{\mathbf{x}}$*, respectively. Then the configuration* $\omega \vee \pi$ *defined by* $(\omega\vee\pi)(e) = \max\{\omega(e), \pi(e)\}$ *has law* $\Phi_{\mathbf{x}}$*. In particular*

$$\mathsf{Loop}_{\mathbf{x}} \le_{st} \Phi_{\mathbf{x}}, \tag{2}$$

where $\le_{st}$ *means stochastic domination, i.e.,* $\mathsf{Loop}_{\mathbf{x}}(A) \le \Phi_{\mathbf{x}}(A)$*, for any increasing event* A.

We give a short proof below. The reader familiar with the Ising model may consult the diagram of Fig. 2 for a more intricate but more natural proof.

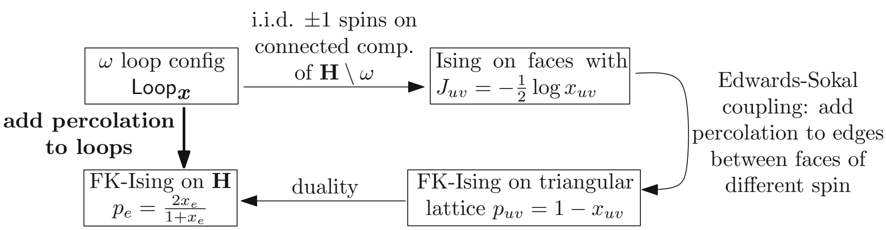


Fig. 2 The coupling of Proposition 1 via the spin-Ising representation

Proof Write $\mathsf{Loop}_{\mathbf{x}} \otimes \mathsf{Perco}_{\mathbf{x}}$ for the measure sampling ω and π independently. Fix $\eta \in \{0,1\}^E$ and let us calculate

$$\begin{aligned}
\mathsf{Loop}_{\mathbf{x}} \otimes \mathsf{Perco}_{\mathbf{x}}(\omega \vee \pi = \eta) &= \sum_{\substack{\omega \subset \eta \\ \omega \text{ loop config}}} \mathsf{Loop}_{\mathscr{D},\mathbf{x}}(\omega) \cdot \mathsf{Perco}_{\mathscr{D}\setminus\omega,\mathbf{x}}(\eta \setminus \omega) \\
&= \sum_{\substack{\omega \subset \eta \\ \omega \text{ loop config}}} \frac{1}{Z_{\text{loop}}(\mathscr{D}, 1, \mathbf{x})} \Big(\prod_{e\in\omega} x_e\Big)\Big(\prod_{e\in\eta\setminus\omega} x_e\Big)\Big(\prod_{e\in E\setminus\eta}(1-x_e)\Big) \\
&= \frac{1}{Z_{\text{loop}}(\mathscr{D}, 1, \mathbf{x})} \Big(\prod_{e\in\eta} x_e\Big)\Big(\prod_{e\in E\setminus\eta}(1-x_e)\Big) \sum_{\substack{\omega \subset \eta \\ \omega \text{ loop config}}} 1. \qquad (3)
\end{aligned}$$

Next we compute the number of loop configurations ω contained in η. Consider η as a graph embedded in the plane and let $F(\eta)$ be the set of connected components of $\mathbb{R}^2 \setminus \eta$; these are the faces of η. The set of loop configurations ω contained in η is in bijection with the set of assignments of spins ± 1 to the faces of η, with the only constraint that the infinite face has spin $+1$. Indeed, given a loop configuration $\omega \subset \eta$, assign spin -1 to the faces of η surrounded by an odd number of loops of ω, and $+1$ to all others. The inverse map is obtained by considering the edges separating faces of distinct spin.

The Euler formula applied to the graph η reads $|V| - |\eta| + |F(\eta)| = 1 + k(\eta)$. Hence, the number of loop configurations contained in η is

$$\sum_{\substack{\omega \subset \eta \\ \omega \text{ loop config}}} 1 = 2^{|F(\eta)|-1} = 2^{k(\eta)+|\eta|-|V|}.$$

Inserting this in (3), we find

$$\begin{aligned}
\mathsf{Loop}_{\mathbf{x}} \otimes \mathsf{Perco}_{\mathbf{x}}\big(\omega \vee \pi = \eta\big) &= \frac{2^{-|V|}}{Z_{\text{loop}}(\mathscr{D}, 1, \mathbf{x})} \Big(\prod_{e\in\eta} 2x_e\Big)\Big(\prod_{e\in E\setminus\eta}(1-x_e)\Big) 2^{k(\eta)} \\
&= \frac{2^{-|V|}\prod_{e\in E}(1+x_e)}{Z_{\text{loop}}(\mathscr{D}, 1, \mathbf{x})} \Big(\prod_{e\in\eta} \tfrac{2x_e}{1+x_e}\Big)\Big(\prod_{e\in E\setminus\eta}(1-\tfrac{2x_e}{1+x_e})\Big) 2^{k(\eta)}.
\end{aligned}$$

Since $\mathsf{Loop}_{\mathbf{x}} \otimes \mathsf{Perco}_{\mathbf{x}}$ is a probability measure, we deduce that it is equal to $\Phi_{\mathbf{x}}$ and that the normalising constants are equal, namely

$$\frac{Z_{\text{loop}}(\mathscr{D}, 1, \mathbf{x})}{2^{-|V|}\prod_{e\in E}(1+x_e)} = Z_{\text{FK}}(\mathscr{D}, \mathbf{x}). \qquad (4)$$

The procedure of sampling $(\omega, \omega \vee \pi)$ provides an increasing coupling between $\mathsf{Loop}_{\mathbf{x}}$ and $\Phi_{\mathbf{x}}$; the stochastic domination follows from the existence of said coupling. □

While the loop model has no apparent monotonicity, the FK-Ising model does. This will be of particular importance.

Proposition 2 (Thm. 3.21 [16]) *Let* $\mathbf{x} = (x_e)_{e\in E} \in [0, 1]^E$ *and* $\tilde{\mathbf{x}} = (\tilde{x}_e)_{e\in E} \in [0, 1]^E$ *be two sets of parameters with* $x_e \le \tilde{x}_e$ *for all* $e \in E$*. Then* $\Phi_{\mathbf{x}} \le_{st} \Phi_{\tilde{\mathbf{x}}}$*.*

The version above is slightly different from [16, Thm 3.21], as it deals with inhomogeneous measures; adapting the proof is straightforward.

Finally, it is well known that the FK-Ising model on the hexagonal lattice exhibits a sharp phase transition at $p_c = \frac{2}{\sqrt{3}+1}$: the critical point for the Ising model was computed by Onsager [21] (see [2] for the explicit formula on the triangular lattice), the sharpness of the phase transition was shown in [1]. For $p = p(x)$ strictly below p_c, which is to say $x < \frac{1}{\sqrt{3}}$, the model exhibits exponential decay of cluster volumes. Indeed, this may be easily deduced using [16, Thm. 5.86].

Theorem 2 *For* $x < \frac{1}{\sqrt{3}}$ *there exist* $c = c(x) > 0$ *and* $C > 0$ *such that, for any domain* $\mathscr{D}$ *and any* $k \in \mathbb{N}$*,*

$$\Phi_{\mathscr{D},x}(|\mathscr{C}_0| \ge k) \le C\, e^{-c\,k},$$

where $\mathscr{C}_0$ *denotes the cluster containing* 0 *and* $|\mathscr{C}_0|$ *its number of vertices.*

3 $n = (n - 1) + 1$

Fix a domain $\mathscr{D} = (V, E)$ and a value $n > 1$. Choose ω according to $\mathsf{Loop}_{\mathscr{D},n,x}$. Colour each loop of ω in blue with probability $1 - \frac{1}{n}$ and red with probability $\frac{1}{n}$. Let ω_b and ω_r be the configurations formed only of the blue and red loops, respectively; extend $\mathsf{Loop}_{\mathscr{D},n,x}$ to incorporate this additional randomness.

Proposition 3 *For any two non-intersecting loop configurations* ω_b *and* ω_r*,*

$$\mathsf{Loop}_{\mathscr{D},n,x}(\omega_r \mid \omega_b) = \mathsf{Loop}_{\mathscr{D}\setminus\omega_b,1,x}(\omega_r) \qquad \textit{and}$$

$$\mathsf{Loop}_{\mathscr{D},n,x}(\omega_b) = \frac{Z_{\text{loop}}(\mathscr{D}\setminus\omega_b, 1, x)}{Z_{\text{loop}}(\mathscr{D}, n, x)}(n-1)^{\ell(\omega_b)} x^{|\omega_b|}.$$

Proof For two non-intersecting loop configurations ω_b and ω_r, if we write $\omega = \omega_b \vee \omega_r$, we have

$$\begin{aligned}\mathsf{Loop}_{\mathscr{D},n,x}(\omega_b,\omega_r) &= \left(\tfrac{n-1}{n}\right)^{\ell(\omega_b)}\left(\tfrac{1}{n}\right)^{\ell(\omega_r)}\mathsf{Loop}_{\mathscr{D},n,x}(\omega)\\ &= \frac{1}{Z_{\text{loop}}(\mathscr{D},n,x)}(n-1)^{\ell(\omega_b)}x^{|\omega_b|+|\omega_r|}\\ &= \frac{Z_{\text{loop}}(\mathscr{D}\setminus\omega_b,1,x)}{Z_{\text{loop}}(\mathscr{D},n,x)}(n-1)^{\ell(\omega_b)}x^{|\omega_b|}\cdot\frac{x^{|\omega_r|}}{Z_{\text{loop}}(\mathscr{D}\setminus\omega_b,1,x)}.\end{aligned}$$

Notice that ω_r only appears in the last fraction. Moreover, if we sum this fraction over all loop configurations ω_r not intersecting ω_b, we obtain 1. This proves both assertions of the proposition. □

Recall that for a percolation configuration, $\mathscr{C}_0$ denotes the connected component containing 0. If ω is a loop configuration, then $\mathscr{C}_0(\omega)$ is simply the loop in ω that passes through 0 (with $\mathscr{C}_0(\omega) := \{0\}$ if no such loop exists).

Corollary 1 *Let $n \geq 1$ and $x < 1/\sqrt{3}$. Then $\mathsf{Loop}_{\mathscr{D},n,x}$ exhibits exponential decay.*

Proof For any domain $\mathscr{D}$ and $k \geq 1$ we have

$$\begin{aligned}\tfrac{1}{n}\mathsf{Loop}_{\mathscr{D},n,x}(|\mathscr{C}_0(\omega)| \geq k) &= \mathsf{Loop}_{\mathscr{D},n,x}(|\mathscr{C}_0(\omega_r)| \geq k)\\ &\leq \mathsf{Loop}_{\mathscr{D},n,x}\big[\Phi_{\mathscr{D}\setminus\omega_b,x}(|\mathscr{C}_0| \geq k)\big] && \text{by Prop. 1 and 3}\\ &\leq \Phi_{\mathscr{D},x}(|\mathscr{C}_0| \geq k) && \text{by Prop. 2}\\ &\leq Ce^{-ck} && \text{by Thm. 2,}\end{aligned}$$

where $c = c(x) > 0$ and $C > 0$ are given by Theorem 2. Thus, the length of the loop of 0 has exponential tail, uniformly in the domain $\mathscr{D}$. In particular, if $\mathscr{D}$ is fixed, the above bound also applies to any translates of $\mathscr{D}$, hence to the loop of any given point in $\mathscr{D}$.

Let $v_0, v_1, v_2 \dots$ be the vertices of $\mathscr{D}$ on the horizontal line to the right of 0, ordered from left to right, starting with $v_0 = 0$. If $\mathsf{R} \geq k$, then the largest loop surrounding 0 either passes through one of the points $v_0, \dots, v_{k-1}$ and has length at least k, or it passes through some v_j with $j \geq k$, and has length at least j, so as to manage to surround 0. Thus, using the bound derived above, we find

$$\mathsf{Loop}_{\mathscr{D},n,x}(\mathsf{R} \geq k) \leq n\Big[C\,ke^{-ck} + \sum_{j\geq k} C\,e^{-cj}\Big] \leq C'e^{-c'k},$$

for some altered constants $c' > 0$ and C' that depend on c, C and n but not on k. □

4 A Little Extra Juice: Enhancement

Fix some domain $\mathscr{D} = (V, E)$ for the whole of this section. Let ω_b be a blue loop configuration. Associate to it the spin configuration $\sigma_b \in \{-1, +1\}^{F(\mathscr{D})}$ obtained by awarding spins -1 to all faces of $\mathscr{D}$ that are surrounded by an odd number of loops, and spins $+1$ to all other faces (the same blue spin configuration was also defined in [15]). Write $\mathsf{D}_+ = \mathsf{D}_+(\sigma_b)$ (and $\mathsf{D}_- = \mathsf{D}_-(\sigma_b)$, respectively) for the set of edges of $\mathscr{D}$ that have σ_b-spin $+1$ (and -1, respectively) on both sides. All faces outside of $\mathscr{D}$ are considered to have spin $+1$ in this definition. Equivalently, D_- is the set of edges of $\mathscr{D} \setminus \omega_b$ surrounded by an odd number of loops of ω_b and $\mathsf{D}_+ = \mathscr{D} \setminus (\omega_b \cup \mathsf{D}_-)$. Both D_+ and D_- will also be regarded as spanning subgraphs of $\mathscr{D}$ with edge-sets D_+ and D_-, respectively.

Since no edge of D_+ is adjacent to any edge of D_-, a sample of the loop $O(1)$ measure $\mathsf{Loop}_{\mathscr{D}\setminus\omega_b,1,x}$ may be obtained by the superposition of two independent samples from $\mathsf{Loop}_{\mathsf{D}_+,1,x}$ and $\mathsf{Loop}_{\mathsf{D}_-,1,x}$, respectively. In particular, using (2),

$$\begin{aligned}\mathsf{Loop}_{\mathscr{D}\setminus\omega_b,1,x}(|\mathscr{C}_0(\omega_r)| \geq k) &= \mathsf{Loop}_{\mathsf{D}_+,1,x}(|\mathscr{C}_0(\omega_r)| \geq k) + \mathsf{Loop}_{\mathsf{D}_-,1,x}(|\mathscr{C}_0(\omega_r)| \geq k)\\ &\leq \Phi_{\mathsf{D}_+,x}(|\mathscr{C}_0| \geq k) + \Phi_{\mathsf{D}_-,x}(|\mathscr{C}_0| \geq k). \qquad (5)\end{aligned}$$

Actually, depending on ω_b, at most one of the terms on the RHS above is non-zero. We nevertheless keep both terms as we will later average on ω_b. The two following lemmas will be helpful in proving Theorem 1.

Lemma 1 *Let $n > 1$, $x < 1$, and set*

$$\alpha = \left[\frac{\max\{(n-1)^2, (n-1)^{-2}\}}{\left(\frac{x}{2(x+1)}\right)^6 + \max\{(n-1)^2, (n-1)^{-2}\}}\right]^{1/6} < 1. \qquad (6)$$

If ω_b has the law of the blue loop configuration of $\mathsf{Loop}_{\mathscr{D},n,x}$, then both laws of D_+ and D_- are stochastically dominated by Perco_α.

Lemma 2 *Fix $x \in (0, 1)$ and $\alpha < 1$. Let $\tilde{x} < x$ be such that*

$$\frac{\tilde{x}}{1-\tilde{x}} = \frac{x}{1-x} \cdot \left(1 + \frac{1+x}{2(1-x)} \cdot \frac{1-\alpha}{\alpha}\right)^{-1}. \qquad (7)$$

Write $\mathsf{Perco}_\alpha(\Phi_{\mathsf{D},x}(.))$ for the law of η chosen using the following two step procedure: choose D according to Perco_α, then choose η according to $\Phi_{\mathsf{D},x}$. Then

$$\mathsf{Perco}_\alpha(\Phi_{\mathsf{D},x}(.)) \leq_{st} \Phi_{\mathscr{D},\tilde{x}}.$$

Before proving the two lemmas above, let us show that they imply the main result.

Proof (of Theorem 1) Fix $n > 1$. An elementary computation proves the existence of some $\varepsilon = \varepsilon(n) > 0$ such that, if $x < \frac{1}{\sqrt{3}} + \varepsilon(n)$ and α and $\tilde{x}$ are defined in terms of n and x via (6) and (7), respectively, then $\tilde{x} < \frac{1}{\sqrt{3}}$. Straightforward computations based on the asymptotics of α established in a footnote in the proof of Lemma 1 lead to the asymptotics of $\varepsilon(n)$ as $n \searrow 1$ given in Theorem 1.

Fix $x < \frac{1}{\sqrt{3}} + \varepsilon(n)$ along with the resulting values α and $\tilde{x} < \frac{1}{\sqrt{3}}$. Then, for any domain $\mathscr{D}$ and $k \geq 1$ we have

$$
\begin{aligned}
&\tfrac{1}{n}\mathsf{Loop}_{\mathscr{D},n,x}(|\mathscr{C}_0(\omega)| \geq k) \\
&\qquad = \mathsf{Loop}_{\mathscr{D},n,x}(|\mathscr{C}_0(\omega_r)| \geq k) \\
&\qquad \leq \mathsf{Loop}_{\mathscr{D},n,x}\big[\Phi_{\mathsf{D}_+,x}(|\mathscr{C}_0| \geq k) + \Phi_{\mathsf{D}_-,x}(|\mathscr{C}_0| \geq k)\big] && \text{by (5)} \\
&\qquad \leq 2\,\mathsf{Perco}_\alpha\big[\Phi_{\mathsf{D},x}(|\mathscr{C}_0| \geq k)\big] && \text{by Lemma 1} \\
&\qquad \leq 2\,\Phi_{\mathscr{D},\tilde{x}}(|\mathscr{C}_0| \geq k) && \text{by Lemma 2} \\
&\qquad \leq 2\,C\,e^{-ck} && \text{by Thm. 2.}
\end{aligned}
$$

In the third line, we have used Lemma 1 and the stochastic monotonicity of Φ in terms of the domain. Indeed, Lemma 1 implies that $\mathsf{Loop}_{\mathscr{D},n,x}$ and Perco_α may be coupled so that the sample D_+ obtained from the former is included in the sample D obtained from the latter. Thus $\Phi_{\mathsf{D}_+,x} \leq \Phi_{\mathsf{D},x}$. The same applies separately for D_-.

To conclude (1), continue in the same way as in the proof of Corollary 1. □

The following computation will be useful for the proofs of both lemmas. Let $D \subset E$ and $e \in E \setminus D$. We will also regard D as a spanning subgraph of $\mathscr{D}$ with edge-set D. Recall that $Z_{FK}(D, x)$ is the partition function of the FK-Ising measure $\Phi_{D,x}$ on D. Then

$$
\begin{aligned}
Z_{FK}(D,x) &= \sum_{\eta \subset D} p^{|\eta|}(1-p)^{|D|-|\eta|}2^{k(\eta)} \\
&= \sum_{\eta \subset D} p^{|\eta|}(1-p)^{|D\cup\{e\}|-|\eta|}\,2^{k(\eta)} + p^{|\eta\cup\{e\}|}(1-p)^{|D|-|\eta|}\,2^{k(\eta)} \\
&\geq \sum_{\eta \subset D\cup\{e\}} p^{|\eta|}(1-p)^{|D\cup\{e\}|-|\eta|}2^{k(\eta)} = Z_{FK}(D\cup\{e\},x), \qquad (8)
\end{aligned}
$$

since $k(\eta) \geq k(\eta \cup \{e\})$. Conversely, $k(\eta) \leq k(\eta \cup \{e\}) + 1$, which implies

$$
Z_{FK}(D,x) \leq 2\,Z_{FK}(D\cup\{e\},x). \qquad (9)
$$

Proof (of Lemma 1) For $\beta \in (0,1)$ let P_β be the Bernoulli percolation on the faces of $\mathscr{D}$ of parameter β:

$$P_\beta(\sigma) = \beta^{\#\{u : \sigma(u)=+1\}}(1-\beta)^{\#\{u : \sigma(u)=-1\}} \qquad \text{for all } \sigma \in \{-1,+1\}^{F_\mathscr{D}}.$$

To start, we will prove that the law induced on σ_b by $\mathsf{Loop}_{\mathscr{D},n,x}$ is dominated by P_β for some β sufficiently close to 1. Both measures are positive, and Holley's inequality [17] states that the stochastic ordering is implied by

$$\frac{\mathsf{Loop}_{\mathscr{D},n,x}(\sigma_b = \varsigma_1)}{\mathsf{Loop}_{\mathscr{D},n,x}(\sigma_b = \varsigma_1 \wedge \varsigma_2)} \le \frac{P_\beta(\varsigma_1 \vee \varsigma_2)}{P_\beta(\varsigma_2)} = \Big(\frac{\beta}{1-\beta}\Big)^{\#\{u : \varsigma_1(u)=+1,\, \varsigma_2(u)=-1\}}$$

for all $\varsigma_1, \varsigma_2 \in \{\pm 1\}^{F_\mathscr{D}}$. The RHS above only depends on the number of faces of spin $+$ in ς_1 and spin $-$ in ς_2. It is then elementary to check that the general inequality above is implied by the restricted case where ς_1 differs at exactly one face u from ς_2, and $\varsigma_1(u) = +1$ but $\varsigma_2(u) = -1$.

Fix two such configurations ς_1, ς_2; write ω_1 and ω_2 for their associated loop configurations. Then, by Lemma 3,

$$\frac{\mathsf{Loop}_{\mathscr{D},n,x}(\sigma_b = \varsigma_1)}{\mathsf{Loop}_{\mathscr{D},n,x}(\sigma_b = \varsigma_2)} = \frac{Z_{\text{loop}}(\mathscr{D} \setminus \omega_1, 1, x)}{Z_{\text{loop}}(\mathscr{D} \setminus \omega_2, 1, x)}\,(n-1)^{\ell(\omega_1)-\ell(\omega_2)} x^{|\omega_1|-|\omega_2|}.$$

Since ς_1 and ς_2 only differ by one face, ω_1 and ω_2 differ only in the states of the edges surrounding that face. In particular $||\omega_1| - |\omega_2|| \le 6$ and $|\ell(\omega_1) - \ell(\omega_2)| \le 2$. Finally, using (4), (8), and (9), we find

$$\frac{Z_{\text{loop}}(\mathscr{D} \setminus \omega_1, 1, x)}{Z_{\text{loop}}(\mathscr{D} \setminus \omega_2, 1, x)} = \frac{Z_{\text{FK}}(\mathscr{D} \setminus \omega_1, 1, x)(1+x)^{|\omega_2|}}{Z_{\text{FK}}(\mathscr{D} \setminus \omega_2, 1, x)(1+x)^{|\omega_1|}} \le [2(x+1)]^{|\omega_2| - |\omega_1 \wedge \omega_2|} \le [2(x+1)]^6.$$

In conclusion

$$\frac{\mathsf{Loop}_{\mathscr{D},n,x}(\sigma_b = \varsigma_1)}{\mathsf{Loop}_{\mathscr{D},n,x}(\sigma_b = \varsigma_2)} \le \Big(\frac{2(x+1)}{x}\Big)^6 \cdot \max\{(n-1)^2, (n-1)^{-2}\}.$$

Then, if we set

$$\beta = \frac{\big(\frac{2(x+1)}{x}\big)^6 \cdot \max\{(n-1)^2, (n-1)^{-2}\}}{1 + \big(\frac{2(x+1)}{x}\big)^6 \cdot \max\{(n-1)^2, (n-1)^{-2}\}},$$

we indeed obtain the desired domination of σ_b by P_β.[1] The same proof shows that $-\sigma_b$ is also dominated by P_β.

Next, let us prove the domination of D_+ by a percolation measure. Set[2] $\alpha = \beta^{1/6}$. Let η_L and η_R be two percolation configurations chosen independently according to Perco_α. Also choose an orientation for every edge of E; for boundary edges, orient them such that the face of $\mathscr{D}$ adjacent to them is on their left.

Define $\tilde{\sigma} \in \{\pm 1\}^{F_\mathscr{D}}$ as follows. Consider some face u. For an edge e adjacent to u, u is either on the left of e or on its right, according to the orientation chosen for e. If it is on the left, retain the number $\eta_L(e)$, otherwise retain $\eta_R(e)$. Consider that u has spin $+1$ under $\tilde{\sigma}$ if and only if all the six numbers retained above are 1. Formally, for each $u \in F_\mathscr{D}$, set $\tilde{\sigma}(u) = +1$ if and only if

$$\prod_{e \text{ adjacent to } u} \big(\eta_L(e)\mathbf{1}_{\{u \text{ is left of } e\}} + \eta_R(e)\mathbf{1}_{\{u \text{ is right of } e\}}\big) = 1.$$

As a consequence, for an edge e in the interior of $\mathscr{D}$ to be in $\mathsf{D}_+(\tilde{\sigma})$, the faces on either side of e need to have $\tilde{\sigma}$-spin $+1$, hence $\eta_L(e) = \eta_R(e) = 1$ is required. For boundary edges e to be in $\mathsf{D}_+(\tilde{\sigma})$, only the restriction $\eta_L(e) = 1$ remains. In conclusion $\eta_L \geq \mathsf{D}_+(\tilde{\sigma})$.

Let us analyse the law of $\tilde{\sigma}$. Each value $\eta_L(e)$ and $\eta_R(e)$ appears in the definition of one $\tilde{\sigma}(u)$. As a consequence, the variables $\big(\tilde{\sigma}(u)\big)_{u \in F}$ are independent. Moreover, $\tilde{\sigma}(u) = 1$ if and only if all the six edges around e are open in one particular configuration η_L or η_R, which occurs with probability $\alpha^6 = \beta$. As a consequence $\tilde{\sigma}$ has law P_β.

By the previously proved domination, $\mathsf{Loop}_{\mathscr{D},n,x}$ may be coupled with P_β so that $\tilde{\sigma} \geq \sigma_b$. If this is the case, we have

$$\eta_L \geq \mathsf{D}_+(\tilde{\sigma}) \geq \mathsf{D}_+(\sigma_b).$$

Thus, η_L indeed dominates $\mathsf{D}_+(\sigma_b)$, as required.

The same proof shows that η_L dominates $\mathsf{D}_-(\sigma_b)$. For clarity, we mention that this does not imply that η_L dominates $\mathsf{D}_+(\sigma_b)$ and $\mathsf{D}_-(\sigma_b)$ simultaneously, which would translate to η_L dominating $\mathscr{D} \setminus \omega_b$. □

Proof (of Lemma 2) Fix x, α and $\tilde{x}$ as in the Lemma. The statement of Holley's inequality applied to our case may easily be reduced to

$$\frac{\mathsf{Perco}_\alpha[\Phi_{\mathsf{D},x}(\eta \cup \{e\})]}{\mathsf{Perco}_\alpha[\Phi_{\mathsf{D},x}(\eta)]} \leq \frac{\Phi_{\mathscr{D},\tilde{x}}(\tilde{\eta} \cup \{e\})}{\Phi_{\mathscr{D},\tilde{x}}(\tilde{\eta})} \qquad \text{for all } \eta \leq \tilde{\eta} \text{ and } e \notin \tilde{\eta}. \tag{10}$$

[1] This domination is of special interest as $n \searrow 1$ and for $x \geq 1/\sqrt{3}$. Then we may simplify the value of β as $\beta = \frac{(2+2\sqrt{3})^6}{(n-1)^2+(2+2\sqrt{3})^6} \sim 1 - \frac{1}{(2+2\sqrt{3})^6}(n-1)^2$.

[2] As $n \searrow 1$ and $x \geq 1/\sqrt{3}$, we may assume that $\alpha \sim 1 - \frac{1}{6\,(2+2\sqrt{3})^6}(n-1)^2$.

Fix η, $\tilde{\eta}$ and $e = (uv)$ as above. For $D \subset E$ with $e \in D$, a standard computation yields

$$\varphi_x(e|\eta) := \frac{\Phi_{D,x}(\eta \cup \{e\})}{\Phi_{D,x}(\eta)} = \begin{cases} \frac{2x}{1-x} & \text{if } u \overset{\eta}{\longleftrightarrow} v \text{ and} \\ \frac{x}{1-x} & \text{otherwise.} \end{cases}$$

The same quantity may be defined for $\tilde{x}$ instead of x and $\tilde{\eta}$ instead of η; it is increasing in both η and x. Moreover $\varphi_x(e|\eta)$ does not depend on D, as long as $e \in D$ and $\eta \subset D$. If the first condition fails, then the numerator is 0; if the second fails then the denominator is null and the ratio is not defined.

Let us perform a helpful computation before proving (10). Fix D with $e \in D$. By (9),

$$\frac{\Phi_{D\setminus\{e\},x}(\eta)}{\Phi_{D,x}(\eta)} = \frac{Z_{FK}(D,x)}{Z_{FK}(D\setminus\{e\},x)} \cdot \frac{1+x}{1-x} \geq \frac{1+x}{2(1-x)}.$$

The factor $\left(\frac{1-x}{1+x}\right)^{-1}$ comes from the fact that the weights of η under $\Phi_{D\setminus\{e\},x}$ and $\Phi_{D,x}$ differ by the contribution of the closed edge e. If follows that

$$\left(1 + \frac{1+x}{2(1-x)} \cdot \frac{1-\alpha}{\alpha}\right) \Phi_{D,x}(\eta) \leq \Phi_{D,x}(\eta) + \frac{1-\alpha}{\alpha}\, \Phi_{D\setminus\{e\},x}(\eta).$$

The choice of $\tilde{x}$ is such that

$$\frac{\varphi_x(e|\eta)}{\varphi_{\tilde{x}}(e|\eta)} = \frac{x}{1-x} \cdot \frac{1-\tilde{x}}{\tilde{x}} = 1 + \frac{1+x}{2(1-x)} \cdot \frac{1-\alpha}{\alpha}.$$

Using the last two displayed equations, we find

$$\begin{aligned}
&\mathsf{Perco}_\alpha[\Phi_{\mathsf{D},x}(\eta \cup \{e\})] = \sum_{D\subset E} \alpha^{|D|}(1-\alpha)^{|E|-|D|}\, \Phi_{D,x}(\eta \cup \{e\}) \\
&= \sum_{\substack{D\subset E \\ \text{with } e\in D}} \alpha^{|D|}(1-\alpha)^{|E|-|D|}\, \varphi_x(e|\eta)\, \Phi_{D,x}(\eta) \\
&\leq \left(1 + \tfrac{1+x}{2(1-x)} \tfrac{1-\alpha}{\alpha}\right)^{-1} \varphi_x(e|\eta) \sum_{\substack{D\subset E \\ \text{with } e\in D}} \alpha^{|D|}(1-\alpha)^{|E|-|D|} \left[\Phi_{D,x}(\eta) + \tfrac{1-\alpha}{\alpha} \Phi_{D\setminus\{e\},x}(\eta)\right] \\
&= \varphi_{\tilde{x}}(e|\tilde{\eta}) \sum_{D\subset E} \alpha^{|D|}(1-\alpha)^{|E|-|D|} \Phi_{D,x}(\eta) \\
&= \varphi_{\tilde{x}}(e|\tilde{\eta})\, \mathsf{Perco}_\alpha[\Phi_{\mathsf{D},x}(\eta)].
\end{aligned}$$

Divide by $\mathsf{Perco}_\alpha[\Phi_{\mathsf{D},x}(\eta)]$ and recall the definition of $\varphi_{\tilde{x}}(e|\tilde{\eta})$ to obtain (10). □

5 Open Questions/Perspectives

Our main theorem shows that if x is such that the model with parameters x and $n = 1$ exhibits exponential decay, then so do all models with the same parameter x and $n \geq 1$. A natural generalisation of this is the following.

Question 1 Show that if x and n are such that the loop $O(n)$ model exhibits exponential decay, then so do all models with parameters x and $\tilde{n}$ for any $\tilde{n} \geq n$.

A positive answer to the above would show that the critical point $x_c(n)$ (assuming it exists) is increasing in n. The same technique as in Sect. 4 may even prove that it is strictly increasing. Moreover, it was recently shown in [12] that, in the regime $n \geq 1$ and $x \leq \frac{1}{\sqrt{n}}$, the loop $O(n)$ model satisfies the following dichotomy: either it exhibits macroscopic loops or exponential decay. In addition, for $n \in [1, 2]$ and $x = \frac{1}{\sqrt{2+\sqrt{2-n}}}$ the loop $O(n)$ model was shown to exhibit macroscopic loops. Thus, assuming Question 1 is solved, we deduce that the loop $O(n)$ model with $n \leq 2$ and $x \in [\frac{1}{\sqrt{2+\sqrt{2-n}}}, \frac{1}{\sqrt{2}}]$ does not exhibit exponential decay. Indeed, Question 1 may be used to compare the model with such parameters n, x to that with parameters $\tilde{n}, x$, where $\tilde{n}$ satisfies $x = \frac{1}{\sqrt{2+\sqrt{2-\tilde{n}}}}$; the latter is known to not have exponential decay. In addition, the dichotomy result would then imply the existence of macroscopic loops for $n \in [1, 2]$ and $x \in [\frac{1}{\sqrt{2+\sqrt{2-n}}}, \frac{1}{\sqrt{2}}]$.

Let us now describe a possible approach to Question 1. The strategy of our proof of Theorem 1 was based on the following observation. The loop $O(1)$ model, or rather its associated FK-Ising model, has a certain monotonicity in x. This translates to a monotonicity in the domain: the larger the domain, the higher the probability that a given point is contained in a large loop. This fact is used to compare the loop $O(1)$ model in a simply connected domain $\mathscr{D}$ with that in the domain obtained from $\mathscr{D}$ after removing certain interior parts. The latter is generally not simply connected, and it is essential that our monotonicity property can handle such domains.

Question 2 Associate to the loop $O(n)$ model with edge weight x in some domain $\mathscr{D}$ a positively associated percolation model $\Psi_{\mathscr{D},n,x}$ with the property that, if one exhibits exponential decay of connection probabilities, then so does the other.

The percolation model $\Psi_{\mathscr{D},n,x}$ actually only needs to have some monotonicity property in the domain, sufficient for our proof to apply. Unfortunately, we only have such an associated model when $n = 1$.

Suppose that one may find such a model Ψ for some value of n. Then our proof may be adapted. Indeed, fix x such that the loop $O(n)$ model exhibits exponential decay. Then $\Psi_{\mathscr{D},n,x}$ also exhibits exponential decay for any domain $\mathscr{D}$. Consider now the loop $O(\tilde{n})$ model with edge-weight x for $\tilde{n} > n$ and colour each loop independently in red with probability $n/\tilde{n}$ and in blue with probability $(\tilde{n} - n)/\tilde{n}$. Then, conditionally on the blue loop configuration ω_b, the red loop configuration has the law of the loop $O(n)$ model with edge-weight x in the domain $\mathscr{D} \setminus \omega_b$. By positive association, since $\Psi_{\mathscr{D},n,x}$ exhibits exponential decay, so does $\Psi_{\mathscr{D}\setminus\omega_b,n,x}$. Then the loop $O(\tilde{n})$ model exhibits exponential decay of lengths of red loops and hence in general of lengths of all loops.

Acknowledgments The authors would like to thank Ron Peled for suggesting to develop in $n = (n-1) + 1$ following Chayes and Machta. Our discussions with Hugo Duminil-Copin, Yinon Spinka and Marcelo Hilario were also very helpful. We acknowledge the hospitality of IMPA (Rio de Janeiro), where this project started.

The first author is supported by the Swiss NSF grant P300P2_177848, and partially supported by the European Research Council starting grant 678520 (LocalOrder). The second author is a member of the NCCR SwissMAP.

References

1. Aizenman, M., Barsky, D.J., Fernández, R.: The phase transition in a general class of Ising-type models is sharp. J. Statist. Phys. **47**(3–4), 343–374 (1987)
2. Beffara, V., Duminil-Copin, H.: The self-dual point of the two-dimensional random-cluster model is critical for $q \geq 1$. Probab. Theory Related Fields **153**(3–4), 511–542 (2012)
3. Benoist, S., Hongler, C.: The scaling limit of critical Ising interfaces is CLE_3. Ann. Probab. **47**(4), 2049–2086 (2019)
4. Blöte, H.W., Nienhuis, B.: The phase diagram of the $O(n)$ model. Phys. A Stat. Mech. Appl. **160**(2), 121–134 (1989)
5. Camia, F., Newman, C.M.: Two-dimensional critical percolation: the full scaling limit. Commun. Math. Phys. **268**(1), 1–38 (2006)
6. Chayes, L., Machta, J.: Graphical representations and cluster algorithms II. Phys. A Stat. Mech. Appl. **254**(3), 477–516 (1998)
7. Chayes, L., Pryadko, L.P., Shtengel, K.: Intersecting loop models on $\mathbb{Z}^d$: rigorous results. Nucl. Phys. B **570**(3), 590–614 (2000)
8. Chelkak, D., Smirnov, S.: Universality in the 2D Ising model and conformal invariance of fermionic observables. Invent. Math. **189**(3), 515–580 (2012)
9. Crawford, N., Glazman, A., Harel, M., Peled, R.: Macroscopic loops in the loop $O(n)$ model via the XOR trick. Preprint arXiv: 2001.11977 (2020)
10. Domany, E., Mukamel, D., Nienhuis, B., Schwimmer, A.: Duality relations and equivalences for models with $O(n)$ and cubic symmetry. Nucl. Phys. B **190**(2), 279–287 (1981)
11. Duminil-Copin, H., Smirnov, S.: The connective constant of the honeycomb lattice equals $\sqrt{2+\sqrt{2}}$. Ann. Math. (2) **175**(3), 1653–1665 (2012)
12. Duminil-Copin, H., Glazman, A., Peled, R., Spinka, Y.: Macroscopic loops in the loop $O(n)$ model at Nienhuis' critical point. Preprint arXiv:1707.09335 (2017)
13. Duminil-Copin, H., Peled, R., Samotij, W., Spinka, Y.: Exponential decay of loop lengths in the loop $O(n)$ model with large n. Commun. Math. Phys. **349**(3), 777–817, 12 (2017)
14. Friedli, S., Velenik, Y.: Statistical Mechanics of Lattice Systems: A Concrete Mathematical Introduction. Cambridge University, Cambridge (2017)
15. Glazman, A., Manolescu, I.: Uniform Lipschitz Functions on the Triangular Lattice have Logarithmic Variations. Preprint arXiv:1810.05592 (2018)
16. Grimmett, G.: The random-cluster model. In: Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 333. Springer, Berlin (2006)
17. Holley, R.: Remarks on the FKG inequalities. Comm. Math. Phys. **36**, 227–231 (1974)

18. Kager, W., Nienhuis, B.: A guide to stochastic Löwner evolution and its applications. J. Statist. Phys. **115**(5–6), 1149–1229 (2004)
19. Lenz, W.: Beitrag zum Verständnis der magnetischen Eigenschaften in festen Körpern. Phys. Zeitschr. **21**, 613–615 (1920)
20. Nienhuis, B.: Exact critical point and critical exponents of $O(n)$ models in two dimensions. Phys. Rev. Lett. **49**(15), 1062–1065 (1982)
21. Onsager, L.: Crystal statistics. I. A two-dimensional model with an order-disorder transition. Phys. Rev. (2) **65**, 117–149 (1944)
22. Peled, R., Spinka, Y.: Lectures on the spin and loop $O(n)$ models. In: Sojourns in Probability Theory and Statistical Physics-I, pp. 246–320 (2019)
23. Smirnov, S.: Critical percolation in the plane: conformal invariance, Cardy's formula, scaling limits. C. R. Acad. Sci. Paris Sér. I Math. **333**(3), 239–244 (2001)
24. Smirnov, S.: Conformal invariance in random cluster models. I. Holomorphic fermions in the Ising model. Ann. Math. (2) **172**(2), 1435–1467 (2010)
25. Taggi, L.: Shifted critical threshold in the loop $O(n)$ model at arbitrary small n. Electron. Commun. Probab. **23**(96), 9 (2018)

Non-Coupling from the Past

Geoffrey R. Grimmett and Mark Holmes

Abstract The method of 'coupling from the past' permits exact sampling from the invariant distribution of a Markov chain on a finite state space. The coupling is successful whenever the stochastic dynamics are such that there is coalescence of all trajectories. The issue of the coalescence or non-coalescence of trajectories of a finite state space Markov chain is investigated in this note. The notion of the 'coalescence number' $k(\mu)$ of a Markovian coupling μ is introduced, and results are presented concerning the set $K(P)$ of coalescence numbers of couplings corresponding to a given transition matrix P.

Keywords Markov chain · Coupling from the past · CFTP · Perfect simulation · Coalescence · Coalescence number

1 Introduction

The method of 'coupling from the past' (CFTP) was introduced by Propp and Wilson [7, 8, 11] in order to sample from the invariant distribution of an irreducible Markov chain on a finite state space. It has attracted great interest amongst theoreticians and practitioners, and there is an extensive associated literature (see, for example [5, 10]).

The general approach of CFTP is as follows. Let X be an irreducible Markov chain on a finite state space S with transition matrix $P = (p_{i,j} : i, j \in S)$, and let

G. R. Grimmett (✉)
Statistical Laboratory, Centre for Mathematical Sciences, Cambridge University, Cambridge, UK

School of Mathematics and Statistics, The University of Melbourne, Parkville, VIC, Australia
e-mail: grg@statslab.cam.ac.uk

M. Holmes
School of Mathematics and Statistics, The University of Melbourne, Parkville, VIC, Australia
e-mail: holmes.m@unimelb.edu.au

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_22

π be the unique invariant distribution (see [4, Chap. 6] for a general account of the theory of Markov chains).

Let $\mathscr{F}_S$ be the set of functions from S to S, and let $\mathscr{P}_S$ be the set of all irreducible stochastic matrices on the finite set S. We write $\mathbb{N}$ for the set $\{1, 2, \dots\}$ of natural numbers, and $\mathbb{P}$ for a generic probability measure.

Definition 1 A probability measure μ on $\mathscr{F}_S$ is *consistent* with $P \in \mathscr{P}_S$, in which case we say that the pair (P, μ) is *consistent*, if

$$p_{i,j} = \mu\big(\{f \in \mathscr{F}_S : f(i) = j\}\big), \qquad i, j \in S. \tag{1}$$

Let $\mathscr{L}(P)$ denote the set of probability measures μ on $\mathscr{F}_S$ that are consistent with $P \in \mathscr{P}_S$.

Let $P \in \mathscr{P}_S$ and $\mu \in \mathscr{L}(P)$. The measure μ is called a *grand coupling* of P. Let $F = (F_s : s \in \mathbb{N})$ be a vector of independent samples from μ, let $\overleftarrow{F}_t$ denote the composition $F_1 \circ F_2 \circ \cdots \circ F_t$, and define the *backward coalescence time*

$$C = \inf\big\{t : \overleftarrow{F}_t(\cdot) \text{ is a constant function}\big\}. \tag{2}$$

We say that *backward coalescence occurs* if $C < \infty$. On the event $\{C < \infty\}$, $\overleftarrow{F}_C$ may be regarded as a random state.

The definition of coupling may seem confusing on first encounter. The function F_1 describes transitions during one step of the chain from time -1 to time 0, as illustrated in Fig. 1. If F_1 is not a constant function, we move back one step in time to -2, and consider the composition $F_1 \circ F_2$. This process is iterated, moving one step back in time at each stage, until the earliest (random) C such that the iterated function $\overleftarrow{F}_C$ is constant. This C (if finite) is the time to backward coalescence.

Propp and Wilson proved the following fundamental theorem.

Theorem 1 ([7]) *Let $P \in \mathscr{P}_S$ and $\mu \in \mathscr{L}(P)$. Either $\mathbb{P}(C < \infty) = 0$ or $\mathbb{P}(C < \infty) = 1$. If it is the case that $\mathbb{P}(C < \infty) = 1$, then the random state $\overleftarrow{F}_C$ has law π.*

Here are two areas of application of CFTP. In the first, one begins with a recipe for a certain probability measure π on S, for example as the posterior distribution

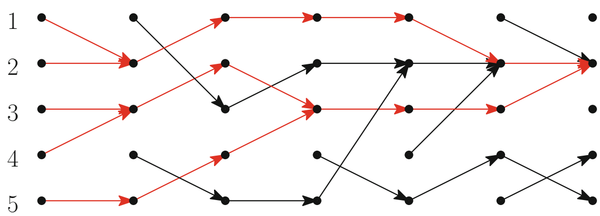


Fig. 1 An illustration of coalescence of trajectories in CFTP with $|S| = 5$

of a Bayesian analysis. In seeking a sample from π, one may find an aperiodic transition matrix P having π as unique invariant distribution, and then run CFTP on the associated Markov chain. In a second situation that may arise in a physical model, one begins with a Markovian dynamics with associated transition matrix $P \in \mathscr{P}_S$, and uses CFTP to sample from the invariant distribution. In the current work, we shall assume that the transition matrix P is specified, and that P is finite and irreducible.

Here is a summary of the work presented here. In Sect. 2, we discuss the phenomena of backward and forward coalescence, and we define the coalescence number of a Markov coupling. Informally, the coalescence number is the (deterministic) limiting number of un-coalesced trajectories of the coupling. Theorem 3 explains the relationship between the coalescence number and the ranks of products of extremal elements in a convex representation of the stochastic matrix P. The question is posed of determining the set $K(P)$ of coalescence numbers of couplings consistent with a given P. A sub-family of couplings, termed 'block measures', is studied in Sect. 4. These are couplings for which there is a fixed set of blocks (partitioning the state space), such that blocks of states are mapped to blocks of states, and such that coalescence occurs within but not between blocks. It is shown in Theorem 4, via Birkhoff's convex representation theorem for doubly stochastic matrices, that $|S| \in K(P)$ if and only if P is doubly stochastic. Some further results about $K(P)$ are presented in Sect. 5.

2 Coalescence of Trajectories

CFTP relies upon almost-sure backward coalescence, which is to say that $\mathbb{P}(C < \infty) = 1$, where C is given in (2). For given $P \in \mathscr{P}_S$, the occurrence (or not) of coalescence depends on the choice of $\mu \in \mathscr{L}(P)$; see for example, Example 1.

We next introduce the notion of 'forward coalescence', which is to be considered as '*coalescence*' but with the difference that time runs forwards rather than backwards. As before, let $P \in \mathscr{P}_S$, $\mu \in \mathscr{L}(P)$, and let $F = (F_s : s \in \mathbb{N})$ be an independent sample from μ. For $i \in S$, define the Markov chain $X^i = (X^i_t : t \geq 0)$ by $X^i_t = \vec{F}_t(i)$ where $\vec{F}_t = F_t \circ F_{t-1} \circ \cdots \circ F_1$. Then $(X^i : i \in S)$ is a family of coupled Markov chains, running forwards in time, each having transition matrix P, and such that X^i starts in state i.

The superscript $\rightarrow$ (respectively, $\leftarrow$) is used to indicate that time is running forwards (respectively, backwards). For $i, j \in S$, we say that i and j *coalesce* if there exists t such that $X^i_t = X^j_t$. We say that *forward coalescence occurs* if, for all pairs $i, j \in S$, i and j coalesce. The *forward coalescence time* is given by

$$T = \inf\{t \geq 0 : X^i_t = X^j_t \text{ for all } i, j \in S\}. \tag{3}$$

Clearly, if P is periodic then $T = \infty$ a.s. for any $\mu \in \mathscr{L}(P)$. A simple but important observation is that C and T have the same distribution.

Theorem 2 *Let $P \in \mathscr{P}_S$ and $\mu \in \mathscr{L}(P)$. The backward coalescence time C and the forward coalescence time T have the same distribution.*

Proof Let $(F_i : i \in \mathbb{N})$ be an independent sample from μ. For $t \geq 0$, we have

$$\mathbb{P}(C \leq t) = \mathbb{P}\big(\overleftarrow{F}_t(\cdot) \text{ is a constant function}\big).$$

By reversing the order of the functions $F_1, F_2, \ldots, F_t$, we see that this equals $\mathbb{P}(T \leq t) = \mathbb{P}(\overrightarrow{F}_t(\cdot) \text{ is a constant function})$.

Example 1 Let $S = \{1, 2, \ldots, n\}$ where $n \geq 2$, and let $P_n = (p_{i,j})$ be the constant matrix with entries $p_{i,j} = 1/n$ for $i, j \in S$. Let $F = (F_i : i \in \mathbb{N})$ be an independent sample from $\mu \in \mathscr{L}(P_n)$.

(a) If each F_i is a uniform random permutation of S, then $T \equiv \infty$ and $\overrightarrow{F}_t(i) \neq \overrightarrow{F}_t(j)$ for all $i \neq j$ and $t \geq 1$.
(b) If $(F_1(i) : i \in S)$ are independent and uniformly distributed on S, then $\mathbb{P}(T < \infty) = 1$.

In this example, there exist measures $\mu \in \mathscr{L}(P_n)$ such that either (a) a.s. no pairs of states coalesce, or (b) a.s. forward coalescence occurs.

For $g \in \mathscr{F}_S$, we let $\overset{g}{\sim}$ be the equivalence relation on S given by $i \overset{g}{\sim} j$ if and only if $g(i) = g(j)$. For $f = (f_t : t \in \mathbb{N}) \subseteq \mathscr{F}_S$ and $t \geq 1$, we write

$$\overleftarrow{f}_t = f_1 \circ f_2 \circ \cdots \circ f_t, \quad \overrightarrow{f}_t = f_t \circ f_{t-1} \circ \cdots \circ f_1.$$

Let $k_t(\overleftarrow{f})$ (respectively, $k_t(\overrightarrow{f})$) denote the number of equivalence classes of the relation $\overset{\overleftarrow{f}_t}{\sim}$ (respectively, $\overset{\overrightarrow{f}_t}{\sim}$). Similarly, we define the equivalence relation $\overset{\overleftarrow{f}}{\sim}$ on S by $i \overset{\overleftarrow{f}}{\sim} j$ if and only if $i \overset{\overleftarrow{f}_t}{\sim} j$ for some $t \in \mathbb{N}$, and we let $k(\overleftarrow{f})$ be the number of equivalence classes of $\overset{\overleftarrow{f}}{\sim}$ (and similarly for $\overrightarrow{f}$). We call $k(\overleftarrow{f})$ the *backward coalescence number* of $\overleftarrow{f}$, and likewise $k(\overrightarrow{f})$ the *forward coalescence number* of $\overrightarrow{f}$. The following lemma is elementary.

Lemma 1

(a) *We have that $k_t(\overleftarrow{f})$ and $k_t(\overrightarrow{f})$ are monotone non-increasing in t. Furthermore, $k_t(\overleftarrow{f}) = k(\overleftarrow{f})$ and $k_t(\overrightarrow{f}) = k(\overrightarrow{f})$ for all large t.*
(b) *Let $F = (F_s : s \in \mathbb{N})$ be independent and identically distributed elements in $\mathscr{F}_S$. Then $k_t(\overleftarrow{F})$ and $k_t(\overrightarrow{F})$ are equidistributed, and similarly $k(\overleftarrow{F})$ and $k(\overrightarrow{F})$ are equidistributed.*

Proof

(a) The first statement holds by consideration of the definition, and the second since $k(\overleftarrow{F})$ and $k(\overrightarrow{F})$ are integer-valued.
(b) This holds as in the proof of Theorem 2.

3 Coalescence Numbers

In light of Theorem 2 and Lemma 1, we henceforth consider only Markov chains running in *increasing positive time*. *Henceforth, expressions involving the word 'coalescence' shall refer to* forward *coalescence*. Let μ be a probability measure on $\mathscr{F}_S$, and let $\text{supp}(\mu)$ denote the support of μ. Let $F = (F_s : s \in \mathbb{N})$ be a vector of independent and identically distributed random functions, each with law μ. The law of F is the product measure $\boldsymbol{\mu} = \prod_{i \in \mathbb{N}} \mu$. The coalescence time T is given by (3), and the term *coalescence number* refers to the quantities $k_t(\overrightarrow{F})$ and $k(\overrightarrow{F})$, which we denote henceforth by $k_t(F)$ and $k(F)$, respectively.

Lemma 2 *Let μ, μ_1, μ_2 be probability measures on $\mathscr{F}_S$.*

(a) *Let $F = (F_s : s \in \mathbb{N})$ be a sequence of independent and identically distributed functions each with law μ. We have that $k(F)$ is $\boldsymbol{\mu}$-a.s. constant, and we write $k(\mu)$ for the almost surely constant value of $k(F)$.*
(b) *If $\text{supp}(\mu_1) \subseteq \text{supp}(\mu_2)$, then $k(\mu_1) \geq k(\mu_2)$.*
(c) *If $\text{supp}(\mu_1) = \text{supp}(\mu_2)$, then $k(\mu_1) = k(\mu_2)$.*

We call $k(\mu)$ the *coalescence number* of μ.

Proof

(a) For $j \in \{1, 2, \ldots, n\}$, let $q_j = \boldsymbol{\mu}(k(F) = j)$, and $k^* = \min\{j : q_j > 0\}$. Then

$$\boldsymbol{\mu}(k(F) \geq k^*) = 1. \tag{4}$$

Moreover, we choose $t \in \mathbb{N}$ such that

$$\kappa := \boldsymbol{\mu}(k_t(F) = k^*) \quad \text{satisfies} \quad \kappa > 0.$$

For $m \in \mathbb{N}$, write $F^m = (F_{mt+s} : s \geq 1)$. The event $E_{t,m} = \{k_t(F^m) = k^*\}$ depends only on $F_{mt+1}, F_{mt+2} \ldots, F_{(m+1)t}$. It follows that the events $\{E_{t,m} : m \in \mathbb{N}\}$ are independent, and each occurs with probability κ. Therefore, almost surely at least one of these events occurs, and hence $\boldsymbol{\mu}(k(F) \leq k^*) = 1$. By (4), this proves the first claim.
(b) Assume $\text{supp}(\mu_1) \subseteq \text{supp}(\mu_2)$, and let k_i^* be the bottom of the $\boldsymbol{\mu}_i$-support of $k(F)$. Since, for large t, $\boldsymbol{\mu}_1(k_t(F) = k_1^*) > 0$, we have also that $\boldsymbol{\mu}_2(k_t(F) = k_1^*) > 0$, whence $k_1^* \geq k_2^*$. Part (c) is immediate.

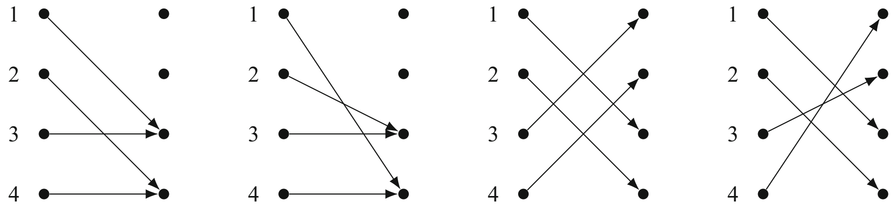


Fig. 2 Diagrammatic representations of the four functions f_i of Example 2. The corresponding equivalence classes are not $\boldsymbol{\mu}$-a.s. constant

Whereas $k(F)$ is a.s. constant (as in Lemma 2(a)), the equivalence classes of $\overset{\vec{F}}{\sim}$ need not themselves be a.s. constant. Here is an example of this, preceded by some notation.

Definition 2 Let $f \in \mathscr{F}_S$ where $S = \{i_1, i_2, \dots, i_n\}$ is a finite ordered set. We write $f = (j_1 j_2 \dots j_n)$ if $f(i_r) = j_r$ for $r = 1, 2, \dots, n$.

Example 2 Take $S = \{1, 2, 3, 4\}$ and any consistent pair (P, μ) with $\text{supp}(\mu) = \{f_1, f_2, f_3, f_4\}$, where

$$f_1 = (3434), \quad f_2 = (4334), \quad f_3 = (3412), \quad f_4 = (3421).$$

Then $k(\mu) = 2$ but the equivalence classes of $\overset{\vec{F}}{\sim}$ may be either $\{1, 3\}$, $\{2, 4\}$ or $\{1, 4\}$, $\{2, 3\}$, each having a strictly positive probability. The four functions f_i are illustrated diagrammatically in Fig. 2.

A probability measure μ on $\mathscr{F}_S$ may be written in the form

$$\mu = \sum_{f \in \text{supp}(\mu)} \alpha_f \delta_f, \tag{5}$$

where α is a probability mass function on $\mathscr{F}_S$ with support $\text{supp}(\mu)$, and δ_f is the Dirac delta-mass on the point $f \in \mathscr{F}_S$. Thus, $\alpha_f > 0$ if and only if $f \in \text{supp}(\mu)$. If $\mu \in \mathscr{L}(P)$, by (1) and (5),

$$P = \sum_{f \in \text{supp}(\mu)} \alpha_f M_f, \tag{6}$$

where M_f denotes the matrix

$$M_f = (1_{\{f(i)=j\}} : i, j \in S), \tag{7}$$

and 1_A is the indicator function of A.

Let Π_S be the set of permutations of S. We denote also by Π_S the set of matrices M_f as f ranges over the permutations of S.

Theorem 3 *Let μ have the representation (5), and $|S| = n$.*

(a) *We have that*

$$k(\mu) = \inf\left\{\operatorname{rank}(M_{f_t} M_{f_{t-1}} \cdots M_{f_1} : f_1, f_2, \dots, f_t \in \operatorname{supp}(\mu),\ t \geq 1\right\}. \tag{8}$$

(b) *There exists $T = T(n)$ such that the infimum in (8) is achieved for some t satisfying $t \leq T$.*

Proof

(a) Let $F = (F_s : s \in \mathbb{N})$ be drawn independently from μ. Then

$$R_t := M_{F_t} M_{F_{t-1}} \cdots M_{F_1}$$

is the matrix with (i, j)th entry $1_{\{\vec{F}_t(i)=j\}}$. Therefore, $k_t(F)$ equals the number of non-zero columns of R_t. Since each row of R_t contains a unique 1, we have that $k_t(F) = \operatorname{rank}(R_t)$. Therefore, $k(\mu)$ is the decreasing limit

$$k(\mu) = \lim_{t\to\infty} \operatorname{rank}(R_t) \qquad \text{a.s.} \tag{9}$$

Equation (8) follows since $k(\mu)$ is integer-valued and deterministic.

(b) Since the rank of a matrix is integer-valued, the infimum in (8) is attained. The claim follows since, for given $|S| = n$, there are boundedly many possible matrices M_f.

Let

$$K(P) = \left\{k : \text{ there exists } \mu \in \mathscr{L}(P) \text{ with } k(\mu) = k\right\}.$$

It is a basic question to ask: what can be said about K as a function of P? We first state a well-known result, based on ideas already in work of Doeblin [2].

Lemma 3 *We have that $1 \in K(P)$ if and only if $P \in \mathscr{P}_S$ is aperiodic.*

Proof For $f \in \mathscr{F}_S$, let $\mu(\{f\}) = \prod_{i\in S} p_{i,f(i)}$. This gives rise to $|S|$ chains with transition matrix P, starting from $1, 2, \dots, n$, respectively, that evolve independently until they meet. If P is aperiodic (and irreducible) then all n chains meet a.s. in finite time.

Conversely, if P is periodic and $p_{i,j} > 0$ then $i \neq j$, and i and j can never coalesce, implying $1 \notin K(P)$.

Remark 1 In a variety of cases of interest including, for example, the Ising and random-cluster models (see [3, Exer. 7.3, Sect. 8.2]), the set S has a partial order, denoted $\leq$. For $P \in \mathscr{P}_S$ satisfying the so-called FKG lattice condition, it is natural to seek $\mu \in \mathscr{L}(P)$ whose transitions preserve this partial order, and such μ may be constructed via the relevant Gibbs sampler (see, for example, [4, Sect. 6.14]). By the irreducibility of P, the trajectory starting at the least state of S passes a.s. through the greatest state of S. This implies that coalescence occurs, so that $k(\mu) = 1$.

4 Block Measures

We introduce next the concept of a block measure, which is a strong form of the *lumpability* of [6] and [4, Exer. 6.1.13].

Definition 3 Let $P \in \mathscr{P}_S$ and $\mu \in \mathscr{L}(P)$. For a partition $\mathscr{S} = \{S_r : r = 1, 2, \dots, l\}$ of S with $l = l(\mathscr{S}) \geq 1$, we call μ an *$\mathscr{S}$-block measure* (or just a *block measure* with l blocks) if

(a) for $f \in \text{supp}(\mu)$, there exists a unique permutation $\pi = \pi_f$ of $I := \{1, 2, \dots, l\}$ such that, for $r \in I$, $fS_r \subseteq S_{\pi(r)}$, and
(b) $k(\mu) = l$.

The action of an $\mathscr{S}$-block measure μ is as follows. Since blocks are mapped a.s. to blocks, the measure μ of (5) induces a random permutation Π of the blocks which may be written as

$$\Pi = \sum_{f \in \text{supp}(\mu)} \alpha_f \delta_{\pi_f}. \tag{10}$$

The condition $k(\mu) = l$ implies that

$$\text{for } r \in I \text{ and } i, j \in S_r, \text{ the pair } i, j \text{ coalesce a.s.,} \tag{11}$$

so that the equivalence classes of $\stackrel{\vec{F}}{\sim}$ are a.s. the blocks $S_1, S_2, \dots, S_l$. If, as the chain evolves, we observe only the evolution of the blocks, we see a Markov chain on I with transition probabilities $\lambda_{r,s} = \mathbb{P}(\Pi(r) = s)$ which, since P is irreducible, is itself irreducible.

Example 2 illustrates the existence of measures μ that are not block measures, when $|S| = 4$. On the other hand, we have the following lemma when $|S| = 3$. For $P \in \mathscr{P}_S$ and $\mu \in \mathscr{L}(P)$, let $\mathscr{C} = \mathscr{C}(\mu)$ be the set of possible coalescing pairs,

$$\mathscr{C} = \big\{\{i, j\} \subseteq S : i \neq j,\ \boldsymbol{\mu}(i, j \text{ coalesce}) > 0\big\}. \tag{12}$$

Lemma 4 *Let $|S| = 3$ and $P \in \mathscr{P}_S$. If (P, μ) is consistent then μ is a block measure.*

Proof Let S, (P, μ) be as given. If $\mathscr{C}$ is empty then $k(\mu) = 3$ and μ is a block measure with 3 blocks.

If $|\mathscr{C}| \geq 2$, we have by the forthcoming Proposition 1(a, b) that $k(\mu) \leq 1$, so that μ is a block measure with 1 block.

Finally, if $\mathscr{C}$ contains exactly one element then we may assume, without loss of generality, that element is $\{1, 2\}$. By Proposition 1(b), we have $k(\mu) = 2$, whence a.s. some pair coalesces. By assumption only $\{1, 2\}$ can coalesce, so in fact a.s. we have that 1 and 2 coalesce, and they do not coalesce with 3. Therefore, μ is a block measure with the two blocks $\{1, 2\}$ and $\{3\}$.

We show next that, for $1 \leq k \leq |S|$, there exists a consistent pair (P, μ) such that μ is a block measure with $k(\mu) = k$.

Lemma 5 *For $|S| = n \geq 2$ and $1 \leq k \leq n$, there exists an aperiodic $P \in \mathscr{P}_S$ such that $k \in K(P)$.*

Proof Let $\mathscr{S} = \{S_r : r = 1, 2, \ldots, l\}$ be a partition of S, and let $\mathscr{G} \subseteq \mathscr{F}_S$ be the set of all functions g satisfying: there exists a permutation π of $\{1, 2, \ldots, l\}$ such that, for $r = 1, 2, \ldots, l$, we have $gS_r \subseteq S_{\pi(r)}$. Any probability measure μ on $\mathscr{F}_S$ with support $\mathscr{G}$ is an $\mathscr{S}$-block measure.

Let μ be such a measure and let P be the associated stochastic matrix on S, given in (1). For $i, j \in S$, there exists $g \in \mathscr{G}$ such that $g(i) = j$. Therefore, P is irreducible and aperiodic.

We identify next the consistent pairs (P, μ) for which either $k(\mu) = |S|$ or $|S| \in K(P)$.

Theorem 4 *Let $|S| = n \geq 2$ and $P \in \mathscr{P}_S$. We have that*

(a) *$k(\mu) = n$ if and only if* $\text{supp}(\mu)$ *contains only permutations of S,*
(b) *$n \in K(P)$ if and only if P is doubly stochastic.*

Before proving this, we remind the reader of Birkhoff's theorem [1] (sometimes attributed also to von Neumann [9]).

Theorem 5 ([1, 9]) *A stochastic matrix P on the finite state space S is doubly stochastic if and only if it lies in the convex hull of the set Π_S of permutation matrices.*

Remark 2 We note that the simulation problem confronted by CFTP is trivial when P is irreducible and doubly stochastic, since such P are characterized as those transition matrices with the uniform invariant distribution $\pi = (\pi_i = n^{-1} : i \in S)$.

Proof of Theorem 4

(a) If $\text{supp}(\mu)$ contains only permutations, then a.s. $k_t(F) = n$ for every $t \in \mathbb{N}$. Hence $n \in K(P)$. If $\text{supp}(\mu)$ contains a non-permutation, then with positive probability $k_1(F) < n$ and hence $k(\mu) < n$.

(b) By Theorem 5, P is doubly stochastic if and only if it may be expressed as a convex combination

$$P = \sum_{f \in \Pi_S} \alpha_f M_f, \tag{13}$$

of permutation matrices M_f (recall (6) and (7)).

If P is doubly stochastic, let the α_f satisfy (13), and let

$$\mu = \sum_{f \in \Pi_S} \alpha_f \delta_f, \tag{14}$$

as in (5). Then $\mu \in \mathscr{L}(P)$, and $k(\mu) = n$ by part (a).

If P is not doubly stochastic and $\mu \in \mathscr{L}(P)$, then μ has no representation of the form (14), so that $k(\mu) < n$ by part (a).

Finally in this section, we present a necessary and sufficient condition for μ to be an $\mathscr{S}$-block measure, Theorem 6 below.

Let $P \in \mathscr{P}_S$, and let $\mathscr{S} = \{S_r : r = 1, 2, \dots, l\}$ be a partition of S with $l \geq 1$. For $r, s \in I := \{1, 2, \dots, l\}$ and $i \in S_r$, let

$$\lambda_{r,s}^{(i)} = \sum_{j \in S_s} p_{i,j}.$$

Since a block measure comprises a transition operator on blocks, combined with a shuffling of states within blocks, it is necessary in order that μ be an $\mathscr{S}$-block measure that

$$\lambda_{r,s}^{(i)} \text{ is constant for } i \in S_r. \tag{15}$$

When (15) holds, we write

$$\lambda_{r,s} = \lambda_{r,s}^{(i)}, \qquad i \in S_r. \tag{16}$$

Under (15), the matrix $\Lambda = (\lambda_{r,s} : r, s \in I)$ is the irreducible transition matrix of the Markov chain derived from P by observing the evolution of blocks, which is to say that

$$\lambda_{r,s} = \mu(\Pi(r) = s), \qquad r, s \in I, \tag{17}$$

where Π is given by (10). Since $l \in K(\Lambda)$, we have by Theorem 4 that Λ is doubly stochastic, which is to say that

$$\sum_{r \in I} \lambda_{r,s} = \sum_{r \in I} \sum_{j \in S_s} p_{i_r,j} = 1, \qquad s \in I, \tag{18}$$

where each i_r is an arbitrarily chosen representative of the block S_r. By (15), Eq. (18) may be written in the form

$$\sum_{i\in S}\sum_{j\in S_s}\frac{1}{|S_{r(i)}|}p_{i,j}=1, \qquad r,s\in I, \tag{19}$$

where $r(i)$ is the index r such that $i\in S_r$. The following theorem is the final result of this section.

Theorem 6 *Let S be a non-empty, finite set, and let $\mathscr{S}=\{S_r : r=1,2,\dots,l\}$ be a partition of S. For $P\in\mathscr{P}_S$, a measure $\mu\in\mathscr{L}(P)$ is an $\mathscr{S}$-block measure if and only if* (15), (19) *hold, and also $k(\mu)=l$.*

Proof The necessity of the conditions holds by the definition of block measure and the above discussion.

Suppose conversely that the stated conditions hold. Let $\Lambda=(\lambda_{r,s})$ be given by (15)–(16). By (16) and (19), Λ is doubly stochastic. By Theorem 4, we may find a measure $\rho\in\mathscr{L}(\Lambda)$ supported on a subset of the set Π_I of permutations of I, and we let Π have law ρ. Conditional on Π, let $Z=(Z_i : i\in S)$ be independent random variables such that

$$\mathbb{P}(Z_i=j\mid \Pi)=\begin{cases} p_{i,j}/\lambda_{r,s} & \text{if } S_r\ni i,\ S_s\ni j,\ \Pi(r)=s,\\ 0 & \text{otherwise.}\end{cases}$$

The law μ of Z is an $\mathscr{S}$-block measure that is consistent with P.

5 The Set $K(P)$

We begin with a triplet of conditions.

Proposition 1 *Let $S=\{1,2,\dots,n\}$ where $n\ge 3$, and let $P\in\mathscr{P}_S$ and $\mu\in\mathscr{L}(P)$. Let $\mathscr{C}=\mathscr{C}(\mu)$ be the set of possible coalescing pairs, as in* (12).

(a) *$k(\mu)=n$ if and only if $|\mathscr{C}|=0$.*
(b) *$k(\mu)=n-1$ if and only if $|\mathscr{C}|=1$.*
(c) *If $|\mathscr{C}|$ comprises the single pair $\{1,2\}$, then P satisfies*

$$\sum_{j=3}^{n}p_{1,j}=\sum_{j=3}^{n}p_{2,j}=\sum_{i=3}^{n}(p_{i,1}+p_{i,2}). \tag{20}$$

Proof

(a) See Theorem 4(a).

(b) By part (a), $k(\mu) \le n-1$ when $|\mathscr{C}| = 1$. It is trivial by definition of k and $\mathscr{C}$ that, if $k(\mu) \le n-2$, then $|\mathscr{C}| \ge 2$. It suffices, therefore, to show that $k(\mu) \le n-2$ when $|\mathscr{C}| \ge 2$. Suppose that $|\mathscr{C}| \ge 2$. Without loss of generality we may assume that $\{1,2\} \in \mathscr{C}$ and either that $\{1,3\} \in \mathscr{C}$ or (in the case $n \ge 4$) that $\{3,4\} \in \mathscr{C}$. Let $F = (F_s : s \in \mathbb{N})$ be an independent sample from μ. Let M be the Markov time $M = \inf\{t > 0 : \vec{F}_t(1) = \vec{F}_t(2) = 1\}$, and write $J = \{M < \infty\}$. By irreducibility, $\boldsymbol{\mu}(J) > 0$, implying that $k(\mu) \le n-1$. Assume that

$$k(\mu) = n-1. \tag{21}$$

We shall obtain a contradiction, and the conclusion of the lemma will follow.

Suppose first that $\{1,2\}, \{1,3\} \in \mathscr{C}$. Let B be the event that there exists $i \ge 3$ such that $\vec{F}_M(i) \in \{1,2,3\}$. On $B \cap J$, we have $k(F) \le n-2$ a.s., since

$$\boldsymbol{\mu}(\text{at least 3 states belong to coalescing pairs}) > 0.$$

Thus $\mu(B \cap J) = 0$ by (21). On $\overline{B} \cap J$, the $\vec{F}_M(i)$, $i \ge 3$, are by (21) a.s. distinct, and in addition take values in $S \setminus \{1,2,3\}$. Thus there exist $n-2$ distinct values of $\vec{F}_M(i)$, $i \ge 3$, but at most $n-3$ values that they can take, which is impossible, whence $\mu(\overline{B} \cap J) = 0$. It follows that

$$0 < \boldsymbol{\mu}(J) = \boldsymbol{\mu}(B \cap J) + \boldsymbol{\mu}(\overline{B} \cap J) = 0, \tag{22}$$

a contradiction.

Suppose secondly that $\{1,2\}, \{3,4\} \in \mathscr{C}$. Let C be the event that either (i) there exists $i \ge 3$ such that $\vec{F}_M(i) \in \{1,2\}$, or (ii) $\{\vec{F}_M(i) : i \ge 3\} \supseteq \{3,4\}$. On $C \cap J$, we have $k(F) \le n-2$ a.s. On $\overline{C} \cap J$, by (21) the $\vec{F}_M(i)$, $i \ge 3$, are a.s. distinct, and in addition take values in $S \setminus \{1,2\}$ and no pair of them equals $\{3,4\}$. This provides a contradiction as in (22).

(c) Let F_1 have law μ. Write $A_i = \{F_1(i) \in \{1,2\}\}$, and

$$M = |\{i \le 2 : A_i \text{ occurs}\}|, \qquad N = |\{i \ge 3 : A_i \text{ occurs}\}|.$$

If $\mu(A_i \cap A_j) > 0$ for some $i \ge 3$ and $j \ne i$, then $\{i,j\} \in \mathscr{C}$, in contradiction of the assumption that $\mathscr{C}$ comprises the singleton $\{1,2\}$. Therefore, $\mu(A_i \cap A_j) = 0$ for all $i \ge 3$ and $j \ne i$, and hence

$$\mu(N \ge 2) = 0, \tag{23}$$

$$\mu(M \ge 1, N = 1) = 0. \tag{24}$$

By similar arguments,

$$\mu(M < 2, N = 0) = 0, \tag{25}$$

$$\mu(M = 1) = 0. \tag{26}$$

It follows that

$$\begin{aligned}\mu(N=1) &= \mu(N=1, M=0) && \text{by (24)}\\ &= \mu(M=0) && \text{by (25) and (23)}\\ &= \mu(\overline{A}_1 \cap \overline{A}_2)\\ &= \mu(\overline{A}_r), \quad r=1,2, && \text{by (26).}\end{aligned}$$

Therefore,

$$\mu(N=1) = \mu(\overline{A}_r) = \mu(F_1(r) \geq 3) = \sum_{j=3}^{n} p_{r,j}, \qquad r = 1,2.$$

By (23),

$$\mu(N=1) = \mu(N) = \sum_{i=3}^{n} \mu(A_i) = \sum_{i=3}^{n} (p_{i,1} + p_{i,2}),$$

where $\mu(N)$ is the mean value of N. This yields (20).

The set $K(P)$ can be fairly sporadic, as illustrated in the next two examples.

Example 3 Consider the matrix

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}. \tag{27}$$

Since P is doubly stochastic, by Theorem 4(a), there exists $\mu \in \mathscr{L}(P)$ such that $k(\mu) = 3$ (one may take $\mu(123) = \mu(231) = \frac{1}{2}$). By Lemma 3, we have that $1 \in K(P)$, and thus $\{1,3\} \subseteq K(P)$. We claim that $2 \notin K(P)$, and we show this as follows.

Let $\mu \in \mathscr{L}(P)$, with $k(\mu) < 3$, so that $|\mathscr{C}| \geq 1$. There exists no permutation of S for which the matrix P satisfies (20), whence $|\mathscr{C}| \geq 2$ by Proposition 1(c). By parts (a, b) of that proposition, $k(\mu) \leq 1$. In conclusion, $K(P) = \{1,3\}$.

Example 4 Consider the matrix

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{pmatrix}. \tag{28}$$

We have, as in Example 3, that $\{1, 4\} \subseteq K(P)$. Taking

$$\mu(1234) = \mu(2244) = \mu(1331) = \mu(2341) = \tfrac{1}{4}$$

reveals that $2 \in K(P)$, and indeed μ is a block measure with blocks $\{1, 2\}$, $\{3, 4\}$. As in Example 3, we have that $3 \notin K(P)$, so that $K(P) = \{1, 2, 4\}$.

We investigate in greater depth the transition matrix on S with equal entries. Let $|S| = n \geq 2$ and let $P_n = (p_{i,j})$ satisfy $p_{i,j} = n^{-1}$ for $i, j \in S = \{1, 2, \ldots, n\}$.

Theorem 7 *For $n \geq 2$ there exists a block measure $\mu \in \mathscr{L}(P_n)$ with $k(\mu) = l$ if and only if $l \mid n$. In particular, $K(P_n) \supseteq \{l : l \mid n\}$. For $n \geq 3$, we have $n - 1 \notin K(P_n)$.*

We do not know whether $K(P_n) = \{l : l \mid n\}$, and neither do we know if there exists $\mu \in \mathscr{L}(P_n)$ that is not a block measure.

Proof Let $n \geq 2$. By Lemma 3 and Theorem 4, we have that $1, n \in K(P_n)$. It is easily seen as follows that $l \in K(P_n)$ whenever $l \mid n$. Suppose $l \mid n$ and $l \neq 1, n$. Let

$$\begin{aligned} S_r &= (r-1)n/l + \{1, 2, \ldots, n/l\} \\ &= \{(r-1)n/l + 1, (r-1)n/l + 2, \ldots, rn/l\}, \qquad r = 1, 2, \ldots, l. \end{aligned} \tag{29}$$

We describe next a measure $\mu \in \mathscr{L}(P_n)$. Let Π be a uniformly chosen permutation of $\{1, 2, \ldots, l\}$. For $i \in S$, let Z_i be chosen uniformly at random from $S_{\Pi(i)}$, where the Z_i are conditionally independent given Π. Let μ be the block measure governing the vector $Z = (Z_i : i \in S)$. By symmetry,

$$q_{i,j} := \mu\big(\{f \in \mathscr{F}_S : f(i) = j\}\big), \qquad i, j \in S,$$

is constant for all pairs $i, j \in S$. Since μ is a probability measure, $Q = (q_{i,j})$ has row sums 1, whence $q_{i,j} = n^{-1} = p_{i,j}$, and therefore $\mu \in \mathscr{L}(P_n)$. By examination of μ, μ is an $\mathscr{S}$-block measure.

Conversely, suppose there exists an $\mathscr{S}$-block measure $\mu \in \mathscr{L}(P_n)$ with corresponding partition $\mathscr{S} = \{S_1, S_2, \ldots, S_l\}$ with index set $I = \{1, 2, \ldots, l\}$. By Theorem 6, Eqs. (15) and (19) hold. By (15), the matrix $\Lambda = (\lambda_{r,s} : r, s \in I)$ satisfies

$$\lambda_{r,s} = \frac{|S_s|}{n}, \qquad r, s \in I. \tag{30}$$

By (19),

$$\frac{|S_s|}{|S_r|} = 1, \qquad s, r \in I,$$

whence $|S_s| = n/l$ for all $s \in I$, and in particular $l \mid n$.

Let $n \geq 3$. We prove next that $k(\mu) \neq n - 1$ for $\mu \in \mathscr{L}(P_n)$. Let $\mathscr{C} = \mathscr{C}(\mu)$ be given as in (12). By Proposition 1(b), it suffices to prove that $|\mathscr{C}| \neq 1$. Assume on the contrary that $|\mathscr{C}| = 1$, and suppose without loss of generality that $\mathscr{C}$ contains the singleton pair $\{1, 2\}$. With $P = P_n$, the necessary condition (20) becomes

$$(n-2)\frac{1}{n} = (n-2)\frac{2}{n},$$

which is false when $n \geq 3$. Therefore, $|\mathscr{C}| \neq 1$, and the proof is complete.

Acknowledgments The authors thank Tim Garoni, Wilfrid Kendall, and an anonymous referee for their comments. MH was supported by Future Fellowship FT160100166 from the Australian Research Council.

References

1. Birkhoff, G.: Three observations on linear algebra. Univ. Nac. Tucumán. Revista A. **5**, 147–151 (1946)
2. Doeblin, W.: Exposé de la théorie des chaînes simples constants de Markoff à un nombre fini d'états. Revue Math. de l'Union Interbalkanique **2**, 77–105 (1938)
3. Grimmett, G.R.: Probability on Graphs, 2nd edn. Cambridge University, Cambridge (2018)
4. Grimmett, G.R., Stirzaker, D.R.: Probability and Random Processes, 4th edn. Oxford University, Oxford (2020)
5. Huber, M.L.: Perfect Simulation. Chapman and Hall/CRC, Boca Raton (2015)
6. Kemeny, J.G., Snell, J.L.: Finite Markov Chains. Van Nostrand, New York (1963)
7. Propp, J.G., Wilson, D.B.: Exact sampling with coupled Markov chains and applications to statistical mechanics. Random Struct. Algoritm. **9**, 223–252 (1996). Proceedings of the Seventh International Conference on Random Structures and Algorithms (Atlanta, GA, 1995)
8. Propp, J.G., Wilson, D.B.: How to get a perfectly random sample from a generic Markov chain and generate a random spanning tree of a directed graph. J. Algorithms **27**, 170–217 (1998). 7th Annual ACM-SIAM Symposium on Discrete Algorithms (Atlanta, GA, 1996)
9. von Neumann, J.: A certain zero-sum two-person game equivalent to the optimal assignment problem. In: Contributions to the Theory of Games, vol. 2. Annals of Mathematics Studies, no. 28. Princeton University, Princeton, pp. 5–12 (1953)
10. Wilson, D.B.: Perfectly Random Sampling with Markov Chains. http://www.dbwilson.com/exact/
11. Wilson, D.B., Propp, J.G.: How to get an exact sample from a generic Markov chain and sample a random spanning tree from a directed graph, both within the cover time. In: Proceedings of the Seventh Annual ACM-SIAM Symposium on Discrete Algorithms (Atlanta, GA, 1996). ACM, New York, pp. 448–457 (1996)

Combinatorial Universality in Three-Speed Ballistic Annihilation

John Haslegrave and Laurent Tournier

Abstract We consider a one-dimensional system of particles, moving at constant velocities chosen independently according to a symmetric distribution on $\{-1, 0, +1\}$, and annihilating upon collision—with, in case of triple collision, a uniformly random choice of survivor among the two moving particles. When the system contains infinitely many particles, whose starting locations are given by a renewal process, a phase transition was proved to happen (see Haslegrave et al., Three-speed ballistic annihilation: phase transition and universality, 2018) as the density of static particles crosses the value $1/4$. Remarkably, this critical value, along with certain other statistics, was observed not to depend on the distribution of interdistances. In the present paper, we investigate further this universality by proving a stronger statement about a finite system of particles with fixed, but randomly shuffled, interdistances. We give two proofs, one by an induction allowing explicit computations, and one by a more direct comparison. This result entails a new nontrivial independence property that in particular gives access to the density of surviving static particles at time t in the infinite model. Finally, in the asymmetric case, further similar independence properties are proved to keep holding, including a striking property of gamma distributed interdistances that contrasts with the general behavior.

Keywords Ballistic annihilation · Interacting particle system · Random permutation · Gamma distribution

AMS MSC 2010 60K35

J. Haslegrave
Mathematics Institute, University of Warwick, Coventry, UK
e-mail: j.haslegrave@warwick.ac.uk

L. Tournier (✉)
Université Sorbonne Paris Nord, LAGA, CNRS, UMR 7539, Villetaneuse, France
e-mail: tournier@math.univ-paris13.fr

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_23

1 Introduction

Annihilating particle systems have been studied extensively in statistical physics since the 1980s. The original motivation for this topic, stemming from the kinetics of chemical reactions, gave rise to models in which particles move diffusively and are removed from the system upon meeting another particle (e.g. [1]), or in some settings another particle of a specified type (e.g. [4]). The study of annihilating systems involving particles which move at constant velocity (that is, ballistic motion) was initiated by Elskens and Frisch [8] and Ben-Naim et al. [3]. In this ballistic annihilation process particles start at random positions on the real line and move at randomly-assigned constant velocities, annihilating on collision. This process displays quite different behavior to diffusive systems and its analysis presents particular challenges.

In order to specify a precise model, we must choose how particles are initially positioned on the real line and how velocities are initially assigned to particles. The most natural choice for the initial positions is arguably the points of a homogeneous Poisson point process, which is the only choice considered in the physics literature. It is also natural to sample i.i.d. velocities from some distribution. In the case of a discrete distribution supported on two values, it is easy to see that almost surely every particle is eventually destroyed when the two velocities have equal probability, but that almost surely infinitely many particles of the more probable velocity survive otherwise. However, this model still displays interesting global behavior; see e.g. [2].

The first case for which the question of survival of individual particles is not trivial is therefore a three-valued discrete distribution. Krapivsky et al. [13] considered the general symmetric distribution on $\{-1, 0, +1\}$, i.e. $\frac{1-p}{2}\delta_{-1} + p\delta_0 + \frac{1-p}{2}\delta_{+1}$ for some $p \in (0, 1)$. They predicted the existence of a critical value p_c such that for $p \leq p_c$ almost surely every particle is eventually destroyed and for $p > p_c$ almost surely infinitely many particles survive, and further that $p_c = 1/4$. Even the existence of such a critical value is far from obvious, given that there is no coupling to imply monotonicity of the annihilation of particles with p.

These predictions were strongly supported by intricate calculations of Droz et al. [6]. More recently, this model attracted significant interest in the mathematics community. Rigorous bounds were established by Sidoravicius and Tournier [14], and independently by Dygert et al. [7], giving survival regimes for $p \geq 0.33$, but a subcritical regime was more elusive. In previous work with the late Vladas Sidoravicius [10], we established the precise phase transition predicted by Krapivsky, Redner and Leyvraz.

A closely-related problem known as the bullet problem was popularised by Kleber and Wilson [12]. In this problem, a series of bullets with random speeds are fired at intervals from a gun, annihilating on collision. Kleber and Wilson [12] asked for the probability that when n bullets are fired, all are destroyed. Broutin and Marckert [5] solved this problem in generality, by showing that the answer does not depend on the choice of speeds or intervals, provided that these are symmetrical.

For any fixed sequence of n speeds and $n-1$ time intervals, consider firing n bullets with a random permutation of the speeds, separated by a random permutation of the intervals. Provided that the speeds and intervals are such that triple collisions cannot occur, they show that the probability that all particles are destroyed, and even the law of the number of particles which are destroyed, does not depend on the precise choice of speeds and intervals. However, they note that this universality property does not extend further, in that the indices of surviving particles does depend on the choice of speeds and intervals.

In proving the phase transition for symmetric three-speed ballistic annihilation, we observed a form of universality applies. Consider a one-sided version where particles are placed on the positive real line. For each n, the probability that the nth particle is the first to reach 0 is universal provided that distances between initial positions of particles are i.i.d. This universality extends to discrete distributions, provided that triple collisions are resolved randomly. These probabilities satisfy a recurrence relation which may be leveraged to prove asymptotics for the decay of particles; see [10] for further details. Note that this universality property encompasses more information than that of Broutin and Marckert, since it relates to the indices of surviving particles, which are not universal in their case; indeed, some further information on the fate of the remaining particles (the "skyline process" of [10]) is universal, although this universality does not extend to the full law of pairings. However, it relies on successive intervals between particles being independent, which is not necessary in [5].

In this article we extend this stronger form of universality for the symmetric three-speed case to the combinatorial setting of Broutin and Marckert. These results are specific to the symmetric three-speed case (that is, where the three speeds are in arithmetic progression and the two extreme speeds have equal probability), and this symmetry is necessary for the quantities we consider to be universal. However, we do prove (see Sect. 5) some unexpected properties of particular interdistance distributions which extend to the asymmetric three-speed case. This case was considered by Junge and Lyu [11], who extended the methods of [10] to give upper and lower bounds on the phase transition.

The so-called "bullet problem", before it was put in relation with the topic of "ballistic annihilation" in the physics literature, gained considerable visibility in the community of probabilists thanks to Vladas' warm descriptions and enthusiasm. It soon became one of Vladas' favorite open problems that he enjoyed sharing around him, and that he relentlessly kept investigating. The second author had innumerable lively discussions with him on this problem over the years, first already in Rio, shortly before his departure, and then in Shanghai. It is mainly thanks to Vladas' never-failing optimism when facing difficult problems that, after years of vain attempts and slow progress, efforts could be joined to lead to a solution in the discrete setting with three speeds and to Paper [10].

2 Definitions and Statements

We define ballistic annihilation with either fixed (shuffled) or random (i.i.d.) lengths.

2.1 Fixed Lengths

Let us first define the model in the combinatorial setting that is specific to this paper. For integers $a \leq b$ we write $[\![a, b]\!]$ for the set of integers in $[a, b]$.

Let n be a positive integer, μ be a symmetric distribution on $\{-1, 0, +1\}$, i.e. for some $p \in [0, 1]$ we have $\mu = \frac{1-p}{2}\delta_{-1} + p\delta_0 + \frac{1-p}{2}\delta_{+1}$, and $\ell = (\ell_1, \dots, \ell_n)$ be an ordered n-uple of positive real numbers:

$$\ell \in \mathbb{L}_n = \left\{(\ell_1, \dots, \ell_n) \in (0, +\infty)^n \,:\, \ell_1 \leq \cdots \leq \ell_n\right\}.$$

On a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, consider independent random variables (v, s, σ) such that

- $v = (v_1, \dots, v_n)$ where $v_1, \dots, v_n$ are independent, with distribution μ;
- $s = (s_1, \dots, s_n)$ where $s_1, \dots, s_n$ are independent, with uniform distribution on $\{-1, +1\}$;
- σ is uniformly distributed on the symmetric group $\mathfrak{S}_n$.

Finally, define the positive random variables $x_1, \dots, x_n$ by

$$x_1 = \ell_{\sigma(1)}, \qquad x_2 = \ell_{\sigma(1)} + \ell_{\sigma(2)}, \qquad \dots, \qquad x_n = \ell_{\sigma(1)} + \cdots + \ell_{\sigma(n)}.$$

We interpret n as a number of particles, $x_1, x_2, \dots, x_n$ as their initial locations, $v_1, v_2, \dots, v_n$ as their initial velocities, and $s_1, s_2, \dots, s_n$ as their “spins”. For any $i \in [\![1, n]\!]$, the spin s_i will only play part in the process if $v_i = 0$, in which case it will be used to resolve a potential *triple collision* at x_i.

In notations, the particles will conveniently be referred to as $\bullet_1, \bullet_2, \dots, \bullet_n$, and particles with velocity 0 will sometimes be called *static* particles, as opposed to *moving* particles.

The evolution of the process of particles describes as follows (see also Fig. 1): at time 0, particles $\bullet_1, \dots, \bullet_n$ respectively start at $x_1, \dots, x_n$, then move at constant velocity $v_1, \dots, v_n$ until, if ever, they collide with another particle. Collisions resolve as follows: where exactly two particles collide, both are annihilated; where three particles, necessarily of different velocities, collide, two are annihilated, and either the right-moving or left-moving particle survives (i.e. continues its motion unperturbed), according to the spin of the static particle involved. Note that each spin affects the resolution of at most one triple collision. Annihilated particles are considered removed from the system and do not take part in any later collision. Thus,

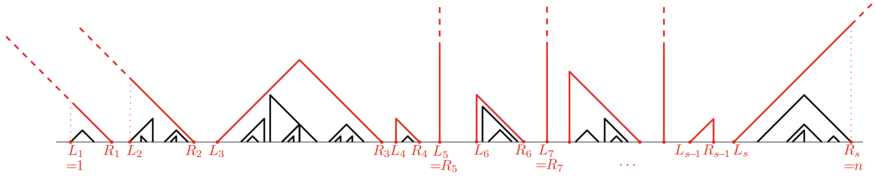


Fig. 1 Graphical space-time interpretation of the skyline of the process (see Definition 1), here highlighted in red

after a finite time, every particle is either annihilated or shall pursue its ballistic trajectory forever.

For a more formal definition, we refer the interested reader to [10].

2.2 *Random Lengths*

We may alternatively, in accordance to the classical setting of ballistic annihilation, consider the distances $\ell_1, \ell_2, \dots$ to be random, independent and identically distributed (i.i.d.), which naturally enables to extend the definition into an infinite number of particles.

Let $\mu = \frac{1-p}{2}\delta_{-1} + p\delta_0 + \frac{1-p}{2}\delta_{+1}$ be a symmetric distribution on $\{-1, 0, +1\}$ and let m be a probability measure on $(0, \infty)$. On $(\Omega, \mathscr{F}, \mathbb{P})$ we consider random variables (ℓ, v, s) where

- $\ell = (\ell_k)_{k\geq 1}$ where $\ell_1, \ell_2, \dots$ are independent, with distribution m;
- $v = (v_k)_{k\geq 1}$ where $v_1, v_2, \dots$ are independent, with distribution μ;
- $s = (s_k)_{k\geq 1}$ where $s_1, s_2, \dots$ are independent, uniformly distributed on $\{-1, +1\}$.

In contrast to the previous setting, we define, for all $n \geq 1$,

$$x_n = \ell_1 + \cdots + \ell_n.$$

In other words, $(x_n)_{n\geq 1}$ is a renewal process on $(0, \infty)$ whose interdistances are m-distributed.

The process is then defined in the same way as in the finite case, now with infinitely many particles respectively starting from $x_1, x_2, \dots$ Note that triple collisions may only happen when the distribution m has atoms, hence in the opposite case the sequence s of spins is irrelevant. Let us nonetheless already mention that some arguments (cf. Sect. 4, most notably) involve the atomic case toward an understanding of the continuous case.

The model in this infinite setting goes through a phase transition as p varies from 0 to 1, specifically at $p = 1/4$, from annihilation of all static particles to survival of a positive density of them. This was the main result of [10]; it is not used in the present paper, although some remarks refer to it.

While most of the paper is concerned with the fixed lengths setting, some consequences about random lengths will be mentioned (Corollary 2), along with a specific property of gamma distributions, Theorem 2. Unless otherwise specified, the model under consideration therefore has fixed lengths, hence finite size.

2.3 Notation

We introduce convenient abbreviations, borrowed from [10], to describe events related to the model.

We use $\bullet_i$ (where $1 \le i \le n$) for the ith particle, $\bullet$ (with no subscript) for an arbitrary particle, and superscripts $\vec{\bullet}$, $\dot{\bullet}$ and $\overleftarrow{\bullet}$ to indicate that those particles have velocity $+1$, 0 and -1 respectively.

We write $\bullet_i \sim_\ell \bullet_j$ to indicate mutual annihilation between $\bullet_i$ and $\bullet_j$, which depends on the fixed lengths $\ell = (\ell_1, \dots, \ell_n)$. Let us indeed emphasize that ℓ plays no role in $\mathbb{P}$, but in the definition of collisions. Still, instead of writing ℓ in subscripts, we will often, for readibility, write $\mathbb{P}_\ell$ to emphasize the choice of ℓ, and drop the notation ℓ from the events.

Every realization of (v, s, σ) induces an involution π on $[\![1, n]\!]$ by $\pi(i) = j$ when $\bullet_i \sim \bullet_j$, and $\pi(i) = i$ if $\bullet_i$ survives. We shall refer to π as the *pairing induced by the annihilations*.

We will usually replace notation $\bullet_i \sim \bullet_j$ by a more precise series of notations: if $\bullet_i \sim \bullet_j$ with $i < j$, we write $\bullet_i \to\leftarrow \bullet_j$, or redundantly $\vec{\bullet}_i \to\leftarrow \overleftarrow{\bullet}_j$, when $v_i = +1$ and $v_j = -1$, $\bullet_i \to \bullet_j$ when $v_i = +1$ and $v_j = 0$, and $\bullet_i \leftarrow \bullet_j$ symmetrically. Note that in all cases this notation refers to annihilation, not merely collision, i.e. it excludes the case where $\bullet_i$ and $\bullet_j$ take part in a triple collision but one of them survives.

Additionally, we write $x \leftarrow \bullet_i$ (for $i \in \mathbb{N}$ and $x \in \mathbb{R}$) to indicate that $\bullet_i$ crosses location x from the right (i.e. $v_i = -1$, $x < x_i$, and $\bullet_i$ is not annihilated when or before it reaches x), and $x \overset{1}{\leftarrow} \bullet_i$ if $\bullet_i$ is *first* to cross location x from the right. Symmetrically, we write $\bullet_i \to x$ and $\bullet_i \overset{1}{\to} x$.

For any interval $I \subset (0, \infty)$, and any condition C on particles, we denote by $(C)_I$ the same condition for the process restricted to the set I, i.e. where all particles outside I are removed at time 0 (however, the indices of remaining particles are unaffected by the restriction). For short, we write $\{C\}_I$ instead of $\{(C)_I\}$, denoting the event that the condition $(C)_I$ is realized.

2.4 Skyline

Finally, we introduce a decomposition of the configuration that plays a key role.

Definition 1 The *skyline* of the configuration (v, s, σ) on $[\![1, n]\!]$ is the family

$$\mathrm{skyline}_\ell(v, s, \sigma) = \big((L_1, R_1, \Sigma_1), \dots, (L_S, R_S, \Sigma_S)\big)$$

characterized by the following properties: $[\![L_1, R_1]\!]$, $[\![L_2, R_2]\!]$,..., $[\![L_S, R_S]\!]$ is a partition of $[\![1, n]\!]$, for $k = 1, \dots, S$, the "shape" Σ_k is one of the 6 elements of the set $\{\uparrow, \nwarrow, \nearrow, \nearrow\uparrow, \uparrow\nwarrow, \nearrow\nwarrow\}$, and satisfies:

- $\Sigma_k = \nwarrow$ if $v_{R_k} = -1$ and all particles indexed in $[\![L_k, R_k - 1]\!]$ are annihilated but $\bullet_{R_k}$ survives;
- $\Sigma_k = \uparrow$ if $v_{L_k} = 0$, $R_k = L_k$, and $\bullet_{R_k}$ survives;
- $\Sigma_k = \nearrow\uparrow$ if $\vec{\bullet}_{L_k} \to \dot{\bullet}_{R_k}$, and no particle hits x_{L_k} from the left, or x_{R_k} from the right (hence $\bullet_{L_k}$ is the last to visit $[x_{L_k}, x_{R_k})$);
- $\Sigma_k = \nearrow\nwarrow$ if $\vec{\bullet}_{L_k} \to\leftarrow \overleftarrow{\bullet}_{R_k}$, and no particle hits x_{L_k} from the left, or x_{R_k} from the right;

and symmetrically for $\uparrow\nwarrow$ and $\nearrow$.

Since this definition amounts to splitting $[\![1, n]\!]$ at indices of left- and right-going survivors and at endpoints of intervals never crossed by a particle (see Fig. 1), the skyline is well-defined.

This definition agrees with the notion of skyline introduced in [10] to study the process on the full line (i.e. with particles indexed by $\mathbb{Z}$), in the supercritical regime. The extra shapes $\nwarrow$ and $\nearrow$ are however specific to the finite setting.

2.5 Main Results and Organization of the Paper

Our main theorem is the following "universality" result.

Theorem 1 *For each n, the distribution of the skyline on $[\![1, n]\!]$ does not depend on ℓ.*

A similar statement was given in [10] in the infinite setting, with independent random lengths, where it only made sense in the supercritical regime. The above result highlights the combinatorial nature of this remarkable universality.

In order to underline the analogy with [5], where universality of the number of surviving (i.e. non-annihilating) particles is proved in the context of generic velocities, let us phrase out an immediate yet sensibly weaker corollary.

Corollary 1 *For each n, the joint law of the number of surviving particles of respective velocity -1, 0 and $+1$, does not depend on ℓ.*

We give two proofs of the main result, of different nature and interest.

- First, in Sect. 3, we show by induction on the number of particles, that the probabilities $\mathbb{P}_\ell(0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_n)$ and $\mathbb{P}_\ell(\overrightarrow{\bullet}_1 \rightarrow\leftarrow \overleftarrow{\bullet}_n)$ do not depend on ℓ, and that this implies the theorem. As a side result, this provides recursive formulae for these probabilities. These formulae were obtained in [10] for random lengths, where they played a key role.
- Then, in Sect. 4, a direct proof is proposed, close in spirit to the proof of [5], in that we study local invariance properties of the law of the skyline on the space $\mathbb{L}_n$ of lengths. Compared to [5], the proof greatly simplifies thanks to the natural definition of the model at singular ℓ, i.e. in a way of dealing with triple collisions that ensures continuity in law.

This result has consequences for the classical case of random lengths. Let us already notice that it readily implies that the distribution of the skyline on $[\![1, n]\!]$ does not depend on the distribution m of interdistances. The same argument actually not only holds for i.i.d. but also for *exchangeable* sequences $(\ell_1, \dots, \ell_n)$. In the infinite setting, the extension of results such as the phase transition at $p_c = 1/4$ (main theorem of [10]) to exchangeable sequences could alternatively already be seen as a consequence of the universality in the i.i.d. case, and of de Finetti's theorem.

The universality of the skyline also implies a previously unnoticed property of independence in the classical case of random lengths, which gives access to the Laplace transform of an interesting quantity:

Corollary 2 *Consider ballistic annihilation with random lengths (on $\mathbb{R}_+$), and define the random variable*

$$A = \min\{n \geq 1 \,:\, 0 \leftarrow \overleftarrow{\bullet}_n\}.$$

(a) The random variables (A, x_A) and $(A, \widetilde{x}_A)$ have same distribution given $\{A < \infty\}$, where $\widetilde{x}$ is a copy of x that is independent of A;
(b) Denote by $\mathscr{L}_\ell$ the Laplace transform of $x_1 = \ell_1$, i.e. $\mathscr{L}_\ell(\lambda) = \mathbb{E}[e^{-\lambda x_1}]$ for all $\lambda > 0$. Then the Laplace transform $\mathscr{L}_D$ of $D := x_A$ satisfies, on $\mathbb{R}_+$,

$$p\mathscr{L}_\ell\mathscr{L}_D^4 - (1+2p)\mathscr{L}_\ell\mathscr{L}_D^2 + 2\mathscr{L}_D - (1-p)\mathscr{L}_\ell = 0. \quad (1)$$

This corollary is proved after Theorem 1 in the upcoming Sect. 3. Note in particular that $\mathscr{L}_\ell = \phi \circ \mathscr{L}_D$ where $\phi(w) = \frac{-2w}{pw^4-(1+2p)w^2-(1-p)}$, hence the distribution of D characterizes the distribution m of ℓ_1, in complete contrast with the absence of dependence of A with respect to m (cf. Proposition 4 of [10], which now follows from the above Theorem 1, cf. also Proposition 1 below).

While we present (a) as a corollary of Theorem 1, from which (b) follows at once, we should also mention that a more direct proof of (b) is possible (see the end of Sect. 3), from which (a) could alternatively be deduced. This approach based on identification of Laplace transforms however gives no insight about the a priori surprising Property (a).

The proof of Property (a) will give the stronger statement of independence between the whole skyline on $[\![1, n]\!]$ and x_n; this fact turns out to also hold in the asymmetric case and is therefore stated in Sect. 5 as Theorem 2, (a).

The interest in the random variable D comes from the fact (mostly a consequence of ergodicity) that, for the process on the full line, the set of indices of static particles remaining at time t has a density (in $\mathbb{Z}$) equal to

$$c_0(t) = p\mathbb{P}(D > t)^2.$$

The value, and in particular the asymptotics for this quantity, which are important to understand the long-term behavior, can in principle be inferred from the Laplace transform of D. Such an analysis was conducted in [10] without access to $\mathscr{L}_D$, under assumption of finite exponential moments for ℓ_1, in order to enable approximating the tail of D using that of A, which in turn could be addressed by combinatorial analytic methods on its generating series. Although the above computation is more explicit than that of [10], we refrain from stating more general asymptotics of $c_0(t)$ as $t \to \infty$, as these are not universal and would depend on a technical choice of further assumptions on the tail of ℓ_1. Let us merely remark that, in the particular case when ℓ_1 is exponentially distributed, i.e. $\mathscr{L}_\ell(\lambda) = (1 + \lambda)^{-1}$, this confirms Equation (31) from [6]:

$$p\mathscr{L}_D(\lambda)^4 - (2p + 1)\mathscr{L}_D(\lambda)^2 + 2(\lambda + 1)\mathscr{L}_D(\lambda) + p - 1 = 0,$$

from which Laplace inversion in the asymptotic regime $\lambda \to 0$ could be conducted, leading to asymptotics of $c_0(t)$, $t \to \infty$.

We dedicate Sect. 5 to a discussion of the contrasting lack of universality as soon as the distribution of velocities is not symmetric any more, thus raising a priori difficulties for explicit computations. Still, we give positive results (Theorem 2), and in particular a remarkable property of gamma distributed interdistances that is insensitive to the asymmetry.

Finally, in Sect. 6, returning to the symmetric case and the universal distribution of the skyline, and in particular of A (cf. Corollary 2 above), which is in a sense explicit, we investigate its monotonicity properties with respect to the parameter p. While intuitively expected, and indeed observed numerically, these turn out to be complicated to establish in spite of the formulae at hand. We state a few conjectures and prove partial results.

3 Combinatorial Universality: Proof of the Main Results

The main result will follow from the particular case in the proposition below, which is a stronger version of Theorem 2 from [10].

Proposition 1 *For all n, and all $\ell \in \mathbb{L}_n$, define probabilities*

$$p_n(\ell) = \mathbb{P}_\ell(0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_n) \qquad \text{and} \qquad \delta_n(\ell) = \mathbb{P}_\ell(\overrightarrow{\bullet}_1 \rightarrow\leftarrow \overleftarrow{\bullet}_n).$$

For all n, p_n and δ_n do not depend on ℓ, and are given by the following recursive equations, for $n \geq 2$,

$$p_n = \Big(p + \frac{1}{2}\Big) \sum_{\substack{k_1+k_2\\=n-1}} p_{k_1} p_{k_2} - \frac{p}{2} \sum_{\substack{k_1+k_2+k_3+k_4\\=n-1}} p_{k_1} p_{k_2} p_{k_3} p_{k_4} \tag{2}$$

$$\delta_n = \frac{1-p}{2} p_{n-1} - \frac{p}{2} \sum_{\substack{k_1+k_2+k_3\\=n-1}} p_{k_1} p_{k_2} p_{k_3}, \tag{3}$$

with base cases $p_1 = (1-p)/2$ and $\delta_1 = 0$.

Note that $p_n = 0$ for even n, and $\delta_n = 0$ for odd n.

Let us stress again that, as is the case in [5], this "universality" with respect to ℓ does not follow from a direct coupling: indeed, the distribution of the full pairing induced by the annihilations is *not* universal. A coupling between given vectors ℓ and ℓ' close enough is however possible, see Sect. 4, but wouldn't extend to a coupling in the *infinite* random length setting, as any fluctuation in the lengths eventually breaks the coupling. Finally, as discussed in Sect. 5, the universality does not extend to the asymmetric case.

Proof The proof follows the main lines of Theorem 2 from [10], although using a stronger invariance. We proceed by induction on n, and show more generally that, for all n, none of the following probabilities depends on ℓ :

$$p_n(\ell) = \mathbb{P}_\ell(0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_n), \quad \alpha_n(\ell) = \mathbb{P}_\ell(\{0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_n\} \cap \{\dot{\bullet}_1\}),$$

$$\beta_n(\ell) = \mathbb{P}_\ell(\{0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_n\} \cap \{\overrightarrow{\bullet}_1 \rightarrow \dot{\bullet}\}), \quad \gamma_n(\ell) = \mathbb{P}_\ell(\{0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_n\} \cap \{\overrightarrow{\bullet}_1 \rightarrow\leftarrow \overleftarrow{\bullet}\}),$$

$$\text{and} \quad \delta_n(\ell) = \mathbb{P}_\ell(\overrightarrow{\bullet}_1 \rightarrow\leftarrow \overleftarrow{\bullet}_n).$$

Since $p_n = \alpha_n + \beta_n + \gamma_n$, the proposition will follow.

We have $p_1 = (1-p)/2$, and $\alpha_1 = \beta_1 = \gamma_1 = \delta_1 = 0$. Let $n \in \mathbb{N}$ be such that $n \geq 2$, and assume that the previous property holds up to the value $n-1$.

First, let us consider α_n and prove more precisely that, for each integer k in $\{2, \ldots, n-1\}$, $\mathbb{P}_\ell(0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_n,\ \dot{\bullet}_1 \leftarrow \overleftarrow{\bullet}_k)$ is universal (i.e., does not depend on ℓ). Let k be such an integer. Note that we have equivalently

$$\{0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_n,\ \dot{\bullet}_1 \leftarrow \overleftarrow{\bullet}_k\} = \{\dot{\bullet}_1\} \cap \{x_1 \overset{1}{\leftarrow} \overleftarrow{\bullet}_k\}_{(x_1, x_k]} \cap \{x_k \overset{1}{\leftarrow} \overleftarrow{\bullet}_n\}_{(x_k, x_n]}. \tag{4}$$

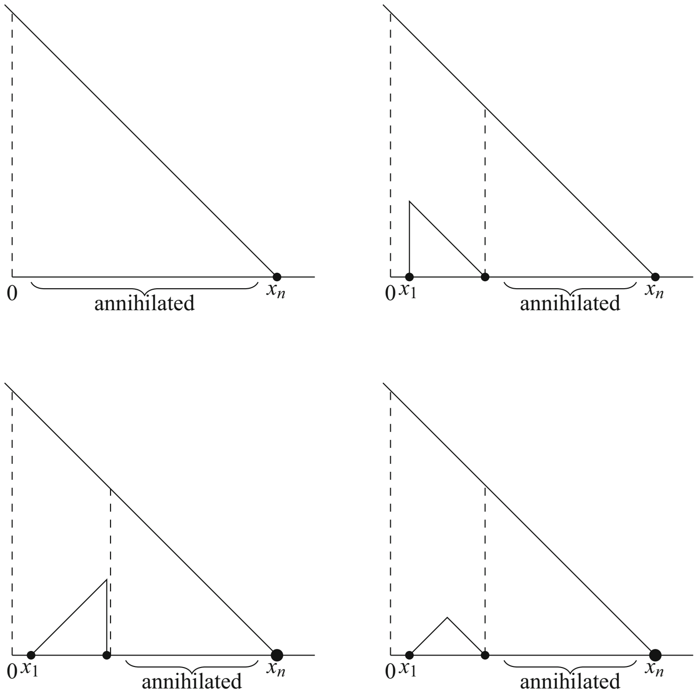


Fig. 2 The event corresponding to p_n (top left) and decomposition into the events corresponding to α_n (top right), β_n (bottom left) and γ_n (bottom right)

Let $\{r_1, \dots, r_{k-1}\}$ and $\{s_1, \dots, s_{n-k}\}$ be any disjoint subsets of $\{1, \dots, n\}$ such that $r_1 < \cdots < r_{k-1}$ and $s_1 < \cdots < s_{n-k}$. Then, by standard properties of uniform permutations, conditional on the event

$$E_{r,s} = \big\{\sigma(\{2, \dots, k\}) = \{r_1, \dots, r_{k-1}\} \text{ and } \sigma(\{k+1, \dots, n\}) = \{s_1, \dots, s_{n-k}\}\big\},$$

the random variables $(\ell_{\sigma(2)}, \dots, \ell_{\sigma(k)})$ and $(\ell_{\sigma(k+1)}, \dots, \ell_{\sigma(n)})$ are independent, and respectively distributed as $\big(\ell^{(r)}_{\tau(1)}, \dots, \ell^{(r)}_{\tau(k-1)}\big)$ and $\big(\ell^{(s)}_{\pi(1)}, \dots, \ell^{(s)}_{\pi(n-k)}\big)$, where $\ell_i^{(r)} = \ell_{r_i}$ and $\ell_i^{(s)} = \ell_{s_i}$, and τ, π are independent uniform permutations of $\{1, \dots, k-1\}$ and $\{1, \dots, n-k\}$ respectively. In particular, from (4),

$$\begin{aligned}\mathbb{P}_\ell(0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_n,\ \dot{\bullet}_1 \leftarrow \overleftarrow{\bullet}_k \mid E_{r,s}) &= p\, \mathbb{P}_{\ell^{(r)}}(0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_{k-1}) \mathbb{P}_{\ell^{(s)}}(0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_{n-k}) \\ &= p\, p_{k-1}\big(\ell^{(r)}\big) p_{n-k}\big(\ell^{(s)}\big),\end{aligned}$$

which by induction does not depend on ℓ. Since $E_{r,s}$ does not depend on ℓ, summing over values of r, s, and of k proves universality for α_n.

Secondly, we consider β_n. Contrary to the previous case, we only show that, for each integer k in $\{2, \ldots, n-1\}$, the sum $\mathbb{P}_\ell(0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_n,\ \overrightarrow{\bullet}_1 \to \dot{\bullet}_k) + \mathbb{P}_\ell(0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_n,\ \overrightarrow{\bullet}_1 \to \dot{\bullet}_{k'})$ is universal, where $k' = n + 1 - k$. Summing over k then gives universality of $2\beta_n$, hence of β_n. Let $k \in \{2, \ldots, n-1\}$, and note that

$$\{0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_n,\ \overrightarrow{\bullet}_1 \to \dot{\bullet}_k\} = \{\overrightarrow{\bullet}_1 \overset{1}{\rightarrow} x_k\}_{[x_1,x_k)} \cap \{\dot{\bullet}_k\} \cap \{x_k \overset{1}{\leftarrow} \overleftarrow{\bullet}_n\}_{(x_k,x_n]}$$
$$\cap \Big(\{x_k - x_1 < x_n - x_k\} \cup \big(\{x_k - x_1 = x_n - x_k\} \cap \{s_k = -1\}\big)\Big).$$

Conditional on the same type of event $E_{r,s}$ as above, the distances $x_k - x_1 = \sum_{1 \le i \le k-1} \ell_i^{(r)}$ and $x_n - x_k = \sum_{1 \le i \le n-k} \ell_i^{(s)}$ become deterministic, and as in the previous case the four events in the above conjunction are independent so that

$$\mathbb{P}_\ell(0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_n,\ \overrightarrow{\bullet}_1 \to \dot{\bullet}_k \mid E_{r,s}) = p_{k-1}\big(\ell^{(r)}\big) p p_{n-k}\big(\ell^{(s)}\big)$$
$$\times \Big(\mathbf{1}_{\left(\sum \ell^{(r)} < \sum \ell^{(s)}\right)} + \tfrac{1}{2}\mathbf{1}_{\left(\sum \ell^{(r)} = \sum \ell^{(s)}\right)}\Big).$$

Symmetrically, recalling that $k' = n + 1 - k$ and denoting by $E'_{s,r}$ the event that $\sigma(\{2, \ldots, k'\}) = \{s_1, \ldots, s_{k'-1}\}$ and $\sigma(\{k'+1, \ldots, n\}) = \{r_1, \ldots, r_{n-k'}\}$, we have

$$\mathbb{P}_\ell(0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_n,\ \overrightarrow{\bullet}_1 \to \dot{\bullet}_{k'} \mid E'_{s,r}) = p_{k'-1}\big(\ell^{(s)}\big) p p_{n-k'}\big(\ell^{(r)}\big)$$
$$\times \Big(\mathbf{1}_{\left(\sum \ell^{(s)} < \sum \ell^{(r)}\right)} + \tfrac{1}{2}\mathbf{1}_{\left(\sum \ell^{(s)} = \sum \ell^{(r)}\right)}\Big).$$

Thus, by summation, for any r, s we have

$$\mathbb{P}_\ell(0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_n,\ \overrightarrow{\bullet}_1 \to \dot{\bullet}_k \mid E_{r,s}) + \mathbb{P}_\ell(0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_n,\ \overrightarrow{\bullet}_1 \to \dot{\bullet}_{k'} \mid E'_{s,r}) = p_{k-1}\big(\ell^{(r)}\big)\, p\, p_{n-k}\big(\ell^{(s)}\big).$$

By induction, this does not depend on ℓ. Summing on values of (r, s) yields the expected conclusion.

Let us now consider γ_n. Here we have again that, for $1 < k < n$, the probability $\mathbb{P}_\ell(0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_n,\ \overrightarrow{\bullet}_1 \to\leftarrow \overleftarrow{\bullet}_k)$ is universal. Note indeed that

$$\{0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_n,\ \overrightarrow{\bullet}_1 \to\leftarrow \overleftarrow{\bullet}_k\} = \{\overrightarrow{\bullet}_1 \to\leftarrow \overleftarrow{\bullet}_k\}_{[x_1,x_k]} \cap \{x_k \overset{1}{\leftarrow} \overleftarrow{\bullet}_n\}_{(x_k,x_n]}, \tag{5}$$

and that, for all (r, s) as before,

$$\mathbb{P}_\ell(0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_n,\ \overrightarrow{\bullet}_1 \to\leftarrow \overleftarrow{\bullet}_k \mid E_{r,s}) = \mathbb{P}_{\ell^{(r)}}(\overrightarrow{\bullet}_1 \to\leftarrow \overleftarrow{\bullet}_k)\mathbb{P}_{\ell^{(s)}}(0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_n - k)$$
$$= \delta_{k-1}\big(\ell^{(r)}\big) p_{n-k}\big(\ell^{(s)}\big),$$

which by induction does not depend on ℓ.

Finally, we are left with δ_n. Note that $\{\vec{\bullet}_1 \to\leftarrow \overleftarrow{\bullet}_n\}$ implies $\{\vec{\bullet}_1\} \cap \{x_1 \overset{1}{\leftarrow} \overleftarrow{\bullet}_n\}_{(x_1,x_n]}$. Furthermore, the difference between these two events is precisely given by configurations where in absence of $\bullet_1$ we would have $0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_n$, but $\vec{\bullet}_1$ actually collides with some static particle $\dot{\bullet}_i$, thereby freeing a particle $\overleftarrow{\bullet}_j$ (compared to the configuration without $\bullet_1$) that becomes the first to hit 0, followed by $\overleftarrow{\bullet}_n$. This can be expressed as follows:

$$\{\vec{\bullet}_1\}\cap\{x_1 \overset{1}{\leftarrow} \overleftarrow{\bullet}_n\}_{(x_1,x_n]} = \{\vec{\bullet}_1 \to\leftarrow \overleftarrow{\bullet}_n\} \cup \bigcup_{1<i<j<n} \Big(\{0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_j\}\cap\{\vec{\bullet}_1 \to \dot{\bullet}_i\}\cap\{x_j \overset{1}{\leftarrow} \overleftarrow{\bullet}_n\}_{(x_j,x_n]}\Big).$$

Since these events are disjoint, we get (furthermore summing over i)

$$\frac{1-p}{2}\mathbb{P}_\ell\big((x_1 \overset{1}{\leftarrow} \overleftarrow{\bullet}_n)_{(x_1,x_n]}\big) = \delta_n + \sum_{1<j<n} \mathbb{P}_\ell\big(0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_j,\ \vec{\bullet}_1 \to \dot{\bullet},\ (x_j \overset{1}{\leftarrow} \overleftarrow{\bullet}_n)_{(x_j,x_n]}\big).$$

Conditioning on $\sigma(1)$ and applying the induction assumption on $(\ell_{\sigma(i)})_{i=2,\ldots,n}$ shows that the probability on the left-hand side equals p_{n-1} and is universal. As for the right-hand side probability, conditional on $\sigma(\{1,\ldots,j\})$, the events $\{0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_j,\ \vec{\bullet}_1 \to \dot{\bullet}\}$ and $\{x_j \overset{1}{\leftarrow} \overleftarrow{\bullet}_n\}_{(x_j,x_n]}$ are independent and we get

$$\mathbb{P}_\ell\big(0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_j,\ \vec{\bullet}_1 \to \dot{\bullet},\ (x_j \overset{1}{\leftarrow} \overleftarrow{\bullet}_n)_{(x_j,x_n]} \bigm| \sigma(\{1,\ldots,j\})\big) = \beta_{j-1}p_{n-j+1},$$

which by induction is universal. As a consequence, δ_n is universal too. Gathering identities proved along the way, we have

$$\alpha_n = p \sum_{1<k<n} p_{k-1}p_{n-k};$$

$$\beta_n = \frac{1}{2}\alpha_n;$$

$$\gamma_n = \sum_{1<k<n} \delta_k p_{n-k};$$

$$\delta_n = \frac{1-p}{2}p_{n-1} - \sum_{1<k<n} \beta_k p_{n-k}.$$

It follows that $p_n = \alpha_n + \beta_n + \gamma_n$ satisfies

$$\begin{aligned} p_n &= \frac{3}{2}p \sum_{1<k<n} p_{k-1}p_{n-k} + \sum_{1<k<n} p_{n-k}\Bigg(\frac{1-p}{2}p_{k-1} - \sum_{1<j<k} \beta_j p_{k-j}\Bigg) \\ &= \Big(p+\frac{1}{2}\Big)\sum_{1<k<n} p_{k-1}p_{n-k} - \frac{p}{2}\sum_{1<k<n}\sum_{1<j<k}\sum_{1<i<j} p_{n-k}p_{k-j}p_{j-i}p_{i-1}, \end{aligned}$$

and the formula for δ_n comes analogously.

From there, the distribution of the skyline follows:

Proof of Theorem 1 Let $s \in \mathbb{N}^*$, and $(l_1, r_1, \varsigma_1), \dots, (l_s, r_s, \varsigma_s)$ be any possible skyline on $\{1, \dots, n\}$, in other words this sequence satisfies

- $1 = l_1 \le r_1 = l_2 - 1 < r_2 = l_3 - 1 < \cdots < r_{s-1} = l_{s-2} + 1 < r_s = n$;
- $r_i - l_i \begin{cases} = 0 & \text{if } \varsigma_i = \uparrow, \\ \text{is odd} & \text{if } \varsigma_i \in \{\nearrow\uparrow, \uparrow\nwarrow\}, \\ \text{is even and } \ge 2 & \text{if } \varsigma_i \in \{\nwarrow, \nearrow\}; \end{cases}$
- $\varsigma_i = \nwarrow$ may only happen at the beginning, i.e. for all $i = 1, \dots, i_0$ (for some $i_0 \ge 0$);
- $\varsigma_i = \nearrow$ may only happen at the end, i.e. for all $i = n - j_0 + 1, \dots, n$ (for some $j_0 \ge 0$).

Then we simply observe that the realization of the skyline reduces to events on the disjoint intervals $[\![l_i, r_i]\!]$, which are independent:

$$\mathbb{P}\big(\text{skyline}_\ell(v, s, \sigma) = (l_i, r_i, \varsigma_i)_{1 \le i \le s}\big) = \prod_{i=1}^{s} q^{(\ell)}_{\varsigma_i}(r_i - l_i),$$

where, for all $m \in \mathbb{N}$,

$$q^{(\ell)}_{\uparrow}(m) = p\mathbf{1}_{(m=0)},$$

$$q^{(\ell)}_{\nearrow\uparrow}(m) = q^{(\ell)}_{\uparrow\nwarrow}(m) = p\mathbb{P}(0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_{m-1}) = p\, p_{m-1},$$

$$q^{(\ell)}_{\nearrow}(m) = q^{(\ell)}_{\nwarrow}(m) = \mathbb{P}(0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_m) = p_m,$$

$$q^{(\ell)}_{\nearrow\nwarrow}(m) = \mathbb{P}(\overrightarrow{\bullet}_1 \rightarrow\leftarrow \overleftarrow{\bullet}_m) = \delta_m,$$

referring to notations p_m and δ_m from the proof of Proposition 1, where they are proved not to depend on ℓ, thereby implying the theorem.

Corollary 2 is finally a direct corollary of Proposition 1:

Proof Property (a) is equivalent to saying that, for all n, the random variables $\mathbf{1}_{(A=n)}$ and x_n are independent. Indeed, for all $n \in \mathbb{N}$ and $t > 0$, $\mathbb{P}(x_A \le t,\ A = n) = \mathbb{P}(x_n \le t,\ A = n)$ while $\mathbb{P}(\widetilde{x}_A \le t,\ A = n) = \mathbb{P}(x_n \le t)\mathbb{P}(A = n)$. Let $n \in \mathbb{N}$. Conditional on $\ell^{(\cdot)} = \big(\ell^{(1)}, \dots, \ell^{(n)}\big)$, which are the order statistics of $(\ell_1, \dots, \ell_n)$, we have $x_n = \ell^{(1)} + \cdots + \ell^{(n)}$ and we reduce to the finite setting of Theorem 1:

$$\mathbb{P}(x_n \le t,\ A = n) = \mathbb{E}\big[\mathbb{P}_{\ell^{(\cdot)}}(0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_n)\mathbf{1}_{\left(\ell^{(1)} + \cdots + \ell^{(n)} \le t\right)}\big].$$

Since, by Theorem 1, the probability on the right-hand side does not depend on $\ell^{(\cdot)}$, and thus equals $\mathbb{P}(A = n)$, this concludes (a).

Property (b) follows: for all $\lambda > 0$, since $x_n = \ell_1 + \cdots + \ell_n$,

$$\mathscr{L}_D(\lambda) = \mathbb{E}[e^{-\lambda x_A}] = \mathbb{E}[e^{-\lambda \widetilde{x}_A}] = \sum_{n=0}^{\infty} \mathbb{E}[e^{-\lambda x_n}]\mathbb{P}(A = n)$$

$$= \sum_{n=0}^{\infty} \mathbb{E}[e^{-\lambda \ell_1}]^n \mathbb{P}(A = n) = f(\mathscr{L}_\ell(\lambda)),$$

where f is the generating function of A. The following relationship satisfied by f was deduced in the restricted setting of i.i.d. interdistances in [10], but we repeat it here for completeness. Introducing the generating series

$$A : x \mapsto \sum_{n=0}^{\infty} \alpha_n x^n, \quad B : x \mapsto \sum_{n=0}^{\infty} \beta_n x^n, \quad C : x \mapsto \sum_{n=0}^{\infty} \gamma_n x^n, \quad \text{and } D : x \mapsto \sum_{n=0}^{\infty} \delta_n x^n,$$

for which the recurrence relations proved above yield the relationships

$$A(x) = pxf(x)^2, \qquad B(x) = \frac{1}{2}A(x),$$

$$C(x) = D(x)f(x), \quad \text{and} \quad D(x) = \frac{1-p}{2}xf(x) - B(x)f(x),$$

and using the fact that $f(x) = \frac{1-p}{2}x + A(x) + B(x) + C(x)$, we obtain

$$f(x) = \frac{1-p}{2}x + \frac{3}{2}pxf(x)^2 + \frac{1-p}{2}xf(x)^2 - \frac{1}{2}pxf(x)^4,$$

from which (b) follows.

Deducing (a) from (b) would actually require computing the joint transform $\mathscr{L}_{(A,x_A)}(s,\lambda) = \mathbb{E}[s^A e^{-\lambda x_A}]$: one gets similarly $\mathscr{L}_{(A,x_A)}(s,\lambda) = f(s\mathscr{L}_\ell(\lambda))$, which coincides with $\mathscr{L}_{(A,\widetilde{x}_A)}(s,\lambda)$.

As advertised after the statement of Corollary 2, a more direct computation of the Laplace transform is possible, that we sketch below.

Alternative Proof of Corollary 2 (b) Let us directly obtain the functional equation (1) for the Laplace transform $\mathscr{L}_D(\lambda)$, using a continuous counterpart to the induction obtained for A. For a change, we shall also partly refer to the mass transport principle, whose use in the context of ballistic annihilation was introduced by Junge and Lyu [11]. Let us warn the reader that, since the arguments rely on similar arguments given elsewhere in the paper, and this is an alternative proof, we give somewhat fewer details. Let $\lambda \geq 0$.

We split the expectation according to the type of collision $\bullet_1$ is involved in. First, if $v_1 = -1$ then $D = x_1 = \ell_1$:

$$\mathbb{E}[e^{-\lambda D}\mathbf{1}_{(\overleftarrow{\bullet}_1)}] = \mathbb{E}[e^{-\lambda \ell_1}\mathbf{1}_{(\overleftarrow{\bullet}_1)}] = \mathscr{L}_\ell(\lambda)\frac{1-p}{2}.$$

Consider now the case when $v_1 = 0$. Define, in general, D', D'' by $x_1 + D' = \min\{x_k \,:\, (x_1 \overset{1}{\longleftarrow} \overleftarrow{\bullet}_k)_{(x_1,x_k]}\}$ and $x_1 + D' + D'' = \min\{x_k \,:\, (x_1 + D' \overset{1}{\longleftarrow} \overleftarrow{\bullet}_k)_{(x_1+D',x_k]}\}$, so that D' and D'' are independent copies of D, and are independent of v_1. Conditional on $v_1 = 0$, we have $D = x_1 + D' + D''$: either $D = \infty$, in which case either $D' = \infty$ or $D'' = \infty$, or $D < \infty$, which implies that $\dot{\bullet}_1$ is first annihilated, and then 0 is hit by a particle that hit $x_1 + D'$ before. Thus,

$$\mathbb{E}[e^{-\lambda D}\mathbf{1}_{(\dot{\bullet}_1)}] = \mathbb{E}[e^{-\lambda(x_1+D+D')}\mathbf{1}_{(\dot{\bullet}_1)}] = \mathbb{E}[e^{-\lambda\ell_1}]\mathbb{E}[e^{-\lambda D}]^2 p = \mathscr{L}_\ell(\lambda)\mathscr{L}_D(\lambda)^2 p.$$

We focus now on $\mathbb{E}[e^{-\lambda D}\mathbf{1}_{(\overrightarrow{\bullet}_1\to\dot{\bullet})}] = \mathbb{E}[e^{-\lambda D}\mathbf{1}_{(\overrightarrow{\bullet}_1\to\dot{\bullet})\wedge(D<\infty)}]$. Let us apply the mass transport principle (cf. [11]) to $f(u,v) = e^{-\lambda D(u,v)}\mathbf{1}_{(\overrightarrow{\bullet}_u\to\dot{\bullet}_v)}$, where $D(u,v)$ is the sum of the distances from x_v to the first particle to cross x_v from the left (which is $x_v - x_u$ on the event $\{\overrightarrow{\bullet}_u \to \dot{\bullet}_v\}$), and to the first particle to cross x_v from the right. Thus, we temporarily extend to process to the full-line $\mathbb{R}$ in order to use the mass transport principle. This gives:

$$\sum_{v\in\mathbb{Z}}\mathbb{E}[e^{-\lambda D(0,v)}\mathbf{1}_{(\overrightarrow{\bullet}_0\to\dot{\bullet}_v)}] = \sum_{u\in\mathbb{Z}}\mathbb{E}[e^{-\lambda D(u,0)}\mathbf{1}_{(\overrightarrow{\bullet}_u\to\dot{\bullet}_0)}],$$

which rewrites as follows (using translation invariance):

$$\mathbb{E}[e^{-\lambda(D-\ell_1)}\mathbf{1}_{(\overrightarrow{\bullet}_1\to\dot{\bullet})\wedge(D<\infty)}] = p\mathbb{E}[e^{-\lambda(D_-+D)}\mathbf{1}_{(D_-<D)}]$$

where D_- is the symmetric counterpart to D on $\mathbb{R}_-$; it is an independent copy of D, hence, since also ℓ_1 is independent of $D - \ell_1$ (actually ℓ_1 plays no role in the collisions),

$$\mathbb{E}[e^{-\lambda D}\mathbf{1}_{(\overrightarrow{\bullet}_1\to\dot{\bullet})}] = \mathbb{E}[e^{-\lambda\ell_1}]\frac{1}{2}\mathbb{E}[e^{-\lambda D}]^2 = \frac{p}{2}\mathscr{L}_\ell(\lambda)\mathscr{L}_D(\lambda)^2.$$

Let us finally consider $\mathbb{E}[e^{-\lambda D}\mathbf{1}_{(\overrightarrow{\bullet}_1\to\leftarrow\overleftarrow{\bullet})}]$. Conditionally on the event $\{\overrightarrow{\bullet}_1\to\leftarrow\overleftarrow{\bullet}\}$, $D = \ell_1 + \varDelta + D'$ where $\varDelta = x_K - x_1$ if K is the index such that $\overrightarrow{\bullet}_1 \to\leftarrow \overleftarrow{\bullet}_K$, and $D' = D - x_K$. Note that ℓ_1, $\varDelta$ and D' are independent on that event, and that D' has same distribution as D so that

$$\begin{aligned}\mathbb{E}[e^{-\lambda D}\mathbf{1}_{(\overrightarrow{\bullet}_1\to\leftarrow\overleftarrow{\bullet})}] &= \mathbb{E}[e^{-\lambda\ell_1}]\mathbb{E}[e^{-\lambda D}]\mathbb{E}[e^{-\lambda\varDelta}\mathbf{1}_{(\overrightarrow{\bullet}_1\to\leftarrow\overleftarrow{\bullet})}]\\ &= \mathscr{L}_\ell(\lambda)\mathscr{L}_D(\lambda)\mathbb{E}[e^{-\lambda\varDelta}\mathbf{1}_{(\overrightarrow{\bullet}_1\to\leftarrow\overleftarrow{\bullet})}].\end{aligned}$$

We are thus left with the Laplace transform of Δ. As in the study of the law of A, we notice that $\{\vec{\bullet}_1 \to \leftarrow \overleftarrow{\bullet}\}$ happens on $\{\vec{\bullet}_1\} \cap \{x_1 \overset{1}{\leftarrow} \overleftarrow{\bullet}\}_{(x_1,+\infty)}$ *unless* $\vec{\bullet}_1$ hits a static particle first, i.e. there exists $j < k$ such that $\{\vec{\bullet}_1 \to \dot{\bullet}_j\} \cap \{0 \overset{1}{\leftarrow} \overleftarrow{\bullet}_k\}_{(0,+\infty)} \cap \{x_k \overset{1}{\leftarrow} \overleftarrow{\bullet}\}_{(x_k,+\infty)}$ is realized. The last condition of that event is independent of the previous ones, and depends on a piece of environment of length distributed as D. Thus we find

$$\begin{aligned}\mathbb{E}[e^{-\lambda\Delta}\mathbf{1}_{(\vec{\bullet}_1\to\leftarrow\overleftarrow{\bullet})}] &= \frac{1-p}{2}\mathscr{L}_D(\lambda) - \mathbb{E}[e^{-\lambda(D-\ell_1)}\mathbf{1}_{(\vec{\bullet}_1\to\dot{\bullet})\wedge(0\overset{1}{\leftarrow}\overleftarrow{\bullet})}]\mathscr{L}_D(\lambda)\\ &= \frac{1-p}{2}\mathscr{L}_D(\lambda) - \frac{p}{2}\mathscr{L}_D(\lambda)^3\end{aligned}$$

where for the last computation we reuse the previous case $\mathbb{E}[e^{-\lambda D}\mathbf{1}_{(\vec{\bullet}_1\to\dot{\bullet})}]$.

All together, this gives

$$\begin{aligned}\mathscr{L}_D(\lambda) = \frac{1-p}{2}\mathscr{L}_\ell(\lambda) + p\mathscr{L}_\ell(\lambda)\mathscr{L}_D(\lambda)^2 + \frac{p}{2}\mathscr{L}_\ell(\lambda)\mathscr{L}_D(\lambda)^2\\ + \mathscr{L}_\ell(\lambda)\mathscr{L}_D(\lambda)\Big(\frac{1-p}{2}\mathscr{L}_D(\lambda) - \frac{p}{2}\mathscr{L}_D(\lambda)^3\Big),\end{aligned}$$

which after simplification is the claimed identity.

Let us merely mention that the proof could also give the joint transform of A and $D = x_A$, from which part (a) of Corollary 2 follows as well.

4 Direct Approach to Universality

We aim here at giving a direct proof of the fact that the law of the skyline does not depend on the sequence of interdistances. This approach unites the model for different values of ℓ and gives a clearer understanding of the universality property but does not however yield explicit distributions. Let n be fixed in this part.

Among the set $\mathbb{L}_n$ of lengths, we distinguish "generic" length sequences, for which no triple collision may happen, in other words no two subsets of lengths have the same sum:

$$\mathbb{L}_{\text{generic}} = \{\ell = (\ell_1, \dots, \ell_n) \in \mathbb{L}_n \;:\; \forall I, J \subset \{1, \dots, n\}, \ell_I \neq \ell_J \text{ unless } I = J\},$$

where for any subset $I \subset \{1, \dots, n\}$, we let $\ell_I = \sum_{i\in I} \ell_i$. Finally, we will need to refer to lengths allowing a *single* triple collision:

$$\mathbb{L}_{\text{single}} = \{\ell \in \mathbb{L}_n \;:\; \exists \text{ a unique pair } I, J \subset \{1, \dots, n\} \text{ such that } I \cap J = \emptyset \text{ and } \ell_I = \ell_J\},$$

and $\mathbb{L}_{\text{multiple}} = \mathbb{L}_n \setminus (\mathbb{L}_{\text{generic}} \cup \mathbb{L}_{\text{single}})$.

Locally Constant on $\mathbb{L}_{\text{generic}}$ Notice first that the joint law of the velocities and pairing among annihilating particles is locally constant on $\mathbb{L}_{\text{generic}}$. Indeed all relative orders among values of ℓ_I, for $I \subset \{1, \ldots, n\}$, are locally constant on $\mathbb{L}_{\text{generic}}$, hence for any velocities $v_1, \ldots, v_n \in \{-1, 0, +1\}$ and any involution $\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}$, the set of permutations σ producing the pairing π (i.e. such that $\bullet_i \sim \bullet_j$ if and only if $\pi(i) = j$, and $\pi(i) = i$ if and only if $\bullet_i$ survives) is itself locally constant in $\mathbb{L}_{\text{generic}}$. As an immediate consequence, the law of the skyline is constant on each connected component of $\mathbb{L}_{\text{generic}}$.

Continuous on $\mathbb{L}_{\text{single}}$ Let us argue that these probabilities are continuous on $\mathbb{L}_{\text{single}}$. Let $\ell \in \mathbb{L}_{\text{single}}$. There is thus a unique pair I, J such that $I \cap J = \emptyset$ and $\ell_I = \ell_J$.

If ε denotes the smallest difference among $|\ell_K - \ell_L|$ for all K, L, $K \cap L = \emptyset$, $\{K, L\} \neq \{I, J\}$, then each vector ℓ' at uniform distance smaller than ε from ℓ either belongs to one of two connected components $\mathbb{L}_+(\ell)$ or $\mathbb{L}_-(\ell)$ of $\mathbb{L}_{\text{generic}}$ according to whether $\ell'_I < \ell'_J$ or $\ell'_I > \ell'_J$, or to $\mathbb{L}_0(\ell) \subset \mathbb{L}_{\text{single}}$ if $\ell'_I = \ell'_J$.

If $\ell' \in \mathbb{L}_0(\ell)$, the joint law of velocities, spins and pairing is preserved as above, hence the probability is the same as for ℓ.

If $\ell' \in \mathbb{L}_+(\ell)$, let us establish that the probability is preserved. Let us describe a one-to-one map Φ on the set of velocities, spins and permutations (v, s, σ) that preserves the number of static (hence of moving) particles and such that $\text{skyline}_\ell(v, s, \sigma) = \text{skyline}_{\ell'}(\Phi(v, s, \sigma))$ (however, contrary to the previous cases, Φ does not a priori preserve the pairing); since the probability of each realization of (v, s, σ) only depends on the number of static particles (thanks to symmetry), this will conclude the argument. The skyline for ℓ does not depend on the pairing of a given subset of $\{1, \ldots, n\}$ given that it totally annihilates and lies below a given annihilating pair or a surviving ± 1 particle. It is therefore sufficient that Φ only alters such subsets.

If (v, s, σ) has no triple collision for ℓ, then $\Phi(v, s, \sigma) = (v, s, \sigma)$, and the pairings for ℓ and ℓ' are the same, as for $\mathbb{L}_{\text{generic}}$, meaning that $\text{skyline}_\ell(v, s, \sigma) = \text{skyline}_{\ell'}(\Phi(v, s, \sigma))$.

If there is a triple collision $\vec{\bullet}_j \to \dot{\bullet}_i \leftarrow \overleftarrow{\bullet}_k$ for (v, s, σ) and ℓ, then we let $(v', s', \sigma') = \text{rev}_{j,k}^{\ell,\ell'}(v, s, \sigma)$. Here, in wider generality, for $1 \le j < k \le n$, $\text{rev}_{j,k}^{\ell,\ell'} : (v, s, \sigma) \mapsto (v', s', \sigma')$ is a "reversing operator around a triple collision between $\vec{\bullet}_j$ and $\overleftarrow{\bullet}_k$", from $\{-1, 0, +1\}^n \times \{-1, +1\}^n \times \mathfrak{S}_n$ to itself, defined by: (see also Fig. 3)

- if (v, s, σ) does not induce $\vec{\bullet}_j \to \dot{\bullet} \leftarrow \overleftarrow{\bullet}_k$ for ℓ, then $(v', s', \sigma') = (v, s, \sigma)$.
- otherwise, i.e. if $\vec{\bullet}_j \to \dot{\bullet}_i \leftarrow \overleftarrow{\bullet}_k$ for some $j < i < k$, for distances ℓ, then denote $I = \sigma(\{j, \ldots, i-1\})$ and $J = \sigma(\{i, \ldots, k-1\})$ (so that $\ell_I = \ell_J$), and
  - if $\ell'_I < \ell'_J$ and $s_i = -1$, or $\ell'_I > \ell'_J$ and $s_i = +1$, then $(v', s', \sigma') = (v, s, \sigma)$;
  - else, define (v', s', σ') from (v, s, σ) by mirroring the interval (x_j, x_k), i.e. reversing the order of interdistances, velocities and spins, and furthermore changing velocities and spins to their opposite, in this interval : for $j < m \le k$,

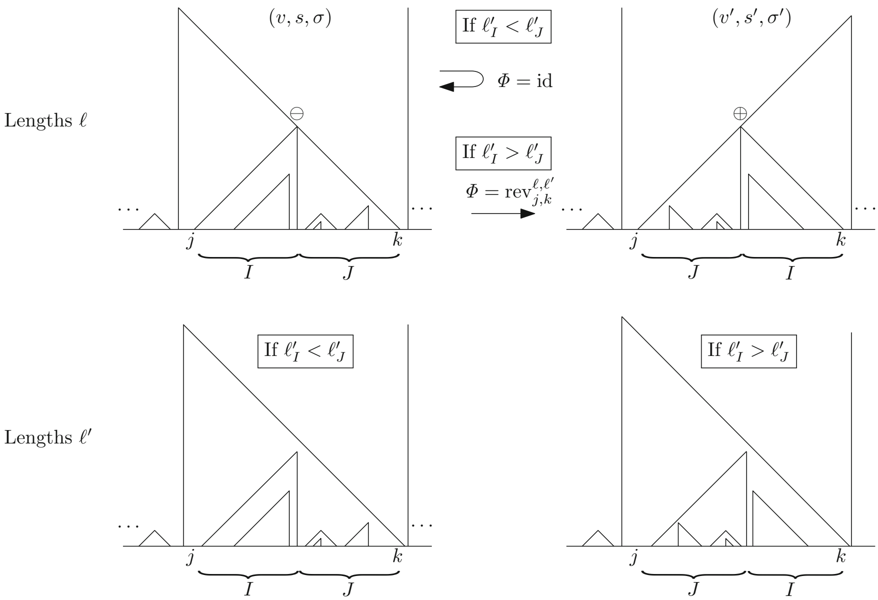


Fig. 3 Illustration of the map Φ in the case of a single triple collision

$\sigma'(m) = \sigma(j + 1 + k - m)$ and for $j < m < k$, $v'_m = -v_{j+k-m}$, $s'_m = -s_{j+k-m}$; and (v', s', σ') and (v, s, σ) coincide elsewhere.

Note that, in the last case, (v', s', σ') still has a triple collision $\vec{\bullet}_j \rightarrow \dot{\bullet}_{i'} \leftarrow \overleftarrow{\bullet}_k$ for ℓ, at $i' = j + k - i = i + |J| - |I|$; the mirroring has the effect of exchanging the roles of I and J, but also of letting $s_{i'} = -s_i$, so that (v', s', σ') would also fall into this last case hence $\mathrm{rev}^{\ell,\ell'}_{j,k}(v', s', \sigma') = (v, s, \sigma)$. Thus this operator is involutive on the subset of all (v, s, σ) that have a triple collision $\vec{\bullet}_j \rightarrow \dot{\bullet} \leftarrow \overleftarrow{\bullet}_k$. For $\ell \in \mathbb{L}_{\text{single}}$, the configuration space is partitioned into those subsets for $1 \leq j < k \leq n$; as a consequence, Φ is involutive, hence bijective, on the whole configuration space $\{-1, 0, +1\}^n \times \{-1, +1\}^n \times \mathfrak{S}_n$.

And by construction, each operator $\mathrm{rev}^{\ell,\ell'}_{j,k}$, and thus Φ, only affects the pairing of indices of particles that annihilate and are lying "under" a moving particle (i.e. whose range is later visited by another particle), which has no consequence on the skyline. In particular, $\mathrm{skyline}_\ell(v, s, \sigma) = \mathrm{skyline}_{\ell'}(v', s', \sigma')$.

Finally, $\mathrm{rev}^{\ell,\ell'}_{j,k}$, hence Φ, clearly preserves the number of static particles.

Connectedness of $\mathbb{L}_{\text{generic}} \cup \mathbb{L}_{\text{single}}$ We have that $\mathbb{L}_{\text{generic}} \cup \mathbb{L}_{\text{single}}$ is a connected subset of $\mathbb{L}$. Indeed, its complement $\mathbb{L}_{\text{multiple}}$ in the positive full-dimensional cone $\mathbb{L}$ is a finite union of subspaces of codimension at least 2, namely induced by at least two different constraints $\ell_{I_1} = \ell_{J_1}$, $\ell_{I_2} = \ell_{J_2}$ (which are not equivalent, hence non-colinear, due to $I_i \cap J_i = \emptyset$, $i = 1, 2$).

From the previous points, we conclude that the probabilities are constant on $\mathbb{L}_{\text{generic}} \cup \mathbb{L}_{\text{single}}$.

Extension to $\mathbb{L}_{\text{multiple}}$ Let us finally extend the argument for $\mathbb{L}_{\text{single}}$ to the general case. We may pick $\ell' \in \mathbb{L}_{\text{generic}}$ close enough to ℓ so that all relative orders among the sums ℓ_I are preserved except for the equality cases. We then describe a permutation Φ of triples (v, s, σ) as above, which only depends on the order within all pairs ℓ'_I and ℓ'_J where $\ell_I = \ell_J$. Let (v, s, σ) be a configuration. For that configuration, and distance ℓ, we may order the triple collision pairs (j, k) (i.e. such that $\vec{\bullet}_j \to \dot{\bullet} \leftarrow \overleftarrow{\bullet}_k$) as $(j_1, k_1), \dots, (j_M, k_M)$ in a way that complies with inclusion: if $1 \le K \le L \le M$, then either $[j_K, k_K] \cap [j_L, k_L] = \emptyset$ or $[j_K, k_K] \subset [j_L, k_L]$. It suffices to first list, in arbitrary order, all intervals that are minimal for inclusion, and then iterate on the remaining ones. Then we define

$$\Phi(v, s, \sigma) = \text{rev}_{j_M,k_M}^{\ell,\ell'} \circ \text{rev}_{j_{M-1},k_{M-1}}^{\ell,\ell'} \circ \cdots \circ \text{rev}_{j_1,k_1}^{\ell,\ell'}(v, s, \sigma).$$

For $K = 1, \dots, M-1$, due to the ordering, the application of $\text{rev}_{j_K,k_K}^{\ell,\ell'}$ in Φ does not alter the indices of the forthcoming triple collisions $(j_{K+1}, k_{K+1}), \dots, (j_M, k_M)$, ensuring that each operator really acts on a triple collision. Also, the ordering among triple collisions with disjoint supports has no effect on Φ since the corresponding rev operators commute. If there is no triple collision for (v, s, σ) and ℓ, we mean to define $\Phi(v, s, \sigma) = (v, s, \sigma)$.

Although the operator Φ does not preserve the set of pairs $(j_1, k_1), \dots, (j_M, k_M)$ of triple collisions, it does preserve the *maximal* ones (for inclusion), which ensures it preserves the skyline when going from $(v, s, \sigma), \ell$ to $\Phi(v, s, \sigma), \ell'$. Also, it is still an involution: this is obtained by induction on the maximum number of nested triple collisions, together with the simple fact that Φ (and even each $\text{rev}_{j,k}^{\ell,\ell'}$) commutes with a mirroring Σ of the whole interval where Φ is acting. Also, Φ still preserves the number of static particles. Altogether, we deduce that the law of skyline_ℓ and $\text{skyline}_{\ell'}$ are equal.

5 Asymmetric Case: Failure of Universality and a Remarkable Property of Gamma Distributions

Let us consider the asymmetric case, where the distribution of velocities is given by $(1-r)(1-p)\delta_{-1} + p\delta_0 + r(1-p)\delta_{+1}$, for some $r \in (0, 1) \setminus \{\frac{1}{2}\}$, and still $p \in (0, 1)$. Many questions remain open in this case, but Junge and Lyu [11] could still prove that some of the identities of the symmetric case can be extended, implying that the model still has a subcritical and a supercritical phase, although the phase transition was not proved unique.

A notable difference, that helps understand why the asymmetric case might be sensibly harder, is the apparent lack of universality. It is indeed simple although tedious to check on the first cases that the distribution of A is distribution-dependent. In particular, one can check that $\mathbb{P}(A = 5)$ depends on the distribution. Writing $z := \mathbb{P}(x_4 - x_1 > x_5 - x_4) + \frac{1}{2}\mathbb{P}(x_4 - x_1 = x_5 - x_4)$, we have

$$\mathbb{P}((A = 5) \wedge (v = (1, 1, 0, 0, -1))) = zr^2(1-r)p^2(1-p)^3/8$$
$$\mathbb{P}((A = 5) \wedge (v = (1, 0, 0, -1, -1))) = (1-z)r(1-r)^2p^2(1-p)^3/8$$
$$\mathbb{P}((A = 5) \wedge (v = (1, 0, -1, 0, -1))) = zr(1-r)^2p^2(1-p)^3r^2(1-r)/16$$
$$\mathbb{P}((A = 5) \wedge (v = (1, 0, 1, 0, -1))) = (1-z)r^2(1-r)p^2(1-p)^3/16,$$

but the total contribution from other possible velocities is universal. Since $r \neq \frac{1}{2}$, the total of these four probabilities depends on z and hence on the distribution of distances.

Let us still state two surprising properties that hold in the asymmetric case. The first one would, in the symmetric case, follow at once from Theorem 1 by a seamless generalization of the proof of Corollary 2 (a). The second one however is new in any case.

Theorem 2 *We consider the random lengths setting.*

(a) In the symmetric or asymmetric cases, for all n, x_n and $\text{skyline}_{(\ell_1,\ldots,\ell_n)}$ *are independent.*
(b) In the symmetric or asymmetric cases and if m is a gamma distribution, for all n, x_n and the whole combinatorial configuration, i.e. (v, π) *(velocities and pairing), are independent.*

Note that, due to the assumed independence between ℓ and v, the property (b) is actually an independence between x_n and π given any velocities $v_1, \ldots, v_n$.

Let us give a simple counterexample illustrating why (b) doesn't hold in general. Consider $m = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_4$ and the configuration on five particles given by velocities $v = (1, 1, 0, 0, -1)$ and pairing $\pi = \begin{pmatrix} 4 & 3 & 2 & 1 & 5 \end{pmatrix}$ (i.e., $\bullet_1 \sim \bullet_4$, etc.). Then, given this realization of (v, π), interdistances *necessarily* are $\ell_2 = \ell_3 = \ell_4 = 1$ and $\ell_5 = 4$ (remember ℓ_1 plays no role), see Fig. 4, hence the distribution of x_5 takes two values depending on ℓ_1, and thus clearly differs from the unconditioned distribution.

Let us remind the reader that the model of ballistic annihilation was studied in the physics literature under the assumption of exponential interdistances (see e.g. [6]), which simplified computations. We don't have knowledge however of a previous result that would rely on that distribution except for technical reasons.

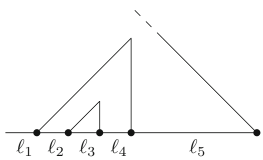


Fig. 4 Simple counterexample to the independence of x_n from the pairing of $[\![1, n]\!]$ for arbitrary interdistance distribution (see above)

Let it finally be mentioned that we can't rule out a different form of universality, which might still await discovery. Still, numerical simulations suggest that the critical probability itself, should it exist, could depend on the distribution of interdistances.

Proof

(a) Note that conditioning on the skyline amounts to a conjunction of independent conditions on disjoint subintervals, that are either of the type $\{x_j \xleftarrow{1} \overleftarrow{\bullet}_k\}$ (including $\dot{\bullet}_j \leftarrow \overleftarrow{\bullet}_k$) or $\{\overrightarrow{\bullet}_j \to\leftarrow \overleftarrow{\bullet}_k\}$, and x_n is the total length of these subintervals, together with unconditioned intervals in-between. It is therefore sufficient to show independence between x_n and both of $\{0 \xleftarrow{1} \overleftarrow{\bullet}_n\}$ and $\{\overrightarrow{\bullet}_1 \to\leftarrow \overleftarrow{\bullet}_n\}$. This property will be obtained via a similar recursion scheme as in the proof of Proposition 1—or rather as in the proof of Theorem 2 from [10], since we are considering random lengths. We actually prove the stronger statement of independence between x_n and each of the events $\{0 \xleftarrow{1} \overleftarrow{\bullet}_n\} \cap \{\dot{\bullet}_1\}$, $\{0 \xleftarrow{1} \overleftarrow{\bullet}_n\} \cap \{\overrightarrow{\bullet}_1 \to \dot{\bullet}\}$, $\{0 \xleftarrow{1} \overleftarrow{\bullet}_n\} \cap \{\overrightarrow{\bullet}_1 \to\leftarrow \overleftarrow{\bullet}\}$ and $\{\overrightarrow{\bullet}_1 \to\leftarrow \overleftarrow{\bullet}_n\}$.

The case $n = 1$ is clear. Assume now $n \geq 2$ and that the independences hold for any number $m < n$ of particles (note that, for each m, depending on parity, only one of the conditions $\{0 \xleftarrow{1} \overleftarrow{\bullet}_m\}$ and $\{\overrightarrow{\bullet}_1 \to\leftarrow \overleftarrow{\bullet}_n\}$ has nonzero probability, so independence is trivial for the other). Symmetrically, we already remark that this assumption implies an independence between x_m (unchanged by left-right symmetry) and the event $\{\overrightarrow{\bullet}_1 \xrightarrow{1} x_{m+1}\}_{[x_1, x_{m+1})}$ for all $m < n$.

In the following, in order to emphasize that we restrict to $[0, x_n]$, we denote $\mathbb{P}^{(n)}$ the probability of the model restricted to particles $\bullet_1, \ldots, \bullet_n$ (remember the random length model was defined for infinitely many particles).

Consider any measurable function $f : \mathbb{R} \to \mathbb{R}_+$. We have

$$\mathbb{E}^{(n)}[f(x_n)\mathbf{1}_{(0 \xleftarrow{1} \overleftarrow{\bullet}_n)}\mathbf{1}_{(\dot{\bullet}_1)}] = \sum_{1<k<n} \mathbb{E}^{(n)}[f(x_n)\mathbf{1}_{(\dot{\bullet}_1 \leftarrow \overleftarrow{\bullet}_k)}\mathbf{1}_{(x_k \xleftarrow{1} \overleftarrow{\bullet}_n)}]$$

and, under the condition appearing on the right hand side, by induction, each of $x_k - x_1$ and $x_n - x_k$ (and trivially x_1) have unconditioned distributions; they are also mutually independent, as in their joint unconditioned distribution, so that the distribution of $x_n = (x_n - x_k) + (x_k - x_1) + x_1$ is unaffected by this condition, hence

$$\mathbb{E}^{(n)}[f(x_n)\mathbf{1}_{(0 \xleftarrow{1} \overleftarrow{\bullet}_n)}\mathbf{1}_{(\dot{\bullet}_1)}] = \mathbb{E}^{(n)}[f(x_n)]\mathbb{P}((0 \xleftarrow{1} \overleftarrow{\bullet}_n) \wedge (\dot{\bullet}_1)),$$

as expected. Next, we have (as in the study of β_n in the proof of Proposition 1), denoting $k' = n + 1 - k$ for any $1 < k < n$,

$$\begin{aligned}\mathbb{E}^{(n)}[f(x_n)\mathbf{1}_{(0 \xleftarrow{1} \overleftarrow{\bullet}_n)}\mathbf{1}_{(\overrightarrow{\bullet}_1 \to \dot{\bullet})}] = \frac{p}{2} \sum_{1<k<n} \Big(& \mathbb{E}^{(n)}[f(x_n)\mathbf{1}_{(\overrightarrow{\bullet}_1 \xrightarrow{1} x_k)}\mathbf{1}_{(x_k \xleftarrow{1} \overleftarrow{\bullet}_n)}\mathbf{1}_{(x_k - x_1 < x_n - x_k)}] \\ & + \mathbb{E}^{(n)}[f(x_n)\mathbf{1}_{(\overrightarrow{\bullet}_1 \xrightarrow{1} x_{k'})}\mathbf{1}_{(x_{k'} \xleftarrow{1} \overleftarrow{\bullet}_n)}\mathbf{1}_{(x_{k'} - x_1 < x_n - x_{k'})}]\end{aligned}$$

$$+\frac{1}{2}\mathbb{E}^{(n)}[f(x_n)\mathbf{1}_{(\vec{\bullet}_1 \xrightarrow{1} x_k)}\mathbf{1}_{(x_k \xleftarrow{1} \overleftarrow{\bullet}_n)}\mathbf{1}_{(x_k-x_1=x_n-x_k)}]$$

$$+\frac{1}{2}\mathbb{E}^{(n)}[f(x_n)\mathbf{1}_{(\vec{\bullet}_1 \xrightarrow{1} x_{k'})}\mathbf{1}_{(x_{k'} \xleftarrow{1} \overleftarrow{\bullet}_n)}\mathbf{1}_{(x_{k'}-x_1=x_n-x_{k'})}]\Big).$$

By the induction assumption, for all k, conditional on the event $\{\vec{\bullet}_1 \xrightarrow{1} x_{k'}\}$, $x_{k'} - x_1$ is unconditioned, and in particular (by the induction again, symmetrically) has same distribution as $x_n - x_{n-k'+1} = x_n - x_k$ conditional on $\{x_k \xleftarrow{1} \overleftarrow{\bullet}_n\}$; similarly, conditional on $\{x_{k'} \xleftarrow{1} \overleftarrow{\bullet}_n\}$, $x_n - x_{k'}$ has same distribution as $x_{n-k'+1} - x_1 = x_k - x_1$ conditional on $\{\vec{\bullet}_1 \xrightarrow{1} x_k\}$; furthermore both are independent given the independent events $\{\vec{\bullet}_1 \xrightarrow{1} x_{k'}\} \cap \{x_k \xleftarrow{1} \overleftarrow{\bullet}_n\}$. Hence, using invariance of x_n by permutation of distances,

$$\mathbb{E}^{(n)}\left[f(x_n)\mathbf{1}_{(\vec{\bullet}_1 \xrightarrow{1} x_{k'})}\mathbf{1}_{(x_{k'} \xleftarrow{1} \overleftarrow{\bullet}_n)}\mathbf{1}_{\substack{(x_{k'}-x_1< \\ x_n-x_{k'})}}\right] = \mathbb{E}^{(n)}\left[f(x_n)\mathbf{1}_{(\vec{\bullet}_1 \xrightarrow{1} x_k)}\mathbf{1}_{(x_k \xleftarrow{1} \overleftarrow{\bullet}_n)}\mathbf{1}_{\substack{(x_n-x_k< \\ x_k-x_1)}}\right].$$

Getting back to the previous summation, the comparison between distances simplifies, leaving independent conditions which by induction are independent of the widths:

$$\begin{aligned}
\mathbb{E}^{(n)}[f(x_n)\mathbf{1}_{(0 \xleftarrow{1} \overleftarrow{\bullet}_n)}\mathbf{1}_{(\vec{\bullet}_1 \to \dot{\bullet})}] &= \frac{p}{2}\sum_{1<k<n}\mathbb{E}^{(n)}[f(x_n)\mathbf{1}_{(\vec{\bullet}_1 \xrightarrow{1} x_k)}\mathbf{1}_{(x_k \xleftarrow{1} \overleftarrow{\bullet}_n)}] \\
&= \frac{p}{2}\sum_{1<k<n}\mathbb{E}^{(n)}[f(x_n)]\mathbb{P}(\vec{\bullet}_1 \xrightarrow{1} x_k)\mathbb{P}(x_k \xleftarrow{1} \overleftarrow{\bullet}_n) \\
&= \mathbb{E}^{(n)}[f(x_n)]\mathbb{P}\big((0 \xleftarrow{1} \overleftarrow{\bullet}_n) \wedge (\vec{\bullet}_1 \to \dot{\bullet})\big).
\end{aligned}$$

Finally,

$$\mathbb{E}^{(n)}[f(x_n)\mathbf{1}_{(\vec{\bullet}_1 \to \leftarrow \overleftarrow{\bullet})}] = \sum_{1<k<n}\mathbb{E}^{(n)}[f(x_n)\mathbf{1}_{(\vec{\bullet}_1 \to \leftarrow \overleftarrow{\bullet}_k)}\mathbf{1}_{(x_k \xleftarrow{1} \overleftarrow{\bullet}_n)}];$$

the last two conditions are independent, and by induction they don't affect the distribution of distances $x_k - x_1$ and $x_n - x_k$, so this case is handled as the first one. This altogether gives independence between x_n and $\{0 \xleftarrow{1} \overleftarrow{\bullet}_n\}$.

It remains to consider x_n and $\{\vec{\bullet}_1 \to \leftarrow \overleftarrow{\bullet}_n\}$. Using the same decomposition as in the proof of Proposition 1 (case of δ_n) or in the alternative proof of Corollary 2 (page 501), we reduce to the independence between x_j and $\{\vec{\bullet}_1 \to \dot{\bullet}\} \cap \{0 \xleftarrow{1} \overleftarrow{\bullet}_j\}$, and conclude as in the previous cases.

(b) Up to scaling, it is enough to prove the result for gamma distributions of scale parameter 1.

We prove, by induction on the number n of particles, that the result holds for a generalized model where some sites may be "devoid of a particle", which we formally handle by considering that the n particles are separated by *sums* of i.i.d. gamma interdistances, i.e. interdistances are again gamma distributed, with possibly different shape parameters (but same scale parameter 1). In the following, when referring to a gamma distribution, it shall always be of scale parameter 1.

The first nontrivial case is $n = 3$, and only in the case of velocities $(+1, 0, -1)$. Then the pairing depends on the comparison between ℓ_2 and ℓ_3, and we need to show that $\ell_2 + \ell_3$ is independent of $\{\ell_2 > \ell_3\}$ (remember ℓ_2 and ℓ_3 may have different distributions). This comes from the following classical property (see [9, Section 4.11]):

Fact 1 *if X and Y are independent random variables with respective distributions $\Gamma(\alpha, 1)$ and $\Gamma(\beta, 1)$, then $X + Y$ is independent of $\left(\frac{X}{X+Y}, \frac{Y}{X+Y}\right)$, hence in particular of $\{X < Y\}$, and has distribution $\Gamma(\alpha + \beta, 1)$.*

Let us now assume $n \geq 4$ and that the result holds for strictly fewer particles. Let $\alpha_1, \ldots, \alpha_n > 0$ be given. We consider n particles, and assume the interdistances $\ell_1, \ldots, \ell_n$ to have respective distributions $\Gamma(\alpha_1, 1), \ldots, \Gamma(\alpha_n, 1)$. Let a configuration (v, π) be given.

Case 1 First consider the case when, in the configuration, there are indices $k < l$, different from $(1, n)$, such that $\vec{\bullet}_k \to\leftarrow \overleftarrow{\bullet}_l$. Then, thanks to the induction applied to the strict subinterval $[\![k, l]\!]$, conditional on the configuration (v, π), the distance $x_l - x_k$ has same distribution as unconditionally (i.e. a gamma distribution, as a sum of independent gamma variables), and in particular same distribution as with particles $\bullet_k, \ldots, \bullet_l$ removed. Since such a pyramid shaped subconfiguration is independent of the configuration outside this subinterval (indeed no collision with particles outside $[\![k, l]\!]$ is possible), we further conclude that the total width x_n, conditioned on (v, π), has same distribution as with particles $\bullet_k, \ldots, \bullet_l$ removed (including, from configuration π). This reduces to a strictly smaller number of particles, enabling to use again the induction to conclude.

It remains to consider the cases when either $\vec{\bullet}_1 \to\leftarrow \overleftarrow{\bullet}_n$ or all collisions are of the type $\vec{\bullet} \to \dot{\bullet}$ or symmetrically.

Case 2 Assume that $\vec{\bullet}_1 \to\leftarrow \overleftarrow{\bullet}_n$. We further consider two subcases.

Case 2a If the configuration contains indices k, l with $l > k + 2$ and $\dot{\bullet}_k \leftarrow \overleftarrow{\bullet}_l$ or $\vec{\bullet}_k \to \dot{\bullet}_l$, we have by the induction applied to the interval $[\![k, l]\!]$ that $x_l - x_k$ is gamma distributed given the configuration. Since the configuration outside $[\![k+1, l-1]\!]$ is independent of the configuration in $[\![k+1, l-1]\!]$, given $\dot{\bullet}_k \leftarrow \overleftarrow{\bullet}_l$ (or symmetrically), we conclude that the total width x_n is distributed as with $\bullet_{k+1}, \ldots, \bullet_{l-1}$ removed. This enables to use the induction and conclude in this subcase.

Case 2b Otherwise, the pairing in $[\![2, n-1]\!]$ must be between neighbors:

$$\pi = \big(n\ 3\ 2\ 5\ 4 \cdots n-1\ n-2\ 1\big), \tag{6}$$

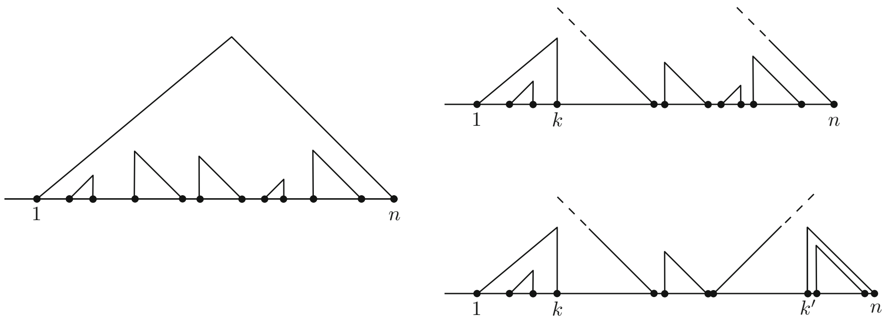


Fig. 5 Pairing π from (6) (left), and general form of the other pairings on the same velocities, up to left-right symmetry (right)

and each pair $(2k, 2k+1)$ has either velocities $(+1, 0)$ or $(0, -1)$, for $2 \leq 2k \leq n-2$. It suffices to show that the law of x_n given any other pairing, and given these same velocities, is unconditioned. Summing over all pairings (multiplied by their probabilities) indeed reduces to the law of x_n given the velocities, which is nothing but the law of x_n since the two are independent of each other. The only possible pairings compatible with these velocities, besides the previous one (6), are of the following type (if any): either for some even index $2 \leq k \leq n-2$, such that $v_k = 0$, $\vec{\bullet}_1 \to \dot{\bullet}_k$, while $\overleftarrow{\bullet}_{k+1}$ does not collide, or symmetrically $\dot{\bullet}_{k'} \leftarrow \overleftarrow{\bullet}_n$ while $\vec{\bullet}_{k'-1}$ does not collide, for some k' such that $v_{k'} = 0$, or both happen, while other neighboring pairs are preserved (see also Fig. 5). We notice that the realization of this configuration on $[\![1, k+1]\!]$ and on $[\![k+2, n]\!]$ (or symmetrically with k') are independent, so that we can apply induction on each of these strict subintervals to show that their width are unaffected by conditioning on the subconfiguration. This concludes this subcase.

Case 3 Finally, let us treat the case of configurations without any collision of the type $\vec{\bullet} \to\leftarrow \overleftarrow{\bullet}$. Similarly to Case 2a, we may apply induction to any configuration that has “nested collisions”, i.e. $\bullet_i \sim \bullet_j$ for some $j \geq i+2$. We may therefore assume that collisions are between neighbors. Some particles may also not collide at all. However, if some particle $\bullet_i$ with $1 < i < n$ does not collide, then the conditions on the configuration on the left and on the right of this particle (including the particle with the side where it is heading to if $v_i = \pm 1$, and with neither if $v_i = 0$) are independent, enabling to use induction as in the end of Case 2b. Also, if $\bullet_1$ or $\bullet_n$ is surviving with velocity 0, or with velocity -1 or $+1$ respectively, then the condition only leans on the other particles, enabling induction again. All in all, either all particles collide, in which case the pairing is necessarily between neighbors hence doesn’t correlate with x_n, or only $\vec{\bullet}_1$ or $\overleftarrow{\bullet}_n$ survives. This last subcase is dealt with exactly as in Case 2b, namely by treating the case of any other pairing on the same velocities, which describes as in Case 2b and brings up conditions that split into independent conditions on subconfigurations, enabling to use induction and finally conclude.

6 Variation of A with Respect to p

In this section we consider, in the setting of independent random lengths, how the (universal) distribution of A (i.e. the index of the leftmost particle that crosses 0) varies with the density p of static particles. Note that A is not monotonic under individual changes to the velocities of particles, and that merely reversing the direction of a single right-moving particle can even alter A from finite to infinite. However, we conjecture that the law of A is affected monotonically by changing p. As in Proposition 1, let us denote, for $n \in \mathbb{N}$ and implicitly $p \in [0, 1]$,

$$p_n = \mathbb{P}(A = n) = \mathbb{P}(0 \xleftarrow{1} \overleftarrow{\bullet}_n).$$

We give three conjectures supported by computer-assisted computations for small values of n. The first one states that in the supercritical region, each individual probability corresponding to a finite value of A is decreasing in p:

Conjecture 1 For each n, the function $p \mapsto p_n$ is monotonically decreasing on the interval $[1/4, 1]$.

This conjecture cannot be extended beyond this region: since $p_1 = (1 - p)/2$ is strictly decreasing on $[0, 1]$, $\mathbb{P}(A = \infty)$ must be strictly increasing wherever the conjectured result holds; however it is constant, equal to 0, on $[0, 1/4]$.

We also conjecture that we have stochastic dominance between the laws of A for any two values of p, even in the subcritical region:

Conjecture 2 For each n, the function $p \mapsto \mathbb{P}(A \le n)$ is monotonically decreasing on the interval $[0, 1]$.

Finally, consider $\mathbb{P}(A = n \mid \overleftarrow{\bullet}_n) = \frac{2}{1-p} p_n$, which may equivalently be thought of as the probability that the first $n - 1$ particles all annihilate one another in such a way that none of them would be in the path of a left-moving particle starting at x_n. We conjecture that this probability peaks at the same value $1/4$ for any n (note that this critical value only appeared in the context of an infinite system so far):

Conjecture 3 For each n, the function $p \mapsto \frac{2}{1-p} p_n$ is maximized at $p = 1/4$.

We now give partial results to support these conjectures. First, Conjecture 1 holds for simple reasons on a restricted range of values of p.

Proposition 2 *For each n, the function $p \mapsto p_n$ is monotonically decreasing on the interval* $[\frac{1}{2} - \frac{1}{2n}, 1]$.

Proof Fix a particular law of interdistances m; recall that this does not affect p_n. The event $\{A = n\}$ may only occur when at most $\frac{n-1}{2}$ particles among the first n are static, hence

$$p_n = \sum_{0 \le k \le \frac{n-1}{2}} \sum_{\mathbf{w} \in \mathcal{V}_k} \mathbb{P}(A = n \mid (v_1, \dots, v_n) = \mathbf{w}) p^k (1 - p)^{n-k} 2^{k-n},$$

where $\mathcal{V}_k$ denotes the set of velocities of the first n particles among which exactly k are 0. For fixed k, $p^k(1-p)^{n-k}$ is monotonically decreasing on $[k/n, 1]$, and so every term in the above sum is monotonically decreasing on the required interval.

We also observe that the function p_n is decreasing around the critical value 1/4.

Proposition 3 *For each n, at* $p = 1/4$ *we have* $p_n' = -\frac{4}{3}p_n$*, where* $p_n' = \frac{\mathrm{d}p_n}{\mathrm{d}p}$

Proof Recall from (2) that $p_1 = (1-p)/2$ and, for all $n \geq 2$,

$$p_n = \left(p + \frac{1}{2}\right) \sum_{\substack{k_1+k_2\\=n-1}} p_{k_1} p_{k_2} - \frac{p}{2} \sum_{\substack{k_1+k_2+k_3+k_4\\=n-1}} p_{k_1} p_{k_2} p_{k_3} p_{k_4}.$$

We prove the claimed statement by induction on n; it is easy to verify for $n = 1$. Suppose it is true for all values less than n. Note that

$$\begin{aligned} p_n' = &\sum_{\substack{k_1+k_2\\=n-1}} \left(p_{k_1} p_{k_2} + \left(p + \frac{1}{2}\right)(p_{k_1}' p_{k_2} + p_{k_1} p_{k_2}') \right) \\ &- \sum_{\substack{k_1+k_2+k_3+k_4\\=n-1}} \left(\frac{1}{2} p_{k_1} p_{k_2} p_{k_3} p_{k_4} + \frac{p}{2}(p_{k_1}' p_{k_2} p_{k_3} p_{k_4} + \cdots + p_{k_1} p_{k_2} p_{k_3} p_{k_4}') \right). \end{aligned}$$

Evaluating at $p = 1/4$, assuming the induction hypothesis, gives

$$\begin{aligned} p_n' &= \sum_{\substack{k_1+k_2\\=n-1}} \left(1 - 2 \times \frac{3}{4} \times \frac{4}{3}\right) p_{k_1} p_{k_2} - \sum_{\substack{k_1+k_2+k_3+k_4\\=n-1}} \left(\frac{1}{2} - 4 \times \frac{1}{8} \times \frac{4}{3}\right) p_{k_1} p_{k_2} p_{k_3} p_{k_4} \\ &= -\frac{4}{3}\left(\frac{3}{4} \sum_{\substack{k_1+k_2\\=n-1}} p_{k_1} p_{k_2} - \frac{1}{8} \sum_{\substack{k_1+k_2+k_3+k_4\\=n-1}} p_{k_1} p_{k_2} p_{k_3} p_{k_4}\right) = -\frac{4}{3} p_n, \end{aligned}$$

as required.

This explicit logarithmic derivative in fact also gives support to Conjecture 3, since it equivalently states that the derivative in p of $\frac{2}{1-p}p_n$ is 0 at $p = 1/4$.

Finally, we can give some additional support to Conjecture 2 by exactly evaluating the (right-hand) derivative of p_n at 0. We may assume $n = 2m + 1$ is odd and at least 3, since for n even $p_n \equiv 0$ and $p_1 = (1-p)/2$. We prove the following.

Theorem 3 *The right-hand derivative of* p_{2m+1} *at* p=0 *is* $\frac{8m-5}{(m+1)(2m+4)}\binom{2m}{m}2^{-2m-1}$.

Since $\mathbb{P}(A \leq n) = \mathbb{P}(A = \infty) - \sum_{k>n} p_k$, and $\mathbb{P}(A = \infty)$ is constant on the subcritical region, we immediately obtain the following consequence.

Corollary 3 *The right-hand derivative of* $\mathbb{P}(A \leq n)$ *at* 0 *is negative for every* $n \geq 1$.

Proof of Theorem 3 Let $n = 2m + 1$. Let S denote the number of static particles among the first n, and observe that the law of $(v_1, \ldots, v_n)$ given S does not depend on p. We have, as $p \to 0$,

$$p_n = \mathbb{P}(A = n) = \sum_{k=0}^{n} \mathbb{P}(A = n \mid S = k)\binom{n}{k} p^k (1-p)^{n-k}$$
$$= \mathbb{P}(A = n \mid S = 0)(1 - np) + \mathbb{P}(A = n \mid S = 1)np + O(p^2),$$

and the probabilities on the right-hand side do not depend on p by the previous remark, hence the derivative of p_n at 0 is given by $n(\mathbb{P}(A = n \mid S = 1) - \mathbb{P}(A = n \mid S = 0))$.

First, we condition on $S = 0$. In this case, the event $\{A = n\}$ means that $v_n = -1$, which has conditional probability $1/2$, and that the first $2m$ particles mutually annihilate. An arrangement of $2m$ particles which mutually annihilate corresponds precisely to an expression of $2m$ correctly-matched parentheses, and so the number of such arrangements is equal to C_m, the mth Catalan number, which is given by $C_m = \frac{1}{m+1}\binom{2m}{m}$. Thus,

$$\mathbb{P}(A = n \mid S = 0) = \frac{1}{2} \cdot \frac{C_m}{2^{2m}}.$$

Next, we turn to the case $S = 1$. For simplicity we consider the case of constant interdistances (with triple collisions resolved at random); by universality, this is sufficient. In this case, $A = n$ means for the last particle to be left-moving (which occurs with conditional probability $\frac{m}{2m+1}$) and for the remaining $2m$ particles, of which one is static, to mutually annihilate in such a way that they do not interfere with the last particle; this includes cases where the last particle survives a triple collision. Thus,

$$\mathbb{P}(A = n \mid S = 1) = \frac{m}{2m + 1} \triangle_m,$$

where we write $\triangle_m$ for the number of such arrangements of $2m$ velocities: we may indeed think of it as a requirement for $2m$ particles (one of which is static) to annihilate and have space-time trajectories contained inside the triangle described by the trajectories of a static particle at 0 and a left-moving particle at $2m + 1$—here we count a collision happening exactly on the right-hand side of this triangle as "inside" only if the spin of the static particle is -1. Note that $\triangle_m$ could be a half integer as some arrangements require a particular spin hence count $1/2$.

By symmetry, $\triangle_m = \triangle_m$, where the latter is the number of arrangements which mutually annihilate inside the reflection of the previous triangle. We consider two other similar quantities: write $\square_m$ for the number of arrangements of $2m$ velocities, of which one is static, which mutually annihilate, and $\triangle_m$ for the number of such arrangements which mutually annihilate inside the space-time triangle described by the trajectories of a right-moving particle at 0 and a left-moving particle at $2m+1$. If a configuration is counted in $\square_m$ but not in $\triangle_m$, then it counts inside exactly one of $⊿_m$ or $◺_m$, otherwise it is in both, thus $⊿_m + ◺_m = (\square_m - \triangle_m) + 2\triangle_m$, i.e. $\square_m + \triangle_m = 2◺_m$.

We claim that $\triangle_m = mC_m$. This is equivalent to the claim that if we take a random set of $2m+2$ moving particles, conditioned on the first and last colliding (this leaves C_m uniform choices), and make a random internal particle static ($2m$ choices), then with probability $1/2$ the first and last still collide together.

We prove this by induction on m. The case $m = 1$ is straightforward; consider $m \geq 2$ and assume the property true in the previous cases. Suppose the particle chosen to become static is not in the "skyline" of the $2m$ internal particles, i.e. it is between two colliding particles other than the first and last. Then the probability that these two particles still collide is $1/2$ by induction. If they do, the outer particles are unaffected, but if not then one of them is released to collide with an outer particle. Thus it suffices to prove the claim for a particle chosen in the skyline and, by symmetry, we may assume this particle is right-moving.

Consider the Dyck paths corresponding to configurations of internal particles, with a step x from 0 to 1 (i.e. corresponding to a right-moving particle in the skyline) marked. Let y be the next step from 1 to 0, $\mathbf{a}$ be the subpath before the marked step, and $\mathbf{b}$ be the subpath of steps strictly between x and y. Making the particle corresponding to x static will cause a collision with one of the external particles if and only if $|\mathbf{a}| + 1 < |\mathbf{b}| + 1$ (or with probability $1/2$ if they are equal), since these are the distances to the two particles which could collide with x. Swapping the subpaths $\mathbf{a}$ and $\mathbf{b}$ gives another Dyck path, so this bijective transformation keeps the particle corresponding to x in the skyline and right-moving; and it maps any configuration where x would be colliding with an external particle if made static, to one where its corresponding particle would not, and vice-versa, so this proves the claim.

We apply a similar argument to calculate $\square_m$: starting from a totally annihilating configuration of $2m$ moving particles, a random one is made static. By the previous claim, if this particle is not in the skyline, the change has chance $1/2$ of preserving total annihilation. However, making a particle in the skyline static always preserves total annihilation. Thus if a random configuration of $2m$ moving particles which mutually annihilate is modified by making a random particle x static, the probability that all particles still annihilate is $\frac{1}{2}\mathbb{P}(x \text{ not in skyline}) + \mathbb{P}(x \text{ in skyline}) = \frac{1}{2} + \frac{1}{2}\mathbb{E}[W]$, where W is the number of particles among $[\![1, 2m]\!]$ that are in the skyline.

Note that $W = 2V - 2$, where V is the number of visits to 0 by the Dyck path (including the start and end of the path). The number of arrangements which visit 0 after $2k$ steps is $C_k C_{m-k}$ for each $k \in [\![0, m]\!]$, and so

$$\mathbb{E}[V] = \sum_{k=0}^{m} \frac{C_k C_{m-k}}{C_n} = \frac{C_{m+1}}{C_m} = \frac{4m+2}{m+2};$$

it follows that $\mathbb{E}[W] = \frac{6m}{m+2}$, giving finally

$$\square_m = 2mC_m\left(\frac{1}{2} + \frac{1}{2} \cdot \frac{3}{m(m+2)}\right) = \frac{m(m+5)}{m+2}C_m.$$

Consequently

$$\triangleright_m = \frac{\square_m + \triangle_m}{2} = \frac{m(2m+7)}{2m+4},$$

hence

$$\mathbb{P}(A = 2m+1 \mid S = 1) = \frac{2m(2m+7)}{(2m+1)(2m+4)}C_m 2^{-2m}.$$

The result follows by gathering the previous computations.

Acknowledgments J.H. was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 639046). L.T. was supported by the French ANR project MALIN (ANR-16-CE93-000).

References

1. Arratia, R.: Site recurrence for annihilating random walks on $\mathbf{Z}_d$. Ann. Probab. **11**(3), 706–713 (1983)
2. Belitsky, V., Ferrari, P.A.: Ballistic annihilation and deterministic surface growth. J. Stat. Phys. **80**(3–4), 517–543 (1995)
3. Ben-Naim, E., Redner, S., Leyvraz, F.: Decay kinetics of ballistic annihilation. Phys. Rev. Lett. **70**(12), 1890–1893 (1993)
4. Bramson, M., Lebowitz, J. L.: Asymptotic behavior of densities for two-particle annihilating random walks. J. Stat. Phys. **62**(1–2), 297–372 (1991)
5. Broutin, N., Marckert, J.-F.: The combinatorics of the colliding bullets. Random Struct. Algoritm. **56**(2), 401–431 (2020)
6. Droz, M., Rey, P.-A., Frachebourg, L., Piasecki, J.: Ballistic-annihilation kinetics for a multivelocity one-dimensional ideal gas. Phys. Rev. E **516**, 5541–5548 (1995)
7. Dygert, B., Kinzel, C., Zhu, J., Junge, M., Raymond, A., Slivken, E.: The bullet problem with discrete speeds. Electron. Commun. Probab. **24**(27), 11 (2019)
8. Elskens, Y., Frisch, H.L.: Annihilation kinetics in the one-dimensional ideal gas. Phys. Rev. A **31**(6), 3812–3816 (1985)
9. Grimmett, G.R., Stirzaker, D.R.: Probability and Random Processes. Oxford University Press, Oxford (1992)
10. Haslegrave, J., Sidoravicius, V., Tournier, L.: Three-speed ballistic annihilation: phase transition and universality (2018). arXiv preprint arXiv:1811.08709
11. Junge, M., Lyu, H.: The phase structure of asymmetric ballistic annihilation (2018). arXiv preprint arXiv:1811.08378
12. Kleber, M., Wilson, D.: "Ponder This" IBM research challenge (2014). https://www.research.ibm.com/haifa/ponderthis/challenges/May2014.html

13. Krapivsky, P.L., Redner, S., Leyvraz, F.: Ballistic annihilation kinetics: the case of discrete velocity distributions. Phys. Rev. E **51**(5), 3977–3987 (1995)
14. Sidoravicius, V., Tournier, L.: Note on a one-dimensional system of annihilating particles. Electron. Commun. Probab. **22**(59), 9 (2017)

Glauber Dynamics on the Erdős-Rényi Random Graph

F. den Hollander and O. Jovanovski

Abstract We investigate the effect of disorder on the Curie-Weiss model with Glauber dynamics. In particular, we study metastability for spin-flip dynamics on the Erdős-Rényi random graph $\mathrm{ER}_n(p)$ with n vertices and with edge retention probability $p \in (0, 1)$. Each vertex carries an Ising spin that can take the values -1 or $+1$. Single spins interact with an external magnetic field $h \in (0, \infty)$, while pairs of spins at vertices connected by an edge interact with each other with ferromagnetic interaction strength $1/n$. Spins flip according to a Metropolis dynamics at inverse temperature β. The standard Curie-Weiss model corresponds to the case $p = 1$, because $\mathrm{ER}_n(1) = K_n$ is the complete graph on n vertices. For $\beta > \beta_c$ and $h \in (0, p\chi(\beta p))$ the system exhibits *metastable behaviour* in the limit as $n \to \infty$, where $\beta_c = 1/p$ is the *critical inverse temperature* and χ is a certain *threshold function* satisfying $\lim_{\lambda\to\infty} \chi(\lambda) = 1$ and $\lim_{\lambda\downarrow 1} \chi(\lambda) = 0$. We compute the average crossover time from the *metastable set* (with magnetization corresponding to the 'minus-phase') to the *stable set* (with magnetization corresponding to the 'plus-phase'). We show that the average crossover time grows exponentially fast with n, with an exponent that is the same as for the Curie-Weiss model with external magnetic field h and with ferromagnetic interaction strength p/n. We show that the correction term to the exponential asymptotics is a multiplicative error term that is *at most polynomial* in n. For the complete graph K_n the correction term is known to be a multiplicative constant. Thus, apparently, $\mathrm{ER}_n(p)$ is so homogeneous for large n that the effect of the fluctuations in the disorder is small, in the sense that the metastable behaviour is controlled by the average of the disorder. Our model is

The research in this paper was supported by NWO Gravitation Grant 024.002.003-NETWORKS.

F. den Hollander (✉)
Mathematical Institute, Leiden University, Leiden, RA, The Netherlands
e-mail: denholla@math.leidenuniv.nl

O. Jovanovski
Manulife, Toronto, ON, Canada
e-mail: oliver_jovanovski@manulife.com

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_24

the first example of a metastable dynamics on a random graph where the correction term is estimated to high precision.

Keywords Erdős-Rényi random graph · Glauber spin-flip dynamics · Metastability · Crossover time

1 Introduction and Main Results

In Sect. 1.1 we provide some background on metastability. In Sect. 1.2 we define our model: spin-flip dynamics on the Erdős-Rényi random graph $\mathrm{ER}_n(p)$. In Sect. 1.3 we identify the metastable pair for the dynamics, corresponding to the 'minus-phase' and the 'plus-phase', respectively. In Sect. 1.4 we recall the definition of spin-flip dynamics on the complete graph K_n, which serves as a comparison object, and recall what is known about the average metastable crossover time for spin-flip dynamics on K_n (Theorem 1.3 below). In Sect. 1.5 we transfer the sharp asymptotics for K_n to a somewhat rougher asymptotics for $\mathrm{ER}_n(p)$ (Theorem 1.4 below). In Sect. 1.6 we close by placing our results in the proper context and giving an outline of the rest of the paper.

1.1 Background

Interacting particle systems, evolving according to a *Metropolis dynamics* associated with an energy functional called the *Hamiltonian*, may end up being trapped for a long time near a state that is a local minimum but not a global minimum. The deepest local minima are called *metastable states*, the global minimum is called the *stable state*. The transition from a metastable state to the stable state marks the relaxation of the system to equilibrium. To describe this relaxation, it is of interest to compute the crossover time and to identify the set of critical configurations the system has to visit in order to achieve the transition. The critical configurations represent the saddle points in the free energy landscape: the set of mini-max configurations that must be hit by any path that achieves the crossover.

Metastability for interacting particle systems on *lattices* has been studied intensively in the past three decades. Various different approaches have been proposed, which are summarised in the monographs by Olivieri and Vares [12], Bovier and den Hollander [4]. Recently, there has been interest in metastability for interacting particle systems on *random graphs*, which is much more challenging because the crossover time typically depends in a delicate manner on the realisation of the graph.

In the present paper we are interested in metastability for spin-flip dynamics on the *Erdős-Rényi random graph*. Our main result is an estimate of the average crossover time from the 'minus-phase' to the 'plus-phase' when the spins feel an external magnetic field at the vertices in the graph as well as a ferromagnetic interaction along the edges in the graph. Our paper is part of a larger enterprise in which the goal is to understand metastability on graphs. Jovanovski [11] analysed the case of the *hypercube*, Dommers [7] the case of the *random regular graph*, Dommers, den Hollander, Jovanovski and Nardi [10] the case of the *configuration model*, and den Hollander and Jovanovski [6] the case of the *hierarchical lattice*. Each case requires carrying out a detailed combinatorial analysis that is model-specific, even though the metastable behaviour is ultimately universal. For lattices like the hypercube and the hierarchical lattice a full identification of the relevant quantities is possible, while for random graphs like the random regular graph and the configuration model so far only the communication height is well understood, while the set of critical configurations and the prefactor remain somewhat elusive.

The equilibrium behaviour of the Ising model on random graphs is well understood. See e.g. Dommers et al. [8, 9].

1.2 Spin-Flip Dynamics on $\mathrm{ER}_n(p)$

Let $\mathrm{ER}_n(p) = (V, E)$ be a realisation of the Erdős-Rényi random graph on $|V| = n \in \mathbb{N}$ vertices with edge retention probability $p \in (0, 1)$, i.e., each edge in the complete graph K_n is present with probability p and absent with probability $1 - p$, independently of other edges (see Fig. 1). We write $\mathbb{P}_{\mathrm{ER}_n(p)}$ to denote the law of $\mathrm{ER}_n(p)$. For typical properties of $\mathrm{ER}_n(p)$, see van der Hofstad [13, Chapters 4–5].

Each vertex carries an Ising spin that can take the values -1 or $+1$. Let $S_n = \{-1, +1\}^V$ denote the set of spin configurations on V, and let H_n be the *Hamiltonian* on S_n defined by

$$H_n(\sigma) = -\frac{1}{n} \sum_{(v,w)\in E} \sigma(v)\sigma(w) - h \sum_{v\in V} \sigma(v), \quad \sigma \in S_n. \tag{1.1}$$

In other words, single spins interact with an *external magnetic field* $h \in (0, \infty)$, while pairs of neighbouring spins interact with each other with a *ferromagnetic coupling strength* $1/n$.

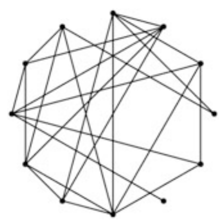


Fig. 1 A realization of the Erdős-Rényi random graph with $n = 12$ and $p = \frac{1}{3}$

Let $\boxminus = \{-1\}^V$ and $\boxplus = \{+1\}^V$ denote the configuration where all spins are -1, respectively, $+1$. Since

$$H_n(\boxminus) = -\frac{|E|}{n} + hn, \tag{1.2}$$

we have the geometric representation

$$H_n(\sigma) = H_n(\boxminus) + \frac{2}{n}|\partial_E \sigma| - 2h|\sigma|, \quad \sigma \in S_n, \tag{1.3}$$

where

$$\partial_E \sigma = \{(v, w) \in E \colon \sigma(v) = -\sigma(w) = +1\} \tag{1.4}$$

is the *edge-boundary* of σ and

$$|\sigma| = \{v \in \mathrm{ER}_n(p) \colon \sigma(v) = +1\} \tag{1.5}$$

is the *vertex-volume* of σ.

In the present paper we consider a spin-flip dynamics on S_n commonly referred to as *Glauber dynamics*, defined as the continuous-time Markov process with transition rates

$$r(\sigma, \sigma') = \begin{cases} \mathrm{e}^{-\beta[H_n(\sigma') - H_n(\sigma)]_+}, & \text{if } \|\sigma - \sigma'\| = 2, \\ 0, & \text{if } \|\sigma - \sigma'\| > 2, \end{cases} \quad \sigma, \sigma' \in S_n, \tag{1.6}$$

where $\|\cdot\|$ is the ℓ_1-norm on S_n. This dynamics has as *reversible* stationary distribution the Gibbs measure

$$\mu_n(\sigma) = \frac{1}{Z_n}\mathrm{e}^{-\beta H_n(\sigma)}, \quad \sigma \in S_n, \tag{1.7}$$

where $\beta \in (0, \infty)$ is the *inverse temperature* and Z_n is the normalizing partition sum. We write

$$\{\xi_t\}_{t \geq 0} \tag{1.8}$$

to denote the path of the random dynamics and $\mathbb{P}_\xi$ to denote its law given $\xi_0 = \xi$. For $\chi \subset S_n$, we write

$$\tau_\chi = \inf\{t \geq 0 \colon \xi_t \in \chi, \ \xi_{t-} \notin \chi\}. \tag{1.9}$$

to denote the *first hitting/return time* of χ.

We define the *magnetization* of σ by

$$m(\sigma) = \frac{1}{n} \sum_{v \in V} \sigma(v), \tag{1.10}$$

and observe the relation

$$m(\sigma) = \frac{2|\sigma|}{n} - 1, \quad \sigma \in S_n. \tag{1.11}$$

We will frequently switch between working with volume and working with magnetization. Equation (1.11) ensures that these are in one-to-one correspondence. Accordingly, we will frequently look at the dynamics from the perspective of the induced *volume process* and *magnetization process*,

$$\{|\xi_t|\}_{t \geq 0}, \quad \{m(\xi_t)\}_{t \geq 0}, \tag{1.12}$$

which are *not* Markov.

1.3 Metastable Pair

For fixed n, the Hamiltonian in (1.1) achieves a *global minimum* at $\boxplus$ and a *local minimum* at $\boxminus$. In fact, $\boxminus$ is the deepest local minimum not equal to $\boxplus$ (at least for h small enough). However, in the limit as $n \to \infty$, these do *not* form a metastable pair of configurations because *entropy* comes into play.

Definition 1.1 (Metastable Regime) The parameters β, h are said to be in the *metastable regime* when

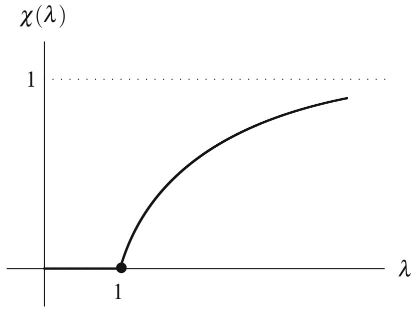
$$\beta \in (1/p, \infty), \qquad h \in \big(0, p\chi(\beta p)\big), \tag{1.13}$$

with (see Fig. 2)

$$\chi(\lambda) = \sqrt{1 - \frac{1}{\lambda}} - \frac{1}{2\lambda} \log\left[\lambda\left(1 + \sqrt{1 - \frac{1}{\lambda}}\right)^2\right], \qquad \lambda \geq 1. \tag{1.14}$$

We have $\lim_{\lambda \to \infty} \chi(\lambda) = 1$ and $\lim_{\lambda \downarrow 1} \chi(\lambda) = 0$ (with slope $\frac{1}{2}$). Hence, for $\beta \to \infty$ any $h \in (0, p)$ is metastable, while for $\beta \downarrow 1/p$ or $p \downarrow 0$ no $h \in (0, \infty)$ is metastable. The latter explains why we do not consider the non-dense Erdős-Rényi random graph with $p = p_n \downarrow 0$ as $n \to \infty$. ■

Fig. 2 Plot of $\lambda \mapsto \chi(\lambda)$



The threshold $\beta_c = 1/p$ is the *critical temperature*: the static model has a phase transition at $h = 0$ when $\beta > \beta_c$ and no phase transition when $\beta \leq \beta_c$ (see e.g. Dommers et al. [9]).

To define the proper metastable pair of configurations, we need the following definitions. Let

$$\begin{aligned} &\Gamma_n = \{-1, -1 + \tfrac{2}{n}, \dots, 1 - \tfrac{2}{n}, 1\}, \\ &I_n(a) = -\tfrac{1}{n} \log \binom{n}{\frac{1+a}{2}n}, \quad J_n(a) = 2\beta(pa + h) - 2I_n'(a). \end{aligned} \tag{1.15}$$

Define

$$\begin{aligned} \mathbf{m}_n &= \min\{a \in \Gamma_n \colon\; J_n(a) \leq 0\}, \\ \mathbf{t}_n &= \min\{a \in \Gamma_n \colon\; a > \mathbf{m}_n,\; J_n(a) \geq 0\}, \\ \mathbf{s}_n &= \min\{a \in \Gamma_n \colon\; a > \mathbf{t}_n,\; J_n(a) \leq 0\}. \end{aligned} \tag{1.16}$$

The numbers in the left-hand side of (1.16) play the role of magnetizations. Further define

$$\mathbf{M}_n = \frac{n}{2}(\mathbf{m}_n + 1), \quad \mathbf{T}_n = \frac{n}{2}(\mathbf{t}_n + 1), \quad \mathbf{S}_n = \frac{n}{2}(\mathbf{s}_n + 1), \tag{1.17}$$

which are the volumes corresponding to (1.16), and

$$A_k = \{\sigma \in S_n \colon\; |\sigma| = k\}, \qquad k \in \{0, 1, \dots, n-1, n\}, \tag{1.18}$$

the set of configurations with volume k. Define

$$R_n(a) = -\frac{1}{2}pa^2 - ha + \frac{1}{\beta} I_n(a) \tag{1.19}$$

and note that

$$R_n'(a) = -pa - h + \frac{1}{\beta} I_n'(a) = -\frac{1}{2\beta} J_n(a). \tag{1.20}$$

The motivation behind the definitions in (1.15), (1.16) and (1.19) will become clear in Sect. 2. Via Stirling's formula it follows that

$$J_n(a) = 2\beta(pa+h) + \log\left(\frac{1-a+\frac{1}{n}}{1+a+\frac{1}{n}}\right) + O(n^{-2}), \quad a \in \Gamma_n. \tag{1.21}$$

We will see that, in the limit as $n \to \infty$ when (β, h) is in the metastable regime defined by (1.13), the numbers in (1.16) are well-defined: $A_{\mathbf{M}_n}$ is the *metastable set*, $A_{\mathbf{S}_n}$ is the *stable set*, $A_{\mathbf{T}_n}$ is the *top set*, i.e., the set of saddle points that lie in between $A_{\mathbf{M}_n}$ and $A_{\mathbf{S}_n}$. Our key object of interest will be the *crossover time* from $A_{\mathbf{M}_n}$ to $A_{\mathbf{S}_n}$ via $A_{\mathbf{T}_n}$.

Note that

$$\Gamma_n \to [-1,1], \quad I_n(a) \to I(a), \quad J_n(a) \to J_{p,\beta,h}(a), \qquad n \to \infty, \tag{1.22}$$

with

$$J_{p,\beta,h}(a) = 2\beta(pa+h) + \log\left(\frac{1-a}{1+a}\right) \tag{1.23}$$

and

$$I(a) = \frac{1-a}{2}\log\left(\frac{1-a}{2}\right) + \frac{1+a}{2}\log\left(\frac{1+a}{2}\right). \tag{1.24}$$

Accordingly,

$$\mathbf{m}_n \to \mathbf{m}, \quad \mathbf{t}_n \to \mathbf{t}, \quad \mathbf{s}_n \to \mathbf{s}, \qquad n \to \infty, \tag{1.25}$$

with $\mathbf{m}$, $\mathbf{t}$, $\mathbf{s}$ the three successive zeroes of J (see Fig. 4 and recall (1.16)). Define

$$R_{p,\beta,h}(a) = -\frac{1}{2}pa^2 - ha + \frac{1}{\beta}I(a). \tag{1.26}$$

Note that $R_{p,\beta,h}(a)$ plays the role of *free energy*: $-\frac{1}{2}pa^2 - ha$ and $\frac{1}{\beta}I(a)$ represent the energy, respectively, entropy at magnetisation a. Note that $I(a)$ equals the relative entropy of the probability measure $\frac{1}{2}(1+a)\delta_{+1} + \frac{1}{2}(1-a)\delta_{-1}$ with respect to the counting measure $\delta_{+1} + \delta_{-1}$. Also note that

$$R'_{p,\beta,h}(a) = -pa - h + \frac{1}{\beta}I'(a) = -\frac{1}{2\beta}J_{p,\beta,h}(a). \tag{1.27}$$

Remark 1.2 As shown in Corollary 3.6 below, if $h \in (p, \infty)$, then (1.6) leads to *non-metastable* behaviour where the dynamics 'drifts' through a sequence of configurations with volume growing from **M** to **S** within time $O(1)$. ■

1.4 Metastability on K_n

Let K_n be the complete graph on n vertices (see Fig. 3). Spin-flip dynamics on K_n, commonly referred to as *Glauber dynamics for the Curie-Weiss model*, is defined as in Sect. 1.2 but with the *Curie-Weiss Hamiltonian*

$$H_n(\sigma) = -\frac{1}{2n} \sum_{1 \le i,j \le n} \sigma(i)\sigma(j) - h \sum_{1 \le i \le n} \sigma(i), \quad \sigma \in S_n. \tag{1.28}$$

This is the same as (1.1) with $p = 1$, except for the diagonal term $-\frac{1}{2n} \sum_{1 \le i \le n} \sigma(i)\sigma(i) = -\frac{1}{2}$, which shifts H_n by a constant and has no effect on the dynamics. The advantage of (1.28) is that we may write

$$H_n(\sigma) = n \left[-\frac{1}{2} m(\sigma)^2 - hm(\sigma) \right], \tag{1.29}$$

which shows that the energy is a function of the magnetization only, i.e., the Curie-Weiss model is a *mean-field* model. Clearly, this property fails on $\mathrm{ER}_n(p)$.

For the Curie-Weiss model it is known that there is a *critical inverse temperature* $\beta_c = 1$ such that, for $\beta > \beta_c$, h small enough and in the limit as $n \to \infty$, the stationary distribution μ_n given by (1.7) and (1.28) has two phases: the 'minus-phase', where the majority of the spins are -1, and the 'plus-phase', where the majority of the spins are $+1$. These two phases are the *metastable state*, respectively, the *stable state* for the dynamics. In the limit as $n \to \infty$, the dynamics of the magnetization introduced in (1.12) (which is Markov) converges to a Brownian motion on $[-1, +1]$ in the *double-well potential* $a \mapsto R_{1,\beta,h}(a)$ (see Fig. 4).

The following theorem can be found in Bovier and den Hollander [4, Chapter 13]. For $p = 1$, the metastable regime in (1.13) becomes

$$\beta \in (1, \infty), \qquad h \in \big(0, \chi(\beta)\big). \tag{1.30}$$

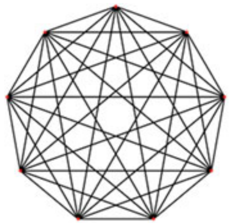


Fig. 3 The complete graph with $n = 9$

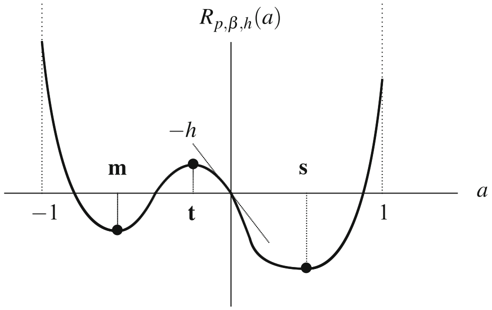


Fig. 4 Plot of $R_{p,\beta,h}(a)$ as a function of the magnetization a. The metastable set $A_{\mathbf{M}}$ has magnetization $\mathbf{m} < 0$, the stable set $A_{\mathbf{S}}$ has magnetization $\mathbf{s} > 0$, the top set has magnetization $\mathbf{t} < 0$. Note that $R_{p,\beta,h}(-1) = -\frac{1}{2}p + h$, $R_{p,\beta,h}(0) = -\beta^{-1}\log 2$, $R_{p,\beta,h}(+1) = -\frac{1}{2}p - h$ and $R'_{p,\beta,h}(-1) = -\infty$, $R'_{p,\beta,h}(0) = -h$, $R'_{p,\beta,h}(+1) = \infty$

Theorem 1.3 (Average Crossover Time on K_n) *Subject to (1.30), as $n \to \infty$, uniformly in $\xi \in A_{\mathbf{M}_n}$,*

$$\mathbb{E}_\xi\left[\tau_{A_{\mathbf{S}_n}}\right] = [1 + o_n(1)]\,\frac{\pi}{1+\mathbf{t}}\sqrt{\frac{1-\mathbf{t}^2}{1-\mathbf{m}^2}}\,\frac{1}{\beta\sqrt{R''_{1,\beta,h}(\mathbf{m})[-R''_{1,\beta,h}(\mathbf{t})]}}\,e^{\beta n[R_{1,\beta,h}(\mathbf{t})-R_{1,\beta,h}(\mathbf{m})]}. \tag{1.31}$$

Figure 4 illustrates the setting: the average crossover time from $A_{\mathbf{M}_n}$ to $A_{\mathbf{S}_n}$ depends on the free energy barrier $R_{1,\beta,h}(\mathbf{t}) - R_{1,\beta,h}(\mathbf{m})$ and on the curvature of $R_{1,\beta,h}$ at $\mathbf{m}$ and $\mathbf{t}$. Note that $\mathbf{m}$, $\mathbf{s}$, $\mathbf{t}$ in Fig. 4 are the limits as $n \to \infty$ of $\mathbf{m}_n$, $\mathbf{s}_n$, $\mathbf{t}_n$ defined in (1.16) for $p = 1$.

1.5 Metastability on $\mathrm{ER}_n(p)$

Unlike for the spin-flip dynamics on K_n, the induced processes defined in (1.12) are *not Markovian*. This is due to the random geometry of $\mathrm{ER}_n(p)$. However, we will see that they are *almost Markovian*, a fact that we will exploit by comparing the dynamics on $\mathrm{ER}_n(p)$ with that on K_n, but with a ferromagnetic coupling strength p/n rather than $1/n$ and with an external magnetic field that is a *small perturbation* of h.

As shown in Lemma 2.2 below, in the metastable regime the function $a \mapsto R_p(a)$ has a double-well structure just like in Fig. 4, so that the metastable state $A_{\mathbf{M}}$ and the stable state $A_{\mathbf{S}}$ are separated by a *free energy barrier* represented by $A_{\mathbf{T}}$.

We are finally in a position to state our main theorem.

Theorem 1.4 (Average Crossover Time on $\mathrm{ER}_n(p)$**)** *Subject to (1.13), with* $\mathbb{P}_{\mathrm{ER}_n(p)}$*-probability tending to* 1 *as* $n \to \infty$*, and uniformly in* $\xi \in A_{\mathbf{M}_n}$*,*

$$\mathbb{E}_\xi\left[\tau_{A_{\mathbf{S}_n}}\right] = n^{E_n}\,\mathrm{e}^{\beta n[R_{p,\beta,h}(\mathbf{t}) - R_{p,\beta,h}(\mathbf{m})]} \tag{1.32}$$

where the random exponent E_n *satisfies*

$$\lim_{n\to\infty} \mathbb{P}_{\mathrm{ER}_n(p)}\left(|E_n| \le \beta(\mathbf{t}-\mathbf{m})\frac{11}{6}\right) = 1. \tag{1.33}$$

Thus, apart from a polynomial error term, the average crossover time is the same as on the complete graph with ferromagnetic interaction strength p/n instead of $1/n$.

1.6 Discussion and Outline

We discuss the significance of our main theorem.

1. Theorem 1.4 provides an estimate on the average crossover time from $A_{\mathbf{M}_n}$ to $A_{\mathbf{S}_n}$ on $\mathrm{ER}_n(p)$ (recall Fig. 4). The estimate is *uniform* in the starting configuration. The exponential term in the estimate is the *same* as on K_n, but with a ferromagnetic interaction strength p/n rather than $1/n$. The multiplicative error term is *at most polynomial* in n. Such an error term is *not* present on K_n, for which the prefactor is known to be a constant up to a multiplicative factor $1 + o(1)$ (as shown in Theorem 1.3). The randomness of $\mathrm{ER}_n(p)$ manifests itself through a more complicated prefactor, which we do not know how to identify. What is interesting is that, apparently, $\mathrm{ER}_n(p)$ is so homogeneous for large n that the prefactor is at most polynomial. We expect the prefactor to be *random* under the law $\mathbb{P}_{\mathrm{ER}_n(p)}$.
2. It is known that on K_n the crossover time divided by it average has an exponential distribution in the limit as $n \to \infty$, as is typical for metastable behaviour. The same is true on $\mathrm{ER}_n(p)$. A proof of this fact can be obtained in a straightforward manner from the comparison properties underlying the proof of Theorem 1.4. These comparison properties, which are based on *coupling* of trajectories, also allow us to identify the *typical set of trajectories* followed by the spin-flip dynamics prior to the crossover. We will not spell out the details.
3. The proof of Theorem 1.4 is based on estimates of transition probabilities and transition times between pairs of configurations with different volume, in combination with a *coupling argument*. Thus we are following the *path-wise* approach to metastability (see [4] for background). Careful estimates are needed because on $\mathrm{ER}_n(p)$ the processes introduced in (1.12) are *not* Markovian, unlike on K_n. The proof is based on a *double* coupling strategy: (1) a sandwich of the Erdős-Rényi dynamics between two small perturbations of the Curie-Weiss dynamics, with the goal to identify the leading order term of the average

crossover time with the help of Theorem 1.3; (2) a two-level coupling (defined in Sect. 6), with the goal to prove asymptotic independence from the starting configuration (see also the beginning of Sects. 5.2 and 7).

4. Bovier et al. [5] use capacity estimates and concentration of measure estimates to show that the prefactors form a *tight* family of random variables under the law $\mathbb{P}_{\mathrm{ER}_n(p)}$ as $n \to \infty$, which constitutes a considerable sharpening of (1.32). The result is valid for $\beta > \beta_c$ and h small enough. The starting configuration is not arbitrary, but is drawn according to the *last-exit-biased distribution* for the transition from $A_{\mathbf{M}_n}$ to $A_{\mathbf{S}_n}$, as is common in the *potential-theoretic* approach to metastability. The exponential limit law is therefore not immediate.

5. Another interesting model is where *the randomness sits in the vertices rather than in the edges*, namely, Glauber spin-flip dynamics with Hamiltonian

$$H_n(\sigma) = -\frac{1}{n} \sum_{1 \leq i,j \leq n} \sigma(i)\sigma(j) - \sum_{1 \leq i \leq n} h_i \sigma(i), \tag{1.34}$$

where h_i, $1 \leq i \leq n$, are i.i.d. random variables drawn from a common probability distribution ν on $\mathbb{R}$. The metastable behaviour of this model was analysed in Bovier et al. [3] (discrete ν) and Bianchi et al. [1] (continuous ν). In particular, the prefactor was computed up to a multiplicative factor $1 + o(1)$, and turns out to be rather involved (see [4, Chapters 14–15]). Our model is even harder because the interaction between the spins runs along the edges of $\mathrm{ER}_n(p)$, which have an *intricate spatial structure*. Consequently, the so-called *lumping technique* (employed in [3] and [1] to monitor the magnetization on the level sets of the magnetic field) can no longer be used. For the dynamics under (1.34) the exponential law was proved in Bianchi et al. [2].

Outline The remainder of the paper is organized as follows. In Sect. 2 we define the perturbed spin-flip dynamics on K_n (Definition 2.1 below) and explain why Definition 1.1 identifies the metastable regime (Lemma 2.2 below). In Sect. 3 we collect a few basic facts about the geometry of $\mathrm{ER}_n(p)$ and the spin-flip dynamics on $\mathrm{ER}_n(p)$. In Sect. 4 we derive rough capacity estimates for the spin-flip dynamics on $\mathrm{ER}_n(p)$. In Sect. 5 we derive refined capacity estimates. In Sect. 6 we show that two copies of the spin-flip dynamics starting near the metastable state can be coupled in a short time. In Sect. 7 we prove Theorem 1.4. In Sect. 8, finally, we do a technical computation of hitting times that is needed in the proof.

2 Preparations

In Sect. 2.1 we define the perturbed spin-flip dynamics on K_n that will be used as comparison object. In Sect. 2.2 we do a rough metastability analysis of the perturbed model. In Sect. 2.3 we show that $R_{p,\beta,h}$ has a double-well structure if and only if (β, h) is in the metastable regime, in the sense of Definition 1.1 (Lemma 2.2 below).

Define

$$J_n^*(a) = 2\beta\left(p\left(a+\frac{2}{n}\right)+h\right)+\log\left(\frac{1-a}{1+a+\frac{2}{n}}\right), \quad a \in \Gamma_n. \tag{2.1}$$

We see from (1.21) that $J_n(a) = J_n^*(a) + O(n^{-2})$ when $\beta p = \frac{1}{1-a^2}$. This will be useful for the analysis of the 'free energy landscape'.

2.1 Perturbed Curie-Weiss

We will compare the dynamics on $\mathrm{ER}_n(p)$ with that on K_n, but with a ferromagnetic coupling strength p/n rather than $1/n$, and with an external magnetic field that is a *small perturbation* of h.

Definition 2.1 (Perturbed Curie-Weiss)

(1) Let

$$H_n^u(\sigma) = -\frac{p}{2n}\sum_{1\le i,j\le n}\sigma(i)\sigma(j) - h_n^u\sum_{1\le i\le n}\sigma(i), \quad \sigma\in S_n, \tag{2.2}$$

$$H_n^l(\sigma) = -\frac{p}{2n}\sum_{1\le i,j\le n}\sigma(i)\sigma(j) - h_n^l\sum_{1\le i\le n}\sigma(i), \quad \sigma\in S_n, \tag{2.3}$$

be the Hamiltonians on S_n corresponding to the Curie-Weiss model on n vertices with ferromagnetic coupling strength p/n, and with external magnetic fields h_n^u and h_n^l given by

$$h_n^u = h + \frac{(1+\epsilon)\log(n^{11/6})}{n}, \qquad h_n^l = h - \frac{(1+\epsilon)\log(n^{11/6})}{n}, \tag{2.4}$$

where $\epsilon > 0$ is arbitrary. The indices u and l stand for upper and lower, and the choice of exponent $\frac{11}{6}$ will become clear in Sect. 4.

(2) The equilibrium measures on S_n corresponding to (2.2) and (2.3) are denoted by μ_n^u and μ_n^l, respectively (recall (1.7)).

(3) The Glauber dynamics based on (2.2) and (2.3) are denoted by

$$\{\xi_t^u\}_{t\ge 0}, \qquad \{\xi_t^l\}_{t\ge 0}, \tag{2.5}$$

respectively.

(4) The analogues of (1.16) and (1.17) are written $\mathbf{m}_n^u, \mathbf{t}_n^u, \mathbf{s}_n^u$, $\mathbf{M}_n^u, \mathbf{T}_n^u, \mathbf{S}_n^u$ and $\mathbf{m}_n^l, \mathbf{t}_n^l, \mathbf{s}_n^l, \mathbf{M}_n^l, \mathbf{T}_n^l, \mathbf{S}_n^l$, respectively. ■

In what follows we will *suppress the n-dependence from most of the notation*. Almost all of the analysis in Sects. 2–7 pertains to the dynamics on $\mathrm{ER}_n(p)$.

2.2 Metastability for Perturbed Curie-Weiss

Recall that $\{\xi_t^u\}_{t\geq 0}$ and $\{\xi_t^l\}_{t\geq 0}$ denote the Glauber dynamics for the Curie-Weiss model driven by (2.2) and (2.3), respectively. An important feature is that their magnetization processes

$$\begin{aligned}\left\{\theta_t^u\right\}_{t\geq 0} &= \{m(\xi^l_{\tau_s^l})\}_{t\geq 0},\\ \left\{\theta_t^l\right\}_{t\geq 0} &= \{m(\xi^u_{\tau_s^u})\}_{t\geq 0},\end{aligned} \tag{2.6}$$

are continuous-time Markov processes themselves (see e.g. Bovier and den Hollander [4, Chaper 13]) with state space $\Gamma_n = \{-1, -1+\frac{2}{n}, \ldots, 1-\frac{2}{n}\}$ and transition rates

$$q^u\left(a,a'\right) = \begin{cases} \frac{n}{2}\left(1-a\right)\mathrm{e}^{-\beta[p(-2a-\frac{2}{n})-2h^u]_+}, & \text{if } a' = a+\frac{2}{n},\\ \frac{n}{2}\left(1+a\right)\mathrm{e}^{-\beta[p(2a+\frac{2}{n})+2h^u]_+}, & \text{if } a' = a-\frac{2}{n},\\ 0, & \text{otherwise,}\end{cases} \tag{2.7}$$

$$q^l\left(a,a'\right) = \begin{cases} \frac{n}{2}\left(1-a\right)\mathrm{e}^{-\beta[p(-2a-\frac{2}{n})-2h^l]_+}, & \text{if } a' = a+\frac{2}{n},\\ \frac{n}{2}\left(1+a\right)\mathrm{e}^{-\beta[p(2a+\frac{2}{n})+2h^l]_+}, & \text{if } a' = a-\frac{2}{n},\\ 0, & \text{otherwise,}\end{cases} \tag{2.8}$$

respectively. The processes in (2.6) are reversible with respect to the Gibbs measures

$$\nu^u\left(a\right) = \frac{1}{z^u}\mathrm{e}^{\beta n(\frac{1}{2}pa^2+h^u a)}\binom{n}{\frac{1+a}{2}n}, \quad a\in\Gamma_n, \tag{2.9}$$

$$\nu^l\left(a\right) = \frac{1}{z^l}\mathrm{e}^{\beta n(\frac{1}{2}pa^2+h^l a)}\binom{n}{\frac{1+a}{2}n}, \quad a\in\Gamma_n, \tag{2.10}$$

respectively.

Define

$$\Psi^u\left(a\right) = -\frac{1}{2}pa^2 - h^u a, \quad a\in\Gamma_n, \tag{2.11}$$

$$\Psi^l(a) = -\frac{1}{2}pa^2 - h^l a, \quad a\in\Gamma_n. \tag{2.12}$$

Note that (2.7) and (2.9) can be written as

$$\begin{aligned} q^u\left(a, a+\frac{2}{n}\right) &= \frac{n}{2}\left(1-a\right)\mathrm{e}^{-\beta n[\Psi^u(a+\frac{2}{n})-\Psi^u(a)]_+},\\ \nu^u\left(a\right) &= \frac{1}{z^u}\mathrm{e}^{-\beta n\Psi^u(a)}\binom{n}{\frac{n}{2}\left(1+a\right)},\end{aligned} \tag{2.13}$$

and similar formulas hold for (2.8) and (2.10). The properties of the function $\nu^u\colon \Gamma_n \to [0,1]$ can be analysed by looking at the ratio of adjacent values:

$$\frac{\nu^u\left(a+\frac{2}{n}\right)}{\nu^u(a)} = \exp\left(2\beta\left(p\left(a+\frac{2}{n}\right)+h^u\right)+\log\left(\frac{1-a}{1+a+\frac{2}{n}}\right)\right), \tag{2.14}$$

which suggests that 'local free energy wells' in ν^u can be found by looking at where the sign of

$$2\beta\left(p\left(a+\frac{2}{n}\right)+h^u\right)+\log\left(\frac{1-a}{1+a+\frac{2}{n}}\right) \tag{2.15}$$

changes from negative to positive. To that end note that, in the limit $n \to \infty$, the second term is positive for $a < 0$, tends to ∞ as $a \to -1$, is negative for $a \geq 0$, tends to $-\infty$ as $a \to 1$, and tends to 0 as $a \to 0$. The first term is linear in a, and for appropriate choices of p, β and h^u (see Definition 1.1) is negative near $a = -1$ and becomes positive at some value $a < 0$. This implies that, for appropriate choices of p, β and h^u, the sum of the two terms in (2.15) can change sign $+ \to - \to +$ on the interval $[-1, 0]$, and can change sign $+ \to -$ on $[0, 1]$. Assuming that our choice of p, β and h^u corresponds to this change-of-signs sequence, we define $\mathbf{m}^u$, $\mathbf{t}^u$ and $\mathbf{s}^u$ as in (1.16) with h replaced by h^u. This observation makes it clear that the sets in the right-hand side of (1.16) indeed are non-empty.

The interval $[\mathbf{m}^u, \mathbf{t}^u]$ poses a barrier for the process $\{\theta_t^u\}_{t\geq 0}$ due to a negative drift, which delays the initiation of the convergence to equilibrium while the process passes through the interval $[\mathbf{t}^u, \mathbf{s}^u]$. The same is true for the process $\{\xi_t^u\}_{t\geq 0}$. Similar observations hold for $\{\theta_t^l\}_{t\geq 0}$ and $\{\xi_t^l\}_{t\geq 0}$. Recall Fig. 4.

2.3 Double-Well Structure

Lemma 2.2 (Metastable Regime) *The potential $R_{p,\beta,h}$ defined in (1.26) has a double-well structure if and only if $\beta p > 1$ and $0 < h < p\chi(\beta p)$, with χ defined in (1.14).*

Proof In order for $R_{p,\beta,h}$ to have a double-well structure, the measure ν must have two distinct maxima on the interval $(-1, 1)$. From (1.22), (1.27) and (2.14) it follows that

$$J_{p,\beta,h}(a) = 2\lambda\left(a+\frac{h}{p}\right)+\log\left(\frac{1-a}{1+a}\right), \qquad \lambda = \beta p, \tag{2.16}$$

must have one local minimum and two zeroes in $(-1, 1)$. Since

$$J'_{p,\beta,h}(a) = 2\left(\lambda - \frac{1}{1-a^2}\right), \qquad a \in [-1, 1], \tag{2.17}$$

it must therefore be that $\lambda > 1$. The local minimum is attained when

$$\lambda = \frac{1}{1-a^2}, \tag{2.18}$$

i.e., when $a = a_\lambda = -\sqrt{1 - \frac{1}{\lambda}}$ (a_λ must be negative because it lies in $(\mathbf{m}, \mathbf{t})$; recall Fig. 4). Since

$$0 > J_{p,\beta,h}(a_\lambda) = 2\lambda\left(a_\lambda + \frac{h}{p}\right) + \log\left(\frac{1-a_\lambda}{1+a_\lambda}\right), \tag{2.19}$$

it must therefore be that

$$\frac{h}{p} < \chi(\lambda) \tag{2.20}$$

with $\chi(\lambda)$ given by (1.14). □

3 Basic Facts

In this section we collect a few facts that will be needed in Sect. 4 to derive capacity estimates for the dynamics on $\mathrm{ER}_n(p)$. In Sect. 3.1 we derive a large deviation bound for the degree of typical vertices $\mathrm{ER}_n(p)$ (Lemma 3.2 below). In Sect. 3.2 we do the same for the edge-boundary of typical configurations (Lemma 3.3 below). In Sect. 3.3 we derive upper and lower bounds for the jump rates of the volume process (Lemmas 3.4–3.5 and Corollary 3.6 below), and show that the return times to the metastable set *conditional* on not hitting the top set are small (Lemma 3.7 below). In Sect. 3.4 we use the various bounds to show that the probability for the volume process to grow by $n^{1/3}$ is almost uniform in the starting configuration (Lemma 3.8 and Corollary 3.9 below).

Definition 3.1 (Notation) For a vertex $v \in V$, we will write $v \in \sigma$ to mean $\sigma(v) = +1$ and $v \notin \sigma$ to mean $\sigma(v) = -1$. Similarly, we will denote by $\overline{\sigma}$ the configuration obtained from σ by flipping the spin at every vertex, i.e., $\sigma(v) = +1$ if and only if $\overline{\sigma}(v) = -1$. For two configurations σ, σ' we will say that $\sigma \subseteq \sigma'$ if and only if $v \in \sigma \Rightarrow v \in \sigma'$. By $\sigma \cup \sigma'$ we denote the configuration satisfying $v \in \sigma \cup \sigma'$ if and only if $v \in \sigma$ or $v \in \sigma'$. A similar definition applies to $\sigma \cap \sigma'$. We will also write $\sigma \sim \sigma'$ when there is a $v \in V$ such that $\sigma = \sigma' \cup \{v\}$ or $\sigma' = \sigma \cup \{v\}$. We will say that σ and σ' are neighbours. We write $\deg(v)$ for the degree of $v \in V$. ■

3.1 Concentration Bounds for $\mathbf{ER}_n(p)$

Recall that $\mathbb{P}_{\mathrm{ER}_n(p)}$ denotes the law $\mathrm{ER}_n(p)$.

Lemma 3.2 (Concentration of Degrees and Energies) *With $\mathbb{P}_{\mathrm{ER}_n(p)}$-probability tending to* 1 *as $n \to \infty$ the following is true. For any $\epsilon > 0$ and any $c > \sqrt{\frac{1}{8}\log 2}$,*

$$pn - (1+\epsilon)\sqrt{n\log n} < \deg(v) < pn + (1+\epsilon)\sqrt{n\log n} \qquad \forall\, v \in V, \tag{3.1}$$

$$\begin{aligned} \frac{1}{n}\left(2p|\xi|(n-|\xi|) - cn^{3/2}\right) - 2h|\xi| &\le H_n(\xi) - H_n(\boxminus) \\ &\le \frac{1}{n}\left(2p|\xi|(n-|\xi|) + cn^{3/2}\right) - 2h|\xi| \qquad \forall\, \xi \in S_n. \end{aligned} \tag{3.2}$$

Proof These bounds are immediate from Hoeffding's inequality and a union bound. □

3.2 Edge Boundaries of $\mathbf{ER}_n(p)$

We partition the configuration space as

$$S_n = \bigcup_{k=0}^{n} A_k, \tag{3.3}$$

where A_k is defined in (1.18). For $0 \le k \le n$ and $-pk\,(n-k) \le i \le (1-p)\,k\,(n-k)$, define

$$\phi_i^k = |\{\sigma \in A_k\colon\ |\partial_E\sigma| = pk\,(n-k) + i\}|\,, \tag{3.4}$$

i.e., ϕ_i^k counts the configurations σ with volume k whose edge-boundary size $|\partial_E\sigma|$ deviates by i from its mean, which is equal to $pk\,(n-k)$. For $0 \le k \le n$, let $\mathbb{P}_k$ denote the uniform distribution on A_k.

Lemma 3.3 (Upper Bound on Edge-Boundary Sizes) *With $\mathbb{P}_{\mathrm{ER}_n(p)}$-probability tending to* 1 *as $n \to \infty$ the following are true. For $-pk(n-k) \le j \le (1-p)k(n-k)$ and $\varrho\colon\ \mathbb{N} \to \mathbb{R}_+$,*

$$\mathbb{P}_k\left[\phi_j^k \ge \varrho\,(n)\binom{n}{k} p^{pk(n-k)+j}(1-p)^{(1-p)k(n-k)-j}\binom{k(n-k)}{pk(n-k)+j}\right] \le \frac{1}{\varrho(n)} \tag{3.5}$$

and

$$\mathbb{P}_k\left[\sum_{j\geq i}\phi_j^k \geq \varrho(n)\binom{n}{k}\mathrm{e}^{-\frac{2i^2}{k(n-k)}}\right] \leq \frac{1}{\varrho(n)},$$
$$\mathbb{P}_k\left[\sum_{j\leq -i}\phi_j^k \geq \varrho(n)\binom{n}{k}\mathrm{e}^{-\frac{2i^2}{k(n-k)}}\right] \leq \frac{1}{\varrho(n)}. \tag{3.6}$$

Proof Write $\simeq$ to denote equality in distribution. Note that if $\sigma \simeq \mathbb{P}_k$, then $|\partial_E\sigma| \simeq \mathrm{Bin}\,(k\,(n-k)\,,\,p)$, and hence

$$\mathbb{P}_k\,[|\partial_E\sigma| = i] = p^i\,(1-p)^{k(n-k)-i}\binom{k\,(n-k)}{i}. \tag{3.7}$$

In particular,

$$\mathbb{E}_k\left[\phi_j^k\right] = \mathbb{E}_k\left[\sum_{\sigma\in A_k}\mathbb{1}_{\{|\partial_E\sigma|=pk(n-k)+j\}}\right]$$
$$= \binom{n}{k}p^{pk(n-k)+j}(1-p)^{(1-p)k(n-k)-j}\binom{k(n-k)}{pk(n-k)+j}. \tag{3.8}$$

Hence, by Markov's inequality, the claim in (3.5) follows. Moreover,

$$\mathbb{E}_k\left[\sum_{j\geq i}\phi_j^k\right] = \mathbb{E}_k\left[\sum_{\sigma\in A_k}\mathbb{1}_{\{|\partial_E\sigma|\geq pk(n-k)+i\}}\right] \leq \binom{n}{k}\mathrm{e}^{-2\frac{i^2}{k(n-k)}}, \tag{3.9}$$

where we again use Hoeffding's inequality. Hence, by Markov's inequality, we get the first line in (3.6). The proof of the second line is similar. □

3.3 Jump Rates for the Volume Process

The following lemma establishes bounds on the rate at which configurations in A_k jump forward to A_{k+1} and backward to A_{k-1}. In Sect. 8 we will sharpen the error in the prefactors in (3.10)–(3.11) from $2n^{2/3}$ to $O(1)$ and the error in the exponents in (3.10)–(3.11) from $3n^{-1/3}$ to $O(n^{-1/2})$. The formulas in (3.13) and (3.14) show that for small and large magnetization the rate forward, respectively, backward are maximal.

Lemma 3.4 (Bounds on Forward Jump Rates) *With $\mathbb{P}_{\mathrm{ER}_n(p)}$-probability tending to 1 as $n \to \infty$ the following are true.*

(a) *For* $2n^{1/3} \leq k \leq n - 2n^{1/3}$,

$$\begin{aligned}&\big(n-k-2n^{2/3}\big)\mathrm{e}^{-2\beta[\vartheta_k+3n^{-1/3}]_+}\\ &\quad\leq \textstyle\sum_{\xi\in A_{k+1}} r(\sigma,\xi) \leq \big(n-k-2n^{2/3}\big)\mathrm{e}^{-2\beta[\vartheta_k-3n^{-1/3}]_+} + 2n^{2/3}, \qquad \sigma\in A_k,\end{aligned} \tag{3.10}$$

and

$$\begin{aligned}&\big(k-2n^{2/3}\big)\mathrm{e}^{-2\beta[-\vartheta_k+3n^{-1/3}]_+}\\ &\quad\leq \textstyle\sum_{\xi\in A_{k-1}} r(\sigma,\xi) \leq \big(k-2n^{2/3}\big)\mathrm{e}^{-2\beta[-\vartheta_k-3n^{-1/3}]_+} + 2n^{2/3}, \qquad \sigma\in A_k,\end{aligned} \tag{3.11}$$

where

$$\vartheta_k = p\left(1-\frac{2k}{n}\right) - h. \tag{3.12}$$

(b) *For* $n - \frac{n}{3}(p+h) \leq k < n$,

$$\sum_{\xi\in A_{k+1}} r(\sigma,\xi) = n-k, \qquad \sigma\in A_k. \tag{3.13}$$

(c) *For* $0 < k \leq \frac{n}{3}(p-h)$,

$$\sum_{\xi\in A_{k-1}} r(\sigma,\xi) = k, \qquad \sigma\in A_k. \tag{3.14}$$

Proof The proof is via probabilistic counting.

(a) Write $\mathbb{P}$ for the law under which $\sigma \in S_n$ is a uniformly random configuration and $v \in \overline{\sigma}$ is a uniformly random vertex. By Hoeffding's inequality, the probability that v has more than $p\,|\sigma| + n^{2/3}$ neighbours in σ (i.e., $w \in V$ such that $(v,w) \in E$ and $\sigma(w) = +1$) is bounded by

$$\mathbb{P}\Big[|E(v,\sigma)| \geq p\,|\overline{\sigma}| + n^{2/3}\Big] \leq \mathrm{e}^{-2n^{1/3}}, \tag{3.15}$$

where

$$E(v,\sigma) = \{w\in\sigma\colon\ (v,w)\in E\}\,. \tag{3.16}$$

Define the event

$$R^+(\sigma) = \Big\{\exists\,\zeta\subseteq\overline{\sigma},\ \zeta\in A_{2n^{2/3}}\colon\ |E(v,\sigma)| \geq p\,|\sigma| + n^{2/3}\ \forall\, v\in\zeta\Big\}, \tag{3.17}$$

i.e., the configuration $\overline{\sigma}$ has at least $2n^{2/3}$ vertices like v, each with at least $p\,|\sigma| + n^{2/3}$ neighbours in σ. Then, for $0 \le k \le n - 2n^{2/3}$,

$$\mathbb{P}\left[R^{+}(\sigma)\right] \le \binom{|\overline{\sigma}|}{2n^{2/3}} \left(\mathrm{e}^{-2n^{1/3}}\right)^{2n^{2/3}} \le 2^{n}\mathrm{e}^{-4n}. \tag{3.18}$$

Hence the probability that some configuration $\sigma \in S_n$ satisfies condition $R^{+}(\sigma)$ is bounded by

$$\mathbb{P}\left[\bigcup_{\sigma \in S_n} R^{+}(\sigma)\right] \le 4^{n}\mathrm{e}^{-4n} \le \mathrm{e}^{-2n}. \tag{3.19}$$

Thus, with $\mathbb{P}_{\mathrm{ER}_n(p)}$-probability tending to 1 as $n \to \infty$ there are no configurations $\sigma \in S_n$ satisfying condition $R^{+}(\sigma)$. The same holds for the event

$$R^{-}(\sigma) = \left\{\exists\, \zeta \subseteq \overline{\sigma},\ \zeta \in A_{2n^{2/3}}\colon\ |E(v,\sigma)| \le p\,|\sigma| - n^{2/3}\ \forall\, v \in \zeta\right\}, \tag{3.20}$$

for which

$$\mathbb{P}\left[\bigcup_{\sigma \in S_n} R^{-}(\sigma)\right] \le \mathrm{e}^{-2n}. \tag{3.21}$$

Now let $\sigma \in A_k$, and observe that σ has $n-k$ neighbours in A_{k+1} and k neighbours in A_{k-1}. But if $\xi = \sigma \cup \{v\} \in A_{k+1}$, then by (1.3),

$$\begin{aligned} H_n(\xi) - H_n(\sigma) &= \frac{2}{n}\Big(\,|E(v,\overline{\sigma})| - |E(v,\sigma)|\,\Big) - 2h \\ &= \frac{2}{n}\big(\deg(v) - 2\,|E(v,\sigma)|\,\big) - 2h \\ &\le \frac{2}{n}\big(pn + n^{1/2}\log n - 2\,|E(v,\sigma)|\,\big) - 2h, \end{aligned} \tag{3.22}$$

where the last inequality uses (3.1) with $\varrho(n) = \log n$. Similarly,

$$H_n(\xi) - H_n(\sigma) \ge \frac{2}{n}\Big(pn - n^{1/2}\log n - 2\,|E(v,\sigma)|\Big) - 2h. \tag{3.23}$$

The events $R^{+}(\sigma)$ in (3.17) and $R^{-}(\sigma)$ in (3.20) guarantee that for any configuration σ at most $2n^{2/3}$ vertices in the configuration $\overline{\sigma}$ can have more than $n^{2/3}$ neighbours in σ. In other words, the configuration σ has at most $2n^{2/3}$ neighbouring configurations in A_{k+1} that differ in energy by more than

$6n^{-1/3} - 2h$. Since on the complement of $R^+(\sigma)$ with $\sigma \in A_k$ we have $|\{w \in \sigma\colon (v,w) \in E\}| \le 2pk + 2n^{1/3}$ (because $n^{1/2}\log n \le n^{2/3}$ for n large enough), from (3.19) and (3.21) we get that, with $\mathbb{P}_{\mathrm{ER}_n(p)}$-probability at least $1 - \mathrm{e}^{-2n}$,

$$\left|\left\{\xi \in A_{k+1}\colon \xi \sim \sigma,\ H_n(\xi) - H_n(\sigma) \ge \tfrac{2}{n}\left(pn - 2pk + 3n^{2/3}\right) - 2h\right\}\right| \le 2n^{2/3},$$
$$\left|\left\{\xi \in A_{k+1}\colon \xi \sim \sigma,\ H_n(\xi) - H_n(\sigma) \le \tfrac{2}{n}\left(pn - 2pk - 3n^{2/3}\right) - 2h\right\}\right| \le 2n^{2/3}, \tag{3.24}$$

and hence, by (1.6), the rate at which the Markov chain starting at $\sigma \in A_k$ jumps to A_{k+1} satisfies

$$\sum_{\xi \in A_{k+1}} r(\sigma,\xi) \ge \left(n - k - 2n^{2/3}\right)\mathrm{e}^{-2\beta[\vartheta_k + 3n^{-1/3}]_+}, \tag{3.25}$$

$$\sum_{\xi \in A_{k+1}} r(\sigma,\xi) \le \left(n - k - 2n^{2/3}\right)\mathrm{e}^{-2\beta[\vartheta_k - 3n^{-1/3}]_+} + 2n^{2/3}. \tag{3.26}$$

Here the term $n - k - 2n^{2/3}$ comes from exclusion of the at most $2n^{2/3}$ neighbours in configurations that differ from σ in energy by more than $6n^{-1/3} - 2h$. Similarly, with $\mathbb{P}_{\mathrm{ER}_n(p)}$-probability at least $1 - \mathrm{e}^{-2n}$,

$$\left|\left\{\xi \in A_{k-1}\colon \xi \sim \sigma,\ H_n(\xi) - H_n(\sigma) \ge \tfrac{2}{n}\left(-pn + 2pk + 3n^{2/3}\right) + 2h\right\}\right| \le 2n^{2/3},$$
$$\le \tfrac{2}{n}\left(-pn + 2pk - 3n^{2/3}\right) + 2h\Big\}\Big| \le 2n^{2/3}, \tag{3.27}$$

and hence, by (1.6), the rate at which the Markov chain starting at $\sigma \in A_k$ jumps to A_{k-1} satisfies

$$\sum_{\xi \in A_{k-1}} r(\sigma,\xi) \le \left(k - 2n^{2/3}\right)\mathrm{e}^{-2\beta[-\vartheta_k - 3n^{-1/3}]_+} + 2n^{2/3}, \tag{3.28}$$

$$\sum_{\xi \in A_{k-1}} r(\sigma,\xi) \ge \left(k - 2n^{2/3}\right)\mathrm{e}^{-2\beta[-\vartheta_k + 3n^{-1/3}]_+}. \tag{3.29}$$

This proves (3.10) and (3.11).

(b) To get (3.13), note that for $\xi = \sigma \cup \{v\}$ with $v \notin \sigma$,

$$\begin{aligned} H_n(\xi) - H_n(\sigma) &= \frac{2}{n}\Big(|E(v,\overline{\sigma})| - |E(v,\sigma)|\Big) - 2h \\ &= \frac{2}{n}\Big(2\,|E(v,\sigma)| - \deg(v)\Big) - 2h \\ &\le 2\left(2(n-k) - p + n^{-1/2}\log n - h\right) \end{aligned} \tag{3.30}$$

for n large enough, which is ≤ 0 when $n-k \leq \frac{n}{3}(p+h)$, so that $r(\sigma, \xi) = 1$ by (1.6).

(c) To get (3.14), note that for $\xi = \sigma \setminus \{v\}$ with $v \notin \sigma$,

$$
\begin{aligned}
H_n(\xi) - H_n(\sigma) &= \frac{2}{n}\Big(|E(v,\sigma)| - |E(v,\overline{\sigma})| \Big) + 2h \qquad (3.31)\\
&= \frac{2}{n}\Big(2\,|E(v,\sigma)| - \deg(v)\Big) + 2h\\
&\leq 2\left(2k - p + n^{-1/2}\log n + h\right)
\end{aligned}
$$

for n large enough, which is ≤ 0 when $k \leq \frac{n}{3}(p-h)$, so that $r(\sigma, \xi) = 1$ by (1.6). □

The following lemma is technical and merely serves to show that near $A_{\mathbf{M}}$ transitions involving a flip from -1 to $+1$ typically occur at rate 1. Write ξ^v to denote the configuration obtained from ξ by flipping the sign at vertex $v \in V$.

Lemma 3.5 (Attraction Towards the Metastable State) *Suppose that* $|\xi| = [1 + o_n(1)]\,\mathbf{M}$. *Then* $r(\xi, \xi^v) = 1$ *for all but* $O(n^{2/3})$ *many* $v \in \xi$.

Proof We want to show that

$$
H_n\left(\xi^v\right) < H_n(\xi) \qquad (3.32)
$$

for all but $O(n^{2/3})$ many $v \in \xi$. Note that by (3.20) and (3.21) there are at most $2n^{2/3}$ many $v \in \xi$ such that $|E(v, \overline{\xi})| \leq p(n-|\xi|) - n^{2/3}$, and at most $2n^{2/3}$ many $v \in \xi$ such that $|E(v,\xi)| \geq p|\xi| + n^{2/3}$. Hence, by (1.3), for all but at most $4n^{2/3}$ many $v \in \xi$ we have that

$$
\begin{aligned}
H_n\left(\xi^v\right) &= H_n(\xi) + \frac{2}{n}\left(|E(v,\xi)| - \left|E\left(v,\overline{\xi}\right)\right|\right) + 2h \qquad (3.33)\\
&= H_n(\xi) + \frac{2p}{n}\left(2\,|\xi| - n\right) + 2h + o_n(1)\\
&= H_n(\xi) + \frac{2p}{n}\left(2\mathbf{M} - n\right) + 2h + o_n(1)\\
&= H_n(\xi) + 2p\mathbf{m} + 2h + o_n(1),
\end{aligned}
$$

where we use (1.17). From the definition of $\mathbf{m}$ in (1.16) it follows that $2p\mathbf{m} + 2h + o_n(1) < 0$, where we recall from the discussion near the end of Sect. 2.2 that $\mathbf{m} < 0$ and hence $\log(\frac{1-\mathbf{m}}{1+\mathbf{m}}) > 0$. Hence (3.32) follows. □

We can now prove the claim made in Remark 1.2, namely, there is no metastable behaviour outside the regime in (1.13). Recall the definition of $\mathbf{S}_n$ in (1.17), which requires the function J in (1.23) to have two zeroes. If it has only one zero, then denote that zero by a' and define $\mathbf{S}_n = \frac{n}{2}(a' + 1)$. Let $A_{\mathbf{S}_n+O(n^{2/3})}$ be the union of all A_k with $|k - \mathbf{S}_n| = O(n^{2/3})$.

Corollary 3.6 (Non-metastable Regime) *Suppose that $\beta \in (1/p, \infty)$ and $h \in (p, \infty)$. Then $\{\xi_t\}_{t\geq 0}$ has a drift towards $A_{\mathbf{S}_n+O(n^{2/3})}$. Consequently, $\mathbb{E}_{\xi_0}[\tau_{\mathbf{S}}] = O(1)$ for any initial configuration $\xi_0 \in S_n$.*

Proof If $\beta \in (1/p, \infty)$ and $h \in (p, \infty)$, then the function $a \mapsto J_{p,\beta,h}(a) = 2\beta(pa+h)+\log(\frac{1-a}{1+a})$ has a unique root in the interval $(0, 1)$. Indeed, $J_{p,\beta,h}(a) > 0$ for $a \in [-1, 0]$, $J'_{p,\beta,h}(0) = 2(\beta p - 1) > 0$, while $a \mapsto \log(\frac{1-a}{1+a})$ is concave and tends to $-\infty$ as $a \uparrow 1$. We claim that the process $\{\xi_t\}_{t\geq 0}$ drifts towards that root, i.e., if we denote the root by a', then the process drifts towards the set $A_{\frac{n}{2}(a'+1)}$, which by convention we identify with A_{S_n}. Note that if $h \in (p, \infty)$, then $\vartheta_k = p(1 - \frac{2k}{n}) - h < 0$ for all $0 \leq k \leq n$ (recall (3.12)) and so, by Lemma 3.4,

$$\sum_{\xi\in A_{k+1}} r(\sigma, \xi) \geq n - k - 2n^{2/3}, \tag{3.34}$$

$$\sum_{\xi\in A_{k-1}} r(\sigma, \xi) \leq \big(k - 2n^{2/3}\big)\mathrm{e}^{-2\beta[-\vartheta_k-3n^{-1/3}]} + 2n^{2/3}.$$

Thus, for $k \leq \frac{n}{2} - 4n^{2/3}$, $\sum_{\xi\in A_{k+1}} r(\sigma, \xi) > \sum_{\xi\in A_{k-1}} r(\sigma, \xi)$. Similarly, for $k \geq \frac{n}{2} + 4n^{2/3}$, the opposite inequality holds. Therefore there is a drift towards $A_{\mathbf{S}_n+O(n^{2/3})}$. □

We close this section with a lemma stating that the average return time to $A_{\mathbf{M}_n}$ conditional on not hitting $A_{\mathbf{T}_n}$ is of order 1 and has an exponential tail. This will be needed to control the time between successive attempts to go from $A_{\mathbf{M}_n}$ to $A_{\mathbf{T}_n}$, until the dynamics crosses $A_{\mathbf{T}_n}$ and moves to $A_{\mathbf{S}_n}$ (recall Fig. 4).

Lemma 3.7 (Conditional Return Time to the Metastable Set) *There exists a $C > 0$ such that, with $\mathbb{P}_{\mathrm{ER}_n(p)}$-probability tending to 1 as $n \to \infty$, uniformly in $\xi \in A_{\mathbf{M}_n}$,*

$$\mathbb{P}_\xi\left[\tau_{A_{\mathbf{M}_n}} \geq k \mid \tau_{A_{\mathbf{M}_n}} < \tau_{A_{\mathbf{T}_n}}\right] \leq \mathrm{e}^{-Ck} \qquad \forall k. \tag{3.35}$$

Proof The proof is given in Sect. 8. □

3.4 Uniformity in the Starting Configuration

The following lemma shows that the probability of the event $\{\tau_{A_{k+o(n^{1/3})}} < \tau_{A_k}\}$ is almost uniform as a function of the starting configuration in A_k.

Lemma 3.8 (Uniformity of Hitting Probability of Volume Level Sets) *With $\mathbb{P}_{\mathrm{ER}_n(p)}$-probability tending to 1 as $n \to \infty$, the following is true. For every $0 \le k < m \le n$,*

$$\frac{\max_{\sigma \in A_k} \mathbb{P}_\sigma \left[\tau_{A_m} < \tau_{A_k}\right]}{\min_{\sigma \in A_k} \mathbb{P}_\sigma \left[\tau_{A_m} < \tau_{A_k}\right]} \le [1 + o_n(1)]\, \mathrm{e}^{K(m-k)n^{-1/3}} \tag{3.36}$$

with $K = K(\beta, h, p) \in (0, \infty)$.

Proof The proof proceeds by estimating the probability of trajectories from A_k to A_m. Observe that

$$\begin{aligned} \mathrm{e}^{-2\beta[\vartheta_k + 3n^{-1/3}]_+} &\ge \mathrm{e}^{-2\beta[\vartheta_k]_+}\left(1 - 6\beta n^{-1/3}\right) \quad \forall n, \\ \mathrm{e}^{-2\beta[\vartheta_k - 3n^{-1/3}]_+} &\le \mathrm{e}^{-2\beta[\vartheta_k]_+}\left(1 + 7\beta n^{-1/3}\right) \quad n \text{ large enough}, \end{aligned} \tag{3.37}$$

and that similar estimates hold for $\mathrm{e}^{-2\beta[-\vartheta_k + 3n^{-1/3}]_+}$ and $\mathrm{e}^{-2\beta[-\vartheta_k - 3n^{-1/3}]_+}$. We will bound the ratio in the left-hand side of (3.36) by looking at two random processes on $\{0, \dots, n\}$, one of which bounds $\max_{\sigma \in A_k} \mathbb{P}_\sigma \left[\tau_{A_m} < \tau_{A_k}\right]$ from above and the other of which bounds $\min_{\sigma \in A_k} \mathbb{P}_\sigma \left[\tau_{A_m} < \tau_{A_k}\right]$ from below. The proof comes in three Steps.

1. We begin with the following observation. Suppose that $\{X_t^+\}_{t \ge 0}$ and $\{X_t^-\}_{t \ge 0}$ are two continuous-time Markov chains taking unit steps in $\{0, \dots, n\}$ at rates $r^-(k, k \pm 1)$ and $r^+(k, k \pm 1)$, respectively. Furthermore, suppose that for every $0 \le k \le n-1$,

$$r^-(k, k+1) \le \min_{\sigma \in A_k} \sum_{\xi \in A_{k+1}} r(\sigma, \xi) \le \max_{\sigma \in A_k} \sum_{\xi \in A_{k+1}} r(\sigma, \xi) \le r^+(k, k+1), \tag{3.38}$$

and for every $1 \le k \le n$,

$$r^-(k, k-1) \ge \max_{\sigma \in A_k} \sum_{\xi \in A_{k-1}} r(\sigma, \xi) \ge \min_{\sigma \in A_k} \sum_{\xi \in A_{k-1}} r(\sigma, \xi) \ge r^+(k, k-1). \tag{3.39}$$

Then

$$\frac{\max_{\sigma \in A_k} \mathbb{P}_\sigma \left[\tau_{A_m} < \tau_{A_k}\right]}{\min_{\sigma \in A_k} \mathbb{P}_\sigma \left[\tau_{A_m} < \tau_{A_k}\right]} \le \frac{\mathbb{P}_k^{X^+}\left[\tau_m < \tau_k\right]}{\mathbb{P}_k^{X^-}\left[\tau_m < \tau_k\right]}. \tag{3.40}$$

Indeed, from (3.38) and (3.39) it follows that we can couple the three Markov chains $\{X_t^+\}_{t \ge 0}$, $\{X_t^-\}_{t \ge 0}$ and $\{\xi_t\}_{t \ge 0}$ in such a way that, for any $0 \le k \le n$ and any $\sigma_0 \in A_k$, if $X_0^- = X_0^+ = |\sigma_0| = k$, then

$$X_t^- \le |\sigma_t| \le X_t^+, \qquad t \ge 0. \tag{3.41}$$

This immediately guarantees that, for any $0 \le k \le m \le n$,

$$\mathbb{P}_k^{X^-}[\tau_m < \tau_k] \le \min_{\sigma \in A_k} \mathbb{P}_\sigma\left[\tau_{A_m} < \tau_{A_k}\right] \le \max_{\sigma \in A_k} \mathbb{P}_\sigma\left[\tau_{A_m} < \tau_{A_k}\right] \le \mathbb{P}_k^{X^+}[\tau_m < \tau_k], \tag{3.42}$$

which proves the claim in (3.40). To get (3.38) and (3.39), we pick $r^-(i,j)$ and $r^+(i,j)$ such that

$$r^-(i,j) = \begin{cases} \left(n - i - (2+6\beta)\,n^{2/3}\right)\mathrm{e}^{-2\beta[\vartheta_i]_+} \\ \qquad + (2+6\beta)\,n^{2/3}\mathbb{1}_{\left\{i \ge n\left(1-\frac{1}{3}(p+h)\right)\right\}}, & j = i+1, \\ \min\left\{i,\ \left(i + (-2+7\beta)\,n^{2/3}\right)\mathrm{e}^{-2\beta[-\vartheta_i]_+} + 2n^{2/3}\right\}, & j = i-1, \\ 0, & \text{otherwise}, \end{cases} \tag{3.43}$$

and

$$r^+(i,j) = \begin{cases} \min\left\{n-i,\ \left(n - i + (-2+7\beta)\,n^{2/3}\right)\mathrm{e}^{-2\beta[\vartheta_i]_+} + 2n^{2/3}\right\}, & j = i+1, \\ \left(i - (2+6\beta)\,n^{2/3}\right)\mathrm{e}^{-2\beta[-\vartheta_i]_+}, & j = i-1, \\ 0, & \text{otherwise}, \end{cases} \tag{3.44}$$

and note that, by Lemma 3.4, (3.37)–(3.39) are indeed satisfied.

2. We continue from (3.40). Our task is to estimate the right-hand side of (3.40). Let $\mathscr{G}$ be the set of all unit-step paths from k to m that only hit m after their final step:

$$\mathscr{G} = \bigcup_{M \in \mathbb{N}} \Big\{ \{\gamma_i\}_{i=0}^{M-1} : \gamma_0 = k,\ \gamma_M = m,\ \gamma_i \in \{0,\ldots,m-1\} \\ \text{and } |\gamma_{i+1} - \gamma_i| = 1 \text{ for } 0 \le i < M \Big\}. \tag{3.45}$$

We will show that

$$\frac{\mathbb{P}_k^{X^+}[X_t^+ \text{ follows trajectory } \gamma]}{\mathbb{P}_k^{X^-}[X_t^- \text{ follows trajectory } \gamma]} \\ \le \exp\left(\left[24\beta + 4\mathrm{e}^{2\beta(p+h+1)}\right](m-k)\,n^{-1/3}\right) \quad \forall\, \gamma \in \mathscr{G}, \tag{3.46}$$

which will settle the claim. (Note that the paths realising $\{\tau_m < \tau_k\}$ form a subset of $\mathscr{G}$.) To that end, let $\gamma^\star \in \mathscr{G}$ be the path $\gamma^\star = \{k, k+1, \ldots, m\}$. We claim that

$$\sup_{\gamma \in \mathscr{G}} \frac{\mathbb{P}_k^{X^+}\left[X_t^+ \text{ follows trajectory } \gamma\right]}{\mathbb{P}_k^{X^-}\left[X_t^- \text{ follows trajectory } \gamma\right]} \le \frac{\mathbb{P}_k^{X^+}\left[X_t^+ \text{ follows trajectory } \gamma^\star\right]}{\mathbb{P}_k^{X^-}\left[X_t^- \text{ follows trajectory } \gamma^\star\right]}. \tag{3.47}$$

Indeed, if $\gamma = (\gamma_1, \ldots, \gamma_M) \in \mathscr{G}$, then by the Markov property we have that

$$\mathbb{P}_k^{X^+}\left[X_t^+ \text{ follows trajectory } \gamma\right] = \prod_{i=0}^{M-1} \mathbb{P}_{\gamma_i}^{X^+}\left[\tau_{\gamma_{i+1}} < \tau_{\gamma_i}\right], \tag{3.48}$$

with a similar expression for $\mathbb{P}_k^{X^-}[X_t^- \text{ follows trajectory } \gamma]$. Therefore, noting that $\gamma_i - 1 = 2\gamma_i - \gamma_{i+1}$ when $\gamma_{i+1} = \gamma_i + 1$ and $\gamma_i + 1 = 2\gamma_i - \gamma_{i+1}$ when $\gamma_{i+1} = \gamma_i - 1$, we have

$$\begin{aligned} &\frac{\mathbb{P}_k^{X^+}\left[X_t^+ \text{ follows trajectory } \gamma\right]}{\mathbb{P}_k^{X^-}\left[X_t^- \text{ follows trajectory } \gamma\right]} = \prod_{i=0}^{M-1} \frac{\mathbb{P}_{\gamma_i}^{X^+}\left[\tau_{\gamma_{i+1}} < \tau_{\gamma_i}\right]}{\mathbb{P}_{\gamma_i}^{X^-}\left[\tau_{\gamma_{i+1}} < \tau_{\gamma_i}\right]} \\ &= \prod_{i=1}^{M} \left(\frac{r^+(\gamma_i, \gamma_{i+1})}{r^+(\gamma_i, \gamma_{i+1}) + r^+(\gamma_i, 2\gamma_i - \gamma_{i+1})}\right)\left(\frac{r^-(\gamma_i, \gamma_{i+1})}{r^-(\gamma_i, \gamma_{i+1}) + r^-(\gamma_i, 2\gamma_i - \gamma_{i+1})}\right)^{-1}. \end{aligned} \tag{3.49}$$

Since, whenever $\gamma_{i+1} = \gamma_i - 1$,

$$\begin{aligned} \frac{r^-(\gamma_i,\gamma_{i+1})}{r^-(\gamma_i,\gamma_{i+1})+r^-(\gamma_i,2\gamma_i-\gamma_{i+1})} &= \frac{r^-(\gamma_i,\gamma_i-1)}{r^-(\gamma_i,\gamma_i-1)+r^-(\gamma_i,\gamma_i+1)} \\ &\geq \frac{r^+(\gamma_i,\gamma_i-1)}{r^+(\gamma_i,\gamma_i-1)+r^+(\gamma_i,\gamma_i+1)} \\ &= \frac{r^+(\gamma_i,\gamma_{i+1})}{r^+(\gamma_i,\gamma_{i+1})+r^+(\gamma_i,2\gamma_i-\gamma_{i+1})}, \end{aligned} \tag{3.50}$$

we get

$$\begin{aligned} &\textstyle\prod_{i=0}^{M-1} \left(\frac{r^+(\gamma_i,\gamma_{i+1})}{r^+(\gamma_i,\gamma_{i+1})+r^+(\gamma_i,2\gamma_i-\gamma_{i+1})}\right)\left(\frac{r^-(\gamma_i,\gamma_{i+1})}{r^-(\gamma_i,\gamma_{i+1})+r^-(\gamma_i,2\gamma_i-\gamma_{i+1})}\right)^{-1} \\ &\quad\leq \textstyle\prod_{i=k}^{m-1} \left(\frac{r^+(i,i+1)}{r^+(i,i+1)+r^+(i,i-1)}\right)\left(\frac{r^-(i,i+1)}{r^-(i,i+1)+r^-(i,i-1)}\right)^{-1} \\ &\quad= \frac{\mathbb{P}_k^{X^+}[X_t^+ \text{ follows trajectory } \gamma^\star]}{\mathbb{P}_k^{X^-}[X_t^- \text{ follows trajectory } \gamma^\star]}. \end{aligned} \tag{3.51}$$

This proves the claim in (3.47).

3. Next, consider the ratio

$$\frac{r^-(i, i+1) + r^-(i, i-1)}{r^+(i, i+1) + r^+(i, i-1)} = \frac{A}{B} \tag{3.52}$$

with

$$\begin{aligned}A &= \big(n - i - (2+6\beta)\, n^{2/3}\big)\, \mathrm{e}^{-2\beta[\vartheta_i]_+} + (2+6\beta)\, n^{2/3}\, \mathbb{1}_{\left\{i \geq n\left(1-\frac{1}{3}(p+h)\right)\right\}} \\ &\quad + \big(i + (-2+7\beta)\, n^{2/3}\big)\, \mathrm{e}^{-2\beta[-\vartheta_i]_+} + 2n^{2/3}, \\ B &= \big(n - i + (-2+7\beta)\, n^{2/3}\big)\, \mathrm{e}^{-2\beta[\vartheta_i]_+} + 2n^{2/3} \\ &\quad + \big(i - (2+6\beta)\, n^{2/3}\big)\, \mathrm{e}^{-2\beta[-\vartheta_i]_+},\end{aligned} \tag{3.53}$$

and the ratio

$$\frac{r^+(i, i+1)}{r^-(i, i+1)} = \frac{C}{D} \tag{3.54}$$

with

$$\begin{aligned}C &= \big(n - i + (-2+7\beta)\, n^{2/3}\big)\, \mathrm{e}^{-2\beta[\vartheta_i]_+} + 2n^{2/3}, \\ D &= \big(n - i - (2+6\beta)\, n^{2/3}\big)\, \mathrm{e}^{-2\beta[\vartheta_i]_+} + (2+6\beta)\, n^{2/3}\, \mathbb{1}_{\left\{i \geq n\left(1-\frac{1}{3}(p+h)\right)\right\}}.\end{aligned} \tag{3.55}$$

Note from (3.52) that for $\vartheta_i \geq 0$ (i.e., $i \leq \frac{n}{2}(1 - p^{-1}h)$, in which case also $i < n(1 - \frac{1}{3}(p+h))$),

$$\frac{r^-(i, i+1) + r^-(i, i-1)}{r^+(i, i+1) + r^+(i, i-1)} \leq 1 + \frac{13\beta \mathrm{e}^{2\beta(p-h)}}{n^{1/3}}, \tag{3.56}$$

and from (3.54) it follows that in this case

$$\frac{r^+(i, i+1)}{r^-(i, i+1)} \leq 1 + \frac{3\,(3+13\beta)\, \mathrm{e}^{2\beta(p-h)}}{n^{1/3}\,(p+h)}. \tag{3.57}$$

Similarly, for $\vartheta_i < 0$ we have that

$$\frac{r^-(i, i+1) + r^-(i, i-1)}{r^+(i, i+1) + r^+(i, i-1)} \leq 1 + \frac{2\mathrm{e}^{2\beta(p+h)}}{n^{1/3}} \tag{3.58}$$

and

$$\frac{r^+(i, i+1)}{r^-(i, i+1)} \leq 1 + \frac{6\,(2+6\beta)}{n^{1/3}\,(p+h)}. \tag{3.59}$$

Combining (3.56)–(3.59), we get that, for all $1 \leq i \leq n-1$,

$$\frac{r^{-}(i, i+1) + r^{-}(i, i-1)}{r^{+}(i, i+1) + r^{+}(i, i-1)} \times \frac{r^{+}(i, i+1)}{r^{-}(i, i+1)} \leq 1 + Kn^{-1/3}, \tag{3.60}$$

where

$$K = \max\left\{ \mathrm{e}^{2\beta(p-h)} \left(\frac{9+39\beta}{p+h} + 13\beta \right), \; 2\mathrm{e}^{2\beta(p+h)} + \frac{12+36\beta}{p+h} \right\}. \tag{3.61}$$

Therefore

$$\frac{\mathbb{P}_k^{X^+}\left[X_t^+ \text{ follows trajectory } \gamma^\star\right]}{\mathbb{P}_k^{X^-}\left[X_t^- \text{ follows trajectory } \gamma^\star\right]} \leq \prod_{i=k}^{m-1} \left(1 + \frac{K}{n^{1/3}}\right) \leq \mathrm{e}^{Kn^{-1/3}(m-k)}. \tag{3.62}$$

□

An application of the path-comparison methods used in Step 2 of the proof of Lemma 3.8 yields the following.

Corollary 3.9 *With $\mathbb{P}_{\mathrm{ER}_n(p)}$-probability tending to 1 as $n \to \infty$ the following is true. For every $0 \leq k < m \leq n$,*

$$\frac{\max_{\sigma \in A_k} \mathbb{E}_\sigma\left[\tau_{A_m} < \tau_{A_k}\right]}{\min_{\sigma \in A_k} \mathbb{E}_\sigma\left[\tau_{A_m} < \tau_{A_k}\right]} \leq [1 + o_n(1)]\, \mathrm{e}^{K(m-k)n^{-1/3}} \tag{3.63}$$

with $K = K(\beta, h, p) \in (0, \infty)$.

4 Capacity Bounds

The goal of this section is to derive various capacity bounds that will be needed to prove Theorem 1.4 in Sects. 6–7. In Sect. 4.1 we derive capacity bounds for the processes $\{\xi_t^l\}_{t\geq 0}$ and $\{\xi_t^u\}_{t\geq 0}$ on K_n introduced in (2.6) (Lemma 4.1 below). In Sect. 4.2 we do the same for the process $\{\xi_t\}_{t\geq 0}$ on $\mathrm{ER}_n(p)$ (Lemma 4.2 below). In Sect. 4.3 we use the bounds to rank-order the mean return times to $A_{\mathbf{M}^\mathbf{l}}$, $A_\mathbf{M}$ and $A_{\mathbf{M}^\mathbf{u}}$, respectively (Lemma 4.3 below). This ordering will be needed in the construction of a coupling in Sect. 6.

Define the Dirichlet form for $\{\xi_t\}_{t\geq 0}$ by

$$\mathscr{E}(f, f) = \frac{1}{2} \sum_{\sigma, \sigma' \in S_n} \mu(\sigma)\, r(\sigma, \sigma') \left[f(\sigma) - f(\sigma')\right]^2, \qquad f\colon S_n \to [0, 1], \tag{4.1}$$

and for $\{\theta_t^u\}_{t\geq 0}$ and $\{\theta_t^l\}_{t\geq 0}$ by

$$\mathscr{E}^u(f,f) = \frac{1}{2}\sum_{a,a'\in\Gamma_n} \nu^u(a)\, q^u\left(a,a'\right)\left[f(a)-f\left(a'\right)\right]^2, \tag{4.2}$$

$$\mathscr{E}^l(f,f) = \frac{1}{2}\sum_{a,a'\in\Gamma_n} \nu^l(a)\, q^l\left(a,a'\right)\left[f(a)-f\left(a'\right)\right]^2, \qquad f\colon\ \Gamma_n \to [0,1].$$

For $A, B \subseteq S_n$, define the capacity between A and B for $\{\xi_t\}_{t\geq 0}$ by

$$\mathrm{cap}(A,B) = \min_{f\in Q(A,B)} \mathscr{E}(f,f), \tag{4.3}$$

where

$$Q(A,B) = \left\{f\colon\ S_n \to [0,1],\ f_{|A}\equiv 1,\ f_{|B}\equiv 0\right\}, \tag{4.4}$$

and similarly for $\mathrm{cap}^u(A,B)$ and $\mathrm{cap}^l(A,B)$.

4.1 Capacity Bounds on K_n

First we derive capacity bounds for $\{\xi_t^l\}_{t\geq 0}$ and $\{\xi_t^u\}_{t\geq 0}$ on K_n. A useful reformulation of (4.3) is given by

$$\mathrm{cap}(A,B) = \sum_{\sigma\in A}\sum_{\sigma'\in S_n} \mu(\sigma)\, r\left(\sigma,\sigma'\right)\mathbb{P}_\sigma(\tau_B < \tau_A). \tag{4.5}$$

Lemma 4.1 (Capacity Bounds for $\{\xi_t^u\}_{t\geq 0}$ and $\{\xi_t^l\}_{t\geq 0}$) *For $a, b \in [\mathbf{m}^u, \mathbf{s}^u]$ with $a < b$,*

$$\frac{\left(1-b+\frac{2}{n}\right)}{2n(b-a)^2} \leq \frac{\mathrm{cap}^u(a,b)}{C^\star(b)} \leq \frac{n(1-b)}{2} \tag{4.6}$$

with

$$C^\star(b) = \frac{1}{z^u}\mathrm{e}^{-\beta n\Psi^u(b)}\binom{n}{\frac{n}{2}(1+\min(b,\mathbf{t}^u))}. \tag{4.7}$$

For $a, b \in \left[\mathbf{m}^l, \mathbf{s}^l\right]$ with $a < b$, analogous bounds hold for $\mathrm{cap}^l(a,b)$.

Proof We will prove the upper and lower bounds only for $\text{cap}^u(a,b)$, the proof for $\text{cap}^l(a,b)$ being identical. Note from the definition in (4.3) that

$$\begin{aligned}&\text{cap}^u(a,b) && (4.8)\\ &= \min_{f\in Q(a,b)} \sum_{i=0}^{\frac{n}{2}(b-a)-1} \nu^u\left(a+\frac{2i}{n}\right) q^u\left(a+\frac{2i}{n}, a+\frac{2(i+1)}{n}\right)\\ &\quad\times\left[f\left(a+\frac{2i}{n}\right)-f\left(a+\frac{2(i+1)}{n}\right)\right]^2,\end{aligned}$$

where it is easy to see that the set $Q(a,b)$ in (4.4) may be reduced to

$$Q(a,b)=\left\{f\colon \Gamma_n\to[0,1],\ f(x)=1 \text{ for } x\le a,\ f(x)=0 \text{ for } x\ge b\right\}. \tag{4.9}$$

Note that for every $f\in Q(a,b)$ there is some $0\le i\le\frac{n}{2}(b-a)-1$ such that

$$\left|f\left(a+\frac{2i}{n}\right)-f\left(a+\frac{2(i+1)}{n}\right)\right|\ge\left(\frac{n}{2}(b-a)\right)^{-1}. \tag{4.10}$$

Also note that, by (2.13),

$$\begin{aligned}&\nu^u\left(a+\frac{2i}{n}\right) q^u\left(a+\frac{2i}{n}, a+\frac{2(i+1)}{n}\right) && (4.11)\\ &=\frac{1}{z^u}\frac{n}{2}\left(1-a-\frac{2i}{n}\right)\mathrm{e}^{-\beta n\max\left\{\Psi^u\left(a+\frac{2i}{n}\right),\Psi^u\left(a+\frac{2(i+1)}{n}\right)\right\}}\binom{n}{\frac{n}{2}(1+a)+i},\end{aligned}$$

and that, for any $\delta\in\mathbb{R}$,

$$\Psi^u(a+\delta)-\Psi^u(a)=-\delta\left(pa+h^u+\frac{p}{2}\delta\right), \tag{4.12}$$

so that

$$\max\left\{\Psi^u\left(a+\frac{2i}{n}\right),\Psi^u\left(a+\frac{2(i+1)}{n}\right)\right\}\le\frac{2}{n}\left(pa+h+\frac{p}{n}\right)+\Psi^u\left(a+\frac{2i}{n}\right). \tag{4.13}$$

Combining, (4.9)–(4.13) with $\delta = \frac{2}{n}$, we get

$$
\begin{aligned}
&\mathrm{cap}^u\,(a,b) \qquad\qquad (4.14)\\
&\geq \min_{0\leq i\leq \frac{n}{2}(b-a)-1} \frac{2\left(1-b+\frac{2}{n}\right)\mathrm{e}^{-2\beta(p+h^u)}}{nz^u\,(b-a)^2}\mathrm{e}^{-\beta n\Psi^u\left(a+\frac{2i}{n}\right)}\binom{n}{\frac{n}{2}\,(1+a)+i}\\
&= \frac{2\left(1-b+\frac{2}{n}\right)\mathrm{e}^{-2\beta(p+h^u+\frac{p}{n})}}{nz^u\,(b-a)^2}\mathrm{e}^{-\beta n\Psi^u(\min(b,\mathbf{t}^u))}\binom{n}{\frac{n}{2}\,(1+\min(b,\mathbf{t}^u))},
\end{aligned}
$$

where we use (4.12) plus the fact that, by the definition of $\mathbf{m}^u$, $\mathbf{t}^u$, $\mathbf{s}^u$, for $a,b\in[\mathbf{m}^u,\mathbf{s}^u]$ with $a<b$, the function $i\mapsto \mathrm{e}^{-\beta n\Psi^u\left(a+\frac{2i}{n}\right)}\binom{n}{\frac{n}{2}(1+a)+i}$ is decreasing on $[\mathbf{m}^u,\mathbf{t}^u]$ and increasing on $[\mathbf{t}^u,\mathbf{s}^u]$. This settles the lower bound in (4.6).

Arguments similar to the ones above give

$$
\begin{aligned}
\mathrm{cap}^u\,(a,b) &\leq \nu^u\left(\min(b,\mathbf{t}^u)-\frac{2}{n}\right)q^u\left(\min(b,\mathbf{t}^u)-\frac{2}{n},\min(b,\mathbf{t}^u)\right)\\
&\leq \frac{n\,(1-b)\,\mathrm{e}^{2\beta(p+h^u)}}{2z^u}\mathrm{e}^{-\beta n\Psi^u(\min(b,\mathbf{t}^u))}\binom{n}{\frac{n}{2}\,(1+\min(b,\mathbf{t}^u))},
\end{aligned}
\tag{4.15}
$$

where for the first equality we use the test function $f\equiv 1$ on $\left[-1,\min(b,\mathbf{t}^u)-\frac{2}{n}\right]$ and $f\equiv 0$ on $[\min(b,\mathbf{t}^u),1]$ in (4.9). □

4.2 *Capacity Bounds on* $\mathrm{ER}_n(p)$

Next we derive capacity bounds for $\{\xi_t\}_{t\geq 0}$ on $\mathrm{ER}_n(p)$. The proof is analogous to what was done in Lemma 4.1 for $\{\theta_t^u\}_{t\geq 0}$ and $\{\theta_t^l\}_{t\geq 0}$ on K_n.

Define the set of direct paths between $A\subseteq S_n$ and $B\subseteq S_n$ by

$$
\mathscr{L}_{A,B}=\left\{\gamma=(\gamma_0,\ldots,\gamma_{|\gamma|})\colon A\to B:\ |\gamma_{i+1}|=|\gamma_i|+1 \text{ for all } \gamma_i\in\gamma\right\},
\tag{4.16}
$$

which may be empty. Recall from (3.12) that $\vartheta_k = p(1-\frac{k}{n})-h$.

Lemma 4.2 (Capacity Bounds for $\{\xi_t\}_{t\geq 0}$**)** *With* $\mathbb{P}_{\mathrm{ER}_n(p)}$*-probability tending to* 1 *as* $n\to\infty$ *the following is true. For every* $0\leq k<k'\leq n$ *and every* $\varrho\colon\mathbb{N}\to\mathbb{R}_+$

satisfying $\lim_{n\to\infty} \varrho(n) = \infty$,

$$\mathrm{cap}\,(A_k, A_{k'}) \le \frac{1}{Z}\,\mathrm{e}^{-\beta H_n(\boxminus)}\; O\big(\varrho(n) n^{11/6}\big)\binom{n}{k_m}\,\mathrm{e}^{-\beta\, 2k_m \vartheta_{k_m}}, \tag{4.17}$$

$$\mathrm{cap}\,(A_k, A_{k'}) \ge \frac{1}{Z}\,\mathrm{e}^{-\beta H_n(\boxminus)}\; \Omega\left(n^{-1}\mathrm{e}^{-\left(\beta+\frac{1}{\sqrt{3}}\right)\sqrt{\log n}}\right)\binom{n}{k_m}\,\mathrm{e}^{-\beta\, 2k_m \vartheta_{k_m}},$$

where

$$k_m = \mathrm{argmin}_{k\le j\le k'}\binom{n}{j}\mathrm{e}^{-\beta\, 2j\theta_j}. \tag{4.18}$$

Proof Recall from (4.1) and (4.3) that

$$\mathrm{cap}\,(A_k, A_{k'}) = \min_{f\in Q} \mathscr{E}\,(f, f) = \min_{f\in Q}\; \frac{1}{2}\sum_{\sigma,\xi\in S_n} \mu\,(\sigma)\, r\,(\sigma,\xi)\,[f\,(\sigma) - f\,(\xi)]^2\,, \tag{4.19}$$

where

$$Q(A_k, A_{k'}) = \left\{f:\ S_n \to [0,1]:\ f|_{A_k} \equiv 1,\ f|_{A_{k'}} \equiv 0\right\}. \tag{4.20}$$

The proof comes in three Steps.

1. We first prove the upper bound in (4.17). Let $B = \bigcup_{j=k}^{k_m-1} A_j$, and note that, by (1.7),

$$\begin{aligned}
\mathrm{cap}\left(A_k, A_{k'}\right) &\le \frac{1}{2}\sum_{\sigma,\xi\in S_n} \mu\,(\sigma)\, r\,(\sigma,\xi)\,[\mathbb{1}_B\,(\sigma) - \mathbb{1}_B\,(\xi)]^2 \\
&= \sum_{\sigma\in A_{k_m-1}}\ \sum_{\xi\in A_{k_m}} \mu\,(\sigma)\, r\,(\sigma,\xi) \\
&= \frac{1}{Z}\sum_{\sigma\in A_{k_m-1}}\ \sum_{\xi\in A_{k_m},\xi\sim\sigma} \mathrm{e}^{-\beta\max\{H_n(\xi), H_n(\sigma)\}} \\
&= \frac{1}{Z}\left(\sum_{\sigma\in A_{k_m-1}}\ \sum_{\substack{\xi\in A_{k_m},\xi\sim\sigma\\ H_n(\sigma)\ge H_n(\xi)}} \mathrm{e}^{-\beta H_n(\sigma)} + \sum_{\sigma\in A_{k_m-1}}\ \sum_{\substack{\xi\in A_{k_m},\xi\sim\sigma\\ H_n(\sigma)< H_n(\xi)}} \mathrm{e}^{-\beta H_n(\xi)}\right) \\
&\le \frac{1}{Z}\max\{k_m, n-k_m\}\left(\sum_{\sigma\in A_{k_m-1}} \mathrm{e}^{-\beta H_n(\sigma)} + \sum_{\xi\in A_{k_m}} \mathrm{e}^{-\beta H_n(\xi)}\right).
\end{aligned} \tag{4.21}$$

Recall from (3.4) that ϕ_i^k denotes the cardinality of the set of all $\sigma \in A_k$ with $|\partial_E \sigma| = pk\,(n-k) + i$. Note from (1.3) that for any $\xi \in A_{k_m}$ such that $|\partial_E \xi| = pk_m\,(n-k_m) + i$,

$$\mathrm{e}^{-\beta H_n(\xi)} = \mathrm{e}^{-\beta H_n(\boxminus)} \mathrm{e}^{-\beta\,(2k_m \vartheta_{k_m} + \frac{2i}{n})}. \tag{4.22}$$

There are $\binom{n}{k_m}$ terms in the sum, and therefore we get

$$\begin{aligned}\textstyle\sum_{\xi \in A_{k_m}} \mathrm{e}^{-\beta H_n(\xi)} &= \mathrm{e}^{-\beta H_n(\boxminus)} \textstyle\sum_{i=-pk_m(n-k_m)}^{(1-p)k_m(n-k_m)} \phi_i^{k_m} \mathrm{e}^{-\beta\left(2k_m\vartheta_{k_m} + \frac{2i}{n}\right)} \\ &= \mathrm{e}^{-\beta H_n(\boxminus)} \left(\textstyle\sum_{i<-Y} \phi_i^{k_m} \mathrm{e}^{-\beta\left(2k_m\vartheta_{k_m} + \frac{2i}{n}\right)} + \textstyle\sum_{i \geq -Y} \mathrm{e}^{-\beta\left(2k_m\vartheta_{k_m} + \frac{2i}{n}\right)} \right) \\ &\leq \mathrm{e}^{-\beta H_n(\boxminus)} \left(\tbinom{n}{k_m} \mathrm{e}^{-\beta(2k_m\vartheta_{k_m} - \frac{2Y}{n})} + \textstyle\sum_{i<-Y} \phi_i^{k_m} \mathrm{e}^{-\beta\left(2k_m\vartheta_{k_m} + \frac{2i}{n}\right)} \right)\end{aligned} \tag{4.23}$$

with $Y = \sqrt{\log(\varrho(n)^2 n^{5/6}) k_m(n-k_m)}$. The choice of Y will become clear shortly. The summand in the right-hand side can be bounded as follows. By the sandwich in (3.2) in Lemma 3.2, the sum over $i < -Y$ can be restricted to $-cn^{3/2} \leq i < -Y$, since with high probability no configuration has a boundary size that deviates by more than $cn^{3/2}$ from the mean. But, using Lemma 3.3, we can also bound from above the number of configurations that deviate by at most Y from the mean, i.e., we can bound $\phi_i^{k_m}$ for $-cn^{3/2} \leq i < -Y$. Taking a union bound over $0 \leq k \leq n$ and $-cn^{3/2} \leq i < -Y$, we get

$$\mathbb{P}\left[\bigcup_{k=0}^{n} \bigcup_{i=-cn^{3/2}}^{-Y} \left\{ \phi_i^{k_m} \geq \varrho\,(n)\, n^{5/2} \binom{n}{k_m} \mathrm{e}^{-\frac{2i^2}{k_m\,(n-k_m)}} \right\} \right] \leq \frac{1}{\varrho\,(n)}. \tag{4.24}$$

Thus, with $\mathbb{P}_{\mathrm{ER}_n(p)}$-probability at least $\geq 1 - \frac{1}{\varrho(n)}$,

$$\begin{aligned}\textstyle\sum_{i<-Y} \phi_i^{k_m} \mathrm{e}^{-\beta\left(2k_m\,\theta_{k_m} + \frac{2i}{n}\right)} &\leq \textstyle\sum_{i>Y} \varrho\,(n)\, n^{5/2} \binom{n}{k_m} \mathrm{e}^{-2i\left(\frac{i}{k_m\,(n-k_m)} - \frac{\beta}{n}\right)} \mathrm{e}^{-\beta\,2k_m\,\theta_{k_m}} \\ &\leq \varrho\,(n)\, n^{5/2} \tbinom{n}{k_m} \mathrm{e}^{-\beta\,2k_m\,\theta_{k_m}} \mathrm{e}^{-2\log(\varrho(n)n^{5/6})} \\ &\leq \frac{n^{5/6}}{\varrho(n)} \tbinom{n}{k_m} \mathrm{e}^{-\beta\,2k_m\,\theta_{k_m}},\end{aligned} \tag{4.25}$$

where we use that, for $i > Y$ and n sufficiently large,

$$\frac{i}{k_m\,(n-k_m)} - \frac{\beta}{n} \geq \sqrt{\frac{\log(\varrho\,(n)^2\, n^{5/6})}{k_m\,(n-k_m)}} - \frac{\beta}{n} \geq \sqrt{\frac{\log(\varrho\,(n)\, n^{5/6})}{k_m\,(n-k_m)}}. \tag{4.26}$$

The above inequality also clarifies our choice of Y. Substituting it into (4.23), we see that

$$\sum_{\xi \in A_{k_m}} \mathrm{e}^{-\beta H_n(\xi)} \leq [1 + o_n(1)]\, \mathrm{e}^{-\beta H_n(\boxminus)} \mathrm{e}^{\frac{2\beta Y}{n}} \mathrm{e}^{-\beta\, 2k_m\, \vartheta_{k_m}} \binom{n}{k_m} \tag{4.27}$$
$$= O\big(\varrho(n) n^{5/6}\big)\, \mathrm{e}^{-\beta H_n(\boxminus)} \mathrm{e}^{-\beta\, 2k_m \vartheta_{k_m}} \binom{n}{k_m}.$$

A similar bound holds for $\sum_{\xi \in A_{k_m - 1}} \mathrm{e}^{-\beta H_n(\xi)}$. A union bound over $1 \leq k_m \leq n$ increases the exponent $\frac{5}{6}$ to $\frac{11}{6}$. Together with (4.21), this proves the upper bound in (4.17).

2. We next derive a combinatorial bound that will be used later for the proof of the lower bound in (4.17). Note that if $f \in Q(A_k, A_{k'})$ and $\gamma \in \mathscr{L}_{A_k, A_{k'}}$ (recall (4.16)), then there must be some $1 \leq i \leq k' - k$ such that

$$|f(\gamma_i) - f(\gamma_{i+1})| \geq \big(k' - k\big)^{-1}. \tag{4.28}$$

A simple counting argument shows that

$$\big|\mathscr{L}_{A_k, A_{k'}}\big| = \binom{n}{k} \frac{(n-k)!}{(n-k')!}, \tag{4.29}$$

since for each $\sigma \in A_k$ there are $(n-k) \times (n-k-1) \times \cdots \times \big(n - k' + 1\big)$ paths in $\mathscr{L}_{A_k, A_{k'}}$ from σ to $A_{k'}$. Let

$$b_i = \Big|\Big\{(\sigma, \xi) \in A_{k+i-1} \times A_{k+i} \colon\ |f(\sigma) - f(\xi)| \geq \big(k' - k\big)^{-1},\ \sigma \sim \xi\Big\}\Big|,$$
$$1 \leq i \leq k' - k. \tag{4.30}$$

We claim that

$$\exists\, 1 \leq i_\blacklozenge \leq k' - k \colon \qquad b_{i_\blacklozenge} \geq \frac{k}{k' - k} \binom{n}{k + i_\blacklozenge}. \tag{4.31}$$

Indeed, the number of paths in $\mathscr{L}_{A_k, A_{k'}}$ that pass through $\sigma \in A_{k + i_\blacklozenge - 1}$ followed by a move to $\xi \in A_{k + i_\blacklozenge}$ equals

$$z_{i_\blacklozenge} = \frac{\big(k + i_\blacklozenge - 1\big)!}{k!} \times \frac{\big(n - k - i_\blacklozenge\big)!}{(n - k')!}, \tag{4.32}$$

where the first term in the product counts the number of paths from $\sigma \in A_{k + i_\blacklozenge + 1}$ to A_k, while the second term counts the number of paths from $\xi \in A_{k + i_\blacklozenge}$ to $A_{k'}$.

Thus, if (4.31) fails, then

$$\sum_{i=1}^{k'-k} b_i z_i < \frac{k}{k'-k} \sum_{i=1}^{k-k'} \frac{1}{k+i} \frac{n!}{k!\,(n-k')!} \leq \frac{n!}{k!\,(n-k')!} = \left|\mathscr{L}_{A_k, A_{k'}}\right|, \tag{4.33}$$

which in turn implies that (4.28) does not hold for some $\gamma \in \mathscr{L}_{A_k,A_{k'}}$ (use that $b_i z_i$ counts the paths that satisfy condition (4.28)), which is a contradiction. Hence the claim in (4.31) holds.

3. In this part we prove the lower bound in (4.17). By Lemma 3.3 we have that, with $\mathbb{P}_{\mathrm{ER}_n(p)}$-probability at least $1 - \frac{1}{\varrho(n)}$, for any $Y \geq 0$,

$$\sum_{j \geq \sqrt{Y(k+i_\blacklozenge)(n-k-i_\blacklozenge)}} \phi_j^{k+i_\blacklozenge} \leq \varrho\,(n) \binom{n}{k+i_\blacklozenge} \mathrm{e}^{-2Y}. \tag{4.34}$$

Picking $Y = \log(\varrho\,(n)\,k^{-1} 2n^{3/2})$, we get that

$$\sum_{j \geq \sqrt{Y(k+i_\blacklozenge)(n-k-i_\blacklozenge)}} \phi_j^{k+i_\blacklozenge} \leq \frac{1}{4} \frac{k^2 \varrho(n)}{\varrho(n)^2 n^3} \binom{n}{k+i_\blacklozenge} \leq \frac{1}{2} \frac{k}{n\,(k'-k)} \binom{n}{k+i_\blacklozenge}, \tag{4.35}$$

and so at least half of the configurations contributing to $b_{i_\blacklozenge}$ have an edge-boundary of size at most

$$p\left(k+i_\blacklozenge\right)\left(n-k-i_\blacklozenge\right) + \sqrt{Y\left(k+i_\blacklozenge\right)\left(n-k-i_\blacklozenge\right)}. \tag{4.36}$$

If $\xi \in A_{k+i_\blacklozenge}$ is such a configuration, then by Lemma 3.2 the same is true for any $\sigma \sim \xi$ (i.e., configurations differing at only one vertex), since

$$|\partial_E \sigma| \leq |\partial_E \xi| + \max_{v \in \sigma \triangle \xi} \deg(v) \leq |\partial_E \xi| + pn + o\left(\rho(n)\sqrt{n \log n}\right). \tag{4.37}$$

This implies

$$\begin{aligned}
\mathscr{E}(f, f) &= \frac{1}{2} \sum_{\sigma,\xi \in S_n} \mu(\sigma)\, r(\sigma,\xi)\, [f(\sigma) - f(\xi)]^2 \\
&\geq \frac{1}{Zn^{2/3}} \sum_{\xi \in A_{k+i_\blacklozenge}} \sum_{\sigma \in A_{k+i_\blacklozenge - 1}} \mathrm{e}^{-\beta \max\{H_n(\sigma), H_n(\xi)\}}
\end{aligned}$$

$$\geq \mathrm{e}^{-\beta H_n(\boxminus)} \frac{k}{2Zn^2} \binom{n}{k+i_\blacklozenge}$$
$$\times \exp\left(-\beta\left(2(k+i_\blacklozenge)\theta_{k+i_\blacklozenge} + 2\frac{\sqrt{Y\left(k+i_\blacklozenge\right)\left(n-k-i_\blacklozenge\right)}}{n}\right)\right). \quad (4.38)$$

Therefore

$$\mathscr{E}(f, f) \geq \mathrm{e}^{-\beta H_n(\boxminus)} \mathrm{e}^{-\beta\sqrt{Y}} \min_{1\leq i\leq k'-k} \frac{k}{2Zn^2}\binom{n}{k+i} \mathrm{e}^{-\beta(2(k+i)\vartheta_{k+i})} \quad (4.39)$$
$$= \mathrm{e}^{-\beta H_n(\boxminus)} \mathrm{e}^{-\beta\sqrt{Y}} \frac{k}{2Zn^2}\binom{n}{k_m} \mathrm{e}^{-\beta(2(k+i)\vartheta_{k-m})}.$$

Since (4.39) is true for any $f \in Q(A_k, A_{k'})$, the lower bound in (4.17) follows, with k_m defined in (4.18). □

4.3 *Hitting Probabilities on* $\mathrm{ER}_n(p)$

Let $\mu_{A_\mathbf{M}}$ be the equilibrium distribution μ conditioned to the set $A_\mathbf{M}$. Write $\mathbb{P}^l$ and $\mathbb{P}^u$ to denote the laws of the processes $\{\xi_t^l\}_{t\geq 0}$ and $\{\xi_t^u\}_{t\geq 0}$, respectively. The following lemma is the crucial sandwich for comparing the crossover times of the dynamics on $\mathrm{ER}_n(p)$ and the perturbed dynamics on K_n.

Lemma 4.3 (Rank Ordering of Hitting Probabilities) *With* $\mathbb{P}_{\mathrm{ER}_n(p)}$*-probability tending to* 1 *as* $n \to \infty$,

$$\begin{aligned} &\max\nolimits_{\xi\in A_{\mathbf{M}^l}} \mathbb{P}_\xi^l\left[\tau_{\mathbf{S}^l} < \tau_{\mathbf{M}^l}\right] \leq \mathbb{P}_{\mu_{A_\mathbf{M}}}\left[\tau_\mathbf{S} < \tau_\mathbf{M}\right] \\ &\leq \min\nolimits_{\sigma\in A_{\mathbf{M}^u}} \mathbb{P}_\sigma^u\left[\tau_{\mathbf{S}^u} < \tau_{\mathbf{M}^u}\right]. \end{aligned} \quad (4.40)$$

Proof The proof comes in three Steps.

1. Recall from (1.10) that the magnetization of $\sigma \in A_k$ is $m(\sigma) = 2\frac{k}{n} - 1$. We first observe that the maximum and the minimum in (4.40) are redundant, because by symmetry

$$\begin{aligned} \max\nolimits_{\xi\in A_k} \mathbb{P}_\xi^l\left[\tau_{A_{k'}} < \tau_{A_k}\right] &= \min\nolimits_{\xi\in A_k} \mathbb{P}_\xi^l\left[\tau_{A_{k'}} < \tau_{A_k}\right], \\ \min\nolimits_{\xi\in A_k} \mathbb{P}_\xi^u\left[\tau_{A_{k'}} < \tau_{A_k}\right] &= \max\nolimits_{\xi\in A_k} \mathbb{P}_\xi^u\left[\tau_{A_{k'}} < \tau_{A_k}\right]. \end{aligned} \quad (4.41)$$

Recall that $\{\xi_t^l\}_{t\geq 0}$ is the Markov process on S_n governed by the Hamiltonian H_n^l in (2.3), and that the associated magnetization process $\{\theta_t^l\}_{t\geq 0} = \{m(\xi_t^l)\}_{t\geq 0}$ is a Markov process on the set Γ_n in (1.15) with transition rates given by q^l in (2.8).

Denoting by $\hat{\mathbb{P}}^l$ the law of $\{\theta_t^l\}_{t\geq 0}$, we get from (4.5) that for any $0 \leq k \leq k' < n$, and with $a = \frac{2k}{n} - 1$ and $b = \frac{2k'}{n} - 1$,

$$\begin{aligned} \mathrm{cap}^l\,(a,b) &= \sum_{u\in\Gamma_n} \nu^l\,(a)\, q^l\,(a,u)\, \hat{\mathbb{P}}_a^l\,[\tau_b < \tau_a] \\ &= \nu^l\,(a) \left[q^l\left(a, a+\frac{2}{n}\right) + q^l\left(a, a-\frac{2}{n}\right)\right] \hat{\mathbb{P}}_a^l\,[\tau_b < \tau_a]\,, \end{aligned} \tag{4.42}$$

and therefore

$$\begin{aligned} \max\nolimits_{\xi\in A_k} \mathbb{P}_\xi^l \left[\tau_{A_{k'}} < \tau_{A_k}\right] &= \hat{\mathbb{P}}_a^l\,[\tau_b < \tau_a] \\ &= \left[\nu^l\,(a)\left(q^l\left(a, a+\tfrac{2}{n}\right) + q^l\left(a, a-\tfrac{2}{n}\right)\right)\right]^{-1} \mathrm{cap}^l\,(a,b)\,. \end{aligned} \tag{4.43}$$

By (2.13), using the abbreviations

$$\Psi_1 = \max\left\{\Psi^l\,(a)\,,\Psi^l\left(a+\frac{2}{n}\right)\right\}, \qquad \Psi_2 = \max\left\{\Psi^l\,(a)\,,\Psi^l\left(a-\frac{2}{n}\right)\right\}, \tag{4.44}$$

we have, with the help of (4.12),

$$\begin{aligned} &\nu^l\,(a)\left(q^l\left(a, a+\frac{2}{n}\right) + q^l\left(a, a-\frac{2}{n}\right)\right) \\ &\quad= \frac{1}{z^l}\frac{n}{2}\binom{n}{\frac{n}{2}\,(1+a)}\left((1-a)\,\mathrm{e}^{-\beta n\Psi_1} + (1+a)\,\mathrm{e}^{-\beta n\Psi_2}\right) \\ &\quad\geq \frac{1}{z^l} n \mathrm{e}^{-2\beta\left(p|a|+h^l+\frac{p}{n}\right)}\mathrm{e}^{-\beta n\Psi^l(a)}\binom{n}{\frac{n}{2}\,(1+a)}. \end{aligned} \tag{4.45}$$

From Lemma 4.1 we have that

$$\mathrm{cap}^l\,(a,b) \leq \frac{n\,(1-a)}{2z^l}\mathrm{e}^{-\beta n\Psi^l(b)}\binom{n}{\frac{n}{2}\,(1+b)}. \tag{4.46}$$

Putting (4.43), (4.45) and (4.46) together, we get

$$\begin{aligned} \max\nolimits_{\xi\in A_k} \mathbb{P}_\xi^l \left[\tau_{A_{k'}} < \tau_{A_k}\right] \leq &\tfrac{(1-a)}{2}\mathrm{e}^{2\beta\left(p|a|+h^l+\frac{p}{n}\right)}\mathrm{e}^{-\beta n\left[\Psi^l(b)-\Psi^l(a)\right]} \\ &\times\tbinom{n}{\frac{n}{2}(1+b)}\tbinom{n}{\frac{n}{2}(1+a)}^{-1}. \end{aligned} \tag{4.47}$$

Similarly, denoting by $\hat{\mathbb{P}}^u$ the law of $\{\theta_t^u\}_{t\geq 0}$, we have

$$\min_{\xi\in A_k} \mathbb{P}_\xi^u\left[\tau_{A_{k'}} < \tau_{A_k}\right] = \hat{\mathbb{P}}_a^u\left[\tau_b < \tau_a\right] \tag{4.48}$$
$$= \left[\nu^u(a)\left(q^u\left(a, a+\frac{2}{n}\right) + q^u\left(a, a-\frac{2}{n}\right)\right)\right]^{-1} \mathrm{cap}^u(a,b),$$

where

$$\begin{aligned}&\left[\nu^u(a)\left(q^u\left(a,a+\tfrac{2}{n}\right)+q^u\left(a,a-\tfrac{2}{n}\right)\right)\right]^{-1}\\ &\geq\left[\tfrac{n}{z^u}\mathrm{e}^{2\beta\left(p|a|+h^l+\frac{p}{n}\right)}\mathrm{e}^{-\beta n\Psi^u(a)}\tbinom{n}{\frac{n}{2}(1+a)}\right]^{-1},\end{aligned} \tag{4.49}$$

and, by Lemma 4.1,

$$\mathrm{cap}^u(a,b) \geq \frac{1}{2nz^u}\mathrm{e}^{-\beta n\Psi^u(b)}\binom{n}{\frac{n}{2}(1+b)}. \tag{4.50}$$

Putting (4.48)–(4.50) together, we get

$$\min_{\xi\in A_k}\mathbb{P}_\xi^u\left[\tau_{A_{k'}}<\tau_{A_k}\right]\geq\frac{1}{n}\mathrm{e}^{-\beta n\left[\Psi^l(b)-\Psi^l(a)\right]}\binom{n}{\frac{n}{2}(1+b)}\binom{n}{\frac{n}{2}(1+a)}^{-1}. \tag{4.51}$$

2. Recall from (4.5) that

$$\mathrm{cap}(A_k, A_{k'}) = \sum_{\sigma\in A_k}\sum_{\xi\in S_n}\mu(\sigma)r(\sigma,\xi)\mathbb{P}_\sigma\left[\tau_{A_{k'}}<\tau_{A_k}\right]. \tag{4.52}$$

Split

$$\begin{aligned}&\sum_{\sigma\in A_k}\sum_{\xi\in S_n}\mu(\sigma)r(\sigma,\xi)\\ &=\sum_{\sigma\in A_k}\sum_{\xi\in A_{k+1}}\mu(\sigma)r(\sigma,\xi)+\sum_{\sigma\in A_k}\sum_{\xi\in A_{k-1}}\mu(\sigma)r(\sigma,\xi)\\ &=\frac{1}{Z}\sum_{\sigma\in A_k}\sum_{\xi\in A_{k+1}}\mathrm{e}^{-\beta\max\{H_n(\sigma),H_n(\xi)\}}+\frac{1}{Z}\sum_{\xi\in A_k}\sum_{\xi' A_{k-1}}\mathrm{e}^{-\beta\max\{H_n(\sigma),H_n(\xi)\}}.\end{aligned} \tag{4.53}$$

By Lemma 3.3 and a reasoning similar to that leading to (4.38),

$$\begin{aligned}\sum_{\xi\in A_k} \mathrm{e}^{-\beta H_n(\xi)} &= \mathrm{e}^{-\beta H_n(\boxminus)} \sum_{i=-pk(n-k)}^{(1-p)k(n-k)} \phi_i^k \mathrm{e}^{-\beta\left(2k\vartheta_k+2\frac{i}{n}\right)} \\ &\geq \frac{1}{2}\binom{n}{k}\mathrm{e}^{-\beta H_n(\boxminus)}\mathrm{e}^{-\beta\left(2k\vartheta_k+2\frac{\sqrt{Yk(n-k)}}{n}\right)} \\ &\geq \frac{1}{2}\binom{n}{k}\mathrm{e}^{-\beta H_n(\boxminus)}\mathrm{e}^{-\beta\sqrt{\log(\sqrt{2\varrho(n)})}}\mathrm{e}^{-\beta\,2k\vartheta_k}\end{aligned} \tag{4.54}$$

with $Y = \log(\sqrt{2\varrho(n)})$. Indeed, by (3.6) fewer than $\frac{1}{2}\binom{n}{k}$ configurations in A_k have an edge-boundary of size $\geq pk\,(n-k) + \sqrt{k\,(n-k)\,Y}$. Moreover, if $\xi \sim \xi'$, then, by Lemma 3.2,

$$\mathrm{e}^{\beta[H_n(\xi')-H_n(\xi)]} \leq [1+o(1)]\,\mathrm{e}^{\beta(p+h)}, \tag{4.55}$$

and since we may absorb this constant inside the error term $\varrho\,(n)$, we get that

$$\begin{aligned}&\sum_{\sigma\in A_k}\sum_{\xi\in A_{k+1}} \mathrm{e}^{-\beta\max\{H_n(\sigma),H_n(\xi)\}} \\ &\geq \mathrm{e}^{-\beta H_n(\boxminus)}\frac{1}{2}\,(n-k)\binom{n}{k}\mathrm{e}^{-\frac{\beta}{2}\sqrt{\log\sqrt{2\varrho(n)}}}\mathrm{e}^{-\beta\,2k\,\vartheta_k},\end{aligned} \tag{4.56}$$

$$\begin{aligned}&\sum_{\sigma\in A_k}\sum_{\xi\in A_{k-1}} \mathrm{e}^{-\beta\max\{H_n(\sigma),H_n(\xi)\}} \\ &\geq \mathrm{e}^{-\beta H_n(\boxminus)}\frac{1}{2}k\binom{n}{k}\mathrm{e}^{-\frac{\beta}{2}\sqrt{\log\sqrt{2\varrho(n)}}}\mathrm{e}^{-\beta\,2k\,\vartheta_k},\end{aligned} \tag{4.57}$$

and hence

$$\begin{aligned}&\textstyle\sum_{\sigma\in A_k}\sum_{\xi\in S_n}\mu(\sigma)r(\sigma,\xi) \\ &\geq \mathrm{e}^{-\beta H_n(\boxminus)}\tfrac{1}{2Z}\tbinom{n}{k}\mathrm{e}^{-\frac{\beta}{2}\sqrt{\log\sqrt{2\varrho(n)}}}\mathrm{e}^{-\beta\,2k\,\vartheta_k}.\end{aligned} \tag{4.58}$$

3. Similar bounds can be derived for $\mathbb{P}_{\mu_{A_k}}\left[\tau_{A_{k'}} < \tau_{A_k}\right]$. Indeed, by Lemma 3.5, $r\,(\sigma,\xi) = 1$ for all $\sigma \in A_k$ and all but $O(n^{2/3})$ many configurations $\xi \in S_n$. Therefore

$$\begin{aligned}\mathrm{cap}(A_k, A_{k'}) &= n\,[1+o(1)]\sum_{\sigma\in A_k}\mu(\sigma)\mathbb{P}_\sigma\left[\tau_{A_{k'}} < \tau_{A_k}\right] \\ &= n\,[1+o(1)]\,\mu(A_k)\,\mathbb{P}_{\mu_{A_k}}\left[\tau_{A_{k'}} < \tau_{A_k}\right]\end{aligned} \tag{4.59}$$

and hence

$$\mathbb{P}_{\mu_{A_k}}\left[\tau_{A_{k'}} < \tau_{A_k}\right] = [1+o(1)]\,\frac{\mathrm{cap}(A_k, A_{k'})}{n\mu(A_k)}. \tag{4.60}$$

Note that

$$\mu(A_k) = \frac{1}{Z}\sum_{\sigma\in A_k} \mathrm{e}^{-\beta H_n(\sigma)}, \tag{4.61}$$

and we have already produced bounds for a sum like (4.61) in Lemma 4.2. Referring to (4.28), we see that

$$\mathbb{P}_{\mu_{A_k}}\left[\tau_{A_{k'}} < \tau_{A_k}\right] \le [1+o(1)]\,\frac{\mathrm{cap}(A_k, A_{k'})}{\frac{1}{Z}\,\mathrm{e}^{-(\beta+\frac{1}{\sqrt{3}})\sqrt{\log n}}\binom{n}{k}\mathrm{e}^{-\beta 2k\vartheta_k}}, \tag{4.62}$$

$$\mathbb{P}_{\mu_{A_k}}\left[\tau_{A_{k'}} < \tau_{A_k}\right] \ge [1+o(1)]\,\frac{\mathrm{cap}(A_k, A_{k'})}{\frac{1}{Z}\,n^{17/6}\mathrm{e}^{-(\beta+\frac{1}{\sqrt{3}})\sqrt{\log n}}\binom{n}{k}\mathrm{e}^{-\beta 2k\vartheta_k}}. \tag{4.63}$$

Finally, we note that if we let $\Delta_h = h - h^u$, then

$$\frac{\mathrm{e}^{-\beta n[\Psi^u(\mathbf{s}^u)-\Psi^u(\mathbf{m}^u)]}}{\mathrm{e}^{-\beta n[\Psi(\mathbf{s})-\Psi(\mathbf{m})]}} = \mathrm{e}^{\beta n C_{\beta,h,p}\Delta_h}, \tag{4.64}$$

where $C_{\beta,h,p}$ is a constant that depends on the parameters β, p and h. A similar expression follows for the ratio

$$\binom{n}{\frac{n}{2}(1+\mathbf{s}^u)}\binom{n}{\frac{n}{2}(1+\mathbf{m}^u)}^{-1}\left[\binom{n}{\frac{n}{2}(1+\mathbf{s})}\binom{n}{\frac{n}{2}(1+\mathbf{m})}^{-1}\right]^{-1}. \tag{4.65}$$

From this the statement of the lemma follows. □

5 Invariance Under Initial States and Refined Capacity Estimates

In this section we use Lemma 4.3 to control the time it takes $\{m(\xi_t)\}_{t\ge 0}$ to cross the interval $[\mathbf{t}^u, \mathbf{s}^u] \cap [\mathbf{t}^l, \mathbf{s}^l]$, which will be a good indicator of the time it takes $\{\xi_t\}_{t\ge 0}$ to reach the basin of the stable state $\mathbf{s}$. In particular, our aim is to control this time by comparing it with the time it takes $\{\theta_t^u\}_{t\ge 0}$ and $\{\theta_t^l\}_{t\ge 0}$ defined in (2.6) to do the same for $\mathbf{s}^u$ and $\mathbf{s}^l$. In Sect. 5.1 we derive bounds on the probability of certain rare events for the dynamics on $\mathrm{ER}_n(p)$ (Lemmas 5.1–5.4 below). In Sect. 5.2 we use these bounds to prove that hitting times are close to being uniform in the starting configuration.

5.1 Estimates for Rare Events

In this section we prove four lemmas that serve as a preparation for the coupling in Sect. 6. Lemma 5.1 shows that the dynamics starting anywhere in $A_{\mathbf{M}}$ is unlikely to stray away from $A_{\mathbf{M}}$ by much during a time window that is comparatively small. Lemma 5.2 bounds the total variation distance between two exponential random variable whose means are close. Lemma 5.3 bounds the tail of the distribution of the first time when all the vertices have been updated. Lemma 5.4 bounds the number of returns to $A_{\mathbf{M}}$ before $A_{\mathbf{S}}$ is hit.

We begin by deriving upper and lower bounds on the number of jumps $N_{\xi}(t)$ taken by the process $\{\xi_t\}_{t\geq 0}$ up to time t. By Lemma 3.4, the jump rate from any $\sigma \in S_n$ is bounded by

$$n\,\mathrm{e}^{-2\beta(p+h)} \leq \sum_{\sigma' \in S_n} r\left(\sigma, \sigma'\right) \leq n. \tag{5.1}$$

Hence $N_{\xi}(t)$ can be stochastically bounded from above by a Poisson random variable with parameter tn, and from below by a Poisson random variable with parameter $tn\mathrm{e}^{-2\beta(p+h)}$. It therefore follows that, for any $M \geq 0$,

$$\begin{aligned} &\mathbb{P}\left[N_{\xi}(t) \geq M\right] \leq \chi_M(nt), \\ &\mathbb{P}\left[N_{\xi}(t) < M\right] \leq 1 - \chi_M\left(nt\,\mathrm{e}^{-2\beta(p+h)}\right), \end{aligned} \tag{5.2}$$

where we abbreviate $\chi_M(u) = \mathrm{e}^{-u}\sum_{k\geq M} u^k/k!$, $u \in \mathbb{R}$, $M \in \mathbb{N}$.

5.1.1 Localisation

The purpose of the next lemma is to show that the probability of $\{\xi_t\}_{t\geq 0}$ straying too far from $A_{\mathbf{M}}$ during its first $n^2 \log n$ jumps is very small. The seemingly arbitrary choice of $n^2 \log n$ is in fact related to the Coupon Collector's problem.

Lemma 5.1 (Localisation) *Let $\xi_0 \in A_{\mathbf{M}}$, $T = \inf\left\{t \geq 0\colon N_{\xi}(t) \geq n^2 \log n\right\}$, and let $C_1 \in \mathbb{R}$ be a sufficiently large constant, possibly dependent on p and h (but not on n). Then*

$$\mathbb{P}_{\xi_0}\left[\xi_t \in A_{\mathbf{M}+C_1 n^{5/6}} \textit{ for some } 0 \leq t \leq T\right] \leq \mathrm{e}^{-n^{2/3}}. \tag{5.3}$$

Proof The idea of the proof is to show that $\{\xi_t\}_{t\geq 0}$ returns many times to $A_{\mathbf{M}}$ before reaching $A_{\mathbf{M}+C_1 n^{5/6}}$. The proof comes in three Steps.

1. We begin by showing that $T \leq n^2 \log n$ with probability $\geq 1 - \mathrm{e}^{-n^3}$ (in other words, it takes less than $n^2 \log n$ time to make $n^2 \log n$ steps). Indeed, by the second line of (5.2),

$$\begin{aligned}
&\mathbb{P}\left[T > n^2 \log n\right] \qquad (5.4)\\
&\quad= \mathbb{P}\left[N_\xi\left(n^2 \log n\right) < n^2 \log n\right]\\
&\quad\leq 1 - \chi_{n^2 \log n}\left((n^3 \log n)\, \mathrm{e}^{-2\beta(p+h)}\right)\\
&\quad\leq \sum_{k=0}^{n^2 \log n} \exp\left(-(n^3 \log n)\, \mathrm{e}^{-2\beta(p+h)} + k \log\left(\frac{e\, n^3 \log n}{k}\right)\right)\\
&\quad\leq (n^2 \log n) \exp\left(-(n^3 \log n)\, \mathrm{e}^{-2\beta(p+h)} + n^{5/2}\right)\\
&\quad\leq \mathrm{e}^{-n^3},
\end{aligned}$$

where for the second inequality we use that $k! \geq (\frac{k}{e})^k$, $k \in \mathbb{N}$, and for the third inequality that, for n sufficiently large,

$$k \log\left(\frac{en^3 \log n}{k}\right) \leq (n^2 \log n) \log\left(e\, n^3 \log n\right) \leq n^{5/2}. \qquad (5.5)$$

Next, observe that

$$\begin{aligned}
&\mathbb{P}_{\xi_0}\left[\xi_t \in A_{\mathbf{M}+C_1 n^{5/6}} \text{ for some } 0 \leq t \leq T\right] \qquad (5.6)\\
&= \mathbb{P}_{\xi_0}\left[\xi_t \in A_{\mathbf{M}+C_1 n^{5/6}} \text{ for some } 0 \leq t \leq T,\ T \leq n^2 \log n\right]\\
&\quad + \mathbb{P}_{\xi_0}\left[\xi_t \in A_{\mathbf{M}+C_1 n^{5/6}} \text{ for some } 0 \leq t \leq T,\ T > n^2 \log n\right]\\
&\leq (n^2 \log n) \max_{\sigma \in A_{\mathbf{M}}} \mathbb{P}_\sigma\left[\tau_{A_{\mathbf{M}+C_1 n^{5/6}}} < \tau_{A_{\mathbf{M}}}\right] + \mathrm{e}^{-n^3}.
\end{aligned}$$

Here, the inequality follows from (5.4) and the observation that the event $\xi_t \in A_{\mathbf{M}+C_1 n^{5/6}}$ for some $0 \leq t \leq T$ with $T \leq n^2 \log n$ is contained in the event that $A_{\mathbf{M}+C_1 n^{5/6}}$ is visited before the $(n^2 \log n)$-th return to $A_{\mathbf{M}}$. From Lemma 4.3 and

(4.47) it follows that

$$\begin{aligned} &\max_{\sigma\in A_{\mathbf{M}}} \mathbb{P}_\sigma\left[\tau_{A_{\mathbf{m}+C_1 n^{5/6}}} < \tau_{A_{\mathbf{M}}}\right] \\ &\leq \tfrac{(1-a)}{2} \mathrm{e}^{2\beta\left(p|a|+h^l+\frac{p}{n}\right)} \mathrm{e}^{-\beta n[\Psi(b)-\Psi(a)]} \binom{n}{\frac{n}{2}(1+b)} \binom{n}{\frac{n}{2}(1+a)}^{-1} \end{aligned} \tag{5.7}$$

with $a = \mathbf{m}/n$ and $b = \left(\mathbf{m} + C_1 n^{5/6}\right)/n$.

2. Our assumption on the parameters β, p and h is that $2\beta(p(a+\frac{2}{n})+h) + \log(\frac{1-a}{1+a+\frac{2}{n}})$ is negative in two disjoint regions. Recall that the first region lies between $a_1 = \frac{2\mathbf{M}}{n} - 1$ and $a_2 = \frac{2\mathbf{T}}{n} - 1$. This, in particular, implies that the derivative of $2\beta(p(a+\frac{2}{n})+h) + \log(\frac{1-a}{1+a+\frac{2}{n}})$ at $a = a_1$ is

$$2\beta p - \frac{1}{1-a_1} - \frac{1}{1+a_1} = -\delta_1 < 0 \tag{5.8}$$

for some $\delta_1 > 0$. Recall that $\Psi(a) = -\frac{p}{2}a^2 - ha$, so that $\Psi(b) - \Psi(a) = (a-b)(\frac{p}{2}(a+b)+h)$, which gives

$$\mathrm{e}^{-\beta n[\Psi(b)-\Psi(a)]} \binom{n}{\frac{n}{2}(1+b)} \binom{n}{\frac{n}{2}(1+a)}^{-1} \tag{5.9}$$

$$\begin{aligned} &= \exp\Bigg(\beta n(b-a)(pa+h) + \beta n(b-a)^2 \frac{p}{2} \\ &\qquad + \frac{n}{2}\log\left(\frac{(1+a)^{(1+a)}(1-a)^{(1-a)}}{(1+b)^{(1+b)}(1-b)^{(1-b)}}\right) + O(\log n)\Bigg), \end{aligned} \tag{5.10}$$

where we use Stirling's approximation in the last line. Since $b = a + C_1 n^{-1/6}$, we have

$$\text{r.h.s. (5.9)} = \exp\left(\beta C_1 n^{5/6}(pa+h) + \frac{p}{2}\beta C_1^2 n^{2/3} + \frac{n}{2}\log F\right) \tag{5.11}$$

with

$$F = (1-U_n(a))^{1+a}(1+V_n(a))^{1-a}(W_n(a))^{C_1 n^{-1/6}}, \tag{5.12}$$

where

$$U_n(a) = \frac{C_1 n^{-1/6}}{1+a+C_1 n^{-1/6}}, \qquad V_n(a) = \frac{C_1 n^{-1/6}}{1-a-C_1 n^{-1/6}}. \tag{5.13}$$

From the Taylor series expansion of $\log(1+x)$ for $0 \le |x| < 1$, we obtain

$$\begin{aligned}&\tfrac{n}{2}(1+a)\log(1-U_n(a)) \le \tfrac{n}{2}(1+a)\left(-U_n(a)-\tfrac{1}{2}(U_n(a))^2\right),\\ &\tfrac{n}{2}(1-a)\log(1+V_n(a)) \le \tfrac{n}{2}(1-a)\left(V_n(a)-\tfrac{1}{2}(V_n(a))^2+O(n^{-1/2})\right),\end{aligned} \tag{5.14}$$

and

$$\begin{aligned}&\frac{1}{2}C_1 n^{5/6}\log\left(\frac{U_n(a)}{V_n(a)}\right)\\ &\quad=\frac{1}{2}C_1 n^{5/6}\log\left(\frac{1-a}{1+a}\frac{1-a-C_1n^{-1/6}}{1-a}\frac{1+a}{1+a+C_1n^{-1/6}}\right)\\ &\quad\le\frac{1}{2}C_1 n^{5/6}\left(\log\left(\frac{1-a}{1+a}\right)-\frac{C_1n^{-1/6}}{1-a}-U_n(a)-O(n^{-2/3})\right).\end{aligned} \tag{5.15}$$

By the definition of **m**, we have

$$C_1 n^{5/6}\left(\beta(pa+h)+\log\left(\frac{1-a}{1+a}\right)\right)\le 0. \tag{5.16}$$

Hence we get

$$\begin{aligned}&\beta C_1 n^{5/6}(pa+h)+\tfrac{p}{2}\beta n^{2/3}C_1^2+\tfrac{n}{2}\log F\\ &\le\tfrac{p}{2}\beta n^{2/3}C_1^2-\tfrac{\frac{1}{2}C_1(1+a)n^{5/6}}{1+a+C_1n^{-1/6}}+\tfrac{\frac{1}{2}C_1(1-a)n^{5/6}}{1-a-C_1n^{-1/6}}-\tfrac{1}{2}C_1^2n^{2/3}\,G\end{aligned} \tag{5.17}$$

with

$$\begin{aligned}G=&\tfrac{1}{1-a}+\tfrac{1}{1+a+C_1n^{-1/6}}\\ &+\left(\tfrac{1-a}{2}\right)\left(\tfrac{1}{1-a-C_1n^{-1/6}}\right)^2+\left(\tfrac{1+a}{2}\right)\left(\tfrac{1}{1+a+C_1n^{-1/6}}\right)^2.\end{aligned} \tag{5.18}$$

Hence

$$\begin{aligned}\text{r.h.s. (5.17)}\le&\tfrac{p}{2}\beta n^{2/3}C_1^2\\ &+\tfrac{1}{2}C_1^2n^{2/3}\left(\tfrac{1}{1-a-cn^{-1/6}}+\tfrac{1}{1+a+C_1n^{-1/6}}\right)-\tfrac{1}{2}C_1^2n^{1/6}\,G\\ \le&\, n^{2/3}\tfrac{1}{2}C_1^2\left(p\beta-\tfrac{1}{2}\tfrac{1}{1-a-C_1n^{-1/6}}-\tfrac{1}{2}\tfrac{1}{1+a+C_1n^{-1/6}}+O(n^{-1/6})\right)\\ =&\, n^{2/3}\tfrac{1}{2}C_1^2\left(p\beta-\tfrac{1}{2}\tfrac{1}{1-a}-\tfrac{1}{2}\tfrac{1}{1+a}+O(n^{-1/3})\right)\le-\tfrac{1}{4}C_1^2\delta_1n^{2/3}.\end{aligned} \tag{5.19}$$

3. Combine (5.7), (5.9) and (5.17), and pick C_1 large enough, to get the claim in (5.3). □

5.1.2 Update Times

The following two lemmas give useful bounds for the coupling scheme. The symbol $\simeq$ stands for equality in distribution.

Lemma 5.2 (Total Variation Between Exponential Distributions) *Let* $X \simeq \mathrm{Exp}(\lambda)$ *and* $Y \simeq \mathrm{Exp}(\lambda+\delta)$. *Then the total variation distance between the distributions of* X *and* Y *is bounded by*

$$d_{TV}(X,Y) \leq \frac{2\delta}{\lambda+\delta}. \tag{5.20}$$

Proof Elementary. □

Lemma 5.3 (Update Times) *Let* T^{ξ}_{update} *be the first time* $\{\xi_t\}_{t\geq 0}$ *has experienced an update at every vertex:*

$$T^{\xi}_{update} = \inf\{t \geq 0\colon\ \forall v \in V\ \exists 0 \leq s \leq t\colon\ \xi_s(v) = -\xi_0(v)\}. \tag{5.21}$$

Then, for any $y > 0$,

$$\mathbb{P}\left[T^{\xi}_{update} \geq y\right] \leq \frac{\exp(-\lambda y + \log n)}{1-\exp(-\lambda y)}, \qquad \lambda = \mathrm{e}^{-\beta(2p+h)}. \tag{5.22}$$

Proof Recall that for $\sigma \in S_n$ and $v \in V$, σ^v denotes the configuration satisfying $\sigma^v(w) = \sigma(w)$ for $w \neq v$, and $\sigma^v(v) = -\sigma(v)$. From (1.3) and (1.6) it follows that

$$r\left(\sigma,\sigma^v\right) \geq \lambda, \tag{5.23}$$

and so T^{ξ}_{update} is dominated by the maximum of n i.i.d. $\mathrm{Exp}(\lambda)$ random variables. Therefore

$$\begin{aligned}\mathbb{P}\left[T^{\xi}_{update} \leq y\right] &\geq \left(1-\mathrm{e}^{-\lambda y}\right)^n = \exp\left(n\log(1-\mathrm{e}^{-\lambda y})\right) \\ &\geq \exp\left(-\frac{n\mathrm{e}^{-\lambda y}}{1-\mathrm{e}^{-\lambda y}}\right) \geq 1-\frac{n\mathrm{e}^{-\lambda y}}{1-\mathrm{e}^{-\lambda y}},\end{aligned} \tag{5.24}$$

which proves the claim. □

5.1.3 Returns

The next lemma establishes a lower bound on the number of returns to $A_{\mathbf{M}}$ before $A_{\mathbf{S}}$ is reached. Let $g_{\xi_0}(A_{\mathbf{M}}, A_{\mathbf{S}})$ denote the number of jumps that $\{\xi_t\}_{t\geq 0}$ makes into the set $A_{\mathbf{M}}$ before reaching $A_{\mathbf{S}}$. More precisely, let $\{s_i\}_{i\in\mathbb{N}_0}$ denote the jump times of the process $\{\xi_t\}_{t\geq 0}$, i.e., $s_0 = 0$ and

$$s_i = \inf\left\{s > s_{i-1} \colon \xi_s \neq \xi_{s_{i-1}}\right\}, \tag{5.25}$$

and define for the process $(\xi_t)_{t\geq 0}$ starting at ξ_0,

$$g_{\xi_0}(A_{\mathbf{M}}, A_{\mathbf{S}}) = \left|\left\{i \in \mathbb{N}_0 \colon \xi_{s_i} \in A_{\mathbf{M}}, \xi_s \notin A_{\mathbf{S}} \ \forall\, s \leq s_i\right\}\right|. \tag{5.26}$$

Lemma 5.4 (Bound on Number of Returns) *For any $\xi_0 \in A_{\mathbf{M}}$ and any $\delta > 0$,*

$$\mathbb{P}_{\xi_0}\left[g_{\xi_0}(A_{\mathbf{M}}, A_{\mathbf{S}}) < \mathrm{e}^{[R_p(t)-R_p(m)]n}\right] \leq \mathrm{e}^{-\delta n + Cn^{2/3}} \tag{5.27}$$

for some constant C that does not depend on n.

Proof Let Y be a geometric random variable with probability of success given by $\mathrm{e}^{-[R_p(t)-R_p(m)]n+Cn^{2/3}}$. Then, by Lemma 4.3, every time the process $\{\xi_t\}_{t\geq 0}$ starts all over from A_M, it has a probability less than $\mathbb{P}^u_\xi[\tau_{S^u} < \tau_{M^u}]$ of making it to A_S. Using the bounds from that lemma, it follows that Y is stochastically dominated by $g_{\xi_0}(A_{\mathbf{M}}, A_{\mathbf{T}})$. Hence

$$\begin{aligned}\mathbb{P}\left[Y \leq \mathrm{e}^{([R_p(t)-R_p(m)]-\delta)n}\right] &\leq \mathrm{e}^{([R_p(t)-R_p(m)]-\delta)n}\,\mathrm{e}^{-[R_p(t)-R_p(m)]n+Cn^{2/3}}\\ &\leq \mathrm{e}^{-\delta n+Cn^{2/3}}.\end{aligned} \tag{5.28}$$

□

5.2 *Uniform Hitting Time*

In this section we show that if Theorem 1.4 holds for *some* initial configuration in $A_{\mathbf{M}}$, then it holds for *all* initial configurations in $A_{\mathbf{M}}$. The proof of this claim, which will be needed in Sect. 7, relies on a *coupling construction* in which the two processes starting anywhere in $A_{\mathbf{M}}$ meet with a sufficiently high probability long before either one reaches $A_{\mathbf{S}}$. Details of the coupling construction are given in Sect. 6.

The idea of the proof is that for $\{\xi_t\}_{t\geq 0}$ starting in $A_{\mathbf{M}}$ the starting configuration is irrelevant for the metastable crossover time because the latter is very large. We will verify this by showing that "local mixing" takes place long before the crossover to $A_{\mathbf{S}}$ occurs. More precisely, we will show that if $\xi_0, \tilde{\xi}_0$ are any two initial configurations in $A_{\mathbf{M}}$, then there is a coupling such that the trajectory $t \mapsto \xi_t$

intersects the trajectory $t \mapsto \tilde{\xi}_t$ well before either strays too far from $A_{\mathbf{M}}$. The coupling is such that there is a small but sufficiently large probability that ξ_t and $\tilde{\xi}_t$ are identical once every spin at every vertex has had a chance to update, which occurs after a time t that is not too large. It follows that after a large number of trials with high probability the two trajectories intersect.

Proof Consider two copies of the process, $\{\xi_t\}_{t\geq 0}$ and $\{\tilde{\xi}_t\}_{t\geq 0}$. Let $\delta > 0$ and $T_0 = \mathrm{e}^{([R_{p,\beta,h}(\mathbf{t})-R_{p,\beta,h}(\mathbf{m})]-\delta)n}$. In order to simplify the notation and differentiate between the two processes, we abbreviate the crossover time $\tau_{A_{\mathbf{S}}}$ by

$$\tau^{\xi} = \inf\{t \geq s \colon \xi_t \in A_{\mathbf{S}}\}, \tag{5.29}$$

with a similar definition for $\tau^{\tilde{\xi}}$. We will show that $\mathbb{E}_{\xi_0}[\tau^{\xi}] \leq [1 + o_n(1)]\mathbb{E}_{\tilde{\xi}_0}[\tau^{\tilde{\xi}}]$, with the proof for the inequality in the other direction being identical. The proof comes in two Steps.

1. We start with the following observation. From Corollary 3.9, we immediately get that

$$\mathbb{E}_{\xi_0}\left[\tau^{\xi}\right]/\mathbb{E}_{\xi_0}\left[\tau^{\tilde{\xi}}\right] = \mathrm{e}^{O(n^{2/3})}. \tag{5.30}$$

Furthermore, the relation in (5.30) together with the initial steps in the proof of Theorem 1.4 implies that, for any initial configuration ξ_0,

$$\mathbb{E}_{\xi_0}\left[\tau^{\xi}\right] = \mathrm{e}^{n[R_{p,\beta,h}(\mathbf{t})-R_{p,\beta,h}(\mathbf{m})]+O(n^{2/3})}. \tag{5.31}$$

Note: Step 2 in Sect. 7 shows that if ξ_0 is distributed according to the law $\mu_{A_{\mathbf{M}}}$, then

$$\mathbb{E}_{\xi_0}[\tau^{\xi}] = \mathrm{e}^{n[R_{p,\beta,h}(\mathbf{t})-R_{p,\beta,h}(\mathbf{m})]+O(\log n)}. \tag{5.32}$$

(Recall from Sect. 4.3 that $\mu_{A_{\mathbf{M}}}$ is the equilibrium distribution μ conditioned on the set $A_{\mathbf{M}}$.) Let $\{\xi_t, \tilde{\xi}_t\}_{t\geq 0}$ be the coupling of the two processes described in Sect. 6, and note that

$$\mathbb{E}_{\xi_0}\left[\tau_{A_{\mathbf{S}}}\right] = \hat{\mathbb{E}}_{\left(\xi_0,\tilde{\xi}_0\right)}[\tau^{\xi}] = \hat{\mathbb{E}}_{\left(\xi_0,\tilde{\xi}_0\right)}\left[\tau^{\xi}\mathbb{1}_{\left\{\xi_{T_0}=\tilde{\xi}_{T_0}\right\}}\right] + \hat{\mathbb{E}}_{\left(\xi_0,\tilde{\xi}_0\right)}\left[\tau^{\xi}\mathbb{1}_{\left\{\xi_{T_0}\neq\tilde{\xi}_{T_0}\right\}}\right], \tag{5.33}$$

where $\hat{\mathbb{E}}$ denotes expectation with respect to the law of the joint process. The above inequality splits the expectation based on whether the coupling has succeeded (in merging the two processes) by time T_0 or not. Note that

$$\tau^{\xi}\mathbb{1}_{\left\{\xi_{T_0}=\tilde{\xi}_{T_0}\right\}} \leq \tau^{\tilde{\xi}}\mathbb{1}_{\left\{\xi_{T_0}=\tilde{\xi}_{T_0},\,\tau^{\tilde{\xi}}\geq T_0\right\}} + \tau^{\xi}\mathbb{1}_{\left\{\xi_{T_0}=\tilde{\xi}_{T_0},\,\tau^{\tilde{\xi}}<T_0\right\}}, \tag{5.34}$$

and

$$\begin{aligned} &\tau^{\xi}\mathbb{1}_{\left\{\xi_{T_0}=\tilde{\xi}_{T_0},\, \tau^{\tilde{\xi}}<T_0\right\}} \\ &\quad = \tau^{\xi}\mathbb{1}_{\left\{\xi_{T_0}=\tilde{\xi}_{T_0},\, \tau^{\tilde{\xi}}<T_0,\, \left|\tilde{\xi}_{T_0}\right|<\mathbf{S}\right\}} + \tau^{\xi}\mathbb{1}_{\left\{\xi_{T_0}=\tilde{\xi}_{T_0},\, \tau^{\tilde{\xi}}<T_0,\, \left|\tilde{\xi}_{T_0}\right|\geq\mathbf{S}\right\}} \\ &\quad \leq \tau^{\xi}\mathbb{1}_{\left\{\xi_{T_0}=\tilde{\xi}_{T_0},\, \tau^{\tilde{\xi}}<T_0,\, \left|\tilde{\xi}_{T_0}\right|<\mathbf{S}\right\}} + T_0. \end{aligned} \tag{5.35}$$

Also note from the definition of the coupling that, for any $\sigma \in S_n$ and any $A \subseteq S_n$, $\hat{\mathbb{E}}_{(\sigma,\sigma)}[\tau_A^{\xi}] = \mathbb{E}_{\sigma}[\tau_A]$ because the two trajectories merge when they start from the same vertex. Hence

$$\begin{aligned} &\hat{\mathbb{E}}_{\left(\xi_0,\tilde{\xi}_0\right)}\left[\tau^{\xi}\mathbb{1}_{\left\{\xi_{T_0}=\tilde{\xi}_{T_0},\, \tau^{\tilde{\xi}}<T_0,\, \left|\tilde{\xi}_{T_0}\right|<\mathbf{S}\right\}}\right] \\ &\quad = \sum_{\sigma\in\bigcup_{i<\chi_1} A_i} \hat{\mathbb{E}}_{\left(\xi_0,\tilde{\xi}_0\right)}\left[\tau^{\xi}\mathbb{1}_{\left\{\xi_{T_0}=\tilde{\xi}_{T_0}=\sigma,\, \tau^{\tilde{\xi}}<T_0\right\}}\right] \\ &\quad \leq \sum_{\sigma\in\bigcup_{i<\chi_1} A_i} \left(\hat{\mathbb{E}}_{(\sigma,\sigma)}\left[\tau^{\xi}\right]+T_0\right)\hat{\mathbb{P}}_{\left(\xi_0,\tilde{\xi}_0\right)}\left[\xi_{T_0}=\tilde{\xi}_{T_0}=\sigma,\, \tau^{\tilde{\xi}}<T_0\right] \\ &\quad \leq \left(T_0+\max_{\sigma\in\bigcup_{i<\chi_1} A_i}\mathbb{E}_{\sigma}\left[\tau^{\sigma}\right]\right)\mathbb{P}_{\tilde{\xi}_0}\left[\tau<T_0\right], \end{aligned} \tag{5.36}$$

where we use the Markov property. Similarly, observe that

$$\begin{aligned} \tau^{\xi}\mathbb{1}_{\left\{\xi_{T_0}\neq\tilde{\xi}_{T_0}\right\}} &= \tau^{\xi}\mathbb{1}_{\left\{\xi_{T_0}\neq\tilde{\xi}_{T_0},\, \tau^{\xi}\leq T_0\right\}} + \tau^{\xi}\mathbb{1}_{\left\{\xi_{T_0}\neq\tilde{\xi}_{T_0},\, \tau^{\xi}>T_0\right\}} \\ &\leq T_0 + \tau^{\xi_{T_0}}\mathbb{1}_{\left\{\xi_{T_0}\neq\tilde{\xi}_{T_0}\, \tau^{\xi}>T_0\right\}}, \end{aligned} \tag{5.37}$$

and

$$\begin{aligned} &\hat{\mathbb{E}}_{\left(\xi_0,\tilde{\xi}_0\right)}\left[\tau^{\xi}\mathbb{1}_{\left\{\xi_{T_0}\neq\tilde{\xi}_{T_0},\, \tau^{\xi}>T_0\right\}}\right] \\ &\quad = \sum_{\sigma:\, |\sigma|<\mathbf{S}}\sum_{\sigma'\neq\sigma}\hat{\mathbb{E}}_{\left(\xi_0,\tilde{\xi}_0\right)}\left[\tau^{\xi}\mathbb{1}_{\left\{\xi_{T_0}=\sigma,\, \tilde{\xi}_{T_0}=\sigma',\, \tau^{\xi}>T_0\right\}}\right] \\ &\quad = \sum_{\sigma:\, |\sigma|<\mathbf{S}}\sum_{\sigma'\neq\sigma}\hat{\mathbb{E}}_{(\sigma,\sigma')}\left[T_0+\tau^{\xi}\right]\hat{\mathbb{E}}_{\left(\xi_0,\tilde{\xi}_0\right)}\left[\mathbb{1}_{\left\{\xi_{T_0}=\sigma,\, \tilde{\xi}_{T_0}=\sigma',\, \tau^{\xi}>T_0\right\}}\right] \\ &\quad \leq \max_{\sigma\in\bigcup_{i<\mathbf{S}} A_i}\left(T_0+\mathbb{E}_{\sigma}\left[\tau^{\xi}\right]\right)\hat{\mathbb{P}}_{\left(\xi_0,\tilde{\xi}_0\right)}\left[\xi_{T_0}\neq\tilde{\xi}_{T_0}\right]. \end{aligned} \tag{5.38}$$

Thus, (5.33) becomes

$$\mathbb{E}_{\xi_0}\left[\tau^{\xi}\right] \le 2T_0 + \mathbb{E}_{\tilde{\xi}_0}\left[\tau_{A_{\mathbf{S}}}\right] + \left(\mathbb{P}_{\tilde{\xi}_0}\left[\tau_{A_{\mathbf{S}}} < T_0\right] + \hat{\mathbb{P}}_{\left(\xi_0,\tilde{\xi}_0\right)}\left[\xi_{T_0} \ne \tilde{\xi}_{T_0}\right]\right) \times \left(T_0 + \max_{\sigma\in\bigcup_{i<\mathbf{S}} A_i} \mathbb{E}_{\sigma}\left[\tau_{A_{\mathbf{S}}}\right]\right). \tag{5.39}$$

We will show that the leading term in the right-hand side is $\mathbb{E}_{\tilde{\xi}_0}[\tau_{A_{\mathbf{S}}}]$, and all other terms are of smaller order. From (5.31) we know that T_0 is of smaller order, and that

$$\max_{\sigma\in\bigcup_{i<\mathbf{S}} A_i} \mathbb{E}_{\sigma}\left[\tau_{A_{\mathbf{S}}}\right] = \mathrm{e}^{O(n^{2/3})}\mathbb{E}_{\tilde{\xi}_0}\left[\tau_{A_{\mathbf{S}}}\right]. \tag{5.40}$$

Hence it suffices to show that the sum $\mathbb{P}_{\tilde{\xi}_0}[\tau_{A_{\mathbf{S}}} < T_0] + \hat{\mathbb{P}}_{(\xi_0,\tilde{\xi}_0)}[\xi_{T_0} \ne \tilde{\xi}_{T_0}]$ is exponentially small. We will show that it is bounded from above by $\mathrm{e}^{-\delta n}$.

2. By Corollary 6.3, the probability $\hat{\mathbb{P}}_{(\xi_0,\tilde{\xi}_0)}[\xi_{T_0} \ne \tilde{\xi}_{T_0}]$ is bounded from above by $\mathrm{e}^{-\delta n+O(n^{2/3})}$. To bound $\mathbb{P}_{\tilde{\xi}_0}[\tau_{A_{\mathbf{S}}} < T_0]$, we first need to limit the number of steps that $\{\tilde{\xi}_t\}_{t\ge0}$ can take until time T_0. From (5.2) and Stirling's approximation we have that

$$\begin{aligned}\mathbb{P}\left[N_{\tilde{\xi}}\left(T_0\right) \ge 3nT_0\right] &\le \sum_{k=0}^{\infty} \mathrm{e}^{nT_0+k}\left(\frac{nT_0}{3nT_0+k}\right)^{3nT_0+k} \\ &\le \mathrm{e}^{nT_0}\left(\frac{1}{3}\right)^{3nT_0}\sum_{k=0}^{\infty}\mathrm{e}^{k}\left(\frac{1}{3}\right)^{k} \le 11\,(0.91)^{3n\mathrm{e}^{\frac{1}{2}[R_{p,\beta,h}(\mathbf{t})-R_{p,\beta,h}(\mathbf{m})]}}.\end{aligned} \tag{5.41}$$

It therefore follows that with high probability $\{\tilde{\xi}_t\}_{t\ge0}$ does not make more than $3nT_0$ steps until time T_0. Hence

$$\mathbb{P}_{\tilde{\xi}_0}\left[\tau_{A_{\mathbf{S}}} < T_0\right] \le \mathbb{P}_{\tilde{\xi}_0}\left[\tau_{A_{\mathbf{S}}} < T_0, N_{\tilde{\xi}}\left(T_0\right) < 3nT_0\right] + 11\,(0.91)^{3n\mathrm{e}^{\frac{1}{2}\,[R_{p,\beta,h}(\mathbf{t})-R_{p,\beta,h}(\mathbf{m})]}}. \tag{5.42}$$

Finally, note that the event $\{\tau^{\tilde{\xi}}_{A_{\mathbf{S}}} < T_0, N_{\tilde{\xi}}(T_0) < 3nT_0\}$ implies that $\{\tilde{\xi}_t\}_{t\ge0}$ makes fewer than $3nT_0$ returns to the set $A_{\mathbf{M}}$ before reaching $A_{\mathbf{S}}$. By Lemma 5.4, the probability of this event is bounded from above by $3nT_0\mathrm{e}^{([R_{p,\beta,h}(\mathbf{t})-R_{p,\beta,h}(\mathbf{m})]-\delta)n+Cn^{2/3}} = 3n\mathrm{e}^{-\frac{1}{2}\delta n+Cn^{2/3}}$, and hence

$$\mathbb{P}_{\tilde{\xi}_0}\left[\tau^{\tilde{\xi}}_{A_{\mathbf{S}}} < T_0\right] \le 4n\,\mathrm{e}^{-\frac{1}{2}\delta n+Cn^{2/3}}. \tag{5.43}$$

Finally, from (5.43) we obtain

$$\mathbb{E}_{\xi_0}\left[\tau^{\xi}_{A_{\mathbf{S}}}\right] = \mathbb{E}\left[\tau^{\tilde{\xi}}_{A_{\mathbf{S}}}\right][1 + o(1)], \tag{5.44}$$

which settles the claim. □

6 A Coupling Scheme

In this section we define a coupling of $(\xi_t)_{t\geq 0}$ and $\{\tilde{\xi}_t\}_{t\geq 0}$ with arbitrary starting configurations in $A_{\mathbf{M}}$. The coupling is divided into a short-term scheme, defined in Sect. 6.1 and analysed in Lemma 6.1 below, followed by a long-term scheme, defined in Sect. 6.2 and analysed Corollary 6.3 below. The goal of the coupling is to keep the process $\{m(\xi_t)\}_{t\geq 0}$ bounded by $\{\theta^u_t\}_{t\geq 0}$ from above and bounded by $\{\theta^l_t\}_{t\geq 0}$ from below (the precise meaning will become clear in the sequel).

6.1 Short-Term Scheme

Lemma 6.1 (Short-Term Coupling) *With $\mathbb{P}_{\mathrm{ER}_n(p)}$-probability tending to 1 as $n \to \infty$, there is a coupling $\{\xi_t, \tilde{\xi}_t\}_{t\geq 0}$ of $\{\xi_t\}_{t\geq 0}$ and $\{\tilde{\xi}_t\}_{t\geq 0}$ such that*

$$\mathbb{P}\left[\xi_{2n} \neq \tilde{\xi}_{2n}\right] \leq O\left(\mathrm{e}^{-n^{-2/3}}\right) \tag{6.1}$$

for any initial states $\xi_0 \in A_{\mathbf{M}}$ and $\tilde{\xi}_0 \in A_{\mathbf{M}}$.

Proof The main idea behind the proof is as follows. Define

$$W_1^t = \{v \in V : \ \xi_t(v) = -\tilde{\xi}_t(v)\} = \xi_t \Delta \tilde{\xi}_t, \tag{6.2}$$

i.e., the symmetric difference between the two configurations ξ_t and $\tilde{\xi}_t$, and

$$W_2^t = \{v \in V : \ \xi_t(v) = \tilde{\xi}_t(v)\} = V \setminus W_1^t. \tag{6.3}$$

The coupling we are about to define will result in the set W_1^t shrinking at a higher rate than the set W_2^t, which will imply that W_1^t contracts to the empty set. The proof comes in eight Steps.

1. We begin with bounds on the relevant transition rates that will be required in the proof. Recall from Lemma 3.4 (in particular, (3.19) and (3.21)) that with $\mathbb{P}_{\mathrm{ER}_n(p)}$-probability at least $1 - \mathrm{e}^{-2n}$ there are at most $2n^{2/3}$ vertices $v \in \overline{\xi_t}$ (i.e., $\xi_t(v) = -1$) such that $|E(v, \xi_t)| = |\{w \in \xi_t : \ (v, w) \in E\}| \geq p\,|\xi_t| + n^{2/3}$, and similarly at most $2n^{2/3}$ vertices $v \in \overline{\xi_t}$ such that $|E(v, \xi_t)| \leq p\,|\xi_t| - n^{2/3}$.

Analogous bounds are true for $\tilde{\xi}_t$, $t \geq 0$. Denote the set of *bad* vertices for ξ_t by

$$B_t = \left\{v \in \overline{\xi_t}\colon \left||E(v,\xi_t)| - p|\xi_t|\right| \geq n^{2/3}\right\}, \tag{6.4}$$

and the set of bad vertices for $\tilde{\xi}_t$ by $\tilde{B}_t$. Let $\hat{B}_t = B_t \cup \tilde{B}_t$. Recall that ξ_t^v denotes the configuration obtained from ξ_t by flipping the sign at vertex $v \in V$. If $v \notin \hat{B}_t$, then from (1.3) and Lemma 3.2 it follows that, for $v \notin \xi_t$,

$$\begin{aligned} H_n\left(\xi_t^v\right) - H_n\left(\xi_t\right) &= \frac{2}{n}\left(\left|\partial_E \xi_t^v\right| - \left|\partial_E \xi_t\right|\right) - 2h \\ &= \frac{2}{n}\left(\deg(v) - 2\left|E(v,\xi_t)\right|\right) - 2h \\ &\leq \frac{2}{n}\left(pn + n^{1/2}\log n - 2p\left|\xi_t\right| + 2n^{2/3}\right) - 2h, \end{aligned} \tag{6.5}$$

and similarly, for $v \in \xi_t$,

$$H_n\left(\xi_t^v\right) - H_n\left(\xi_t\right) \leq \frac{2}{n}\left(pn + n^{1/2}\log n - 2p\left(n - |\xi_t|\right) + 2n^{2/3}\right) + 2h. \tag{6.6}$$

Again, by (1.3) and Lemma 3.2, we have similar lower bounds, namely, if $v \notin \hat{B}_t$, then, for $v \notin \xi_t$,

$$H_n(\xi_t^v) - H_n(\xi_t) \geq \frac{2}{n}\left(pn - n^{1/2}\log n - 2p\left|\xi_t\right| - 2n^{2/3}\right) - 2h, \tag{6.7}$$

and, for $v \in \xi_t$,

$$H_n(\xi_t^v) - H_n(\xi_t) \geq \frac{2}{n}\left(pn - n^{1/2}\log n - 2p\left(n - |\xi_t|\right) - 2n^{2/3}\right) + 2h. \tag{6.8}$$

Identical bounds hold for $H_n(\tilde{\xi}_t^v) - H_n(\tilde{\xi}_t)$. Therefore, if $v \notin \hat{B}_t$, and if either $v \in \xi_t \cap \tilde{\xi}_t$ or $v \notin \xi_t \cup \tilde{\xi}_t$, then

$$\begin{aligned} &\left|r(\xi_t, \xi_t^v) - r(\tilde{\xi}_t, \tilde{\xi}_t^v)\right| \\ &= \left|\mathrm{e}^{-\beta[H_n(\xi_t^v) - H_n(\xi_t)]_+} - \mathrm{e}^{-\beta[H_n(\tilde{\xi}_t^v) - H_n(\tilde{\xi}_t)]_+}\right| \\ &= \mathrm{e}^{-\beta[H_n(\xi_t^v) - H_n(\xi_t)]_+}\left|1 - \mathrm{e}^{\beta\left([H_n(\xi_t^v) - H_n(\xi_t)]_+ - [H_n(\tilde{\xi}_t^v) - H_n(\tilde{\xi}_t)]_+\right)}\right| \\ &\leq [1 + o_n(1)]\,\mathrm{e}^{-\beta[H_n(\xi_t^v) - H_n(\xi_t)]_+}\left(\mathrm{e}^{8\beta n^{-1/3} + \frac{4p}{n}\left(|\xi_t| - |\tilde{\xi}_t|\right)} - 1\right) \\ &\leq [1 + o_n(1)]\left(\mathrm{e}^{8\beta n^{-1/3} + \frac{4p}{n}\left(|\xi_t| - |\tilde{\xi}_t|\right)} - 1\right) \\ &\leq [1 + o_n(1)]\left(8\beta n^{-1/3} + \tfrac{4p}{n}\left(|\xi_t| - |\tilde{\xi}_t|\right)\right). \end{aligned} \tag{6.9}$$

2. Having established the above bounds on the transition rates, we give an explicit construction of the coupling $\{\xi_t, \tilde{\xi}_t\}_{t\geq 0}$. □

Definition 6.2

(I) We first define the coupling for time $t = 0$. For $t > 0$ this coupling will be *renewed* after each renewal of $\{\xi_t, \tilde{\xi}_t\}_{t\geq 0}$, i.e., whenever either of the two processes jumps to a new state. To that end, for every $v \in W_2^0$ (i.e., $\xi_0(v) = \tilde{\xi}_0(v)$), couple the exponential random variables $e_0^v \simeq \text{Exp}(r(\xi_0, \xi_0^v))$ and $\tilde{e}_0^v \simeq \text{Exp}(r(\tilde{\xi}_0, \tilde{\xi}_0^v))$ associated with the transitions $\xi_0 \to \xi_0^v$ and $\tilde{\xi}_0 \to \tilde{\xi}_0^v$ according to the following scheme:

1. Choose a point

$$(x, y) \in \{(x', y') : 0 \leq x' < \infty, \ 0 \leq y' \leq r\left(\xi_0, \xi_0^v\right) \mathrm{e}^{-r(\xi_0,\xi_0^v)x'}\}$$

uniformly and set $e_0^v = x$. Note that, indeed, this gives $e_0^v \simeq \text{Exp}(r(\xi_0, \xi_0^v))$.

2. If the value y from step 1 satisfies $y \leq r(\tilde{\xi}_0, \tilde{\xi}_0^v) \exp(-r(\tilde{\xi}_0, \tilde{\xi}_0^v)x)$, then set $\tilde{e}_0^v = e_0^v = x$. Else, choose

$$(x*, y*) \in \left\{(x', y') : 0 \leq x' < \infty, \ r\left(\xi_0, \xi_0^v\right) \mathrm{e}^{-r(\xi_0,\xi_0^v)x'} < y' \leq r(\tilde{\xi}_0, \tilde{\xi}_0^v)\mathrm{e}^{-r(\tilde{\xi}_0,\tilde{\xi}_0^v)x'}\right\}$$

uniformly and independently from the sampling in step 1, and set $\tilde{e}_0^v = x*$. Note that this too gives $e_0^v \simeq \text{Exp}(r(\tilde{\xi}_0, \tilde{\xi}_0^v))$.

(II) For every $v \in W_1^0$, sample the random variables $e_0^v \simeq \text{Exp}(r(\xi_0, \xi_0^v))$ and $\tilde{e}_0^v \simeq \text{Exp}(r(\tilde{\xi}_0, \tilde{\xi}_0^v))$ associated with the transitions $\xi_0 \to \xi_0^v$ and $\tilde{\xi}_0 \to \tilde{\xi}_0^v$ independently. At time $t = 0$, we use the above rules to define the *jump* times associated with any vertex $v \in V$. Recall that W_2^0 is the set of vertices where the two configurations agree in sign. The aim of the coupling defined above is to preserve that agreement. Following every renewal, we re-sample all transition times anew (i.e., we choose new copies of the exponential variables as was done above). We proceed in this way until the first of the following two events happens: either $\xi_t = \tilde{\xi}_t$, or $n \log n$ transitions have been made by either one of the two processes.

3. Note that the purpose of limiting the number of jumps to $n \log n$ is to permit us to employ Lemma 5.1, which in turn we use to maintain control on the two processes being similar in volume. Further down we will also show that, with high probability, in time $2n$ no more than $n \log n$ transitions occur. By (6.9) and Lemma 5.2, if $v \notin \hat{B}_t$, then

$$\mathbb{P}\left[e_t^v \neq e_t^v\right] \leq \frac{2(8\beta n^{-1/3} + \frac{4p}{n}(|\xi_t| - |\tilde{\xi}_t|))}{\mathrm{e}^{-2\beta(p+h)}}. \tag{6.10}$$

On the other hand, if $v \in \hat{B}_t$ and we let $z = \frac{2\beta(p+h)}{1-\mathrm{e}^{-2\beta(p+h)}}$, then

$$\mathbb{P}\left[e_t^v \neq e_t^v\right] = d_{TV}\left(e_t^v, e_t^v\right) \leq \mathrm{e}^{-2\beta(p+h)} \int_0^z \mathrm{d}x \, \exp\left(-x\mathrm{e}^{-2\beta(p+h)}\right)$$
$$= 1 - \exp\left(-\frac{2\beta\,(p+h)\,\mathrm{e}^{-2\beta(p+h)}}{1-\mathrm{e}^{-2\beta(p+h)}}\right). \tag{6.11}$$

Observe that, for $v \in W_1^t$, with $\mathbb{P}_{\mathrm{ER}_n(p)}$-high probability

$$\sum_{v \in W_1^t} \left[r(\xi_t, \xi_t^v) + r(\tilde{\xi}_t, \tilde{\xi}_t^v)\right] \geq [1 + o_n(1)]\left|W_1^t\right|. \tag{6.12}$$

Indeed, by the concentration inequalities of Lemma 3.2 and the bound in Lemma 5.1, it follows that $|\xi_t|$ and $|\tilde{\xi}_t|$ are of similar magnitude:

$$\mathbb{P}\big[||\xi_t| - |\tilde{\xi}_t|| \geq n^{5/6}\big] \leq \mathrm{e}^{-n^{2/3}}. \tag{6.13}$$

Therefore, with $\mathbb{P}_{\mathrm{ER}_n(p)}$-high probability, for all but $O(n^{2/3})$ such v,

$$H(\xi_t) - H(\xi_t^v) = [1 + o_n(1)]\left[H\left(\tilde{\xi}_t^v\right) - H\left(\tilde{\xi}_t\right)\right], \tag{6.14}$$

from which (6.12) follows. The rate at which the set W_2^t shrinks is equal to the rate at which it loses $v \in W_2^t$ such that $v \notin \hat{B}_t$, plus the rate at which it loses $v \in W_2^t$ such that $v \in \hat{B}_t$. From (6.9) it follows that the former is bounded from above by $|W_2^t|(8\beta n^{-1/3} + \frac{4p}{n}(|\xi_t| - |\tilde{\xi}_t|))$, while by (3.19) the latter is bounded by $4n^{2/3}$. Therefore, defining the stopping time

$$\upsilon_i = \inf\left\{t\colon\ |W_1^t| = i\right\}, \tag{6.15}$$

we have that

$$\mathbb{P}_{(\xi_t,\tilde{\xi}_t)}\big[\upsilon_{|W_1^t|-1} < \upsilon_{|W_1^t|+1}\big] \geq \frac{|W_1^t|}{|W_2^t|[8\beta n^{-1/3} + \frac{4p}{n}(|\xi_t| - |\tilde{\xi}_t|)] + 4n^{2/3} + |W_1^t|}. \tag{6.16}$$

From Lemma 5.1 we know that (with probability $\geq 1 - \mathrm{e}^{-n^{2/3}}$) neither $|\xi_t|$ nor $|\tilde{\xi}_t|$ will stray beyond $\mathbf{M} + Cn^{5/6}$ and $\mathbf{M} - Cn^{5/6}$ within $n^2 \log n$ steps. Thus,

$$\Big||\xi_t| - |\tilde{\xi}_t|\Big| \leq Cn^{5/6}. \tag{6.17}$$

Hence, for $\left|W_1^t\right| \geq n^{6/7}$ we have that (6.16) is equal to $1 - o_n(1)$.

4. Next suppose that $\left|W_1^t\right| < n^{6/7}$. To bound the rate at which the set W_2^t shrinks, we argue as follows. The rate at which a matching vertex v becomes non-matching equals

$$|r(\xi_t, \xi_t^v) - r(\tilde{\xi}_t, \tilde{\xi}_t^v)|. \tag{6.18}$$

Let

$$\begin{aligned} B_1 &= -2h + \tfrac{2}{n}\big(\deg(v) - 2|E(v, \xi_t)|\big), \\ B_2 &= -2h + \tfrac{2}{n}\big(\deg(v) - 2|E(v, \tilde{\xi}_t)|\big), \\ B_3 &= -h + \tfrac{1}{n}\big(\deg(v) - 2|E(v, \xi_t \cap \tilde{\xi}_t)|\big). \end{aligned} \tag{6.19}$$

For $v \notin \xi_t \cup \tilde{\xi}_t$, we can estimate

$$\begin{aligned} |r(\xi_t, \xi_t^v) - r(\tilde{\xi}_t, \tilde{\xi}_t^v)| &= \left|\mathrm{e}^{-\beta[B_1]_+} - \mathrm{e}^{-\beta[B_2]_+}\right| \\ &\le \mathrm{e}^{-2\beta[B_3]_+} \left|\mathrm{e}^{-\frac{4\beta}{n}|E(v,\xi_t\backslash\tilde{\xi}_t)|} - \mathrm{e}^{-\frac{4\beta}{n}|E(v,\tilde{\xi}_t\backslash\xi_t)|}\right| \\ &\le \mathrm{e}^{-2\beta[B_3]_+} \frac{4\beta}{n} \Big||E(v, \xi_t\backslash\tilde{\xi}_t)| - |E(v, \tilde{\xi}_t\backslash\xi_t)|\Big| \\ &\le [1 + o_n(1)]\, \mathrm{e}^{-2\beta[-p\mathbf{m}-h]_+} \frac{4\beta}{n} \left|E\left(v, W_1^t\right)\right|, \end{aligned} \tag{6.20}$$

where we note that $W_1^t = \xi_t\backslash\tilde{\xi}_t \cup \tilde{\xi}_t\backslash\xi_t$ and use that, by Lemma 3.2 and the bound $|\xi_t\backslash\tilde{\xi}_t| \le |W_1^t| \le n^{6/7}$,

$$\begin{aligned} &\tfrac{1}{n}\left(\deg(v) - 2|E(v, \xi_t \cap \tilde{\xi}_t)|\right) = [1 + o_n(1)]\, p\left(1 - \tfrac{2|\xi_t\cap\tilde{\xi}_t|}{n}\right) \\ &= [1 + o_n(1)]\, p\left(1 - \tfrac{2|\xi_t|}{n}\right) = [1 + o_n(1)]\, p\left(1 - \tfrac{2\mathbf{M}}{n}\right) = -[1 + o_n(1)]\, p\mathbf{m}. \end{aligned} \tag{6.21}$$

Note that since ξ_t and $\tilde{\xi}_t$ disagree at most at $n^{6/7}$ vertices, and since $|\xi_t| = [1 + o_n(1)]\mathbf{M} = [1+o_n(1)]\frac{n}{2}(1+\mathbf{m})$, we have that $|v \in \xi_t\cup\tilde{\xi}_t| = [1+o_n(1)]\frac{n}{2}(1-\mathbf{m})$. Furthermore, since $|E(v, W_1^t)| \le [1+o_n(1)]p|W_1^t|$ and $|V| = n$, we have that

$$\sum_{v\notin\xi_t\cup\tilde{\xi}_t} |r(\xi_t, \xi_t^v) - r(\tilde{\xi}_t, \tilde{\xi}_t^v)| \le [1 + o_n(1)]\, \mathrm{e}^{-2\beta[-p\mathbf{m}-h]_+}\, 2\beta p(1 - \mathbf{m})\, |W_1^t|. \tag{6.22}$$

For $v \in \xi_t \cap \tilde{\xi}_t$, on the other hand, Lemma 3.5 gives that

$$r(\xi_t, \xi_t^v) = r(\tilde{\xi}_t, \tilde{\xi}_t^v) = 1 \tag{6.23}$$

for all but $O(n^{2/3})$ many such v. If v is such that $r(\xi_t, \xi_t^v) \neq r(\tilde{\xi}_t, \tilde{\xi}_t^v)$, then a computation identical to the one leading to (6.22) gives that

$$\sum_{v\in\xi_t\cap\tilde{\xi}_t} |r\left(\xi_t, \xi_t^v\right) - r(\tilde{\xi}_t, \tilde{\xi}_t^v)| = O(n^{-1/6})\left|W_1^t\right|. \tag{6.24}$$

Combining (6.22) and (6.24), we obtain

$$\sum_{v\in W_t^2} |r\left(\xi_t, \xi_t^v\right) - r(\tilde{\xi}_t, \tilde{\xi}_t^v)| \leq [1 + o_n(1)]\, \mathrm{e}^{-2\beta[-p\mathbf{m}-h]_+}\, 2\beta p(1-\mathbf{m})\left|W_1^t\right|, \tag{6.25}$$

which bounds the rate at which W_2^t shrinks.

5. To bound the rate at which W_1^t shrinks, we argue as follows. The rate at which a non-matching vertex v becomes matching equals

$$r(\xi_t, \xi_t^v) + r(\tilde{\xi}_t, \tilde{\xi}_t^v). \tag{6.26}$$

Note that, for every $v \in W_1^t$,

$$H\left(\xi_t^v\right) - H\left(\xi_t\right) = -[1 + o_n(1)]\left[H(\tilde{\xi}_t^v) - H(\tilde{\xi}_t)\right], \tag{6.27}$$

since, up to an arithmetic correction of magnitude $|W_1^t| = O(n^{6/7})$, v has the same number of neighbours in ξ_t as in $\tilde{\xi}_t$. Hence it follows that

$$\sum_{v\in W_1^t}\left[r(\xi_t, \xi_t^v) + r(\tilde{\xi}_t, \tilde{\xi}_t^v)\right] = [1 + o_n(1)]\left(\mathrm{e}^{-2\beta[-p\mathbf{m}-h]_+} + \mathrm{e}^{-2\beta[p\mathbf{m}+h]_+}\right)|W_1^t|, \tag{6.28}$$

which bounds the rate at which W_1^t shrinks.

6. Combining (6.25) and (6.28), and noting that $p\mathbf{m} + h < 0$, we see that $\left|W_1^t\right|$ is contracting when

$$[1 + o_n(1)]\left(\mathrm{e}^{2\beta(p\mathbf{m}+h)} + 1\right)\left|W_1^t\right| > [1 + o_n(1)]\left(\mathrm{e}^{2\beta(p\mathbf{m}+h)}\, 4\beta p\right)\left|W_1^t\right|. \tag{6.29}$$

For this in turn it suffices that

$$\mathrm{e}^{2\beta(p\mathbf{m}+h)} + 1 > \mathrm{e}^{2\beta(p\mathbf{m}+h)} 2\beta p(1-\mathbf{m}). \tag{6.30}$$

7. Note from the definition of $\mathbf{m}$ in (1.16) that, up to a correction factor of $1+o_n(1)$, $\mathbf{m}$ solves the equation $J(\mathbf{m}) = 0$ with

$$J_{p,\beta,h}(a) = 2\lambda\left(a+\frac{h}{p}\right)+\log\left(\frac{1-a}{1+a}\right), \qquad \lambda = \beta p, \tag{6.31}$$

i.e.,

$$\frac{1+\mathbf{m}}{1-\mathbf{m}} = \mathrm{e}^{2\lambda\left(\mathbf{m}+\frac{h}{p}\right)}. \tag{6.32}$$

Hence from (6.30) it follows that $|W_1^t|$ is contracting whenever we are in the metastable regime and the inequality

$$\lambda < \frac{1}{1-\mathbf{m}^2} \tag{6.33}$$

is satisfied. From (2.17) it follows that the equality

$$\lambda = \frac{1}{1-a^2} \tag{6.34}$$

holds for $a = a_\lambda = -\sqrt{1-1/\lambda}$, which in turn is bounded between the values $\mathbf{m} < a_\lambda < \mathbf{t} < 0$, and therefore

$$\frac{1}{1-\mathbf{m}^2} > \frac{1}{1-a_\lambda^2} = \lambda. \tag{6.35}$$

This shows that $|W_1^t|$ is contracting whenever we are in the metastable regime.

8. To conclude, we summarise the implication of the contraction of the process $|W_1^t|$. The probability in (6.16) is equal to $1 - O_n(n^{\frac{5}{6}-\frac{6}{7}})$ for $|W_1^t| > n^{6/7}$, and is strictly larger than $\frac{1}{2}$ for $|W_1^t| \leq n^{6/7}$. Furthermore, from (6.12) we know that the rate at which W_1^t shrinks is ≥ 1. This allows us to ensure that sufficiently many steps are made by time $2n$ to allow W_1^t to contract to the empty set. In particular, the number steps taken by W_1^t up to time $2n$ is bounded from below by a Poisson point process $N(t)$ with unit rate, for which we have

$$\mathbb{P}\left[N(2n) \leq \frac{3n}{2}\right] \leq 2n\frac{(2n)^{\left(\frac{3n}{2}\right)}\mathrm{e}^{-2n}}{\left(\frac{3n}{2}\right)!} \leq 2n\left(\frac{4n}{3}\right)^{\left(\frac{3n}{2}\right)}\mathrm{e}^{-\frac{n}{2}} \leq 1.07^{\left(-\frac{n}{2}\right)}. \tag{6.36}$$

In other words, with probability exponentially close to 1, we have that at least $3n/2$ jumps are made in time $2n$. To bound the probability that W_1^t has not converged to the empty set, note that this probability decreases in the number of transitions made by W_1^t. Therefore, without loss of generality, we may assume that $\frac{3n}{2}$ transitions were made, and that we start with $|W_1^0| = n$. We claim that, with high probability, in time $2n$, W_1^t takes at most $\frac{100n}{\log n}$ increasing steps (i.e., $i \to i+1$) in the interval $[n^{5/6}, n]$. Indeed, note that the probability of the latter occurring is less than

$$2^M O\big(n^{-1/42}\big)^{\frac{100n}{\log n}} = O\big(\mathrm{e}^{-n}\big). \tag{6.37}$$

It follows that at least $\frac{n}{2}[1 + o_n(1)]$ steps are taken in the interval $[0, n^{5/6}]$. But then, using (6.16), we have that the probability of an increasing step is at most $\frac{1}{2} - \epsilon$ for some $\epsilon > 0$, and therefore the probability of that event is at most

$$\begin{aligned} &2^{\frac{n}{2}[1+o_n(1)]} \left(\tfrac{1}{2} + \epsilon\right)^{\frac{n}{4}[1+o_n(1)]} \left(\tfrac{1}{2} - \epsilon\right)^{\frac{n}{4}[1+o_n(1)]} \\ &= 4^{\frac{n}{4}[1+o_n(1)]} \left(\tfrac{1}{4} - \epsilon^2\right)^{\frac{n}{4}[1+o_n(1)]} = \left(1 - 4\epsilon^2\right)^{\frac{n}{4}[1+o_n(1)]}. \end{aligned} \tag{6.38}$$

Finally, observing that in the entire proof so far the largest probability for any of the bounds not to hold is $O(\mathrm{e}^{-n^{-2/3}})$ (see (6.13) and the paragraph following (6.16)), we get

$$\mathbb{P}\left[|W_1^{2n}| > 0\right] \le O\big(\mathrm{e}^{-n^{2/3}}\big) \tag{6.39}$$

and so the claim of the lemma follows. □

6.2 Long-Term Scheme

Corollary 6.3 (Long-Term Coupling) *Let $\delta > 0$. With $\mathbb{P}_{\mathrm{ER}_n(p)}$-probability tending to 1 as $n \to \infty$, there is a coupling of $\{\xi_t\}_{t\ge 0}$ and $\{\tilde{\xi}_{\tilde{t}}\}_{t\ge 0}$, and there are times t and $\tilde{t}$ with $\max(t, \tilde{t}) < \mathrm{e}^{n\Gamma^\star - \delta n}$, such that*

$$\mathbb{P}\big[\xi_t \ne \tilde{\xi}_{\tilde{t}}\big] \le \mathrm{e}^{-n\delta + O(n^{2/3})}. \tag{6.40}$$

Proof Let $\mathbf{s}_i$ be the ith return-time of $\{\xi_t\}_{t\ge 0}$ to $A_{\mathbf{M}}$. Define $\{\tilde{\mathbf{s}}_i\}_{i\ge 0}$ in an analogous manner for $\{\tilde{\xi}_t\}_{t\ge 0}$. Then we can define a coupling of $\{\xi_t\}_{t\ge 0}$ and $\{\tilde{\xi}_t\}_{t\ge 0}$ as follows. For $i \ge 0$ and $0 \le s \le 2n$, couple $\xi_{\mathbf{s_i}+s}$ and $\tilde{\xi}_{\tilde{\mathbf{s}}_\mathbf{i}+s}$ as described in Lemma 6.1. For times $t \in (\mathbf{s_i} + 2n, \mathbf{s_{i+1}})$ and $\tilde{t} \in (\tilde{\mathbf{s}}_\mathbf{i} + 2n, \tilde{\mathbf{s}}_\mathbf{i+1})$, let $\{\xi_t\}_{t\ge 0}$ and $\{\tilde{\xi}_{\tilde{t}}\}_{t\ge 0}$ run independently of each other. Terminate this coupling at the first time t such that $t = \mathbf{s_i} + s$ for some $s \le 2n$ and $\xi_t = \tilde{\xi}_{\tilde{t}}$ with $\tilde{t} = \tilde{\mathbf{s}}_\mathbf{i} + s$, from which point onward we simply let $\xi_t = \tilde{\xi}_{\tilde{t}}$. It is easy to see that the coupling above is an attempt at repeating the coupling scheme of Lemma 6.1 until the paths of the two processes have crossed. To avoid having to wait until both processes are in $A_{\mathbf{M}}$ at the same time, the coupling defines a joint distribution of ξ_t and $\tilde{\xi}_{\tilde{t}}$.

Note that, by Lemma 5.4, with probability of at least $1 - \mathrm{e}^{-\delta n + O(n^{2/3})}$, $\{\xi_t\}_{t \geq 0}$ will visit $A_{\mathbf{M}}$ at least $\mathrm{e}^{(\Gamma^\star - \delta)n}$ times before reaching $A_{\mathbf{S}}$, for any $\delta > 0$. The same statement is true for $\{\tilde{\xi}_{\tilde{t}}\}_{\tilde{t} \geq 0}$. Assuming that the aforementioned event holds for both ξ_t and $\tilde{\xi}_{\tilde{t}}$, we see that the probability that the coupling does not succeed (i.e., the two trajectories do not intersect as described earlier) is at most

$$\left[O\left(\mathrm{e}^{-n^{-2/3}}\right) \right]^{\mathrm{e}^{(\Gamma^\star - \delta)n}}. \tag{6.41}$$

Therefore, the probability that the coupling does not succeed before either of $\{\xi_t\}_{t \geq 0}$ or $\{\tilde{\xi}_{\tilde{t}}\}_{\tilde{t} \geq 0}$ reaches $A_{\mathbf{S}}$ is at most $\mathrm{e}^{-\delta n + O(n^{2/3})}$. □

7 Proof of the Main Metastability Theorem

In this section we prove Theorem 1.4.

Proof The key is to show that with $\mathbb{P}_{\mathrm{ER}_n(p)}$-probability tending to 1 as $n \to \infty$, for any $\xi_0^u \in A_{\mathbf{M}^u}$, $\xi_0 \in A_{\mathbf{M}}$ and $\xi_0^l \in A_{\mathbf{M}^l}$,

$$\mathbb{E}_{\xi_0^u}\left[\tau_{\mathbf{S}^u}\right] \leq \mathbb{E}_{\xi_0}\left[\tau_{\mathbf{S}}\right] \leq \mathbb{E}_{\xi_0^l}\left[\tau_{\mathbf{S}^l}\right]. \tag{7.1}$$

Note that $\mathbb{E}_{\xi_0}[\tau_{\mathbf{S}}]$ is the same for all $\xi_0 \in A_{\mathbf{M}}$ up to a multiplicative factor of $1 + o_n(1)$, as was shown in Sects. 5.2 and 6. Therefore it is suffices to find *some* convenient $\xi \in A_{\mathbf{M}}$ for which we can prove the aforementioned theorem.

1. Our proof follows four steps:

(1) Recall that for $A \subset S_n$, μ_A is the probability distribution μ conditioned to the set A. Starting from the initial distribution $\mu_{A_{\mathbf{M}}}$ on the set $A_{\mathbf{M}}$, the trajectory segment taken by ξ_t from ξ_0 to ξ_τ, with $\tau = \min\{\tau_{\mathbf{M}}, \tau_{\mathbf{S}}\}$, can be coupled to the analogous trajectory segments taken by ξ_t^l and ξ_t^u, starting in $A_{\mathbf{M}^l}$ and $A_{\mathbf{M}^u}$, respectively, and this coupling can be done in such a way that the following two conditions hold:

   (a) If ξ_t reaches $A_{\mathbf{S}}$ before returning to $A_{\mathbf{M}}$ (i.e., $\tau_{\mathbf{S}} < \tau_{\mathbf{M}}$), then ξ_t^u reaches $A_{\mathbf{S}^u}$ before returning to $A_{\mathbf{M}^u}$.
   (b) If ξ_t returns to $A_{\mathbf{M}}$ before reaching $A_{\mathbf{S}}$ (i.e., $\tau_{\mathbf{M}} < \tau_{\mathbf{S}}$), then ξ_t^l returns to $A_{\mathbf{M}^l}$ before reaching $A_{\mathbf{S}^l}$.

(2) We show that if ξ_t has initial distribution $\mu_{A_{\mathbf{M}}}$ and $\tau_{\mathbf{M}} < \tau_{\mathbf{S}}$, then upon returning to $A_{\mathbf{M}}$ the distribution of ξ_t is once again given by $\mu_{A_{\mathbf{M}}}$. This implies that the argument in Step (1) can be applied repeatedly, and that

the number of returns ξ_t makes to $A_{\mathbf{M}}$ before reaching $A_{\mathbf{S}}$ is bounded from below by the number of returns ξ_t^u makes to $A_{\mathbf{M}^u}$ before reaching $A_{\mathbf{S}^u}$, and is bounded from above by the number of returns ξ_t^l makes to $A_{\mathbf{M}^l}$ before reaching $A_{\mathbf{S}^l}$.

(3) Using Lemma 3.7, we bound the time between unsuccessful excursions, i.e., the expected time it takes for ξ_t, when starting from $\mu_{A_{\mathbf{M}}}$, to return to $A_{\mathbf{M}}$, given that $\tau_{\mathbf{M}} < \tau_{\mathbf{S}}$. This bound is used together with the outcome of Step (2) to obtain the bound

$$\mathbb{E}^u_{\mu_{A_{\mathbf{M}^u}}}\left[\tau^u_{\mathbf{S}}\right] \leq \mathbb{E}_{\mu_{A_{\mathbf{M}}}}\left[\tau_{\mathbf{S}}\right] \leq \mathbb{E}^l_{\mu_{A_{\mathbf{M}^l}}}\left[\tau^l_{\mathbf{S}}\right]. \tag{7.2}$$

Here, the fact that the conditional average return time is bounded by some large constant rather than 1 does not affect the sandwich in (7.2), because the errors coming from the perturbation of the magnetic field in the Curie-Weiss model are polynomial in n (see below).

(4) We complete the proof by showing that, for any distribution μ_0 restricted to $A_{\mathbf{M}}$,

$$\mathbb{E}_{\mu_0}\left[\tau_{\mathbf{S}}\right] = [1 + o_n(1)]\, \mathbb{E}_{\mu_{A_{\mathbf{M}}}}\left[\tau_{\mathbf{S}}\right]. \tag{7.3}$$

2. Before we turn to the proof of these steps, we explain how the bound on the exponent in the prefactor of Theorem 1.4 comes about. Return to (2.4). The magnetic field h is perturbed to $h \pm (1+\epsilon)\log(n^{11/6})/n$. We need to show how this affects the formulas for the average crossover time in the Curie-Weiss model. For this we use the computations carried out in [4, Chapter 13]. According to [4, Eq. (13.2.4)] we have, for any $\xi \in A_{\mathbf{M}_n}$ and any $\epsilon > 0$,

$$\mathbb{E}_{\xi}\left[\tau_{A_{\mathbf{S}_n}}\right] = [1 + o_n(1)]\, \frac{2}{1-\mathbf{t}}\, \mathrm{e}^{\beta n[R_n(\mathbf{t}) - R_n(\mathbf{m})]}\, \frac{1}{n}\, \mathbf{S}_n \tag{7.4}$$

with

$$\mathbf{S}_n = \sum_{\substack{a,a' \in \Gamma_n \\ |a-\mathbf{t}|<\epsilon,\, |a'-\mathbf{m}|<\epsilon}} \mathrm{e}^{\beta n[R_n(\mathbf{a}) - R_n(\mathbf{t})] - \beta n[R_n(a') - R_n(\mathbf{m})]}, \tag{7.5}$$

where R_n is the free energy defined by $R_n' = -J_n/2\beta$ (recall (1.20)). (Here we suppress the dependence on β, h and note that (7.4) carries an extra factor $\frac{1}{n}$ because [4, Chapter 13] considers a discrete-time dynamics where at every unit of time a single spin is drawn uniformly at random and is flipped with a probability that is given by the right-hand side of (1.6).) According to [4, Eq. (13.2.5)–(13.2.6)] we have

$$I_n(a) - I(a) = [1 + o_n(1)]\frac{1}{2n}\log\left(\frac{1}{2}\pi n(1-a^2)\right), \qquad a \in [-1, 1], \tag{7.6}$$

so that

$$\mathrm{e}^{\beta n[R_n(a)-R(a)]} = [1 + o_n(1)]\sqrt{\frac{1}{2}\pi n(1-a^2)}, \qquad a \in [-1,1], \tag{7.7}$$

where R is the limiting free energy defined by $R' = -J/2\beta$ (recall (1.27)). Inserting (7.7) into (7.4), we get

$$\mathbb{E}_\xi\left[\tau_{A_{\mathbf{S}_n}}\right] = [1 + o_n(1)]\,\frac{2}{1+\mathbf{t}}\sqrt{\frac{1-\mathbf{t}^2}{1-\mathbf{m}^2}}\,\mathrm{e}^{\beta n[R(\mathbf{t})-R(\mathbf{m})]}\,\frac{1}{n}\,\mathbf{S}_n^* \tag{7.8}$$

with

$$\mathbf{S}_n^* = \sum_{\substack{a,a' \in \Gamma_n \\ |a-\mathbf{t}|<\epsilon,\ |a'-\mathbf{m}|<\epsilon}} \mathrm{e}^{\beta n[R(\mathbf{a})-R(\mathbf{t})]-\beta n[R(a')-R(\mathbf{m})]}. \tag{7.9}$$

Finally, according to [4, Eq. (13.2.9)–(13.2.11)] we have, with the help of a Gaussian approximation,

$$\lim_{n\to\infty}\frac{1}{n}\,\mathbf{S}_n^* = \frac{\pi}{2\beta\sqrt{[R''(\mathbf{m})[-R''(\mathbf{t})]}}. \tag{7.10}$$

Putting together (7.8) and (7.10), we see how Theorem 1.3 arises as the correct formula for the Curie-Weiss model.

3. The above computations are for β, h fixed and $p = 1$. We need to investigate what changes when $p \in (0,1)$, β is fixed, but h depends on n:

$$h_n = h \pm (1+\epsilon)\,\frac{\log(n^{11/6})}{n}. \tag{7.11}$$

We write R_n^n to denote R_n when h is replaced by h_n. In the argument in [4, Chapter 13] leading up to (7.4), the approximation only enters through the prefactor. But since $h_n \to h$ as $n \to \infty$, the perturbation affects the prefactor only by a factor $1 + o_n(1)$. Since h plays no role in (7.6) and $R_n^n(a) - R^n(a) = \frac{1}{\beta}[I_n(a) - I(a)]$ (recall (1.19) and (1.26)), we get (7.8) with exponent $\beta n[R^n(\mathbf{t}) - R^n(\mathbf{m})]$ and (7.9) with exponent $\beta n[R^n(\mathbf{a}) - R^n(\mathbf{t})] - \beta n[R^n(a') - R^n(\mathbf{m})]$. The latter affects the Gaussian approximation behind (7.10) only by a factor $1+o_n(1)$. However, the former leads to an error term in the exponent, compared to the Curie Weiss model, that equals

$$\begin{aligned}
\beta n[R^n(\mathbf{t}) - R^n(\mathbf{m})] - \beta n[R(\mathbf{t}) - R(\mathbf{m})] &= \beta n \textstyle\int_{\mathbf{m}}^{\mathbf{t}} \mathrm{d}a\,[(R^n)'(a) - R'(a)] \\
&= \beta n \textstyle\int_{\mathbf{m}}^{\mathbf{t}} \mathrm{d}a\,[-(h_n - h)] = \beta(\mathbf{t}-\mathbf{m})\,n(h-h_n) \\
&= \mp\beta(\mathbf{t}-\mathbf{m})\,(1+\epsilon)\log(n^{11/6}).
\end{aligned} \tag{7.12}$$

The exponential of this equals $n^{\mp\beta(\mathbf{t}-\mathbf{m})\,(1+\epsilon)(11/6)}$, which proves Theorem 1.4 with the bound in (1.33) because ϵ is arbitrary.

Proof of Step (1)
This step is a direct application of Lemma 4.3.

Proof of Step (2)
Write $=^d$ to denote equality in distribution. Let $\xi_0 =^d \mu_{A_\mathbf{M}}$, and recall that $\tau_\mathbf{M}$ is the first return time of ξ_t to $A_\mathbf{M}$ once the initial state ξ_0 has been left. We want to show that $\xi_{\tau_\mathbf{M}} =^d \mu_{A_\mathbf{M}}$ or, in other words, that $\mathbb{P}_{\mu_{A_\mathbf{M}}}[\xi_{\tau_\mathbf{M}} = \sigma] = \mu_{A_\mathbf{M}}(\sigma)$ for any $\sigma \in A_\mathbf{M}$. To facilitate the argument, we begin by defining the set of all finite permissible trajectories $\mathscr{T}$, i.e.,

$$\mathscr{T} = \bigcup_{N\in\mathbb{N}} \left\{ \gamma = \{\gamma_i\}_{i=0}^N \in S_n^N :\ ||\gamma_i| - |\gamma_{i+1}|| = 1\ \forall\, 0 \le i \le N-1 \right\}. \tag{7.13}$$

Let $\gamma \in \mathscr{T}$ be any finite trajectory beginning at $\gamma_0 \in A_\mathbf{M}$, ending at $\gamma_{|\gamma|-1} = \sigma \in A_\mathbf{M}$, and satisfying $\gamma_i \notin A_\mathbf{M}$ for $0 < i < |\gamma| - 1$. Then the probability that ξ_t follows the trajectory γ is given by

$$\begin{aligned}
\mathbb{P}\left[\xi_t \text{ follows } \gamma\right] &= \mu_{A_\mathbf{M}}(\gamma_0)\, P(\gamma_0, \gamma_1) \times \cdots \times P\left(\gamma_{|\gamma|-2}, \sigma\right) \\
&= \tfrac{1}{\mu(A_\mathbf{M})} \mu(\gamma_0)\, P(\gamma_0, \gamma_1) \times \cdots \times P\left(\gamma_{|\gamma|-2}, \sigma\right) \\
&= \tfrac{1}{\mu(A_\mathbf{M})} \mu(\sigma)\, P\left(\sigma, \gamma_{|\gamma|-2}\right) \times \cdots \times P(\gamma_1, \gamma_0) \\
&= \mu_{A_\mathbf{M}}(\sigma)\, P\left(\sigma, \gamma_{|\gamma|-2}\right) \times \cdots \times P(\gamma_1, \gamma_0),
\end{aligned} \tag{7.14}$$

where the third line follows from reversibility. Thus, if we let $\mathscr{T}(\sigma)$ be the set of all trajectories in $\mathscr{T}$ that begin in $A_\mathbf{M}$, end at σ, and do not visit $A_\mathbf{M}$ in between, then we get

$$\begin{aligned}
\mathbb{P}_{\mu_{A_\mathbf{M}}}\left[\xi_{\tau_\mathbf{M}} = \sigma\right] &= \textstyle\sum_{\gamma\in\mathscr{T}(\sigma)} \mu_{A_\mathbf{M}}(\sigma)\, P\left(\sigma, \gamma_{|\gamma|-2}\right) \times \cdots \times P(\gamma_1, \gamma_0) \\
&= \mu_{A_\mathbf{M}}(\sigma)\, \mathbb{P}_\sigma\left[\tau_\mathbf{M} < \infty\right] \\
&= \mu_{A_\mathbf{M}}(\sigma),
\end{aligned} \tag{7.15}$$

where we use recurrence and the law of total probability, since the trajectories in $\mathscr{T}(\sigma)$, when reversed, give all possible trajectories that start at $\sigma \in A_\mathbf{M}$ and return to $A_\mathbf{M}$ in a finite number of steps. This shows that if ξ_t has initial distribution $\mu_{A_\mathbf{M}}$, then it also has this distribution upon every return to $A_\mathbf{M}$.

We can now define a segment-wise coupling of the trajectory taken by ξ_t with the trajectories taken by ξ_t^u and ξ_t^l. First, we define the subsets of trajectories that start and end in particular regions of the state space: (1) $\mathscr{T}_{\sigma,L,K}$ is the set of trajectories that start at a particular configuration σ and end in A_K without ever visiting A_K or A_L in between, for some $K < L$; (2) $\mathscr{T}_{\sigma,L,L}$ is the set of trajectories that start at some σ and end in A_L without ever visiting A_K or A_L in between; (3) $\mathscr{T}_{\sigma,L}$ is the union of the two aforementioned sets. In explicit form,

$$\begin{aligned}
\mathscr{T}_{\sigma,L,K} &= \left\{\gamma \in \mathscr{T}\colon \gamma_0 = \sigma, \gamma_{|\gamma|-1} \in A_K, K < \left|\gamma_j\right| < L \ \forall 0 < j < |\gamma| - 1\right\}, \\
\mathscr{T}_{\sigma,L,L} &= \left\{\gamma \in \mathscr{T}\colon \gamma_0 = \sigma, \gamma_{|\gamma|-1} \in A_L, K < \left|\gamma_j\right| < L \ \forall 0 < j < |\gamma| - 1\right\}, \\
\mathscr{T}_{\sigma,L} &= \mathscr{T}_{\sigma,L,K} \cup \mathscr{T}_{\sigma,L,L}.
\end{aligned} \tag{7.16}$$

By Step (1), for any $\xi_0^l \in A_{\mathbf{M}^l}$ and $\xi_0^u \in A_{\mathbf{M}^u}$,

$$\mathbb{P}^l_{\xi_0^l}\left[\mathscr{T}_{\xi_0^l,\mathbf{S}^l,\mathbf{S}^l}\right] \le \mathbb{P}_{\xi_0}\left[\mathscr{T}_{\xi_0,\mathbf{S},\mathbf{S}}\right] \le \mathbb{P}^u_{\xi_0^u}\left[\mathscr{T}_{\xi_0^u,\mathbf{S}^u,\mathbf{S}^u}\right]. \tag{7.17}$$

It is clear that the two probabilities at either end of (7.17) are independent of the starting points ξ_0^l and ξ_0^u. By the argument given above, if for the probability in the middle $\xi_0 =^d \mu_{A_\mathbf{M}}$, then each subsequent return to $A_\mathbf{M}$ also has this distribution. For this reason, we may define a coupling of the trajectories as follows.

Sample a trajectory segment γ^l from $\mathscr{T}_{\xi_0^l,\mathbf{S}^l}$ for the process ξ_t^l. If γ^l happens to be in $\mathscr{T}_{\xi_0^l,\mathbf{S}^l,\mathbf{S}^l}$, then by (7.17) we may sample a trajectory segment γ from $\mathscr{T}_{\xi_0,\mathbf{S},\mathbf{S}}$ for the process ξ_t, and a trajectory segment γ^u from $\mathscr{T}_{\xi_0^u,\mathbf{S}^u,\mathbf{S}^u}$ for the process ξ^u. Otherwise, $\gamma^l \in \mathscr{T}_{\xi_0^l,\mathbf{S}^l,\mathbf{M}^l}$, and we independently take $\gamma \in \mathscr{T}_{\xi_0,\mathbf{S},\mathbf{S}}$ with probability $\mathbb{P}_{\xi_0}[\mathscr{T}_{\xi_0,\mathbf{S},\mathbf{S}}] - \mathbb{P}^l_{\xi_0^l}[\mathscr{T}_{\xi_0^l,\mathbf{S}^l,\mathbf{S}^l}]$, and $\gamma \in \mathscr{T}_{\xi_0,\mathbf{S},\mathbf{M}}$ otherwise. If $\gamma \in \mathscr{T}_{\xi_0,\mathbf{S},\mathbf{S}}$, then sample γ^u from $\mathscr{T}_{\xi_0^u,\mathbf{S}^u,\mathbf{S}^u}$. Otherwise $\gamma \in \mathscr{T}_{\xi_0,\mathbf{S},\mathbf{M}}$, and so take independently $\gamma^u \in \mathscr{T}_{\xi_0^u,\mathbf{S}^u,\mathbf{S}^u}$ with probability $\mathbb{P}^u_{\xi_0^u}[\mathscr{T}_{\xi_0^u,\mathbf{S}^u,\mathbf{S}^u}] - \mathbb{P}_{\xi_0}[\mathscr{T}_{\xi_0,\mathbf{S},\mathbf{S}}]$, and $\gamma^u \in \mathscr{T}_{\xi_0^u,\mathbf{S}^u,\mathbf{M}^u}$ with the remaining probability. We glue together the sampling of segments leaving and returning to $A_{\mathbf{M}^l}/A_\mathbf{M}/A_{\mathbf{M}^u}$ with the next sampling of such segments. This results in trajectories for ξ^u, ξ, and ξ^l that reach $A_{\mathbf{S}^u}/A_\mathbf{S}/A_{\mathbf{S}^l}$, in that particular order.

Proof of Step (3) and Step (4)
These two steps are immediate from Lemma 3.7.

□

8 Conditional Average Return Time for Inhomogeneous Random Walk

In this section we prove Lemma 3.7. In Sects. 8.1–8.2 we compute the harmonic function and the conditional average return time for an arbitrary nearest-neighbour random walk on a finite interval. In Sect. 8.3 we use these computations to prove the lemma.

8.1 Harmonic Function

Consider a *nearest-neighbour* random walk on the set $\{0, \dots, N\}$ with strictly positive transition probabilities $p(x, x+1)$ and $p(x, x-1)$, $0 < x < N$, and with 0 and N acting as absorbing boundaries. Let τ_0 and τ_N denote the first hitting times of 0 and N. The *harmonic function* is defined as

$$h_N(x) = \mathbb{P}_x(\tau_N < \tau_0), \quad 0 \le x \le N, \tag{8.1}$$

where $\mathbb{P}_x$ is the law of the random walk starting from x. This is the unique solution of the recursion relation

$$h_N(x) = p(x, x+1)h_N(x+1) + p(x, x-1)h_N(x-1), \quad 0 < x < N, \tag{8.2}$$

with boundary conditions

$$h_N(0) = 0, \qquad h_N(N) = 1. \tag{8.3}$$

Since $p(x, x+1) + p(x, x-1) = 1$, the recursion can be written as

$$p(x, x+1)[h_N(x+1) - h_N(x)] = p(x, x-1)[h_N(x) - h_N(x-1)]. \tag{8.4}$$

Define $\Delta h_N(x) = h_N(x+1) - h_N(x)$, $0 \le x < N$. Iteration gives

$$\Delta h_N(x) = \pi[1, x]\, \Delta h_N(0), \quad 0 \le x < N, \tag{8.5}$$

where we define

$$\pi(I) = \prod_{z \in I} \frac{p(z, z-1)}{p(z, z+1)}, \quad I \subseteq \{1, \dots, N-1\}, \tag{8.6}$$

with the convention that the empty product equals 1. Since $h_N(0) = 0$, we have

$$h_N(x) = \sum_{z=0}^{x-1} \Delta h_N(z) = \left(\sum_{z=0}^{x-1} \pi[1, z] \right) \Delta h_N(0), \qquad 0 < x \le N. \tag{8.7}$$

Put $C = \Delta h_N(0)$, and abbreviate

$$R(x) = \sum_{z=0}^{x-1} \pi[1, z], \quad 0 \le x \le N. \tag{8.8}$$

Since $h_N(N) = 1$, we have $C = 1/R(N)$. Therefore we arrive at

$$h_N(x) = \frac{R(x)}{R(N)}, \quad 0 \le x \le N. \tag{8.9}$$

Remark 8.1 For simple random walk we have $p(x, x \pm 1) = \frac{1}{2}$, hence $\pi[1, x] = 1$ and $R(x) = x$, and so

$$h_N(x) = \frac{x}{N}, \quad 0 \le x \le N, \tag{8.10}$$

which is the standard gambler's ruin formula.

8.2 Conditional Average Hitting Time

We are interested in the quantity

$$e_N(x) = \mathbb{E}_x(\tau_N \mid \tau_N < \tau_0), \qquad 0 < x \le N. \tag{8.11}$$

The conditioning amounts to taking the *Doob transformed* random walk, which has transition probabilities

$$q(x, x \pm 1) = p(x, x \pm 1)\,\frac{h_N(x \pm 1)}{h_N(x)}. \tag{8.12}$$

We have the recursion relation

$$e_N(x) = 1 + q(x, x+1)e_N(x+1) + q(x, x-1)e_N(x-1), \quad 0 < x < N, \tag{8.13}$$

in this section we prove with boundary conditions

$$e_N(N) = 0, \quad e_N(1) = 1 + e_N(2). \tag{8.14}$$

Putting $f_N(x) = h_N(x)e_N(x)$, we get the recursion

$$f_N(x) = h_N(x) + p(x, x+1) f_N(x+1) + p(x, x-1) f_N(x-1), \quad 0 < x < N, \tag{8.15}$$

which can be rewritten as

$$p(x, x+1)[f_N(x+1) - f_N(x)] = p(x, x-1)[f_N(x) - f_N(x-1)] - h_N(x). \tag{8.16}$$

Define $\Delta f_N(x) = f_N(x+1) - f_N(x), 0 < x < N$. Iteration gives

$$\Delta f_N(x) = \pi(1,x]\,\Delta f_N(1) - \chi(1,x], \quad 0 < x < N, \tag{8.17}$$

with

$$\chi(1,x] = \sum_{y=2}^{x} \pi(y,x]\,\frac{h_N(y)}{p(y,y+1)}, \quad 0 < x < N. \tag{8.18}$$

Since $f_N(N) = 0$, we have

$$f_N(x) = -\sum_{z=x}^{N-1} \Delta f_N(z) = \sum_{z=x}^{N-1} \chi(1,z] - \left(\sum_{z=x}^{N-1} \pi(1,z]\right) \Delta f_N(1), \quad 0 < x < N, \tag{8.19}$$

or

$$e_N(x) = \frac{1}{h_N(x)} \sum_{z=x}^{N-1} \chi(1,z] - \frac{1}{h_N(x)} \left(\sum_{z=x}^{N-1} \pi(1,z]\right) \Delta f_N(1), \quad 0 < x < N. \tag{8.20}$$

Put $C = \Delta f_N(1)$, and abbreviate

$$A(x) = \sum_{z=x}^{N-1} \pi(1,z], \quad B(x) = \sum_{z=x}^{N-1} \chi(1,z], \quad 0 < x \leq N. \tag{8.21}$$

Then

$$e_N(x) = \frac{1}{h_N(x)}\big[B(x) - CA(x)\big]. \tag{8.22}$$

Since $e_N(1) = 1 + e_N(2)$, we have

$$C = \frac{[h_N(2)B(1) - h_N(1)B(2)] - h_N(1)h_N(2)}{h_N(2)A(1) - h_N(1)A(2)}. \tag{8.23}$$

Abbreviate

$$\bar{R}(x) = \sum_{z=0}^{x-1} \pi(1,z], \quad \bar{S}(x) = \sum_{z=0}^{x-1} \chi(1,z], \quad 0 < x \leq N. \tag{8.24}$$

Then

$$A(x) = \bar{R}(N) - \bar{R}(x), \quad B(x) = \bar{S}(N) - \bar{S}(x), \quad 0 < x < N. \tag{8.25}$$

Note that $h_N(x) = R(x)/R(N) = \bar{R}(x)/\bar{R}(N)$, because $\pi[1,z] = \pi(1)\pi(1,z]$ and a common factor $\pi(1)$ drops out. Note further that $\bar{R}(1) = 1$, $\bar{R}(2) = 2$, while $\bar{S}(1) = \bar{S}(2) = 0$. Therefore

$$C = \frac{\bar{S}(N)}{\bar{R}(N)} - \frac{2}{\bar{R}(N)^2}. \tag{8.26}$$

Therefore we arrive at

$$e_N(x) = \bar{S}(N) - \frac{\bar{R}(N)}{\bar{R}(x)}\,\bar{S}(x) + \frac{2}{\bar{R}(x)} - \frac{2}{\bar{R}(N)}, \quad 0 < x \le N. \tag{8.27}$$

Abbreviating

$$\bar{T}(x) = \bar{S}(x)\bar{R}(N) = \sum_{z=0}^{x-1}\sum_{y=2}^{z} \frac{\pi(y,z]}{p(y,y+1)}\,\bar{R}(y), \quad \bar{U}(x) = \frac{\bar{T}(x)-2}{\bar{R}(x)}, \tag{8.28}$$

we can write

$$e_N(x) = \bar{U}(N) - \bar{U}(x), \quad 0 < x \le N. \tag{8.29}$$

Remark 8.2 For simple random walk we have $p(x, x \pm 1) = \frac{1}{2}$, $\pi(y,z] = 1$, $\bar{R}(x) = x$, $\bar{S}(x) = \frac{1}{3N}(x^3 - 7x + 6)$ and $\bar{U}(x) = \frac{1}{3}(x^2 - 7)$, and so

$$e_N(x) = \frac{1}{3}\,(N^2 - x^2), \qquad 0 < x \le N. \tag{8.30}$$

This is to be compared with the *unconditional* average hitting time $\mathbb{E}_x(\tau) = x(N - x)$, $0 \le x \le N$, where $\tau = \tau_0 \wedge \tau_N$ is the first hitting time of $\{0, N\}$.

8.3 Application to Spin-Flip Dynamics

We will use the formulas in (8.6), (8.24) and (8.28)–(8.29) to obtain an upper bound on the conditional return time to the metastable state. This bound will be sharp enough to prove Lemma 3.7. We first do the computation for the complete graph (Curie-Weiss model). Afterwards we turn to the Erdős-Rényi Random Graph (our spin-flip dynamics).

8.3.1 Complete Graph

We monitor the magnetization of the *continuous-time* Curie-Weiss model by looking at the magnetization at the times of the spin-flips. This gives a *discrete-time* random walk on the set Γ_n defined in (1.15). This set consists of $n+1$ sites. We first consider the excursions to the *left* of $\mathbf{m}_n$ (recall (1.16)). After that we consider the excursions to the *right*.

1. For the Curie-Weiss model we have (use the formulas in Lemma 3.4 without the error terms)

$$\sigma \in A_k\colon \quad \sum_{\xi \in A_{k+1}} r(\sigma,\xi) = (n-k)\,\mathrm{e}^{-2\beta[\vartheta_k]_+}, \qquad \sum_{\xi \in A_{k-1}} r(\sigma,\xi) = k\,\mathrm{e}^{-2\beta[-\vartheta_k]_+}, \tag{8.31}$$

where $\vartheta_k = p(1-\frac{2k}{n}) - h$. Hence, the quotient of the rate to move downwards, respectively, upwards in magnetization equals

$$Q(k) = \frac{\sum_{\xi \in A_{k-1}} r(\sigma,\xi)}{\sum_{\xi \in A_{k+1}} r(\sigma,\xi)} = \frac{k}{n-k}\,\mathrm{e}^{2\beta([\vartheta_k]_+ - [-\vartheta_k]_+)}. \tag{8.32}$$

It is convenient to change variables by writing $k = \frac{n}{2}(a_k+1)$, so that $\vartheta_k = -pa_k - h$. The metastable state corresponds to $k = \mathbf{M}_n = \frac{n}{2}(\mathbf{m}_n+1)$, i.e., $a_k = \mathbf{m}_n$. We know from (1.16)–(1.15) that $\mathbf{m}_n$ is the smallest solution of the equation $J_n(\mathbf{m}_n) = 0$ (rounded off by $1/n$ to fall in Γ_n). Hence $\mathbf{m}_n = \mathbf{m} + O(1/n)$ with $\mathbf{m}$ the smallest solution of the equation $J_{p,\beta,h}(\mathbf{m}) = 0$, satisfying $\frac{1-\mathbf{m}}{1+\mathbf{m}} = \mathrm{e}^{-2\beta(p\mathbf{m}+h)}$ (recall (1.23)). Hence we can write (for ease of notation we henceforth ignore the error $O(1/n)$)

$$Q(k) = \frac{F(\mathbf{m}_n)}{F(a_k)}, \qquad F(a) = \frac{1-a}{1+a}\,\mathrm{e}^{2\beta pa}. \tag{8.33}$$

Here, we use that $[\vartheta_k]_+ - [-\vartheta_k]_+ = \vartheta_k$, which holds because $0 = R'_{p,\beta,h}(\mathbf{m}) = -p\mathbf{m} - h + \beta^{-1}I'(\mathbf{m})$ with $I'(\mathbf{m}) < 0$ because $\mathbf{m} < 0$ (recall (1.27)), so that $-p\mathbf{m}_n - h > 0$ for n large enough, which implies that also $-pa - h > 0$ for all $a < \mathbf{m}_n$ for n large enough. We next note that (recall (1.27) and (2.17))

$$\frac{d}{da}\log\left[\frac{F(\mathbf{m}_n)}{F(a)}\right] = -2\left(\beta p - \frac{1}{1-a^2}\right) = -J'_{p,\beta,h}(a) = 2\beta R''_{p,\beta,h}(a) \geq \delta \quad \text{for some } \delta > 0, \tag{8.34}$$

where the inequality comes from the fact that $a \mapsto R_{p,\beta,h}(a)$ has a positive curvature that is bounded away from zero on $[-1,\mathbf{m}]$ (recall Fig. 4).

2. We view the excursions to the left of $\mathbf{m}_n$ as starting from site N in the set $\{0, \ldots, N\}$ with $N = \mathbf{M}_n = \frac{n}{2}(\mathbf{m}_n + 1)$. From (8.28)–(8.29), we get

$$\begin{aligned} e_N(x) &= \sum_{z=0}^{N-1} \sum_{y=2}^{z} \frac{\pi(y,z]}{p(y,y+1)} \frac{\bar{R}(y)}{\bar{R}(N)} - \sum_{z=0}^{x-1} \sum_{y=2}^{z} \frac{\pi(y,z]}{p(y,y+1)} \frac{\bar{R}(y)}{\bar{R}(x)} \\ &\quad + \frac{2}{\bar{R}(N)\bar{R}(x)}[\bar{R}(N) - \bar{R}(x)] \\ &\leq \sum_{z=x}^{N-1} \sum_{y=2}^{z} \frac{\pi(y,z]}{p(y,y+1)} \frac{\bar{R}(y)}{\bar{R}(N)} + \frac{2}{\bar{R}(x)} \\ &\leq 2 \sum_{z=1}^{N-1} \sum_{y=2}^{z} \pi(y,z] + 2. \end{aligned} \tag{8.35}$$

Here, we use that $p(y, y+1) \geq \frac{1}{2}$ and $1 = \bar{R}(0) \leq \bar{R}(y) \leq \bar{R}(N)$ for all $0 < y < N$ (recall (8.24) and note that $x \mapsto \bar{R}(x)$ is non-decreasing). The bound is independent of x. Using the estimate

$$Q(x) = \frac{p(x, x-1)}{p(x, x+1)} \leq \mathrm{e}^{-\epsilon(N-x)/N}, \quad 0 < x < N, \quad \text{for some } \epsilon = \epsilon(\delta) > 0, \tag{8.36}$$

which comes from (8.34), we can estimate

$$\pi(y, z] \leq \prod_{x=y+1}^{z} \mathrm{e}^{-\epsilon(N-x)/N} = \exp\left[-\epsilon \sum_{x=y+1}^{z} (N-x)/N\right], \quad 0 \leq y \leq z < N, \tag{8.37}$$

from which it follows that

$$\sum_{z=1}^{N-1} \sum_{y=2}^{z} \pi(y, z] = O(N/\epsilon), \qquad N \to \infty. \tag{8.38}$$

Thus we arrive at

$$e_N(x) = O(N), \qquad N \to \infty, \qquad \text{uniformly in } 0 < x < N. \tag{8.39}$$

To turn (8.39) into a tail estimate, we use the Chebyshev inequality: (8.39) implies that every N time units there is a probability at least c to hit N, for some $c > 0$ and uniformly in $0 < x < N$. Hence

$$\mathbb{P}_x(\tau_N \geq kN \mid \tau_N < \tau_0) \leq (1-c)^k \qquad \forall k \in \mathbb{N}_0. \tag{8.40}$$

3. For excursions to the right of $\mathbf{m}_n$ the argument is similar. Now $N = \mathbf{T}_n - \mathbf{M}_n = \frac{n}{2}(\mathbf{t}_n - \mathbf{m}_n)$ (recall (1.17)), and the role of 0 and N is interchanged. Both near 0 and near N the drift towards $\mathbf{M}_n$ vanishes *linearly* (because of the non-zero curvature). If we condition the random walk not to hit N, then the average hitting time of 0 starting from x is again $O(N)$, uniformly in x.

4. Returning from the discrete-time random walk to the continuous-time Curie-Weiss model, we note that order n spin-flips occur per unit of time. Since $N \asymp n$ as $n \to \infty$, (8.40) and its analogue for excursions to the right give that, uniformly in $\xi \in A_{\mathbf{M}_n}$,

$$\mathbb{P}_\xi\left[\tau_{A_{\mathbf{M}_n}} \geq k \mid \tau_{A_{\mathbf{M}_n}} < \tau_{A_{\mathbf{T}_n}}\right] \leq \mathrm{e}^{-Ck} \qquad \forall\, k \in \mathbb{N}_0. \tag{8.41}$$

for some $C > 0$, which is the bound in (3.35).

8.3.2 Erdős-Rényi Random Graph

We next argue that the above argument can be extended to our spin-flip dynamics after taking into account that the rates to move downwards and upwards in magnetization are *perturbed by small errors*. In what follows we will write $p^{\mathrm{CW}}(x, x \pm 1)$ for the transition probabilities in the Curie-Weiss model and $p^{\mathrm{ER}}(x, x \pm 1)$ for the transition probabilities that serve as *uniform upper and lower bounds* for the transition probabilities in our spin-flip model. Recall that the latter actually depend on the configuration and not just on the magnetization, but Lemma 3.4 provides us with uniform bounds that allow us to *sandwich* the magnetization between the magnetizations of *two perturbed Curie-Weiss models*.

1. Suppose that

$$\frac{p^{\mathrm{ER}}(x, x-1)}{p^{\mathrm{ER}}(x, x+1)} = \frac{p^{\mathrm{CW}}(x, x-1)}{p^{\mathrm{CW}}(x, x+1)}[1 + O(N^{-1/2})]. \tag{8.42}$$

Then there exists a $C > 0$ large enough such that

$$\pi^{\mathrm{ER}}(y, z] \leq C\pi^{\mathrm{CW}}(y, z], \qquad 0 \leq y \leq z < N. \tag{8.43}$$

Indeed, as long as $z - y \leq C_1 N^{1/2}$ we have the bound in (8.43) (with C depending on C_1). On the other hand, if $z - y > C_1 N^{1/2}$ with C_1 large enough, then *the drift of the Curie-Weiss model sets in and overrules the error*: recall from (8.36) that the drift at distance $N^{1/2}$ from N is of order $N^{1/2}/N = N^{-1/2}$. It follows from (8.43) that (8.38)–(8.40) carry over, with suitably adapted constants, and hence so does (8.41).

2. To prove (8.42), we must show that (8.32) holds up to a multiplicative error $1 + O(n^{-1/2})$. In the argument that follows we assume that k is such that $\theta_k \geq \delta$ for some fixed $\delta > 0$. We comment later on how to extend the argument to other k values. Recall that $\theta_k = -pa_k - h$ and that $\theta_k \geq \delta > 0$ for all $a_k \in [-1, \mathbf{m}]$ for n large enough.

3. Let $\sigma \in A_k$ and $\sigma^v \in A_{k-1}$, where σ^v is obtained from σ by flipping the sign at vertex $v \in \sigma$ from $+1$ to -1. Write the transition rate from σ to σ^v as

$$\begin{aligned} r(\sigma, \sigma^v) &= \exp\left(-\beta\left[2p(\tfrac{2k}{n} - 1) + 2h + \tfrac{2}{n}\big(\epsilon(\sigma, v) - \epsilon(\overline{\sigma}, v)\big)\right]_+\right) \\ &= \exp\left(-2\beta\left[-\vartheta_k + \tfrac{1}{n}\big(\epsilon(\sigma, v) - \epsilon(\overline{\sigma}, v)\big)\right]_+\right). \end{aligned} \tag{8.44}$$

Here, $2p(\frac{2k}{n} - 1) = \frac{2}{n}p[k - (n-k)]$ equals $\frac{2}{n}$ times the average under $\mathbb{P}_{\mathrm{ER}_n(p)}$ of $E(\sigma, v) - E(\overline{\sigma}, v)$, with $E(\sigma, v)$ the number of edges between the support of σ and v and $E(\overline{\sigma}, v)$ the number of edges between the support of $\overline{\sigma}$ and v (recall the notation in Definition 3.1), and $\epsilon(\sigma, v) - \epsilon(\overline{\sigma}, v)$ is an *error term* that arises from deviations of this average. Since $-\vartheta_k \le -\delta$, the error terms are *not seen except* when they represent a large deviation of size at least δn. A union bound over all the vertices and all the configurations, in combination with Hoeffding's inequality, guarantees that, with $\mathbb{P}_{\mathrm{ER}_n(p)}$-probability tending to 1 as $n \to \infty$, for any σ there are at most $(\log 2)/2\delta^2 = O(1)$ many vertices that can lead to a large deviation of size at least δn. Since $r(\sigma, \sigma^v) \le 1$, we obtain

$$\sum_{v \in \sigma} r(\sigma, \sigma^v) = O(1) + [n - k - O(1)]\, \mathrm{e}^{-2\beta[-\vartheta_k]_+}. \tag{8.45}$$

This is a refinement of (3.10).

4. Similarly, let $\sigma \in A_k$ and $\sigma^v \in A_{k+1}$, where σ^v is obtained from σ by flipping the sign at vertex $v \notin \sigma$ from -1 to $+1$. Write the transition rate from σ to σ^v as

$$\begin{aligned} r(\sigma, \sigma^v) &= \exp\left(-\beta\left[2p(1 - \tfrac{2k}{n}) - 2h + \tfrac{2}{n}\big(\epsilon(\overline{\sigma}, v) - \epsilon(\sigma, v)\big)\right]_+\right) \\ &= \exp\left(-2\beta\left[\vartheta_k + \tfrac{1}{n}\big(\epsilon(\overline{\sigma}, v) - \epsilon(\sigma, v)\big)\right]_+\right). \end{aligned} \tag{8.46}$$

We *cannot remove* $[\cdot]_+$ when the error terms represent a large deviation of order δn. By the same argument as above, this happens for all but $(\log 2)/2\delta^2 = O(1)$ many vertices v. For all other vertices, we *can remove* $[\cdot]_+$ and write

$$r(\sigma, \sigma^v) = \mathrm{e}^{-2\beta\vartheta_k} \exp\left(\frac{1}{n}\big(\epsilon(\overline{\sigma}, v) - \epsilon(\sigma, v)\big)\right). \tag{8.47}$$

Next, we sum over v and use the inequality, valid for δ small enough,

$$\mathrm{e}^{-(1+\delta)\frac{1}{M}|\sum_{i=1}^{M} a_i|} \leq \frac{1}{M}\sum_{i=1}^{M} \mathrm{e}^{a_i} \leq \mathrm{e}^{(1+\delta)\frac{1}{M}|\sum_{i=1}^{M} a_i|} \qquad \forall\, 0 \leq |a_i| \leq \delta,\ 1 \leq i \leq M. \tag{8.48}$$

This gives

$$\begin{aligned} &\textstyle\sum_{v\notin\sigma} r(\sigma,\sigma^v) = O(1) + [k - O(1)]\,\mathrm{e}^{-2\beta\vartheta_k}\,\mathrm{e}^{O(|S_n|)},\\ &S_n = \tfrac{1}{[k-O(1)]}\tfrac{1}{n}\textstyle\sum_{v\notin\sigma}\big(\epsilon(\overline{\sigma},v) - \epsilon(\sigma,v)\big). \end{aligned} \tag{8.49}$$

We know from Lemma 3.2 that, with $\mathbb{P}_{\mathrm{ER}_n(p)}$-probability tending to 1 as $n \to \infty$,

$$|S_n| \leq \frac{cn^{3/2}}{[k - O(1)]n} \qquad \forall\, c > \sqrt{\frac{1}{8}\log 2}. \tag{8.50}$$

Since we may take $k \geq \frac{n}{3}(p - h)$ (recall (3.14)), we obtain

$$\sum_{v\notin\sigma} r(\sigma,\sigma^v) = O(1) + [k - O(1)]\,\mathrm{e}^{-2\beta\vartheta_k}\,\mathrm{e}^{O(n^{-1/2})}. \tag{8.51}$$

This is a refinement of (3.11).

5. The same argument works when we assume that k is such that $\vartheta_k \leq -\delta$ for some fixed $\delta > 0$: simply reverse the arguments in Steps 3 and 4. It therefore remains to explain what happens when $\vartheta_k \approx 0$, i.e., $a_k \approx -\frac{h}{p}$. We then see from (1.27) that $R'_{p,\beta,h}(a_k) \approx \beta^{-1} I'(a_k) < 0$, so that a_k lies in the interval $[\mathbf{t}, 0]$, which is *beyond* the top state (recall Fig. 4).

References

1. Bianchi, A., Bovier, A., Ioffe, D.: Sharp asymptotics for metastability in the random field Curie-Weiss model. Electron. J. Probab. **14**, 1541–1603 (2009)
2. Bianchi, A., Bovier, A., Ioffe, D.: Pointwise estimates and exponential laws in metastable systems via coupling methods. Ann. Probab. **40**, 339–379 (2012)
3. Bovier, A., Eckhoff, M., Gayrard, V., Klein, M.: Metastability in stochastic dynamics of disordered mean-field models. Probab. Theory Relat. Fields **119**, 99–161 (2001)
4. Bovier, A., den Hollander, F.: Metastability – A Potential-Theoretic Approach. Grundlehren der mathematischen Wissenschaften, vol. 351. Springer, Berlin (2015)
5. Bovier, A., Marello, S., Pulvirenti, E.: Metastability for the dilute Curie-Weiss model with Glauber dynamics, Dec 2019. arXiv:1912.10699
6. den Hollander, F., Jovanovski, O.: Metastability on the hierarchical lattice. J. Phys. A. Math. Theor. **50**, 305001 (2017)

7. Dommers, S.: Metastability of the Ising model on random regular graphs at zero temperature. Probab. Theory Relat. Fields **167**, 305–324 (2017)
8. Dommers, S., Giardinà, C., van der Hofstad, R.: Ising models on power-law random graphs. J. Stat. Phys. **141**, 638–660 (2010)
9. Dommers, S., Giardinà, C., van der Hofstad, R.: Ising critical exponents on random trees and graphs. Commun. Math. Phys. **328**, 355–395 (2014)
10. Dommers, S., den Hollander, F., Jovanovski, O., Nardi, F.R.: Metastability for Glauber dynamics on random graphs. Ann. Appl. Probab. **27**, 2130–2158 (2017)
11. Jovanovski, O.: Metastability for the Ising model on the hypercube. J. Stat. Phys. **167**, 135–159 (2017)
12. Olivieri, E., Vares, M.E.: Large Deviations and Metastability. Encyclopedia of Mathematics and Its Applications, vol. 100. Cambridge University Press, Cambridge (2005)
13. van der Hofstad, R.: Random Graphs and Complex Networks, vol. I. Cambridge University Press, Cambridge (2017)

The Parabolic Anderson Model on a Galton-Watson Tree

Frank den Hollander, Wolfgang König, and Renato S. dos Santos

We dedicate this work to Vladas,
who was a leading light in our understanding of disordered systems.

Abstract We study the long-time asymptotics of the total mass of the solution to the parabolic Anderson model (PAM) on a supercritical Galton-Watson random tree with bounded degrees. We identify the second-order contribution to this asymptotics in terms of a variational formula that gives information about the local structure of the region where the solution is concentrated. The analysis behind this formula suggests that, under mild conditions on the model parameters, concentration takes place on a tree with minimal degree. Our approach can be applied to locally tree-like finite random graphs, in a coupled limit where both time and graph size tend to infinity. As an example, we consider the configuration model, i.e., uniform simple random graphs with a prescribed degree sequence.

Keywords Galton-Watson tree · Sparse random graph · Parabolic Anderson model · Double-exponential distribution · Quenched Lyapunov exponent

MSC 2010 60H25, 82B44, 05C80

F. den Hollander
Mathematical Institute, University of Leiden, Leiden, The Netherlands
e-mail: denholla@math.leidenuniuv.nl

W. König
Technische Universität Berlin, Berlin, Germany

Weierstrass Institute for Applied Analysis and Stochastics (WIAS), Berlin, Germany
e-mail: koenig@wias-berlin.de

R. S. dos Santos (✉)
Department of Mathematics, Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil
e-mail: rsantos@mat.ufmg.br

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_25

1 Introduction and Main Results

In Sect. 1.1 we give a brief introduction to the parabolic Anderson model. In Sect. 1.2 we give the basic notation. In Sects. 1.3 and 1.4 we present our results for Galton-Watson trees and for the configuration model, respectively. In Sect. 1.5 we discuss these results.

1.1 The PAM and Intermittency

The *parabolic Anderson model (PAM)* concerns the Cauchy problem for the heat equation with a random potential, i.e., solutions u to the equation

$$\partial_t u(t,x) = \Delta u(t,x) + \xi(x)u(t,x), \qquad t > 0,\ x \in \mathscr{X}, \tag{1.1}$$

where $\mathscr{X}$ is a space equipped with a Laplacian Δ, and ξ is a random potential on $\mathscr{X}$. The operator $\Delta + \xi$ is called the *Anderson operator*. Although $\mathbb{Z}^d$ and $\mathbb{R}^d$ are the most common choices for $\mathscr{X}$, other spaces are interesting as well, such as Riemannian manifolds or discrete graphs. In the present paper we study the PAM on *random graphs*. For surveys on the mathematical literature on the PAM until 2016, we refer the reader to [1, 12].

The main question of interest in the PAM is a detailed description of the concentration effect called *intermittency*: in the limit of large time the solution u concentrates on small and well-separated regions in space, called *intermittent islands*. This concentration effect can be studied particularly well in the PAM because efficient mathematical tools are available, such as eigenvalue expansions and the Feynman-Kac formula. In particular, these lead to a detailed description of the *locations* of the intermittent islands, as well as the *profiles* of the potential ξ and the solution u inside these islands.

The analysis of intermittency usually starts with a computation of the logarithmic large-time asymptotics of the total mass, encapsulated in *Lyapunov exponents*. There is an important distinction between the *annealed* setting (i.e., averaged over the random potential) and the *quenched* setting (i.e., almost surely with respect to the random potential). Often both types of Lyapunov exponents admit explicit descriptions in terms of *characteristic variational formulas* that contain information about how the mass concentrates in space, and serve as starting points for deeper investigations. The 'annealed' and the 'quenched' variational formula are typically connected, but take two different points of view. They contain two parts: a rate function term that identifies which profiles of the potential are most favourable for mass concentration, and a spectral term that identifies which profiles the solution takes inside the intermittent islands.

From now on, we restrict to *discrete* spaces and to random potentials that consist of i.i.d. variables. For $\mathbb{Z}^d$, the above intermittent picture was verified for several classes of marginal distributions. It turned out that the *double-exponential distribution* with parameter $\varrho \in (0, \infty)$, given by

$$\mathrm{P}(\xi(0) > u) = \mathrm{e}^{-\mathrm{e}^{u/\varrho}}, \qquad u \in \mathbb{R}, \tag{1.2}$$

is particularly interesting, because it leads to non-trivial intermittent islands and to interesting profiles of both potential and solution inside. There are four different classes of potentials, distinguished by the type of variational formula that emerges and the scale of the diameter of the intermittent island (cf. [17]). The double-exponential distribution is critical in the sense that the intermittent islands neither grow nor shrink with time, and therefore represents a class of its own.

The setup of the present paper contains two features that are novel in the study of the PAM: (1) we consider a *random* discrete space, thereby introducing another layer of randomness into the model; (2) this space has a *non-Euclidean* topology, in the form of an *exponential growth* of the volume of balls as a function of their radius. As far as we are aware, the discrete-space PAM has so far been studied only on $\mathbb{Z}^d$ and on two examples of finite deterministic graphs: the *complete graph* with n vertices [6] and the N-dimensional *hypercube* with $n = 2^N$ vertices [2]. These graphs have unbounded degrees as $n \to \infty$, and therefore the Laplace operator was equipped with a prefactor that is equal to the inverse of the degree, unlike the Laplace operator considered here.

Our main target is the PAM on a Galton-Watson tree with bounded degrees. However, our approach also applies to large finite graphs that are *sparse* (e.g. bounded degrees) and *locally tree-like* (rare loops). As an illustration, we consider here the *configuration model* or, more precisely, the *uniform simple random graph* with prescribed degree sequence. We choose to work in the almost-sure (or large-probability) setting with respect to the randomnesses of both graph *and* potential, and we take as initial condition a unit mass at the root of the graph. We identify the leading order large-time asymptotics of the total mass, and derive a variational formula for the correction term. This formula contains a *spatial part* (identifying the subgraph on which the concentration takes place) and a *profile part* (identifying the shape on that subgraph of both the potential and the solution). Both parts are new. In some cases we can identify the minimiser of the variational formula. As in the case of $\mathbb{Z}^d$, the structure of the islands does not depend on time: no spatial scaling is necessary.

1.2 The PAM on a Graph

We begin with some definitions and notations, and refer the reader to [1, 12] for more background on the PAM in the case of $\mathbb{Z}^d$.

Let $G = (V, E)$ be a *simple undirected* graph, either finite or countably infinite. Let Δ_G be the Laplacian on G, i.e.,

$$(\Delta_G f)(x) := \sum_{\substack{y \in V: \\ \{x,y\} \in E}} [f(y) - f(x)], \qquad x \in V, \ f\colon \ V \to \mathbb{R}. \tag{1.3}$$

Our object of interest is the non-negative solution of the Cauchy problem for the heat equation with potential $\xi\colon \ V \to \mathbb{R}$ and localised initial condition,

$$\begin{aligned} \partial_t u(x,t) &= (\Delta_G u)(x,t) + \xi(x) u(x,t), \qquad && x \in V, \ t > 0, \\ u(x,0) &= \delta_{\mathcal{O}}(x), && x \in V, \end{aligned} \tag{1.4}$$

where $\mathcal{O} \in V$ is referred to as the *origin* or *root* of G. We say that G is *rooted at* $\mathcal{O}$ and call $G = (V, E, \mathcal{O})$ a *rooted graph*. The quantity $u(t,x)$ can be interpreted as the amount of mass present at time t at site x when initially there is unit mass at $\mathcal{O}$.

Criteria for existence and uniqueness of the non-negative solution to (1.4) are well-known for the case $G = \mathbb{Z}^d$ (see [8]), and rely on the *Feynman-Kac formula*

$$u(x,t) = \mathbb{E}_{\mathcal{O}}\left[\exp\left\{\int_0^t \xi(X_s)\mathrm{d}s\right\} 1\!\!1\{X_t = x\}\right], \tag{1.5}$$

where $X = (X_t)_{t \geq 0}$ is the continuous-time random walk on the vertices V with jump rate 1 along the edges E, and $\mathbb{P}_{\mathcal{O}}$ denotes the law of X given $X_0 = \mathcal{O}$. We will be interested in the *total mass* of the solution,

$$U(t) := \sum_{x \in V} u(x,t) = \mathbb{E}_{\mathcal{O}}\left[\exp\left\{\int_0^t \xi(X_s)\mathrm{d}s\right\}\right]. \tag{1.6}$$

Often we suppress the dependence on G, ξ from the notation. Note that, by time reversal and the linearity of (1.4), $U(t) = \hat{u}(\mathcal{O}, t)$ with $\hat{u}$ the solution with a different initial condition, namely, constant and equal to 1.

Throughout the paper, we assume that the random potential $\xi = (\xi(x))_{x \in V}$ consists of i.i.d. random variables satisfying:

Assumption (DE) For some $\varrho \in (0, \infty)$,

$$\mathrm{P}(\xi(0) \geq 0) = 1, \qquad \mathrm{P}(\xi(0) > u) = \mathrm{e}^{-\mathrm{e}^{u/\varrho}} \ \text{ for } u \text{ large enough.} \tag{1.7}$$

Under Assumption (DE), $\xi(0) \geq 0$ almost surely and $\xi(x)$ has an eventually exact double-exponential upper tail. The latter restrictions are helpful to avoid certain technicalities that are unrelated to the main message of the paper and that require no new ideas. In particular, (1.7) is enough to guarantee existence and uniqueness of the non-negative solution to (1.4) on any discrete graph with at most exponential growth, as can be inferred from the proof of the $\mathbb{Z}^d$-case in [9]. All our results remain valid under (1.2) or even milder conditions, e.g. [9, Assumption (F)] plus an integrability condition on the lower tail of $\xi(0)$.

The following *characteristic variational problem* will turn out to be important for the description of the asymptotics of $U(t)$ when ξ has a double-exponential tail. Denote by $\mathcal{P}(V)$ the set of probability measures on V. For $p \in \mathcal{P}(V)$, define

$$I_E(p) := \sum_{\{x,y\}\in E} \left(\sqrt{p(x)} - \sqrt{p(y)}\right)^2, \qquad J_V(p) := -\sum_{x\in V} p(x)\log p(x), \tag{1.8}$$

and set

$$\chi_G(\varrho) := \inf_{p\in\mathcal{P}(V)} [I_E(p) + \varrho J_V(p)], \qquad \varrho \in (0,\infty). \tag{1.9}$$

The first term in (1.9) is the quadratic form associated with the Laplacian, describing the solution $u(\cdot, t)$ in the intermittent islands, while the second term in (1.9) is the Legendre transform of the rate function for the potential, describing the highest peaks of $\xi(\cdot)$ inside the intermittent islands. See Sect. 1.5 for its relevance and interpretation, and Sect. 2.3 for alternate representations.

1.3 Results: Galton-Watson Trees

In this section we focus on our first example of a random graph.

Let D_0, D_g be random variables taking values in $\mathbb{N} = \{1, 2, 3, \dots\}$. The *Galton-Watson tree* with initial degree distribution D_0 and general degree distribution D_g is constructed as follows. Start with a root vertex $\mathcal{O}$, and attach edges from $\mathcal{O}$ to D_0 first-generation vertices. Proceed recursively: after having attached the n-th generation of vertices, attach to each one of them an independent $(D_g - 1)$-distributed number of new vertices, whose union gives the $(n+1)$-th generation of vertices. Denote by $\mathcal{GW} = (V, E)$ the graph obtained, by $\mathfrak{P}$ its probability law, and by E the corresponding expectation. The law of $D_g - 1$ is the offspring distribution of $\mathcal{GW}$, and the law of D_g is the degree distribution. Write $\mathrm{supp}(D_g)$ to denote the set of degrees that are taken by D_g with positive probability.

We will work under the following bounded-degree assumption:

Assumption (BD)

$$d_{\min} := \min \mathrm{supp}(D_g) \geq 2, \qquad E[D_g] > 2, \tag{1.10}$$

and, for some $d_{\max} \in \mathbb{N}$ with $d_{\max} \geq d_{\min}$,

$$\max \mathrm{supp}(D_g) \leq d_{\max}. \tag{1.11}$$

Under Assumption (BD), $\mathcal{GW}$ is almost surely an infinite tree. Moreover,

$$\lim_{r\to\infty} \frac{\log|B_r(\mathcal{O})|}{r} = \log E[D_g - 1] =: \vartheta > 0 \qquad \mathfrak{P} - a.s., \tag{1.12}$$

where $B_r(\mathcal{O})$ is the ball of radius r around $\mathcal{O}$ in the graph distance (see e.g. [13, pp. 134–135]). Note that Assumption (BD) allows deterministic trees with constant offspring $d_{\min} - 1$ (provided $d_{\min} \geq 3$).

To state our main result, we define the constant

$$\widetilde{\chi}(\varrho) := \inf\left\{\chi_T(\varrho)\colon\ T \text{ infinite tree with degrees in } \mathrm{supp}(D_g)\right\} \tag{1.13}$$

with $\chi_G(\varrho)$ defined in (1.9).

Theorem 1.1 (Quenched Lyapunov Exponent for the PAM on $\mathcal{GW}$) *Let $G = \mathcal{GW} = (V, E, \mathcal{O})$ be the rooted Galton-Watson random tree satisfying Assumption (BD), and let ϑ be as in (1.12). Let $\xi = (\xi(x))_{x\in V}$ be an i.i.d. potential satisfying Assumption (DE). Let $U(t)$ denote the total mass at time t of the solution u to the PAM on $\mathcal{GW}$. Then, as $t \to \infty$,*

$$\frac{1}{t}\log U(t) = \varrho \log\left(\frac{\varrho t \vartheta}{\log\log t}\right) - \varrho - \widetilde{\chi}(\varrho) + o(1), \qquad (\mathrm{P}\times\mathfrak{P})\text{-}a.s. \tag{1.14}$$

The proof of Theorem 1.1 is given in Sect. 4.

For ϱ sufficiently large we can identify the infimum in (1.13). For $d \geq 2$, denote by $\mathcal{T}_d$ the *infinite homogeneous tree* with degree equal to d at every node.

Theorem 1.2 (Identification of the Minimiser) *If $\varrho \geq 1/\log(d_{\min} + 1)$, then $\widetilde{\chi}(\varrho) = \chi_{\mathcal{T}_{d_{\min}}}(\varrho)$.*

The proof of Theorem 1.2 is given in Sect. 6 with the help of a comparison argument that appends copies of the infinite $d_{\min}$-tree to itself. We believe $\mathcal{T}_{d_{\min}}$ to be the *unique* minimiser of (1.13) under the same assumptions, but proving so would require more work.

1.4 Results: Configuration Model

In this section we focus on our second example of a random graph.

For $n \in \mathbb{N}$, let $\mathfrak{d}^{(n)} = (d_i^{(n)})_{i=1}^n$ be a collection of positive integers. The *configuration model with degree sequence* $\mathfrak{d}^{(n)}$ is a random *multigraph* (i.e., a graph that may have self-loops and multiple edges) on the vertex set $V_n := \{1, \ldots, n\}$ defined as follows. To each $i \in V_n$, attach $d_i^{(n)}$ 'half-edges'. After that, construct edges by successively attaching each half-edge uniformly at random to a remaining half-edge. For this procedure to be successful, we must require that

$$d_1^{(n)} + \cdots + d_n^{(n)} \text{ is even for every } n \in \mathbb{N}. \tag{1.15}$$

Draw a root $\mathcal{O}_n$ uniformly at random from V_n. Denote by $\mathcal{CM}_n = (V_n, E_n, \mathcal{O}_n)$ the rooted multigraph thus obtained, and by $\mathfrak{P}_n$ its probability law. For further details, we refer the reader to [15, Chapter 7].

We will work under the following assumption on $\mathfrak{d}^{(n)}$:

Assumption (CM) The degree sequences $\mathfrak{d}^{(n)} = (d_i^{(n)})_{i=1}^n$, $n \in \mathbb{N}$, satisfy (1.15). Moreover,

1. There exists an $\mathbb{N}$-valued random variable D such that $d_{\mathcal{O}_n}^{(n)} \Rightarrow D$ as $n \to \infty$.
2. $d_{\min} := \min \operatorname{supp}(D) \geq 3$.
3. There exists a $d_{\max} \in \mathbb{N}$ such that $2 \leq d_i^{(n)} \leq d_{\max}$ for all $n \in \mathbb{N}$ and $1 \leq i \leq n$.

In particular, $3 \leq d_{\min} \leq d_{\max} < \infty$ and $D \leq d_{\max}$ almost surely. It is possible to take $\mathfrak{d}^{(n)}$ random. In that case Assumption (CM) must be required almost surely or in probability with respect to the law of $\mathfrak{d}^{(n)}$, and our results below must be interpreted accordingly.

Proposition 1.3 (Connectivity and Simplicity of $\mathcal{CM}_n$) *Under Assumption (CM),*

$$\lim_{n\to\infty} \mathfrak{P}_n(\mathcal{CM}_n \textit{ is a simple graph}) = \mathrm{e}^{-\frac{\nu}{2} - \frac{\nu^2}{4}}, \tag{1.16}$$

where

$$\nu := \frac{E[D(D-1)]}{E[D]} \in [2, \infty). \tag{1.17}$$

Moreover,

$$\lim_{n\to\infty} \mathfrak{P}_n\big(\mathcal{CM}_n \textit{ is connected} \mid \mathcal{CM}_n \textit{ is simple}\big) = 1. \tag{1.18}$$

Proof See [15, Theorem 7.12] and [5, Theorem 2.3]. □

Item (1.16) in Proposition 1.3 tells us that for large n the set

$$\mathscr{U}_n(\mathfrak{d}^{(n)}) := \big\{\text{simple graphs on } \{1, \ldots, n\} \text{ with degrees } d_1^{(n)}, \ldots, d_n^{(n)}\big\} \tag{1.19}$$

is non-empty. Hence, we may consider the *uniform simple random graph* $\mathcal{UG}_n$ that is drawn uniformly at random from $\mathscr{U}_n(\mathfrak{d}^{(n)})$.

Proposition 1.4 (Conditional Law of $\mathcal{CM}_n$ Given Simplicity) *Under the conditional law $\mathfrak{P}_n(\,\cdot \mid \mathcal{CM}_n$ is simple), $\mathcal{CM}_n$ has the same law as $\mathcal{UG}_n$.*

Proof See [15, Proposition 7.15]. □

As usual, for a sequence of events $(A_n)_{n\in\mathbb{N}}$, we say that A_n occurs *with high probability (whp)* as $n \to \infty$ if the probability of A_n tends to 1 as $n \to \infty$. This notion does not require the events to be defined on the same probability space. We denote by $\mathrm{dist}_{\mathrm{TV}}(X, Y)$ the total variation distance between two random variables X and Y (i.e., between their laws). Let

$$\Phi_n := \left(\frac{1}{n} \vee \mathrm{dist}_{\mathrm{TV}}(d^{(n)}_{\mathcal{O}_n}, D)\right)^{-1}, \tag{1.20}$$

and note that, by Assumption (CM), $\Phi_n \to \infty$ as $n \to \infty$.

Theorem 1.5 (Quenched Lyapunov Exponent for the PAM on $\mathcal{UG}_n$) *For any $n \in \mathbb{N}$, let $G = \mathcal{UG}_n$ be the uniform simple random graph with degree sequence $\mathfrak{d}^{(n)}$ satisfying Assumption (CM). For any $n \in \mathbb{N}$, let ξ be an i.i.d. potential on V_n satisfying Assumption (DE). Let $U_n(t)$ denote the total mass of the solution to the PAM on $G = \mathcal{UG}_n$ as defined in Sect. 1.2. Fix a sequence of times $(t_n)_{n\in\mathbb{N}}$ with $t_n \to \infty$ and $t_n \log t_n = o(\log \Phi_n)$ as $n \to \infty$. Then, with high $\mathrm{P} \times \mathfrak{P}_n$-probability as $n \to \infty$,*

$$\frac{1}{t_n} \log U_n(t_n) = \varrho \log\left(\frac{\varrho t_n \vartheta}{\log\log t_n}\right) - \varrho - \widetilde{\chi}(\varrho) + o(1), \tag{1.21}$$

where $\vartheta := \log \nu > 0$ with ν as in (1.17), and $\widetilde{\chi}(\varrho)$ is as in (1.13).

The proof of Theorem 1.5 is given in Sect. 5. The main ingredients in the proof are Theorem 1.1 and a well-known comparison between the configuration model and an associated Galton-Watson tree inside a slowly-growing ball, from which the condition on t_n originates.

Condition (1) in Assumption (CM) is a standard regularity condition. Conditions (2) and (3) provide easy access to results such as Propositions 1.3 and 1.4 above. As examples of degree sequences satisfying Assumption (CM) we mention:

- *Constant degrees.* In the case where $d_i = d \geq 3$ for a deterministic $d \in \mathbb{N}$ and all $1 \leq i \leq n$, we have $d_{\mathcal{O}_n} = D = d$ almost surely, and $\mathcal{UG}_n$ is a uniform regular random graph. To respect (1.15), it is enough to restrict to n such that nd is even. In this case $\mathrm{dist}_{\mathrm{TV}}(d_{\mathcal{O}_n}, D) = 0$, and so $\Phi_n = n$ in (1.20).
- *Random degrees.* In the case where $(d_i)_{i\in\mathbb{N}}$ forms an i.i.d. sequence taking values in $\{3, \ldots, d_{\max}\}$, classical concentration bounds (e.g. Azuma's inequality) can be used to show that, for any $\gamma \in (0, \frac{1}{2})$,

$$d_{\mathrm{TV}}(d_{\mathcal{O}_n}, D) = o(n^{-\gamma}) \quad \text{almost surely as } n \to \infty, \tag{1.22}$$

  and so $\Phi_n \gg n^{\gamma}$. The condition in (1.15) can be easily satisfied after replacing d_n by $d_n + 1$ when $d_1 + \cdots + d_n$ is odd, which does not affect (1.22). With this change, Assumption (CM) is satisfied. For more information about $\mathcal{CM}_n$ with i.i.d. degrees, see [15, Chapter 7].

1.5 Discussion

Our main results, Theorems 1.1 and 1.5, identify the quenched logarithmic asymptotics of the total mass of the PAM. Our proofs show that the first term in the asymptotics comes from the height of the potential in an intermittent island, the second term $-\varrho$ from the probability of a quick sprint by the random walk in the Feynman-Kac formula from $\mathcal{O}$ to the island, and the third term $\widetilde{\chi}(\varrho)$ from the structure of the island and the profile of the potential inside. Below we explain how each of these three terms comes about. Much of what follows is well-known from the study of the PAM on $\mathbb{Z}^d$ (see also [12]), but certain aspects are new and derive from the randomness of the ambient space and its exponential growth.

1.5.1 Galton-Watson Tree

First and Second Terms

The large-t asymptotics of the Feynman-Kac formula (2.10) for $U(t)$ comes from those random walk paths $(X_s)_{s\in[0,t]}$ that run within $\mathfrak{s}_t$ time units to some favorable local region of the graph (the intermittent island) and subsequently stay in that region for the rest of the time. In order to find the scale $\mathfrak{r}_t$ of the distance to the region and the time $\mathfrak{s}_t$ of the sprint, we have to balance and optimise a number of crucial quantities: the number of sites in the ball $B_{\mathfrak{r}_t}(\mathcal{O})$ around $\mathcal{O}$ with radius $\mathfrak{r}_t$, the scale of the maximal value of the potential within that ball, the probability to reach that ball within time $\mathfrak{s}_t$, and the gain from the Feynman-Kac formula from staying in that ball during $t-\mathfrak{s}_t$ time units. One key ingredient is the well-known fact that the maximum of m independent random variables satisfying Assumption (DE) is asymptotically equal to $h_m \approx \varrho \log\log m$ for large m. Another key ingredient is that $B_{\mathfrak{r}_t}(\mathcal{O})$ has approximately $\mathrm{e}^{\mathfrak{r}_t\vartheta}$ vertices (see (1.12)). Hence, this ball contains values of the potential of height $\approx h_{\mathrm{e}^{\mathfrak{r}_t\vartheta}} \approx \varrho \log(r_t\vartheta)$, not just at one vertex but on a cluster of vertices of arbitrary finite size. The contribution from staying in such as cluster during $\approx t$ time units yields the first term of the asymptotics, where we still need to identify $\mathfrak{r}_t$. A slightly more precise calculation, involving the probabilistic cost to run within $\mathfrak{s}_t$ time units over $\mathfrak{r}_t$ space units and to afterwards gain a mass of size $(t-\mathfrak{s}_t)\varrho\log(r_t\vartheta)$, reveals that the optimal time is $\mathfrak{s}_t \approx \mathfrak{r}_t/\varrho\log\mathfrak{r}_t$. Optimising this together with the first term $\varrho\log(r_t\vartheta)$ over $\mathfrak{r}_t$, we see that the optimal distance is $\mathfrak{r}_t = \varrho t/\log\log t$. The term $-\varrho$ comes from the probability of making $\mathfrak{r}_t$ steps within $\mathfrak{s}_t = \mathfrak{r}_t/\varrho\log\mathfrak{r}_t$ time units.

Third Term

The variational formula $\widetilde{\chi}_G(\varrho)$ describes the second-order asymptotics of the gain of the random walk from staying $\approx t$ time units in an optimal local region (the first-order term has already been identified as $\varrho \log(r_t \vartheta)$). Indeed, pick some finite tree T that is admissible, i.e., has positive probability to occur locally in the graph $G = \mathcal{GW}$. Many copies of T occur disjointly with positive density in G. In particular, they appear within the ball $B_{\mathfrak{r}_t}(\mathcal{O})$ a number of times that is proportional to the volume of the ball. By standard extreme-value analysis, on one of these many copies of T the random potential achieves an approximately optimal height ($\approx \varrho \log(r_t \vartheta)$) and shape. The optimality of the shape is measured in terms of the negative local Dirichlet eigenvalue $-\lambda_T(\xi)$ of $\Delta_G + \xi$ inside T. The shapes q that ξ can assume locally are those that have a large-deviation rate value $\mathcal{L}(q) = \sum_x \mathrm{e}^{q(x)/\varrho}$ at most 1 (note that $\mathcal{L}(q)$ measures the probabilistic cost of the shape q on an exponential scale). All allowed shapes q are present locally at some location inside the ball $B_{\mathfrak{r}_t}(\mathcal{O})$ for large t. Each of these locations can be used by the random walk as an intermittent island. Optimising over all allowed shapes q, we see that the second-order term of the long stay in that island must indeed be expressed by the term

$$\sup_{q:\, \mathcal{L}(q) \le 1} [-\lambda_T(q)]. \tag{1.23}$$

When T is appropriately chosen, this number is close to the number $\widetilde{\chi}(\varrho)$ defined in (1.13) (cf. Proposition 2.3). This completes the heuristic explanation of the asymptotics in (1.14).

1.5.2 Configuration Model

The analogous assertion for the configuration model in (1.21) is understood in the same way, ignoring the fact that the graph is now finite, and that size and time are coupled. As to the additional growth constraint on $t_n \log t_n$ in Theorem 1.5: its role is to guarantee that the ball $B_{\mathfrak{r}_{t_n}}(\mathcal{O})$ is small enough to contain no loop with high probability. In fact, this ball is very close in distribution to the same ball in an associated Galton-Watson tree (cf. Proposition 5.1), which allows us to carry over our result.

1.5.3 Minimal Degree Tree Is Optimal

What is a heuristic explanation for our result in Theorem 1.2 that the optimal tree is an infinitely large homogeneous tree of minimal degree $d_{\min}$ at every vertex? The first term in (1.9), the quadratic form associated with the Laplacian, has a spread-out effect. Apparently, the self-attractive effect of the second term is not strong enough to cope with this, as the super-linear function $p \mapsto p \log p$ in the definition of J_V in (1.8) is 'weakly superlinear'. This suggests that the optimal structure should be infinitely large (also on $\mathbb{Z}^d$ the optimal profile is positive anywhere in the ambient space $\mathbb{Z}^d$). The first term is obviously monotone in the degree, which explains why the infinite tree with minimal degree optimises the formula.

1.5.4 Hurdles

The exponential growth of the graph poses a number of technical difficulties that are not present for the PAM on $\mathbb{Z}^d$ or $\mathbb{R}^d$. Indeed, one of the crucial points in the proof of the upper bound for the large-time asymptotics is to restrict the infinite graph G to some finite but time-dependent subgraph (in our case the ball $B_{\mathfrak{r}_t}(\mathcal{O})$). On $\mathbb{Z}^d$, a reflection technique that folds $\mathbb{Z}^d$ into a box of an appropriate size gives an upper bound at the cost of a negligible boundary term. For exponentially growing graphs, however, this technique can no longer be used because the boundary of a large ball is comparable in size to the volume of the ball. Therefore we need to employ and adapt an intricate method developed on $\mathbb{Z}^d$ for deriving deeper properties of the PAM, namely, Poisson point process convergence of all the top eigenvalue-eigenvector pairs and asymptotic concentration in a single island. This method relies on certain *path expansions* which are developed in Sect. 3 and rely on ideas from [4, 14].

1.6 Open Questions

We discuss next a few natural questions for future investigation.

1.6.1 Unbounded Degrees

A central assumption used virtually throughout in the paper is that of a uniformly bounded degree for the vertices of the graph. While this assumption can certainly be weakened, doing so would require a careful analysis of many interconnected technical arguments involving both the geometry of the graph and the behaviour the random walk. An inspection of our proofs will reveal that some mild growth of the maximal degree with the volume is allowed, although this would not address the real issues at hand and would therefore be far from optimal. For this reason we prefer to leave unbounded degrees for future work.

1.6.2 Small ϱ

The question of whether Theorem 1.2 is still true when $\varrho < 1/\log(d_{\min} + 1)$ seems to us not clear at all, and in fact interesting. Indeed, the analogous variational problem in $\mathbb{Z}^d$ was analysed in [7] and was shown to be highly non-trivial for small ϱ.

1.6.3 Different Time Scales

In a fixed finite graph, the PAM can be shown to localise for large times on the site that maximises the potential. It is reasonable to expect the same when the graph is allowed to grow but only very slowly in comparison to the time window considered, leading to a behaviour very different from that shown in Theorem 1.5. A more exciting and still widely open question is whether there could be other growth regimes between graph size and time that would lead to new asymptotic behaviours. We expect that Theorem 1.5 would still hold for times well above the time cutoff given. For investigations of a similar flavour we direct the reader to [2, 6].

1.6.4 Annealing

In the present paper we only consider the *quenched* setting, i.e., statements that hold almost-surely or with high probability with respect to the law of both the random graph and the random potential. There are three possible *annealed* settings, where we would average over one or both of these laws. Such settings would certainly lead to different growth scales for the total mass, corresponding to new probabilities to observe local structures in the graph and/or the potential. The variational problems could be potentially different, but for double-exponential tails comparison with the $\mathbb{Z}^d$ case suggests that they would coincide.

1.7 Outline

The remainder of the paper is organised as follows. In Sect. 2 we collect some basic notations and facts about graphs, spectral objects, alternate representations of the characteristic formula $\chi_G(\varrho)$, and the potential landscape. In Sect. 3 we employ a path expansion technique to estimate the contribution to the Feynman-Kac formula coming from certain specific classes of paths. In Sect. 4 we prove Theorem 1.1. In Sect. 5 we prove Theorem 1.5. In Sect. 6 we analyse the behaviour of the variational formula χ_T for trees T under certain glueing operations, and prove Theorem 1.2.

2 Preliminaries

In this section we gather some facts that will be useful in the remainder of the paper. In particular, we transfer some basic properties of the potential landscape derived in [3] and [4] for the Euclidean-lattice setting to the *sparse-random-graph* setting. In Sect. 2.1 we describe the classes of graphs we will work with. In Sect. 2.2 we derive spectral bounds on the Feynman-Kac formula. In Sect. 2.3 we provide alternative representations for the constant χ in (1.9). In Sect. 2.4 we obtain estimates on the maximal height of the potential in large balls as well as on the sizes and local eigenvalues of the islands where the potential is close to maximal. In Sect. 2.5 we obtain estimates on the heights of the potential seen along self-avoiding paths and on the number of islands where the potential is close to maximal.

2.1 Graphs

All graphs considered in Sect. 2 are simple, connected and undirected, and are either finite or countably infinite. For a graph $G = (V, E)$, we denote by $\mathrm{dist}(x, y) = \mathrm{dist}_G(x, y)$ the graph distance between $x, y \in V$, and by

$$\deg(x) = \deg_G(x) := \#\{y \in V\colon\ \{y, x\} \in E\}, \tag{2.1}$$

the degree of the vertex $x \in V$. The ball of radius $\ell > 0$ around x is defined as

$$B_\ell(x) = B_\ell^G(x) := \{y \in V\colon\ \mathrm{dist}_G(y, x) \le \ell\}. \tag{2.2}$$

For a rooted graph $G = (V, E, \mathcal{O})$, the distance to the root is defined as

$$|x| := \mathrm{dist}_G(x, \mathcal{O}), \qquad x \in V, \tag{2.3}$$

and we set $B_\ell := B_\ell(\mathcal{O})$, $L_\ell := |B_\ell|$.

The classes of graphs that we will consider are as follows. Fix a parameter $d_{\max} \in \mathbb{N}$. For $r \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$, define

$$\mathfrak{G}_r := \left\{ \begin{matrix} \text{simple connected undirected rooted graphs } G = (V, E, \mathcal{O}) \text{ with} \\ V \text{ finite or countable, } |V| \ge r + 1 \text{ and } \max_{x\in V} \deg_G(x) \le d_{\max} \end{matrix} \right\}. \tag{2.4}$$

Note that if $G \in \mathfrak{G}_r$, then $L_r = |B_r| \ge r + 1$. Also define

$$\begin{aligned} \mathfrak{G}_\infty &= \textstyle\bigcap_{r\in\mathbb{N}_0} \mathfrak{G}_r \\ &= \left\{ \begin{matrix} \text{simple connected undirected rooted graphs } G=(V,E,\mathcal{O}) \text{ with} \\ V \text{ countable, } |V|=\infty \text{ and } \max_{x\in V} \deg_G(x)\le d_{\max} \end{matrix} \right\}. \end{aligned} \tag{2.5}$$

When dealing with infinite graphs, we will be interested in those that have an *exponential growth*. Thus we define, for $\vartheta > 0$,

$$\mathfrak{G}_\infty^{(\vartheta)} = \left\{ G \in \mathfrak{G}_\infty\colon\ \lim_{r\to\infty} \frac{\log L_r}{r} = \vartheta \right\}. \tag{2.6}$$

Note that $\mathcal{GW} \in \mathfrak{G}_\infty^{(\vartheta)}$ almost surely, with ϑ as in (1.12).

2.2 Spectral Bounds

Let $G = (V, E)$ be a simple connected graph with maximal degree $d_{\max} \in \mathbb{N}$, where the vertex set V may be finite or countably infinite.

We recall the Rayleigh-Ritz formula for the principal eigenvalue of the Anderson Hamiltonian. For $\Lambda \subset V$ and $q\colon V \to [-\infty, \infty)$, let $\lambda_\Lambda^{(1)}(q; G)$ denote the largest eigenvalue of the operator $\Delta_G + q$ in Λ with Dirichlet boundary conditions on $V \backslash \Lambda$. More precisely,

$$\lambda_\Lambda^{(1)}(q; G) := \sup\left\{\langle(\Delta_G + q)\phi, \phi\rangle_{\ell^2(V)}\colon \phi \in \mathbb{R}^V,\ \mathrm{supp}\phi \subset \Lambda,\ \|\phi\|_{\ell^2(V)} = 1\right\}. \tag{2.7}$$

We will often omit the superscript "(1)", i.e., write $\lambda_\Lambda(q; G) = \lambda_\Lambda^{(1)}(q; G)$, and abbreviate $\lambda_G(q) := \lambda_V(q; G)$. When there is no risk of confusion, we may also suppress G from the notation, and omit q when $q = \xi$.

Here are some straightforward consequences of the Rayleigh-Ritz formula:

1. For any $\Gamma \subset \Lambda$,

$$\max_{z\in\Gamma} q(z) - d_{\max} \le \lambda_\Gamma^{(1)}(q; G) \le \lambda_\Lambda^{(1)}(q; G) \le \max_{z\in\Lambda} q(z). \tag{2.8}$$

2. The eigenfunction corresponding to $\lambda_\Lambda^{(1)}(q; G)$ can be taken to be non-negative.
3. If q is real-valued and $\Gamma \subsetneq \Lambda$ are finite and connected in G, then the middle inequality in (2.8) is strict and the non-negative eigenfunction corresponding to $\lambda_\Lambda^{(1)}(q; G)$ is strictly positive.

In what follows we state some spectral bounds for the Feynman-Kac formula. These bounds are deterministic, i.e., they hold for any fixed realisation of the potential $\xi \in \mathbb{R}^V$.

Inside G, fix a finite connected subset $\Lambda \subset V$, and let H_Λ denote the Anderson Hamiltonian in Λ with zero Dirichlet boundary conditions on $\Lambda^c = V \backslash \Lambda$ (i.e., the restriction of the operator $H_G = \Delta_G + \xi$ to the class of functions supported on Λ). For $y \in \Lambda$, let u_Λ^y be the solution of

$$\begin{aligned} \partial_t u(x, t) &= (H_\Lambda u)(x, t), && x \in \Lambda,\ t > 0,\\ u(x, 0) &= \mathbb{1}_y(x), && x \in \Lambda, \end{aligned} \tag{2.9}$$

and set $U_\Lambda^y(t) := \sum_{x\in\Lambda} u_\Lambda^y(x, t)$. The solution admits the Feynman-Kac representation

$$u_\Lambda^y(x, t) = \mathbb{E}_y\left[\exp\left\{\int_0^t \xi(X_s)\mathrm{d}s\right\} \mathbb{1}\{\tau_{\Lambda^c} > t,\ X_t = x\}\right], \tag{2.10}$$

where τ_{Λ^c} is the hitting time of Λ^c. It also admits the spectral representation

$$u_\Lambda^y(x, t) = \sum_{k=1}^{|\Lambda|} \mathrm{e}^{t\lambda_\Lambda^{(k)}} \phi_\Lambda^{(k)}(y)\phi_\Lambda^{(k)}(x), \tag{2.11}$$

where $\lambda_\Lambda^{(1)} \geq \lambda_\Lambda^{(2)} \geq \cdots \geq \lambda_\Lambda^{(|\Lambda|)}$ and $\phi_\Lambda^{(1)}, \phi_\Lambda^{(2)}, \dots, \phi_\Lambda^{(|\Lambda|)}$ are, respectively, the eigenvalues and the corresponding orthonormal eigenfunctions of H_Λ. These two representations may be exploited to obtain bounds for one in terms of the other, as shown by the following lemma.

Lemma 2.1 (Bounds on the Solution) *For any $y \in \Lambda$ and any $t > 0$,*

$$\begin{aligned} &\mathrm{e}^{t\lambda_\Lambda^{(1)}} \phi_\Lambda^{(1)}(y)^2 \leq \mathbb{E}_y\left[\mathrm{e}^{\int_0^t \xi(X_s)\mathrm{d}s} \mathbb{1}_{\{\tau_{\Lambda^c} > t, X_t = y\}}\right] \\ &\leq \mathbb{E}_y\left[\mathrm{e}^{\int_0^t \xi(X_s)\mathrm{d}s} \mathbb{1}_{\{\tau_{\Lambda^c} > t\}}\right] \leq \mathrm{e}^{t\lambda_\Lambda^{(1)}} |\Lambda|^{1/2}. \end{aligned} \tag{2.12}$$

Proof The first and third inequalities follow from (2.10)–(2.11) after a suitable application of Parseval's identity. The second inequality is elementary. □

The next lemma bounds the Feynman-Kac formula integrated up to an exit time.

Lemma 2.2 (Mass up to an Exit Time) *For any $y \in \Lambda$ and $\gamma > \lambda_\Lambda^{(1)}$,*

$$\mathbb{E}_y\left[\exp\left\{\int_0^{\tau_{\Lambda^c}} (\xi(X_s) - \gamma)\,\mathrm{d}s\right\}\right] \leq 1 + \frac{d_{\max}|\Lambda|}{\gamma - \lambda_\Lambda^{(1)}}. \tag{2.13}$$

Proof See [10, Lemma 4.2]. □

2.3 *About the Constant χ*

We next introduce alternative representations for χ in (1.9) in terms of a 'dual' variational formula. Fix $\varrho \in (0, \infty)$ and a graph $G = (V, E)$. The functional

$$\mathcal{L}_V(q; \varrho) := \sum_{x \in V} \mathrm{e}^{q(x)/\varrho} \in [0, \infty], \qquad q\colon\ V \to [-\infty, \infty), \tag{2.14}$$

plays the role of a large deviation rate function for the potential ξ in V (compare with (1.7)). Henceforth we suppress the superscript "(1)" from the notation for the principal eigenvalue (2.7), i.e., we write

$$\lambda_\Lambda(q; G) = \lambda_\Lambda^{(1)}(q; G), \qquad \Lambda \subset V, \tag{2.15}$$

and abbreviate $\lambda_G(q) = \lambda_V(q; G)$. We also define

$$\widehat{\chi}_\Lambda(\varrho; G) := -\sup_{\substack{q\colon V \to [-\infty, \infty), \\ \mathcal{L}_V(q;\varrho) \leq 1}} \lambda_\Lambda(q; G) \in [0, \infty), \qquad \widehat{\chi}_G(\varrho) := \widehat{\chi}_V(\varrho; G). \tag{2.16}$$

The condition $\mathcal{L}_V(q;\varrho) \le 1$ on the supremum above ensures that the potentials q have a fair probability under the i.i.d. double-exponential distribution. Finally, for an infinite rooted graph $G = (V, E, \mathcal{O})$, we define

$$\chi_G^{(0)}(\varrho) := \inf_{r>0} \widehat{\chi}_{B_r}(\varrho; G). \tag{2.17}$$

Both $\chi^{(0)}$ and $\widehat{\chi}$ give different representations for χ.

Proposition 2.3 (Alternative Representations for χ) *For any graph $G = (V, E)$ and any $\Lambda \subset V$,*

$$\widehat{\chi}_\Lambda(\varrho; G) \ge \widehat{\chi}_V(\varrho; G) = \widehat{\chi}_G(\varrho) = \chi_G(\varrho). \tag{2.18}$$

If $G = (V, E, \mathcal{O}) \in \mathfrak{G}_\infty$, then

$$\chi_G^{(0)}(\varrho) = \lim_{r\to\infty} \widehat{\chi}_{B_r}(\varrho; G) = \chi_G(\varrho). \tag{2.19}$$

Proposition 2.3 will be proved in Sect. 6.1.

2.4 *Potentials and Islands*

We next consider properties of the potential landscape. Recall that $(\xi(x))_{x\in V}$ are i.i.d. double-exponential random variables. Set

$$a_L := \varrho \log\log(L \vee \mathrm{e}^{\mathrm{e}}). \tag{2.20}$$

The next lemma shows that a_{L_r} is the leading order of the maximum of ξ in B_r.

Lemma 2.4 (Maximum of the Potential) *Fix $r \mapsto g_r > 0$ with $\lim_{r\to\infty} g_r = \infty$. Then*

$$\sup_{G\in\mathfrak{G}_r} \mathrm{P}\left(\left|\max_{x\in B_r} \xi(x) - a_{L_r}\right| \ge \frac{g_r}{\log L_r}\right) \le \max\left\{\frac{1}{r^2}, \mathrm{e}^{-\frac{g_r}{\varrho}}\right\} \qquad \forall r > 2e^2. \tag{2.21}$$

Moreover, for any $\vartheta > 0$ and any $G \in \mathfrak{G}_\infty^{(\vartheta)}$, P*-almost surely eventually as $r \to \infty$,*

$$\left|\max_{x\in B_r} \xi(x) - a_{L_r}\right| \le \frac{2\varrho \log r}{\vartheta r}. \tag{2.22}$$

Proof Without loss of generality, we may assume that $g_r \le 2\varrho \log r$. Fix $G \in \mathfrak{G}_r$ and estimate

$$\mathrm{P}\left(\max_{x\in B_n} \xi(x) \le a_{L_r} - \frac{g_r}{\log L_r}\right) = \mathrm{e}^{-\frac{1}{\varrho} L_r (\log L_r) \mathrm{e}^{-\frac{g_r}{\varrho \log L_r}}} \le \mathrm{e}^{-\frac{r \log r}{\mathrm{e}^2 \varrho}} \le \mathrm{e}^{-\frac{g_r}{\varrho}}, \tag{2.23}$$

provided $r > 2\mathrm{e}^2$. On the other hand, using $\mathrm{e}^x \ge 1 + x$, $x \in \mathbb{R}$, we estimate

$$\mathrm{P}\left(\max_{x\in B_n} \xi(x) \ge a_{L_r} + \frac{g_r}{\log r}\right) = 1 - \left(1 - \mathrm{e}^{-\mathrm{e}^{\log\log L_r + \frac{g_r}{\varrho \log r}}}\right)^{L_r} \le \mathrm{e}^{-\frac{g_r}{\varrho}}. \tag{2.24}$$

Since the bounds above do not depend on $G \in \mathfrak{G}_r$, (2.21) follows.

For the case $G \in \mathfrak{G}_\infty^{(\vartheta)}$, let $g_r := \frac{3}{2}\varrho \log r$. Note that the right-hand side of (2.21) is summable over $r \in \mathbb{N}$, so that, by the Borel-Cantelli lemma,

$$\left|\max_{x\in B_r} \xi(x) - a_{L_r}\right| < \frac{g_r}{\log L_r} < \frac{2\varrho \log r}{\vartheta r} \tag{2.25}$$

P-almost surely eventually as $r \to \infty$, proving (2.22). □

For a fixed rooted graph $G = (V, E, \mathcal{O}) \in \mathfrak{G}_r$, we define sets of high excedances of the potential in B_r as follows. Given $A > 0$, let

$$\Pi_{r,A} = \Pi_{r,A}(\xi) := \{z \in B_r \colon \ \xi(z) > a_{L_r} - 2A\} \tag{2.26}$$

be the set vertices in B_r where the potential is close to maximal. For a fixed $\alpha \in (0, 1)$, define

$$S_r := (\log r)^\alpha \tag{2.27}$$

and set

$$D_{r,A} = D_{r,A}(\xi) := \{z \in B_r \colon \ \mathrm{dist}_G(z, \Pi_{r,A}) \le S_r\} \supset \Pi_{r,A}, \tag{2.28}$$

i.e., $D_{r,A}$ is the S_r-neighbourhood of $\Pi_{r,A}$. Let $\mathfrak{C}_{r,A}$ denote the set of all connected components of $D_{r,A}$ in G, which we call *islands*. For $\mathcal{C} \in \mathfrak{C}_{r,A}$, let

$$z_{\mathcal{C}} := \mathrm{argmax}\{\xi(z) \colon \ z \in \mathcal{C}\} \tag{2.29}$$

be the point with highest potential within $\mathcal{C}$. Since $\xi(0)$ has a continuous law, $z_{\mathcal{C}}$ is P-a.s. well defined for all $\mathcal{C} \in \mathfrak{C}_{r,A}$.

The next lemma gathers some useful properties of $\mathfrak{C}_{r,A}$.

Lemma 2.5 (Maximum Size of the Islands) *For every* $A > 0$*, there exists* $M_A \in \mathbb{N}$ *such that the following holds. For a graph* $G \in \mathfrak{G}_r$*, define the event*

$$\mathcal{B}_r := \left\{ \exists\, \mathcal{C} \in \mathfrak{C}_{r,A} \text{ with } |\mathcal{C} \cap \Pi_{r,A}| > M_A \right\}. \tag{2.30}$$

Then $\sum_{r\in\mathbb{N}_0} \sup_{G\in\mathfrak{G}_r} \mathrm{P}(\mathcal{B}_r) < \infty$*. In particular,*

$$\lim_{r\to\infty} \sup_{G\in\mathfrak{G}_r} \mathrm{P}(\mathcal{B}_r) = 0, \tag{2.31}$$

and, for any fixed $G \in \mathfrak{G}_\infty$*,* P*-almost surely eventually as* $r \to \infty$*,* $\mathcal{B}_r$ *does not occur. Note that*

$$\text{on } \mathcal{B}_r^{\mathrm{c}}, \forall\, \mathcal{C} \in \mathfrak{C}_{r,A}, |\mathcal{C} \cap \Pi_{r,A}| \le M_A, \mathrm{diam}_G(\mathcal{C}) \le 2M_A S_r, |\mathcal{C}| \le M_A d_{\max}^{S_r}. \tag{2.32}$$

Proof The claim follows from a straightforward estimate based on (1.7) (see [3, Lemma 6.6]). □

Apart from the dimensions, it will be also important to control the principal eigenvalues of islands in $\mathfrak{C}_{r,A}$. For this we restrict to graphs in $\mathfrak{G}_\infty^{(\vartheta)}$.

Lemma 2.6 (Principal Eigenvalues of the Islands) *For any* $\vartheta > 0$*, any* $G \in \mathfrak{G}_\infty^{(\vartheta)}$*, and any* $\varepsilon > 0$*,* P*-almost surely eventually as* $r \to \infty$*,*

$$\text{all } \mathcal{C} \in \mathfrak{C}_{r,A} \text{ satisfy: } \quad \lambda_{\mathcal{C}}^{(1)}(\xi; G) \le a_{L_r} - \widehat{\chi}_{\mathcal{C}}(\varrho; G) + \varepsilon. \tag{2.33}$$

Proof We follow [9, Lemma 2.11]. Let $\varepsilon > 0$, $G = (V, E, \mathcal{O}) \in \mathfrak{G}_\infty^{(\vartheta)}$, and define the event

$$\mathcal{B}_r := \left\{ \begin{array}{c} \text{there exists a connected subset } \Lambda \subset V \text{ with } \Lambda \cap B_r \neq \emptyset, \\ |\Lambda| \le M_A d_{\max}^{S_r} \text{ and } \lambda_\Lambda^{(1)}(\xi; G) > a_{L_r} - \widehat{\chi}_\Lambda(\varrho; G) + \varepsilon \end{array} \right\} \tag{2.34}$$

with M_A as in Lemma 2.5. Note that, by (1.7), $\mathrm{e}^{\xi(x)/\varrho}$ is stochastically dominated by $C \vee E$, where E is an Exp(1) random variable and $C > 0$ is a constant. Thus, for any $\Lambda \subset V$, using (2.16), taking $\gamma := \sqrt{\mathrm{e}^{\varepsilon/\varrho}} > 1$ and applying Markov's inequality, we may estimate

$$\begin{aligned} &\mathrm{P}\left(\lambda_\Lambda^{(1)}(\xi; G) > a_{L_r} - \widehat{\chi}_\Lambda(\varrho; G) + \varepsilon\right) \le \mathrm{P}\left(\mathcal{L}_\Lambda(\xi - a_{L_r} - \varepsilon) > 1\right) \\ &= \mathrm{P}\left(\gamma^{-1}\mathcal{L}_\Lambda(\xi) > \gamma \log L_r\right) \le \mathrm{e}^{-\gamma \log L_r} E[\mathrm{e}^{\gamma^{-1}\mathcal{L}_\Lambda(\xi)}] \le \mathrm{e}^{-\gamma \log L_r} K_\gamma^{|\Lambda|} \end{aligned} \tag{2.35}$$

for some constant $K_\gamma \in (1, \infty)$. Next note that, for any $x \in B_r$, $n \in \mathbb{N}$, the number of connected subsets $\Lambda \subset V$ with $x \in \Lambda$ and $|\Lambda| = n$ is at most $\mathrm{e}^{c_\circ n}$ for some

$c_\circ = c_\circ(d_{\max}) > 0$ (see e.g. [11, Proof of Theorem (4.20)]). Using a union bound and applying $\log L_r \sim \vartheta r$, we estimate, for some constants $c_1, c_2, c_3 > 0$,

$$\mathrm{P}(\mathcal{B}_r) \le \mathrm{e}^{-(\gamma-1)\log L_r} \sum_{n=1}^{\lfloor M_A d_{\max}^{S_r} \rfloor} \mathrm{e}^{c_\circ n} K_\gamma^n \le c_1 \exp\left\{-c_2 r + c_3 d_{\max}^{(\log r)^\alpha}\right\} \le \mathrm{e}^{-\frac{1}{2}c_2 r} \tag{2.36}$$

when r is large. Now the Borel-Cantelli lemma implies that, P-a.s. eventually as $r \to \infty$, $\mathcal{B}_r$ does not occur. The proof is completed by invoking Lemma 2.5. □

For later use, we state the consequence for $\mathcal{GW}$ in terms of $\widetilde{\chi}(\rho)$ in (1.13).

Corollary 2.7 (Uniform Bound on Principal Eigenvalue of the Islands) *For $G = \mathcal{GW}$ as in Sect. 1.3, $\vartheta >$ as in (1.12), and any $\varepsilon > 0$, $\mathrm{P} \times \mathfrak{P}$-almost surely eventually as $r \to \infty$,*

$$\max_{\mathcal{C} \in \mathfrak{C}_{r,A}} \lambda_{\mathcal{C}}^{(1)}(\xi; G) \le a_{L_r} - \widetilde{\chi}(\varrho) + \varepsilon. \tag{2.37}$$

Proof Note that $\mathcal{GW} \in \mathfrak{G}_\infty^{(\vartheta)}$ almost surely, so Lemma 2.6 applies. By Lemma 2.4, for any constant $C > 0$, the maximum of ξ in a ball of radius CS_r around $\mathcal{O}$ is of order $O(\log\log r)$. This means that $\mathcal{O}$ is distant from $\Pi_{r,A}$, in particular, $\mathrm{dist}(\mathcal{O}, D_{r,A}) \ge 2$ almost surely eventually as $r \to \infty$. For $\mathcal{C} \in \mathfrak{C}_{r,A}$, let $T_{\mathcal{C}}$ be the infinite tree obtained by attaching to each $x \in \partial\mathcal{C} := \{y \notin \mathcal{C} \colon \exists z \in \mathcal{C} \text{ with } z \sim y\} \not\ni \mathcal{O}$ an infinite tree with constant offspring $d_{\min} - 1$. Then $T_{\mathcal{C}}$ is an infinite tree with degrees in $\mathrm{supp}(D_g)$ and, by Proposition 2.3,

$$\widehat{\chi}_{\mathcal{C}}(\varrho; \mathcal{GW}) = \widehat{\chi}_{\mathcal{C}}(\varrho; T_{\mathcal{C}}) \ge \chi_{T_{\mathcal{C}}}(\varrho) \ge \widetilde{\chi}(\varrho), \tag{2.38}$$

so the claim follows by Lemma 2.6. □

2.5 Connectivity

We again work in the setting of Sect. 2.1. We recall the following Chernoff bound for a Binomial random variable $\mathrm{Bin}(n, p)$ with parameters n, p (see e.g. [4, Lemma 5.9]):

$$P\left(\mathrm{Bin}(n, p) \ge u\right) \le \exp\left\{-u\left(\log\frac{u}{np} - 1\right)\right\} \qquad \forall u > 0. \tag{2.39}$$

Lemma 2.8 (Number of Intermediate Peaks of the Potential) *For any $\beta \in (0, 1)$ and any $\varepsilon \in (0, \beta/2)$, the following holds. For $G \in \mathfrak{G}_r$ and a self-avoiding path π in G, set*

$$N_\pi = N_\pi(\xi) := |\{z \in \mathrm{supp}(\pi)\colon\ \xi(z) > (1-\varepsilon)a_{L_r}\}|. \tag{2.40}$$

Define the event

$$\mathcal{B}_r := \left\{ \begin{array}{c} \textit{there exists a self-avoiding path } \pi \textit{ in } G \textit{ with} \\ \mathrm{supp}(\pi) \cap B_r \neq \emptyset,\ |\mathrm{supp}(\pi)| \geq (\log L_r)^\beta \textit{ and } N_\pi > \frac{|\mathrm{supp}(\pi)|}{(\log L_r)^\varepsilon} \end{array} \right\}. \tag{2.41}$$

Then $\sum_{r\in\mathbb{N}_0} \sup_{G\in\mathfrak{G}_r} \mathrm{P}(\mathcal{B}_r) < \infty$. In particular,

$$\lim_{r\to\infty} \sup_{G\in\mathfrak{G}_r} \mathrm{P}(\mathcal{B}_r) = 0 \tag{2.42}$$

and, for any fixed $G \in \mathfrak{G}_\infty$, P-almost surely eventually as $r \to \infty$, all self-avoiding paths π in G with $\mathrm{supp}(\pi) \cap B_r \neq \emptyset$ and $|\mathrm{supp}(\pi)| \geq (\log L_r)^\beta$ satisfy $N_\pi \leq \frac{|\mathrm{supp}(\pi)|}{(\log L_r)^\varepsilon}$.

Proof Fix $\beta \in (0, 1)$ and $\varepsilon \in (0, \beta/2)$. For any $G \in \mathfrak{G}_r$, (1.7) implies

$$p_r := \mathrm{P}(\xi(0) > (1-\varepsilon)a_{L_r}) = \exp\left\{-(\log L_r)^{1-\varepsilon}\right\}. \tag{2.43}$$

Fix $x \in B_n$ and $k \in \mathbb{N}$. The number of self-avoiding paths π in B_r with $|\mathrm{supp}(\pi)| = k$ and $\pi_0 = x$ is at most $d_{\max}^k$. For such a π, the random variable N_π has a Bin(p_r, k)-distribution. Using (2.39) and a union bound, we obtain

$$\begin{aligned} &\mathrm{P}\Big(\exists \text{ self-avoiding } \pi \text{ with } |\mathrm{supp}(\pi)| = k, \pi_0 = x \text{ and } N_\pi > k/(\log L_r)^\varepsilon\Big) \\ &\leq \exp\left\{-k\left((\log L_r)^{1-2\varepsilon} - \log d_{\max} - \tfrac{1+\varepsilon \log\log L_r}{(\log L_r)^\varepsilon}\right)\right\}. \end{aligned} \tag{2.44}$$

Note that, since $L_r > r$ and the function $x \mapsto \log\log x/(\log x)^\varepsilon$ is eventually decreasing, for r large enough and uniformly over $G \in \mathfrak{G}_r$, the expression in parentheses above is at least $\frac{1}{2}(\log L_r)^{1-2\varepsilon}$. Summing over $k \geq (\log L_r)^\beta$ and $x \in B_r$, we get

$$\begin{aligned} &\mathrm{P}\left(\exists \text{ self-avoiding } \pi \text{ such that } |\mathrm{supp}(\pi)| \geq (\log L_r)^\beta \text{ and (2.40 does not hold}\right) \\ &\quad \leq 2L_r \exp\left\{-\tfrac{1}{2}(\log L_r)^{1+\beta-2\varepsilon}\right\} \leq c_1 \exp\left\{-c_2(\log L_r)^{1+\delta}\right\} \end{aligned} \tag{2.45}$$

for some positive constants c_1, c_2, δ, uniformly over $G \in \mathfrak{G}_r$. Since $L_r > r$, (2.45) is summable in r (uniformly over $G \in \mathfrak{G}_r$). The proof is concluded invoking the Borel-Cantelli lemma. □

A similar computation bounds the number of high exceedances of the potential.

Lemma 2.9 (Number of High Exceedances of the Potential) *For any $A > 0$ there is a $C \geq 1$ such that, for all $\delta \in (0, 1)$, the following holds. For $G \in \mathfrak{G}_r$ and a self-avoiding path π in G, let*

$$N_\pi := |\{x \in \operatorname{supp}(\pi)\colon\ \xi(x) > a_{L_r} - 2A\}|. \tag{2.46}$$

Define the event

$$\mathcal{B}_r := \left\{ \begin{array}{c} \textit{there exists a self-avoiding path } \pi \textit{ in } G \textit{ with} \\ \operatorname{supp}(\pi) \cap B_r \neq \emptyset,\ |\operatorname{supp}(\pi)| \geq C(\log L_r)^\delta \textit{ and } N_\pi > \frac{|\operatorname{supp}(\pi)|}{(\log L_r)^\delta} \end{array} \right\}. \tag{2.47}$$

Then $\sum_{r \in \mathbb{N}_0} \sup_{G \in \mathfrak{G}_r} \mathrm{P}(\mathcal{B}_r) < \infty$. In particular,

$$\lim_{r \to \infty} \sup_{G \in \mathfrak{G}_r} \mathrm{P}(\mathcal{B}_r) = 0 \tag{2.48}$$

and, for any fixed $G \in \mathfrak{G}_\infty$, P-almost surely eventually as $r \to \infty$, all self-avoiding paths π in G with $\operatorname{supp}(\pi) \cap B_r \neq \emptyset$ and $|\operatorname{supp}(\pi)| \geq C(\log L_r)^\delta$ satisfy

$$N_\pi = |\{x \in \operatorname{supp}(\pi)\colon\ \xi(x) > a_{L_r} - 2A\}| \leq \frac{|\operatorname{supp}(\pi)|}{(\log L_r)^\delta}. \tag{2.49}$$

Proof Proceed as for Lemma 2.8, noting that this time

$$p_r := \mathrm{P}\big(\xi(0) > a_{L_r} - 2A\big) = L_r^{-\epsilon} \tag{2.50}$$

where $\epsilon = \mathrm{e}^{-\frac{2A}{\varrho}}$, and taking $C > 2/\epsilon$. □

3 Path Expansions

We again work in the setting of Sect. 2.1. In the following, we develop a way to bound the contribution of certain specific classes of paths to the Feynman-Kac formula, similar to what is done in [4] in the $\mathbb{Z}^d$-case. In Sect. 3.1 we state a key proposition reducing the entropy of paths. This proposition is proved in Sect. 3.4 with the help of a lemma bounding the mass of an equivalence class of paths, which is stated and proved in Sect. 3.3 and is based on ideas from [14]. The proof of this lemma requires two further lemmas controlling the mass of the solution along excursions, which are stated and proved in Sect. 3.2.

3.1 Key Proposition

Fix a graph $G = (V, E, \mathcal{O}) \in \mathfrak{G}_r$. We define various sets of nearest-neighbour paths in G as follows. For $\ell \in \mathbb{N}_0$ and subsets $\Lambda, \Lambda' \subset V$, put

$$\begin{aligned} \mathscr{P}_\ell(\Lambda, \Lambda') &:= \left\{ (\pi_0, \ldots, \pi_\ell) \in V^{\ell+1} \colon \begin{array}{l} \pi_0 \in \Lambda, \pi_\ell \in \Lambda', \\ \{\pi_i, \pi_{i-1}\} \in E \;\forall\, 1 \le i \le \ell \end{array} \right\}, \\ \mathscr{P}(\Lambda, \Lambda') &:= \textstyle\bigcup_{\ell \in \mathbb{N}_0} \mathscr{P}_\ell(\Lambda, \Lambda'), \end{aligned} \tag{3.1}$$

and set

$$\mathscr{P}_\ell := \mathscr{P}_\ell(V, V), \qquad \mathscr{P} := \mathscr{P}(V, V). \tag{3.2}$$

When Λ or Λ' consists of a single point, we write x instead of $\{x\}$. For $\pi \in \mathscr{P}_\ell$, we set $|\pi| := \ell$. We write $\mathrm{supp}(\pi) := \{\pi_0, \ldots, \pi_{|\pi|}\}$ to denote the set of points visited by π.

Let $X = (X_t)_{t \ge 0}$ be the continuous-time random walk on G that jumps from $x \in V$ to any neighbour $y \sim x$ with rate 1. We denote by $(T_k)_{k \in \mathbb{N}_0}$ the sequence of jump times (with $T_0 := 0$). For $\ell \in \mathbb{N}_0$, let

$$\pi^{(\ell)}(X) := (X_0, \ldots, X_{T_\ell}) \tag{3.3}$$

be the path in $\mathscr{P}_\ell$ consisting of the first ℓ steps of X and, for $t \ge 0$, let

$$\pi(X_{[0,t]}) = \pi^{(\ell_t)}(X), \qquad \text{where } \ell_t \in \mathbb{N}_0 \text{ satisfies } T_{\ell_t} \le t < T_{\ell_t + 1}, \tag{3.4}$$

denote the path in $\mathscr{P}$ consisting of all the steps taken by X between times 0 and t.

Recall the definitions from Sect. 2.4. For $G \in \mathfrak{G}_r$, $\pi \in \mathscr{P}$ and $A > 0$, define

$$\lambda_{r,A}(\pi) := \sup\left\{ \lambda^{(1)}_{\mathcal{C}}(\xi; G) \colon \mathcal{C} \in \mathfrak{C}_{r,A},\ \mathrm{supp}(\pi) \cap \mathcal{C} \cap \Pi_{r,A} \ne \emptyset \right\}, \tag{3.5}$$

with the convention $\sup \emptyset = -\infty$. This is the largest principal eigenvalue among the components of $\mathfrak{C}_{r,A}$ in G that have a point of high exceedance visited by the path π.

The main result of this section is the following proposition. Hereafter we abbreviate $\log^{(3)} x := \log\log\log x$.

Proposition 3.1 (Entropy Reduction) *For every fixed $d_{\max} \in \mathbb{N}$, there exists an $A_0 = A_0(d_{\max}) > 0$ such that the following holds. Let $\alpha \in (0, 1)$ be as in (2.27) and let $\kappa \in (\alpha, 1)$. For all $A > A_0$, there exists a constant $c_A = c_A(d_{\max}) > 0$ such that, with probability tending to one as $r \to \infty$ uniformly over $G \in \mathfrak{G}_r$, the following statement is true: For each $x \in B_r$, each $\mathcal{N} \subset \mathscr{P}(x, B_r)$ satisfying*

$\mathrm{supp}(\pi) \subset B_r$ *and* $\max_{1\le \ell\le|\pi|} \mathrm{dist}_G(\pi_\ell, x) \ge (\log L_r)^\kappa$ *for all* $\pi \in \mathcal{N}$, *and each assignment* $\pi \mapsto (\gamma_\pi, z_\pi) \in \mathbb{R} \times V$ *satisfying*

$$\gamma_\pi \ge \Big(\lambda_{r,A}(\pi) + \mathrm{e}^{-S_r}\Big) \vee (a_{L_r} - A) \qquad \textit{for all } \pi \in \mathcal{N} \tag{3.6}$$

and

$$z_\pi \in \mathrm{supp}(\pi) \cup \bigcup_{\substack{\mathcal{C}\in\mathfrak{C}_{r,A}:\\ \mathrm{supp}(\pi)\cap\mathcal{C}\cap\Pi_{r,A}\neq\emptyset}} \mathcal{C} \qquad \textit{for all } \pi \in \mathcal{N}, \tag{3.7}$$

the following inequality holds for all $t \ge 0$:

$$\log \mathbb{E}_x \Big[\mathrm{e}^{\int_0^t \xi(X_s)\mathrm{d}s} 1\!\!1_{\{\pi(X_{[0,t]})\in\mathcal{N}\}}\Big] \le \sup_{\pi\in\mathcal{N}} \Big\{t\gamma_\pi - \big(\log^{(3)} L_r - c_A\big)\, \mathrm{dist}_G(x, z_\pi)\Big\}. \tag{3.8}$$

Moreover, for any $G \in \mathfrak{G}_\infty$, P-*almost surely eventually as* $r \to \infty$, *the same statement is true.*

The key to the proof of Proposition 3.1 in Sect. 3.4 is Lemma 3.5 in Sect. 3.3, whose proof depends on Lemmas 3.2–3.3 in Sect. 3.2. We note that all these results are deterministic, i.e., they hold for any realisation of the potential ξ.

3.2 Mass of the Solution Along Excursions

Fix $G = (V, E, \mathcal{O}) \in \mathfrak{G}_r$. The first step to control the contribution of a path to the total mass is to control the contribution of excursions outside $\Pi_{r,A}$ (recall (2.26)).

Lemma 3.2 (Path Evaluation) *For* $\ell \in \mathbb{N}_0$, $\pi \in \mathscr{P}_\ell$ *and* $\gamma > \max_{0\le i<|\pi|}\{\xi(\pi_i) - \deg(\pi_i)\}$,

$$\mathbb{E}_{\pi_0}\left[\exp\left\{\int_0^{T_\ell} (\xi(X_s) - \gamma)\,\mathrm{d}s\right\} \,\middle|\, \pi^{(\ell)}(X) = \pi\right] = \prod_{i=0}^{\ell-1} \frac{\deg(\pi_i)}{\gamma - [\xi(\pi_i) - \deg(\pi_i)]}. \tag{3.9}$$

Proof The left-hand side of (3.9) can be evaluated using the fact that T_ℓ is the sum of ℓ independent Exp(deg(π_i)) random variables that are independent of $\pi^{(\ell)}(X)$. The condition on γ ensures that all integrals are finite. □

For a path $\pi \in \mathscr{P}$ and $\varepsilon \in (0, 1)$, we write

$$M_\pi^{r,\varepsilon} := \big|\big\{0 \le i < |\pi|\colon\ \xi(\pi_i) \le (1-\varepsilon)a_{L_r}\big\}\big|, \tag{3.10}$$

with the interpretation that $M_\pi^{r,\varepsilon} = 0$ if $|\pi| = 0$.

Lemma 3.3 (Mass of Excursions) *For every $A, \varepsilon > 0$ there exist $c > 0$ and $n_0 \in \mathbb{N}$ such that, for all $r \geq n_0$, all $\gamma > a_{L_r} - A$ and all $\pi \in \mathscr{P}$ satisfying $\pi_i \notin \Pi_{r,A}$ for all $0 \leq i < \ell := |\pi|$,*

$$\mathbb{E}_{\pi_0}\left[\exp\left\{\int_0^{T_\ell} (\xi(X_t) - \gamma)\,\mathrm{d}s\right\} \middle| \pi^{(\ell)}(X) = \pi\right] \leq q_A^\ell \mathrm{e}^{(c - \log^{(3)} L_r) M_\pi^{r,\varepsilon}}, \tag{3.11}$$

where $q_A := (1 + A/d_{\max})^{-1}$. Note that $\pi_\ell \in \Pi_{r,A}$ is allowed.

Proof By our assumptions on π and γ, we can use Lemma 3.2. Splitting the product in the right-hand side of (3.9) according to whether $\xi(\pi_i) \geq (1-\varepsilon)a_{L_r}$ or not, and using that $\xi(\pi_i) \leq a_{L_r} - 2A$ for $0 \leq i < |\pi|$, we bound the left-hand side of (3.11) by

$$q_A^\ell \left[q_A \frac{\varepsilon a_{L_r} - A}{d_{\max}}\right]^{-|\{0 \leq i < \ell \colon \xi(\pi_i) \leq (1-\varepsilon) a_{L_r}\}|}. \tag{3.12}$$

Since $a_{L_r} = \varrho \log\log L_r \geq \varrho \log\log r$, for large r the number within square brackets in (3.12) is at least $q_A \varepsilon \varrho (\log\log L_r)/2d_{\max} > 1$. Hence (3.11) holds with $c := \log(1 \vee 2d_{\max}(q_A \varepsilon \varrho)^{-1})$. □

3.3 Equivalence Classes of Paths

We follow [4, Section 6.2]. Note that the distance between $\Pi_{r,A}$ and $D_{r,A}^{\mathrm{c}}$ in G is at least $S_r = (\log L_r)^\alpha$.

Definition 3.4 (Concatenation of Paths)

(a) When π and π' are two paths in $\mathscr{P}$ with $\pi_{|\pi|} = \pi'_0$, we define their *concatenation* as

$$\pi \circ \pi' := (\pi_0, \ldots, \pi_{|\pi|}, \pi'_1, \ldots, \pi'_{|\pi'|}) \in \mathscr{P}. \tag{3.13}$$

Note that $|\pi \circ \pi'| = |\pi| + |\pi'|$.

(b) When $\pi_{|\pi|} \neq \pi'_0$, we can still define the *shifted concatenation* of π and π' as $\pi \circ \hat{\pi}'$, where $\hat{\pi}' := (\pi_{|\pi|}, \pi_{|\pi|} + \pi'_1 - \pi'_0, \ldots, \pi_{|\pi|} + \pi'_{|\pi'|} - \pi'_0)$. The shifted concatenation of multiple paths is defined inductively via associativity.

Now, if a path $\pi \in \mathscr{P}$ intersects $\Pi_{r,A}$, then it can be decomposed into an initial path, a sequence of excursions between $\Pi_{r,A}$ and $D_{r,A}^{\mathrm{c}}$, and a terminal path. More precisely, there exists $m_\pi \in \mathbb{N}$ such that

$$\pi = \check{\pi}^{(1)} \circ \hat{\pi}^{(1)} \circ \cdots \circ \check{\pi}^{(m_\pi)} \circ \hat{\pi}^{(m_\pi)} \circ \bar{\pi}, \tag{3.14}$$

where the paths in (3.14) satisfy

$$\begin{array}{llll}\check{\pi}^{(1)} \in \mathscr{P}(V, \Pi_{r,A}) & \text{with } \check{\pi}^{(1)}_i \notin \Pi_{r,A}, & 0 \le i < |\check{\pi}^{(1)}|, & \\ \hat{\pi}^{(k)} \in \mathscr{P}(\Pi_{r,A}, D^{\mathrm{c}}_{r,A}) & \text{with } \hat{\pi}^{(k)}_i \in D_{r,A}, & 0 \le i < |\hat{\pi}^{(k)}|, & 1 \le k \le m_\pi - 1, \\ \check{\pi}^{(k)} \in \mathscr{P}(D^{\mathrm{c}}_{r,A}, \Pi_{r,A}) & \text{with } \check{\pi}^{(k)}_i \notin \Pi_{r,A}, & 0 \le i < |\check{\pi}^{(k)}|, & 2 \le k \le m_\pi, \\ \hat{\pi}^{(m_\pi)} \in \mathscr{P}(\Pi_{r,A}, V) & \text{with } \hat{\pi}^{(m_\pi)}_i \in D_{r,A}, & 0 \le i < |\hat{\pi}^{(m_\pi)}|, & \end{array} \tag{3.15}$$

while

$$\begin{array}{ll}\bar{\pi} \in \mathscr{P}(D^{\mathrm{c}}_{r,A}, V) \text{ and } \bar{\pi}_i \notin \Pi_{r,A}\ \forall i \ge 0 & \text{if } \hat{\pi}^{(m_\pi)} \in \mathscr{P}(\Pi_{r,A}, D^{\mathrm{c}}_{r,A}), \\ \bar{\pi}_0 \in D_{r,A}, |\bar{\pi}| = 0 & \text{otherwise.}\end{array} \tag{3.16}$$

Note that the decomposition in (3.14)–(3.16) is unique, and that the paths $\check{\pi}^{(1)}$, $\hat{\pi}^{(m_\pi)}$ and $\bar{\pi}$ can have zero length. If π is contained in B_r, then so are all the paths in the decomposition.

Whenever $\mathrm{supp}(\pi) \cap \Pi_{r,A} \neq \emptyset$ and $\varepsilon > 0$, we define

$$s_\pi := \sum_{i=1}^{m_\pi} |\check{\pi}^{(i)}| + |\bar{\pi}|, \qquad k^{r,\varepsilon}_\pi := \sum_{i=1}^{m_\pi} M^{r,\varepsilon}_{\check{\pi}^{(i)}} + M^{r,\varepsilon}_{\bar{\pi}} \tag{3.17}$$

to be the total time spent in exterior excursions, respectively, on moderately low points of the potential visited by exterior excursions (without their last point).

In case $\mathrm{supp}(\pi) \cap \Pi_{r,A} = \emptyset$, we set $m_\pi := 0$, $s_\pi := |\pi|$ and $k^{r,\varepsilon}_\pi := M^{r,\varepsilon}_\pi$. Recall from (3.5) that, in this case, $\lambda_{r,A}(\pi) = -\infty$.

We say that $\pi, \pi' \in \mathscr{P}$ are *equivalent*, written $\pi' \sim \pi$, if $m_\pi = m_{\pi'}$, $\check{\pi}'^{(i)} = \check{\pi}^{(i)}$ for all $i = 1, \ldots, m_\pi$, and $\bar{\pi}' = \bar{\pi}$. If $\pi' \sim \pi$, then $s_{\pi'}$, $k^{r,\varepsilon}_{\pi'}$ and $\lambda_{r,A}(\pi')$ are all equal to the counterparts for π.

To state our key lemma, we define, for $m, s \in \mathbb{N}_0$,

$$\mathscr{P}^{(m,s)} = \{\pi \in \mathscr{P} \colon\ m_\pi = m, s_\pi = s\}\,, \tag{3.18}$$

and denote by

$$C_{r,A} := \max\{|\mathcal{C}| \colon\ \mathcal{C} \in \mathfrak{C}_{r,A}\} \tag{3.19}$$

the maximal size of the islands in $\mathfrak{C}_{r,A}$.

Lemma 3.5 (Mass of an Equivalence Class) *For every $A, \varepsilon > 0$ there exist $c > 0$ and $r_0 \in \mathbb{N}$ such that, for all $r \ge r_0$, all $m, s \in \mathbb{N}_0$, all $\pi \in \mathscr{P}^{(m,s)}$ with $\mathrm{supp}(\pi) \subset B_r$, all $\gamma > \lambda_{r,A}(\pi) \vee (a_{L_r} - A)$ and all $t \ge 0$,*

$$\begin{aligned}&\mathbb{E}_{\pi_0}\left[\mathrm{e}^{\int_0^t (\xi(X_u)-\gamma)\,\mathrm{d}u}\, \mathbb{1}_{\{\pi(X_{[0,t]})\sim\pi\}}\right] \\ &\le \left(C^{1/2}_{r,A}\right)^{\mathbb{1}_{\{m>0\}}} \left(1 + \tfrac{d_{\max} C_{r,A}}{\gamma - \lambda_{r,A}(\pi)}\right)^m \left(\tfrac{q_A}{d_{\max}}\right)^s \mathrm{e}^{\left(c - \log^{(3)} L_r\right) k^{r,\varepsilon}_\pi}.\end{aligned} \tag{3.20}$$

Proof Fix $A, \varepsilon > 0$ and let $c > 0$, $n_0 \in \mathbb{N}$ be as given by Lemma 3.3. Set

$$I_a^b := \mathrm{e}^{\int_a^b (\xi(X_u)-\gamma)\mathrm{d}u}, \qquad 0 \le a \le b < \infty. \tag{3.21}$$

We use induction on m. Suppose that $m = 1$, let $\ell := |\breve{\pi}^{(1)}|$. There are two possibilities: either $\bar{\pi}_0$ belongs to $D_{r,A}$ or not. First we consider the case $\bar{\pi}_0 \in D_{r,A}$, which implies that $|\bar{\pi}| = 0$. By the strong Markov property,

$$\begin{aligned}\mathbb{E}_{\pi_0}\Big[I_0^t\, 1\!\!1_{\{\pi(X_{[0,t]})\sim\pi\}}\Big] &\le \mathbb{E}_{\pi_0}\left[I_0^{T_\ell} I_{T_\ell}^t\, 1\!\!1_{\{\pi^{(\ell)}(X)=\breve{\pi}^{(1)}\}} 1\!\!1_{\{T_\ell<t\}} 1\!\!1_{\{X_{u+T_\ell}\in D_{r,A}\ \forall u\in[0,t-T_\ell]\}}\right] \\ &= \mathbb{E}_{\pi_0}\left[I_0^{T_\ell}\, 1\!\!1_{\{\pi^{(\ell)}(X)=\breve{\pi}^{(1)}\}} 1\!\!1_{\{T_\ell<t\}} \left(\mathbb{E}_{\breve{\pi}_\ell^{(1)}}\left[I_0^{t-u}\, 1\!\!1_{\{\tau_{D^{\mathrm{c}}_{r,A}}>t-u\}}\right]\right)_{u=T_\ell}\right].\end{aligned} \tag{3.22}$$

Put $z = \breve{\pi}_\ell^{(1)}$. Since $z \in \Pi_{r,A}$, we may write $\mathcal{C}_z$ to denote the island in $\mathfrak{C}_{r,A}$ containing z. Since $\tau_{D^{\mathrm{c}}_{r,A}} = \tau_{\mathcal{C}_z^{\mathrm{c}}}$ $\mathbb{P}_z$-a.s., Lemma 2.1 and the hypothesis on γ allow us to bound the inner expectation in (3.22) by $|\mathcal{C}_z|^{1/2}$. Applying Lemma 3.3, we further bound (3.22) by

$$|\mathcal{C}_z|^{1/2}\mathbb{E}_{\pi_0}\left[I_0^{T_\ell}\, 1\!\!1_{\{\pi^{(\ell)}(X)=\breve{\pi}^{(1)}\}}\right] \le C_{r,A}^{1/2}\left(\frac{q_A}{d_{\max}}\right)^\ell \mathrm{e}^{\left(c-\log^{(3)} L_r\right) M^{r,\varepsilon}_{\breve{\pi}^{(1)}}}, \tag{3.23}$$

which proves (3.20) for $m = 1$ and $\bar{\pi}_0 \in D_{r,A}$.

Next consider the case $\bar{\pi}_0 \in D^{\mathrm{c}}_{r,A}$. Abbreviating $\sigma := \inf\{u > T_\ell\colon\ X_u \notin D_{r,A}\}$, write

$$\mathbb{E}_{\pi_0}\left[I_0^t\, 1\!\!1_{\{\pi(X_{[0,t]})\sim\pi\}}\right] \le \mathbb{E}_{\pi_0}\left[I_0^\sigma\, 1\!\!1_{\{\pi^{(\ell)}(X)=\breve{\pi}^{(1)},\ \sigma<t\}}\Big(\mathbb{E}_{\bar{\pi}_0}\left[I_0^{t-u}\, 1\!\!1_{\{\pi(X_{[0,t-u]})=\bar{\pi}\}}\right]\Big)_{u=\sigma}\right]. \tag{3.24}$$

Let $\ell_* := |\bar{\pi}|$ and note that, since $\bar{\pi}_{\ell_*} \notin \Pi_{r,A}$, by the hypothesis on γ we have

$$\mathbb{E}_{\bar{\pi}_0}\left[I_0^{t-u}\, 1\!\!1_{\{\pi(X_{[0,t-u]})=\bar{\pi}\}}\right] \le \mathbb{E}_{\bar{\pi}_0}\left[I_0^{T_{\ell_*}}\, 1\!\!1_{\{\pi^{(\ell_*)}(X)=\bar{\pi}\}}\right] \le \left(\frac{q_A}{d_{\max}}\right)^{\ell_*} \mathrm{e}^{\left(c-\log^{(3)} L_r\right) M^{r,\varepsilon}_{\bar{\pi}}} \tag{3.25}$$

where the second inequality holds by Lemma 3.3. On the other hand, by Lemmas 2.2 and 3.3,

$$\begin{aligned}\mathbb{E}_{\pi_0}\left[I_0^\sigma\, 1\!\!1_{\{\pi^{(\ell)}(X)=\breve{\pi}^{(1)}\}}\right] &= \mathbb{E}_{\pi_0}\left[I_0^{T_\ell}\, 1\!\!1_{\{\pi^{(\ell)}(X)=\breve{\pi}^{(1)}\}}\right]\mathbb{E}_z\left[I_0^{\tau_{\mathcal{C}_z^{\mathrm{c}}}}\right] \\ &\le \left(1+\tfrac{d_{\max}\, C_{r,A}}{\gamma-\lambda_{r,A}(\pi)}\right)\left(\tfrac{q_A}{d_{\max}}\right)^\ell \mathrm{e}^{\left(c-\log^{(3)} L_r\right) M^{r,\varepsilon}_{\breve{\pi}^{(1)}}}.\end{aligned} \tag{3.26}$$

Putting together (3.24)–(3.26), we complete the proof of the case $m = 1$. The case $m = 0$ follows from (3.25) after we replace $\bar{\pi}$ by π and $t - u$ by t.

Suppose now that the claim is proved for some $m \geq 1$, and let $\pi \in \mathscr{P}^{(m+1,s)}$. Define $\pi' := \check{\pi}^{(2)} \circ \hat{\pi}^{(2)} \circ \cdots \circ \check{\pi}^{(m+1)} \circ \hat{\pi}^{(m+1)} \circ \bar{\pi}$. Then $\pi' \in \mathscr{P}^{(m,s')}$, where $s = s' + |\check{\pi}^{(1)}|$ and $k_{\pi}^{r,\varepsilon} = M_{\check{\pi}^{(1)}}^{r,\varepsilon} + k_{\pi'}^{r,\varepsilon}$. Setting $\ell := |\check{\pi}^{(1)}|$, $\sigma := \inf\{u > T_\ell \colon X_u \notin D_{r,A}\}$ and $x := \check{\pi}_0^{(2)}$, we get

$$\mathbb{E}_{\pi_0}\left[I_0^t 1\!\!1_{\{\pi(X_{0,t}) \sim \pi\}}\right] \leq \mathbb{E}_{\pi_0}\left[I_0^{\sigma} 1\!\!1_{\{\pi^{(\ell)}(X) = \check{\pi}^{(1)},\, \sigma < t\}} \Big(\mathbb{E}_x\left[I_0^{t-u} 1\!\!1_{\{\pi(X_{0,t-u}) \sim \pi'\}}\right]\Big)_{s=\sigma}\right], \tag{3.27}$$

from which (3.20) follows via the induction hypothesis and (3.26). □

3.4 *Proof of Proposition 3.1*

Proof The proof is based on Lemma 3.5. First define

$$c_0 := 1 + 3 \log\log d_{\max}, \qquad A_0 := d_{\max}\left(e^{3c_0} - 1\right). \tag{3.28}$$

Fix $A > A_0$, $\beta < \alpha$ and $\varepsilon \in (0, \beta/2)$ as in Lemma 2.8. Let $r_0 \in \mathbb{N}$ be as given by Lemma 3.5, and take $r \geq r_0$ so large that the conclusions of Lemmas 2.5–2.8 hold, i.e., assume that the events $\mathcal{B}_r$ from both lemmas do not occur with either $G = (V, E, \mathcal{O}) \in \mathfrak{G}_r$ or $G \in \mathfrak{G}_\infty$ accordingly. Fix $x \in B_r$. Recall the definitions of $C_{r,A}$ and $\mathscr{P}^{(m,s)}$. Noting that the relation $\sim$ defined below (3.17) is an equivalence relation in $\mathscr{P}^{(m,s)}$, we define

$$\widetilde{\mathscr{P}}_x^{(m,s)} := \left\{\text{equivalence classes of the paths in } \mathscr{P}(x, V) \cap \mathscr{P}^{(m,s)}\right\}. \tag{3.29}$$

Lemma 3.6 (Bound Equivalence Classes)
$|\widetilde{\mathscr{P}}_x^{(m,s)}| \leq [2d_{\max} C_{r,A}]^m d_{\max}^s$ *for all* $m, s \in \mathbb{N}_0$.

Proof The estimate is clear when $m = 0$. To prove that it holds for $m \geq 1$, write $\partial\Lambda := \{z \notin \Lambda \colon \operatorname{dist}_G(z, \Lambda) = 1\}$ for $\Lambda \subset V$. Then $|\partial\mathcal{C} \cup \mathcal{C}| \leq (d_{\max} + 1)|\mathcal{C}| \leq 2d_{\max} C_{r,A}$. We define a map $\Phi \colon \widetilde{\mathscr{P}}_x^{(m,s)} \to \mathscr{P}_s(x, V) \times \{1, \ldots, 2d_{\max} C_{r,A}\}^m$ as follows. For each $\Lambda \subset V$ with $1 \leq |\Lambda| \leq 2d_{\max} C_{r,A}$, fix an injective function $f_\Lambda \colon \Lambda \to \{1, \ldots, 2d_{\max} C_{r,A}\}$. Given a path $\pi \in \mathscr{P}^{(m,s)} \cap \mathscr{P}(x, V)$, decompose π as in (3.14), and denote by $\widetilde{\pi} \in \mathscr{P}_s(x, V)$ the shifted concatenation (cf. Definition 3.4) of $\check{\pi}^{(1)}, \ldots, \check{\pi}^{(m)}, \bar{\pi}$. Note that, for $2 \leq k \leq m$, the point $\check{\pi}_0^{(k)}$ lies in $\partial\mathcal{C}_k$ for some $\mathcal{C}_k \in \mathfrak{C}_{r,A}$, while $\bar{\pi}_0 \in \partial\overline{\mathcal{C}} \cup \overline{\mathcal{C}}$ for some $\overline{\mathcal{C}} \in \mathfrak{C}_{r,A}$. Thus, we may set

$$\Phi(\pi) := \left(\widetilde{\pi}, f_{\partial\mathcal{C}_2}(\check{\pi}_0^{(2)}), \ldots, f_{\partial\mathcal{C}_m}(\check{\pi}_0^{(m)}), f_{\partial\bar{\mathcal{C}} \cup \bar{\mathcal{C}}}(\bar{\pi}_0)\right). \tag{3.30}$$

As is readily checked, $\Phi(\pi)$ depends only on the equivalence class of π and, when restricted to equivalence classes, Φ is injective. Hence the claim follows. □

Now take $\mathcal{N} \subset \mathscr{P}(x, V)$ as in the statement, and set

$$\widetilde{\mathcal{N}}^{(m,s)} := \left\{\text{equivalence classes of paths in } \mathcal{N} \cap \mathscr{P}^{(m,s)}\right\} \subset \widetilde{\mathscr{P}}_x^{(m,s)}. \tag{3.31}$$

For each $\mathcal{M} \in \widetilde{\mathcal{N}}^{(m,s)}$, choose a representative $\pi_{\mathcal{M}} \in \mathcal{M}$, and use Lemma 3.6 to write

$$\begin{aligned}\mathbb{E}_x\left[\mathrm{e}^{\int_0^t \xi(X_u)\mathrm{d}u}\,\mathbb{1}_{\{\pi(X_{[0,t]})\in\mathcal{N}\}}\right] &= \textstyle\sum_{m,s\in\mathbb{N}_0}\sum_{\mathcal{M}\in\widetilde{\mathcal{N}}^{(m,s)}} \mathbb{E}_x\left[\mathrm{e}^{\int_0^t \xi(X_u)\mathrm{d}u}\,\mathbb{1}_{\{\pi(X_{[0,t]})\sim\pi_{\mathcal{M}}\}}\right]\\ &\le \textstyle\sum_{m,s\in\mathbb{N}_0}(2d_{\max}C_{r,A})^m d_{\max}^s \sup_{\pi\in\mathcal{N}^{(m,s)}} \mathbb{E}_x\left[\mathrm{e}^{\int_0^t \xi(X_u)\mathrm{d}u}\,\mathbb{1}_{\{\pi(X_{[0,t]})\sim\pi\}}\right],\end{aligned} \tag{3.32}$$

where we use the convention $\sup\emptyset = 0$. For fixed $\pi \in \mathcal{N}^{(m,s)}$, by (3.6), we may apply (3.20) and Lemma 2.5 to obtain, for all r large enough and with c_0 as in (3.28),

$$(2d_{\max})^m d_{\max}^s \mathbb{E}_x\left[\mathrm{e}^{\int_0^t \xi(X_u)\mathrm{d}u}\,\mathbb{1}_{\{\pi(X_{[0,t]})\sim\pi\}}\right] \le \mathrm{e}^{t\gamma_\pi}\,\mathrm{e}^{c_0 m S_r}\, q_A^s\, \mathrm{e}^{\left(c-\log^{(3)} L_r\right)k_\pi^{r,\varepsilon}}. \tag{3.33}$$

We next claim that, for r large enough and $\pi \in \mathcal{N}^{(m,s)}$,

$$s \ge [(m-1)\vee 1]\,S_r. \tag{3.34}$$

Indeed, when $m \ge 2$, $|\mathrm{supp}(\check{\pi}^{(i)})| \ge S_r$ for all $2 \le i \le m$. When $m = 0$, $|\mathrm{supp}(\pi)| \ge \max_{1\le\ell\le|\pi|}|\pi_\ell - x| \ge (\log L_r)^\kappa \gg S_r$ by assumption. When $m = 1$, the latter assumption and Lemma 2.5 together imply that $\mathrm{supp}(\pi) \cap D^{\mathrm{c}}_{r,A} \ne \emptyset$, and so either $|\mathrm{supp}(\check{\pi}^{(1)})| \ge S_r$ or $|\mathrm{supp}(\check{\pi}^{(1)})| \ge S_r$. Thus, (3.34) holds by (3.17) and (2.27).

Note that $q_A < \mathrm{e}^{-3c_0}$, so

$$\sum_{m\ge 0}\sum_{s\ge[(m-1)\vee 1]S_r} \mathrm{e}^{c_0 m S_r} q_A^s = \frac{q_A^{S_r} + \mathrm{e}^{c_0 S_r} q_A^{S_r} + \sum_{m\ge 2}\mathrm{e}^{c_0 S_r m} q_A^{(m-1)S_r}}{1-q_A} \le \frac{4\mathrm{e}^{-c_0 S_r}}{1-q_A} < 1 \tag{3.35}$$

for r large enough. Inserting this back into (3.32), we obtain

$$\log \mathbb{E}_x\left[\mathrm{e}^{\int_0^t \xi(X_s)\mathrm{d}s}\,\mathbb{1}_{\{\pi(X_{0,t})\in\mathcal{N}\}}\right] \le \sup_{\pi\in\mathcal{N}}\left\{t\gamma_\pi + \left(c - \log^{(3)} L_r\right) k_\pi^{r,\varepsilon}\right\}. \tag{3.36}$$

Thus the proof will be finished once we show that, for some $\varepsilon' > 0$, whp (respectively, almost surely eventually) as $n \to \infty$, all $\pi \in \mathcal{N}$ satisfy

$$k_\pi^{r,\varepsilon} \ge \mathrm{dist}_G(x, z_\pi)(1 - 2(\log L_r)^{-\varepsilon'}). \tag{3.37}$$

To that end, we define for each $\pi \in \mathcal{N}$ an auxiliary path $\pi_\star$ as follows. First note that by using our assumptions we can find points $z', z'' \in \mathrm{supp}(\pi)$ (not necessarily distinct) such that

$$\mathrm{dist}_G(x, z') \geq (\log L_r)^\kappa, \qquad \mathrm{dist}_G(z'', z_\pi) \leq 2M_A S_r, \tag{3.38}$$

where the latter holds by Lemma 2.5. Write $\{z_1, z_2\} = \{z', z''\}$ with z_1, z_2 ordered according to their hitting times by π, i.e., $\inf\{\ell : \pi_\ell = z_1\} \leq \inf\{\ell : \pi_\ell = z_2\}$. Define π_e as the concatenation of the loop erasure of π between x and z_1 and the loop erasure of π between z_1 and z_2. Since π_e is the concatenation of two self-avoiding paths, it visits each point at most twice. Finally, define $\pi_\star \sim \pi_e$ by substituting the excursions of π_e from $\Pi_{r,A}$ to $D^{\mathrm{c}}_{r,A}$ by direct paths between the corresponding endpoints, i.e., substitute each $\hat{\pi}^{(i)}_e$ with $|\hat{\pi}^{(i)}_e| = \ell_i$, $(\hat{\pi}^{(i)}_e)_0 = x_i \in \Pi_{r,A}$ and $(\hat{\pi}^{(i)}_e)_{\ell_i} = y_i \in D^{\mathrm{c}}_{r,A}$ by a shortest-distance path $\widetilde{\pi}^{(i)}_\star$ with the same endpoints and $|\widetilde{\pi}^{(i)}_\star| = \mathrm{dist}_G(x_i, y_i)$. Since $\pi_\star$ visits each $x \in \Pi_{r,A}$ at most 2 times,

$$k^{r,\varepsilon}_\pi \geq k^{r,\varepsilon}_{\pi_\star} \geq M^{r,\varepsilon}_{\pi_\star} - 2|\mathrm{supp}(\pi_\star) \cap \Pi_{r,A}|(S_r + 1) \geq M^{r,\varepsilon}_{\pi_\star} - 4|\mathrm{supp}(\pi_\star) \cap \Pi_{r,A}|S_r. \tag{3.39}$$

Note that $M^{r,\varepsilon}_{\pi_\star} \geq \big|\{x \in \mathrm{supp}(\pi_\star) : \xi(x) \leq (1-\varepsilon)a_{L_r}\}\big| - 1$ and, by (3.38), $|\mathrm{supp}(\pi_\star)| \geq \mathrm{dist}_G(x, z') \geq (\log L_r)^\kappa \gg (\log L_r)^{\alpha+2\varepsilon'}$ for some $0 < \varepsilon' < \varepsilon$. Applying Lemmas 2.8–2.9 and using (2.27) and $L_r > r$, we obtain, for r large enough,

$$k^{r,\varepsilon}_\pi \geq |\mathrm{supp}(\pi_\star)| \left(1 - \tfrac{2}{(\log L_r)^\varepsilon} - \tfrac{4S_r}{(\log L_r)^{\alpha+2\varepsilon'}}\right) \geq |\mathrm{supp}(\pi_\star)| \left(1 - \tfrac{1}{(\log L_r)^{\varepsilon'}}\right). \tag{3.40}$$

On the other hand, since $|\mathrm{supp}(\pi_\star)| \geq (\log L_r)^\kappa$ and by (3.38) again,

$$\begin{aligned}|\mathrm{supp}(\pi_\star)| &= \big(|\mathrm{supp}(\pi_\star)| + 2M_A S_r\big) - 2M_A S_r \\ &\geq \big(\mathrm{dist}_G(x, z'') + 2M_A S_r\big)\left(1 - \tfrac{2M_A S_r}{(\log L_r)^\kappa}\right) \\ &\geq \mathrm{dist}_G(x, z_\pi)\left(1 - \tfrac{1}{(\log L_r)^{\varepsilon'}}\right).\end{aligned} \tag{3.41}$$

Now (3.37) follows from (3.40)–(3.41). □

4 Proof of Theorem 1.1

This section is devoted to the proof of Theorem 1.1. We note that, after replacing $d_{\max}$ by $d_{\max} \vee D_0$ if necessary, we may assume without loss of generality that

$$\mathcal{GW} \in \mathfrak{G}^{(\vartheta)}_\infty. \tag{4.1}$$

4.1 Lower Bound

In this section we give the proof of the lower bound for the large-t asymptotics of the total mass. This proof already explains the random mechanism that produces the main contribution to the total mass. This mechanism comes from an *optimisation* of the behaviour of the random path in the Feynman-Kac formula, which in turn comes from the existence of a *favorite region* in the random graph, both in terms of the local graph structure and the high values of the potential in this local graph structure. The optimality is expressed in terms of a distance to the starting point $\mathcal{O}$ that can be reached in a time $o(t)$ with a sufficiently high probability, such that time $t - o(t)$ is left for staying inside the favorite region, thus yielding a maximal contribution to the Feynman-Kac formula. The latter is measured in terms of the *local eigenvalue* of the Anderson operator $\Delta + \xi$, which in turn comes from high values and optimal shape of the potential ξ in the local region.

We write the total mass of the solution of (2.9) in terms of the Feynman-Kac formula as

$$U(t) = \mathbb{E}_{\mathcal{O}}\Big[\exp\Big\{\int_0^t \xi(X_s)\,\mathrm{d}s\Big\}\Big], \tag{4.2}$$

where $(X_s)_{s\geq 0}$ is the continuous-time random walk on $\mathcal{GW}$, i.e., the Markov chain with generator $\Delta_{\mathcal{GW}} = \Delta$, the Laplacian on $\mathcal{GW}$, starting from the origin $\mathcal{O}$. As usual in the literature of the PAM, this formula is the main point of departure for our proof.

Fix $\varepsilon > 0$. By the definition of $\widetilde{\chi}$, there exists an infinite rooted tree $T = (V', E', \mathcal{Y})$ with degrees in $\mathrm{supp}(D_g)$ such that $\chi_T(\varrho) < \widetilde{\chi}(\varrho) + \frac{1}{4}\varepsilon$. Let $Q_r = B_r^T(\mathcal{Y})$ be the ball of radius r around $\mathcal{Y}$ in T. By Proposition 2.3 and (2.16), there exist a radius $R \in \mathbb{N}$ and a potential profile $q\colon B_R^T \to \mathbb{R}$ with $\mathcal{L}_{Q_R}(q;\varrho) < 1$ (in particular, $q \leq 0$) such that

$$\lambda_{Q_R}(q;T) \geq -\widehat{\chi}_{Q_R}(\varrho;T) - \frac{1}{2}\varepsilon > -\widetilde{\chi}(\varrho) - \varepsilon. \tag{4.3}$$

For $\ell \in \mathbb{N}$, let $B_\ell = B_\ell(\mathcal{O})$ denote the ball of radius ℓ around $\mathcal{O}$ in $\mathcal{GW}$. We will show next that, almost surely eventually as $\ell \to \infty$, B_ℓ contains a copy of the ball Q_R where ξ is lower bounded by $\varrho \log\log|B_\ell| + q$.

Proposition 4.1 (Balls with High Exceedances) $\mathfrak{P} \times \mathrm{P}$-*almost surely eventually as* $\ell \to \infty$, *there exists a vertex* $z \in B_\ell$ *with* $B_{R+1}(z) \subset B_\ell$ *and an isomorphism* $\varphi : B_{R+1}(z) \to Q_{R+1}$ *such that* $\xi \geq \varrho \log\log|B_\ell| + q \circ \varphi$ *in* $B_R(z)$. *In particular,*

$$\lambda_{B_R(z)}(\xi;\mathcal{GW}) > \varrho \log\log|B_\ell| - \widetilde{\chi}(\varrho) - \varepsilon. \tag{4.4}$$

Any such z *necessarily satisfies* $|z| \geq c\ell$ $\mathfrak{P} \times \mathrm{P}$-*almost surely eventually as* $\ell \to \infty$ *for some constant* $c = c(\varrho, \vartheta, \widetilde{\chi}(\varrho), \varepsilon) > 0$.

Proof First note that, as a consequence of the definition of $\mathcal{GW}$, it may be shown straightforwardly that, for some $p = p(T, R) \in (0, 1)$ and $\mathfrak{P}$-almost surely eventually as $\ell \to \infty$, there exist $N \in \mathbb{N}$, $N \geq p|B_\ell|$ and distinct $z_1, \dots, z_N \in B_\ell$ such that $B_{R+1}(z_i) \cap B_{R+1}(z_j) = \emptyset$ for $1 \leq i \neq j \leq N$ and, for each $1 \leq i \leq N$, $B_{R+1}(z_i) \subset B_\ell$ and $B_{R+1}(z_i)$ is isomorphic to Q_{R+1}. Now, by (1.7), for each $i \in \{1, \dots, N\}$,

$$\mathrm{P}\big(\xi \geq \varrho \log\log |B_\ell| + q \text{ in } B_R(z_i)\big) = |B_\ell|^{-\mathcal{L}_{Q_R}(q)}. \tag{4.5}$$

Using additionally that $|B_\ell| \geq \ell$ and $1 - x \leq \mathrm{e}^{-x}$, $x \in \mathbb{R}$, we obtain

$$\begin{aligned} &\mathrm{P}(\nexists i \in \{1, \dots, N\} \colon \xi \geq \varrho \log\log |B_\ell| + q \text{ in } B_R(z_i)) \\ &= \Big(1 - |B_\ell|^{-\mathcal{L}_{Q_R}(q)}\Big)^N \leq \mathrm{e}^{-p\ell^{1-\mathcal{L}_{Q_R}(q)}}, \end{aligned} \tag{4.6}$$

which is summable in $\ell \in \mathbb{N}$, so the proof of the first statement is completed using the Borel-Cantelli lemma. As for the last statement, note that, by (2.8), Lemma 2.4 and $L_r \sim \vartheta r$,

$$\lambda_{B_{c\ell}}(\xi; \mathcal{GW}) \leq \max_{x \in B_{c\ell}} \xi(x) < a_{L_{c\ell}} + o(1) < a_{L_\ell} + \varrho \log c\vartheta + o(1) < a_{L_\ell} - \widetilde{\chi}(\varrho) - \varepsilon \tag{4.7}$$

provided $c > 0$ is small enough. □

Proof (Of the Lower Bound in (1.14)) Let z be as in Proposition 4.1. Write τ_z for the hitting time of z by the random walk X. For any $s \in (0, t)$, we obtain a lower bound for $U(t)$ as follows:

$$\begin{aligned} U(t) &\geq \mathbb{E}_{\mathcal{O}}\Big[\exp\Big\{\int_0^t \xi(X_u)\,\mathrm{d}u\Big\} \mathbb{1}_{\{\tau_z \leq s\}} \mathbb{1}_{\{X_u \in B_R(z)\, \forall u \in [\tau_z, t]\}}\Big] \\ &= \mathbb{E}_{\mathcal{O}}\Big[\mathrm{e}^{\int_0^{\tau_z} \xi(X_u)\,\mathrm{d}u} \mathbb{1}_{\{\tau_z \leq s\}} \mathbb{E}_z\Big[\mathrm{e}^{\int_0^v \xi(X_u)\,\mathrm{d}u} \mathbb{1}_{\{X_u \in T\, \forall u \in [0,v]\}}\Big]\Big|_{v = t - \tau_z}\Big], \end{aligned} \tag{4.8}$$

where we use the strong Markov property at time τ_z. We first bound the last term in the integrand in (4.8). Since $\xi \geq \varrho \log\log |B_\ell| + q$ in $B_R(z)$,

$$\begin{aligned} \mathbb{E}_z\Big[\mathrm{e}^{\int_0^v \xi(X_u)\,\mathrm{d}u} \mathbb{1}_{\{X_u \in B_R(z)\, \forall u \in [0,v]\}}\Big] &\geq \mathrm{e}^{v\varrho \log\log |B_\ell|} \mathbb{E}_{\mathcal{Y}}\Big[\mathrm{e}^{\int_0^v q(X_u)\,\mathrm{d}u} \mathbb{1}_{\{X_u \in Q_R\, \forall u \in [0,v]\}}\Big] \\ &\geq \mathrm{e}^{v\varrho \log\log |B_\ell|} \mathrm{e}^{v\lambda_{Q_R}(q;T)} \phi^{(1)}_{Q_R}(\mathcal{Y})^2 \\ &> \exp\big\{v\,(\varrho \log\log |B_\ell| - \widetilde{\chi}(\varrho) - \varepsilon)\big\}, \end{aligned} \tag{4.9}$$

for large v, where we used that $B_{R+1}(z)$ is isomorphic to Q_{R+1} and we applied Lemma 2.1 and (4.3). On the other hand, since $\xi \geq 0$,

$$\mathbb{E}_{\mathcal{O}}\Big[\exp\Big\{\int_0^{\tau_z} \xi(X_u)\,\mathrm{d}u\Big\}\mathbb{1}\{\tau_z \leq s\}\Big] \geq \mathbb{P}_{\mathcal{O}}(\tau_z \leq s), \tag{4.10}$$

and we can bound the latter probability from below by the probability that the random walk runs along a shortest path from the root $\mathcal{O}$ to z within a time at most s. Such a path $(y_i)_{i=0}^{|z|}$ has $y_0 = \mathcal{O}$, $y_{|z|} = z$, $y_i \sim y_{i-1}$ for $i = 1, \dots, |z|$, has at each step from y_i precisely $\deg(y_i)$ choices for the next step with equal probability, and the step is carried out after an exponential time E_i with parameter $\deg(y_i)$. This gives

$$\mathrm{P}_{\mathcal{O}}(\tau_z \leq s) \geq \Big(\prod_{i=1}^{|z|} \frac{1}{\deg(y_i)}\Big) P\Big(\sum_{i=1}^{|z|} E_i \leq s\Big) \geq d_{\max}^{-|z|}\mathrm{Poi}_{d_{\min}s}([|z|,\infty)), \tag{4.11}$$

where Poi_γ is the Poisson distribution with parameter γ, and P is the generic symbol for probability. Summarising, we obtain

$$\begin{aligned} U(t) &\geq d_{\max}^{-|z|}\mathrm{e}^{-d_{\min}s}\,\tfrac{(d_{\min}s)^{|z|}}{|z|!}\,\mathrm{e}^{(t-s)\left[\varrho\log\log|B_\ell| - \widetilde{\chi}(\varrho) - \varepsilon\right]} \\ &\geq \exp\Big\{-d_{\min}s + (t-s)\left[\varrho\log\log|B_\ell| - \widetilde{\chi}(\varrho) - \varepsilon\right] - |z|\log\Big(\tfrac{d_{\max}}{d_{\min}}\tfrac{|z|}{s}\Big)\Big\} \\ &\geq \exp\Big\{-d_{\min}s + (t-s)\left[\varrho\log\log|B_\ell| - \widetilde{\chi}(\varrho) - \varepsilon\right] - \ell\log\Big(\tfrac{d_{\max}}{d_{\min}}\tfrac{\ell}{s}\Big)\Big\}, \end{aligned} \tag{4.12}$$

where for the last inequality we assume $s \leq |z|$ and use $\ell \geq |z|$. Further assuming that $\ell = o(t)$, we see that the optimum over s is obtained at

$$s = \frac{\ell}{d_{\min} + \varrho\log\log|B_\ell| - \widetilde{\chi}(\varrho) - \varepsilon} = o(t). \tag{4.13}$$

Note that, by Proposition 4.1, this s indeed satisfies $s \leq |z|$. Applying (1.12) we get, after a straightforward computation, almost surely eventually as $t \to \infty$,

$$\frac{1}{t}\log U(t) \geq \varrho\log\log|B_\ell| - \frac{\ell}{t}\log\log\ell - \widetilde{\chi}(\varrho) - \varepsilon + O\left(\frac{\ell}{t}\right). \tag{4.14}$$

Analysing the main terms above and using $\log|B_\ell| \sim \vartheta\ell$, we find that the optimal ℓ satisfies $\ell\log\log\ell - \frac{\ell}{\log\ell} \sim t\varrho$, i.e., $\ell \sim \varrho t/\log\log t = \mathfrak{r}_t$. For this choice we obtain

$$\frac{1}{t}\log U(t) \geq \varrho\log\log|B_{\mathfrak{r}_t}| - \mathfrak{r}_t\log\log\mathfrak{r}_t - \widetilde{\chi}(\varrho) - \varepsilon + O\left(\frac{1}{\log\log t}\right). \tag{4.15}$$

Substituting $\log|B_r| \sim \vartheta r$ and the definition of $\mathfrak{r}_t$, we obtain, $\mathfrak{P} \times \mathrm{P}$-almost surely,

$$\liminf_{t\to\infty} \left\{ \frac{1}{t} \log U(t) - \varrho \log\left(\frac{\varrho\vartheta t}{\log\log t} \right) \right\} \geq -\varrho - \widetilde{\chi}(\varrho) - \varepsilon. \tag{4.16}$$

Since $\varepsilon > 0$ is arbitrary, the proof of the lower bound in (1.14) is complete. □

4.2 Upper Bound

In this section we prove the upper bound in (1.14). A first step is to reduce the problem to a ball of radius $t \log t$. Here we include more general graphs.

Lemma 4.2 (Spatial Truncation) *For any $c > 0$ and any $\ell_t \in \mathbb{N}$, $\ell_t \geq ct \log t$,*

$$\sup_{G\in\mathfrak{G}_{\ell_t}} \mathbb{E}_{\mathcal{O}}\left[\mathrm{e}^{\int_0^t \xi(X_s)\mathrm{d}s} \mathbb{1}_{\{\tau_{B^{\mathrm{c}}_{\ell_t}} < t\}} \right] \leq \mathrm{e}^{-\ell_t} \quad \textit{whp as } t\to\infty. \tag{4.17}$$

Moreover, for any $G \in \mathfrak{G}^{(\vartheta)}_\infty$,

$$\mathbb{E}_{\mathcal{O}}\left[\mathrm{e}^{\int_0^t \xi(X_s)\mathrm{d}s} \mathbb{1}_{\{\tau_{B^{\mathrm{c}}_{\ell_t}} < t\}} \right] \leq \mathrm{e}^{-\ell_t} \quad \mathrm{P}\textit{-a.s. eventually as } t\to\infty. \tag{4.18}$$

Proof For $r \geq \ell_t$ and $G \in \mathfrak{G}_{\ell_t}$, let

$$\mathcal{B}_r := \left\{ \max_{x\in B_r} \xi(x) \geq a_{L_r} + 2\varrho \right\}. \tag{4.19}$$

By Lemma 2.4 and a union bound, we see that

$$\sup_{G\in\mathfrak{G}_{\ell_t}} \mathrm{P}\left(\bigcup_{r\geq\ell_t} \mathcal{B}_r \right) \leq \sum_{r\geq\ell_t} \sup_{G\in\mathfrak{G}_{\ell_t}} \mathrm{P}(\mathcal{B}_r) \to 0, \qquad t\to\infty, \tag{4.20}$$

while, for $G \in \mathfrak{G}^{(\vartheta)}_\infty$, by the Borel-Cantelli lemma,

$$\bigcup_{r\geq\ell_t} \mathcal{B}_r \text{ does not occur P-a.s. eventually as } t\to\infty. \tag{4.21}$$

We may therefore work on the event $\bigcap_{r\geq\ell_t} \mathcal{B}^{\mathrm{c}}_r$. On this event, we may write

$$\begin{aligned} \mathbb{E}_{\mathcal{O}}\left[\mathrm{e}^{\int_0^t \xi(X_s)\mathrm{d}s} \mathbb{1}_{\{\tau_{B^{\mathrm{c}}_{\ell_t}} < t\}} \right] &= \textstyle\sum_{r\geq\ell_t} \mathbb{E}_{\mathcal{O}}\left[\mathrm{e}^{\int_0^t \xi(X_s)\mathrm{d}s} \mathbb{1}_{\{\sup_{s\in[0,t]}|X_s|=r\}} \right] \\ &\leq \mathrm{e}^{Ct} \textstyle\sum_{r\geq\ell_t} \mathrm{e}^{\varrho t \log r}\, \mathbb{P}_{\mathcal{O}}(J_t \geq r), \end{aligned} \tag{4.22}$$

where J_t is the number of jumps of X up to time t, $C = \varrho(2 + \log\log d_{\max})$, and we use that $|B_r| \le d_{\max}^r$. Note that J_t is stochastically dominated by a Poisson random variable with parameter $td_{\max}$. Hence

$$\mathbb{P}_{\mathcal{O}}(J_t \ge r) \le \frac{(td_{\max})^r}{r!} \le \exp\left\{-r\log\left(\frac{r}{\mathrm{e}td_{\max}}\right)\right\} \tag{4.23}$$

for large r. Using $\ell_t \ge ct\log t$, we can check that, for $r \ge \ell_t$ and t large enough,

$$r\log\left(\frac{r}{\mathrm{e}td_{\max}}\right) - \varrho t\log r > 2r \tag{4.24}$$

and thus (4.22) is at most $\mathrm{e}^{-\ell_t}\mathrm{e}^{-\ell_t+Ct+2} < \mathrm{e}^{-\ell_t}$. □

In order to be able to apply Proposition 3.1 in the following, we need to make sure that all paths considered exit a ball with a slowly growing radius.

Lemma 4.3 (No Short Paths) *For any $\gamma \in (0,1)$,*

$$\sup_{G\in\mathfrak{G}_{\lceil t^\gamma\rceil}} \frac{\mathbb{E}_{\mathcal{O}}\left[\mathrm{e}^{\int_0^t \xi(X_s)\mathrm{d}s}\mathbb{1}_{\{\tau_{B^{\mathrm{c}}_{\lceil t^\gamma\rceil}}>t\}}\right]}{U(t)} = o(1) \quad \textit{whp as } t\to\infty. \tag{4.25}$$

Moreover, for any $G \in \mathfrak{G}_\infty$,

$$\lim_{t\to\infty} \frac{\mathbb{E}_{\mathcal{O}}\left[\mathrm{e}^{\int_0^t \xi(X_s)\mathrm{d}s}\mathbb{1}_{\{\tau_{B^{\mathrm{c}}_{\lceil t^\gamma\rceil}}>t\}}\right]}{U(t)} = 0 \quad \mathrm{P}\text{-}\textit{a.s. almost surely.} \tag{4.26}$$

Proof By Lemma 2.4 with $g_r = 2\varrho\log r$, we may assume that

$$\max_{x\in B_{\lceil t^\gamma\rceil}} \xi(x) \le \varrho\log\log L_{\lceil t^\gamma\rceil} + 2\varrho = \gamma\varrho\log t + 2\varrho + o(1) \text{ as } t\to\infty. \tag{4.27}$$

By (4.16), for some constant $C > 0$,

$$\frac{\mathbb{E}_{\mathcal{O}}\left[\mathrm{e}^{\int_0^t \xi(X_s)\mathrm{d}s}\mathbb{1}_{\{\tau_{B^{\mathrm{c}}_{\lceil t^\gamma\rceil}}>t\}}\right]}{U(t)} \le \mathrm{e}^{Ct\log^{(3)}t}\mathrm{e}^{-(1-\gamma)\varrho t\log t} \to 0, \qquad t\to\infty. \tag{4.28}$$

□

For the remainder of the proof we fix $\gamma \in (\alpha, 1)$ with α as in (2.27). Let

$$K_t := \lceil t^{1-\gamma}\log t\rceil, \qquad r_t^{(k)} := k\lceil t^\gamma\rceil,\ 1\le k\le K_t \qquad \text{and} \qquad \ell_t := K_t\lceil t^\gamma\rceil \ge t\log t. \tag{4.29}$$

For $1 \le k \le K_t$ and $G \in \mathfrak{G}_\infty^{(\vartheta)}$, define

$$\mathcal{N}_t^{(k)} := \left\{ \pi \in \mathscr{P}(\mathcal{O}, V) \colon \ \mathrm{supp}(\pi) \subset B_{r_t^{(k+1)}}, \mathrm{supp}(\pi) \cap B^{\mathrm{c}}_{r_t^{(k)}} \neq \emptyset \right\} \tag{4.30}$$

and set

$$U_t^{(k)} := \mathbb{E}_{\mathcal{O}} \left[\mathrm{e}^{\int_0^t \xi(X_s) \mathrm{d}s} 1\!\!1_{\{\pi_{[0,t]}(X) \in \mathcal{N}_t^{(k)}\}} \right]. \tag{4.31}$$

Recall the scale $\mathfrak{r}_t = \varrho t / \log\log t$.

Lemma 4.4 (Upper Bound on $U_t^{(k)}$) *For any $\varepsilon > 0$ and any $G \in \mathfrak{G}_\infty^{(\vartheta)}$, P-almost surely eventually as $t \to \infty$,*

$$\sup_{1 \le k \le K_t} \frac{1}{t} \log U_t^{(k)} \le \varrho \log(\vartheta \mathfrak{r}_t) - \varrho - \widetilde{\chi}(\varrho) + \varepsilon. \tag{4.32}$$

Proof Before we apply Proposition 3.1, we first do a bit of analysis. For $c > 0$, let

$$F_{c,t}(r) := \varrho \log(\vartheta r) - \frac{r}{t} \left(\log\log r - c \right), \qquad r > 0. \tag{4.33}$$

Note that $F_{c,t}$ is maximised at a point $r_{c,t}$ satisfying

$$\varrho t = r_{c,t} \log\log r_{c,t} - c r_{c,t} + \frac{r_{c,t}}{\log r_{c,t}}. \tag{4.34}$$

In particular, $r_{c,t} \sim \mathfrak{r}_t$, which implies

$$\sup_{r > 0} F_{c,t}(r) \le \varrho \log(\vartheta \mathfrak{r}_t) - \varrho + o(1) \qquad \text{as } t \to \infty. \tag{4.35}$$

Next, fix $k \in \{1, \dots, K_t\}$. For $\pi \in \mathcal{N}_t^{(k)}$, let

$$\gamma_\pi := \lambda_{r_t^{(k+1)}, A}(\pi) + \exp\{-S_{\lceil t^\gamma \rceil}\}, \qquad z_\pi \in \mathrm{supp}(\pi), |z_\pi| > r_t^{(k)}. \tag{4.36}$$

By Proposition 3.1, almost surely eventually as $t \to \infty$,

$$\frac{1}{t} \log U_t^{(k)} \le \gamma_\pi + \frac{|z_\pi|}{t} \left(\log\log r_t^{(k+1)} - c_A + o(1) \right). \tag{4.37}$$

Using Corollary 2.7 and $\log L_r \sim \vartheta r$, we bound

$$\gamma_\pi \le \varrho \log(\vartheta r_t^{(k+1)}) - \widetilde{\chi}(\varrho) + \frac{1}{2} \varepsilon + o(1). \tag{4.38}$$

Moreover, $|z_\pi| > r_t^{(k+1)} - \lceil t^\gamma \rceil$ and

$$\frac{\lceil t^\gamma \rceil}{t}\left(\log\log r_t^{(k+1)} - c_A\right) \le \frac{2}{t^{1-\gamma}} \log\log(2t\log t) = o(1), \tag{4.39}$$

which allows us to further bound (4.37) by

$$\varrho \log(\vartheta r_t^{(k+1)}) - \frac{r_t^{(k+1)}}{t}\left(\log\log r_t^{(k+1)} - 2c_A\right) - \widetilde{\chi}(\varrho) + \frac{1}{2}\varepsilon + o(1). \tag{4.40}$$

Applying (4.35) we obtain $\frac{1}{t}\log U_t^{(k)} < \varrho\log(\vartheta \mathfrak{r}_t) - \varrho - \widetilde{\chi}(\varrho) + \varepsilon$. □

Proof (Of Upper Bound in (1.14)) To avoid repetition, all statements are assumed to be made $\mathfrak{P} \times$ P-almost surely eventually as $t \to \infty$. Let $G = \mathcal{GW}$ and note that $\mathcal{GW} \in \mathfrak{G}_\infty^{(\vartheta)}$ almost surely, where ϑ is as in (1.12). Define

$$U_t^{(0)} := \mathbb{E}_\mathcal{O}\left[\mathrm{e}^{\int_0^t \xi(X_s)\mathrm{d}s}\, 1\!\!1_{\{\tau_{B^c_{\lceil t^\gamma\rceil}} > t\}}\right], \qquad U_t^{(\infty)} := \mathbb{E}_\mathcal{O}\left[\mathrm{e}^{\int_0^t \xi(X_s)\mathrm{d}s}\, 1\!\!1_{\{\tau_{B^c_{\lceil t\log t\rceil}} \le t\}}\right]. \tag{4.41}$$

Note that

$$U(t) \le U_t^{(0)} + U_t^{(\infty)} + K_t \max_{1\le k\le K_t} U_t^{(k)} \tag{4.42}$$

and, since $U_t^{(0)} + U_t^{(\infty)} \le o(1)U(t)$ by Lemmas 4.2–4.3 and (4.15),

$$U(t) \le 2K_t \max_{1\le k\le K_t} U_t^{(k)} \quad \text{and so} \quad \frac{1}{t}\log U(t) \le \frac{\log(2K_t)}{t} + \max_{1\le k\le K_t} \frac{1}{t}\log U_t^{(k)}. \tag{4.43}$$

By Lemma 4.4 and (4.29), for any $\varepsilon > 0$,

$$\tfrac{1}{t}\log U(t) \le \varrho\log(\vartheta\mathfrak{r}_t) - \varrho - \widetilde{\chi}(\varrho) + \varepsilon + o(1) \tag{4.44}$$

therefore, $\mathfrak{P} \times$ P-almost surely,

$$\limsup_{t\to\infty}\left\{\frac{1}{t}\log U(t) - \varrho\log\left(\frac{\vartheta\varrho t}{\log\log t}\right)\right\} \le -\varrho - \widetilde{\chi}(\varrho) + \varepsilon. \tag{4.45}$$

Since $\varepsilon > 0$ is arbitrary, this completes the proof of the lower bound in (1.14). □

5 Proof of Theorem 1.5

In this section we give the proof of Theorem 1.5. The proof is based on the fact that, up to a radius growing slower than $\log \Phi_n$ (cf. (1.20)), the configuration model equals a Galton-Watson tree with high probability. From this the result will follow via Theorem 1.1 and Lemma 4.2.

To describe the associated Galton-Watson tree, we define a random variable $D_\star$ as the size-biased version of D in Assumption (CM)(1), i.e.,

$$P(D_\star = k) = \frac{kP(D=k)}{E[D]}. \tag{5.1}$$

Proposition 5.1 (Coupling of $\mathcal{UG}_n$ and $\mathcal{GW}$) *Let $\mathcal{UG}_n = (V_n, E_n, \mathcal{O}_n)$ be the uniform simple random graph with degree sequence $\mathfrak{d}^{(n)}$ satisfying Assumption (CM), and let $\mathcal{GW} = (V, E, \mathcal{O})$ be a Galton-Watson tree with initial degree distribution $D_0 = D$ and general degree distribution $D_g = D_\star$. There exists a coupling $\widetilde{\mathrm{P}}$ of $\mathcal{UG}_n$ and $\mathcal{GW}$ such that, for any $m_n \in \mathbb{N}$ satisfying $1 \ll m_n \ll \log \Phi_n$,*

$$\lim_{n\to\infty} \widetilde{\mathrm{P}}\left(B^{\mathcal{UG}_n}_{m_n}(\mathcal{O}_n) = B^{\mathcal{GW}}_{m_n}(\mathcal{O})\right) = 1. \tag{5.2}$$

Proof For $\mathcal{CM}_n$ in place of $\mathcal{UG}_n$, this is a consequence of the proof of [16, Proposition 5.4]: the statement there only covers coupling $|B_{m_n}|$, but the proof actually gives B_{m_n}. The fact that m_n may be taken up to $o(\log \Phi_n)$ can be inferred from the proof. In fact, m_n could be taken up to $c \log \Phi_n$ with some $c = c(\nu) > 0$. The result is then passed to $\mathcal{UG}_n$ by (1.16) (see e.g. [15, Corollary 7.17]). □

Proof (Of Theorem 1.5) First note that, by Propositions 1.3–1.4, we may assume that $\mathcal{UG}_n$ is connected, thus fitting the setup of Sect. 2. Let $U_n(t)$ be the total mass for $\mathcal{UG}_n$ and $U(t)$ the total mass for $\mathcal{GW}$ as in Proposition 5.1. Define

$$U_n^\circ(t) := \mathbb{E}_{\mathcal{O}_n}\left[\mathrm{e}^{\int_0^t \xi(X_s)\mathrm{d}s} \mathbb{1}_{\{\tau_{B^c_{t\log t}} > t\}} \right], \tag{5.3}$$

and analogously $U^\circ(t)$. By Lemma 4.2 and Proposition 5.1, whp as $n \to \infty$,

$$U_n(t_n) = U_n^\circ(t_n) + o(1) = U^\circ(t_n) + o(1) = U(t_n) + o(1), \tag{5.4}$$

so (1.21) follows from Theorem 1.1 after noting that ν in (1.17) is equal to $E[D_\star - 1]$. □

6 Appendix: Analysis of $\chi(\rho)$

In this appendix we study the variational problem in (1.9). In particular, we prove the alternative representations in Proposition 2.3, and we prove Theorem 1.2, i.e., we identify for $\varrho \geq 1/\log(d_{\min}+1)$ the quantity $\widetilde{\chi}(\varrho)$ that appears in Theorems 1.1 and 1.5 as χ_G with G the infinite tree with homogeneous degree $d_{\min} \in \mathbb{N}\backslash\{1\}$, the smallest degree that has a positive probability in our random graphs. In other words, we show that the infimum in (1.13) is attained on the infinite tree with the smallest admissible degrees.

It is not hard to understand heuristically why the optimal tree is infinite and has the smallest degree: the first part in (1.9) (the quadratic energy term coming from the Laplace operator) has a spreading effect and is the smaller the less bonds there are. However, proving this property is not so easy, since the other term (the Legendre transform from the large-deviation term of the random potential) has an opposite effect. In the setting where the underlying graph is $\mathbb{Z}^d$ instead of a tree, this problem is similar to the question whether or not the minimiser has compact support. However, our setting is different because of the exponential growth of balls on trees. We must therefore develop new methods.

Indeed, we will not study the effect on the principal eigenvalue due to the restriction of a large graph to a subgraph, but rather due to an opposite manipulation, namely, the glueing of two graphs obtained by adding one single edge (or possibly a joining vertex). The effect of such a glueing is examined in Sect. 6.2. The result will be used in Sect. 6.3 to finish the proof of Theorem 1.2. Before that, we discuss in Sect. 6.1 alternative representations for χ and prove Proposition 2.3.

In this section, no probability is involved. We drop ϱ from the notation at many places.

6.1 Alternative Representations

Fix a graph $G = (V, E)$. Recall that $\mathcal{P}(V)$ denotes the set of probability measures on V, and recall that the constant $\chi_G = \chi_G(\varrho)$ in (1.9) is defined as $\inf_{p\in\mathcal{P}(V)}[I_E(p) + \varrho J_V(p)]$ with I, J as in (1.8). As the next lemma shows, the constant $\widehat{\chi}$ in (2.16) can be also represented in terms of I,J.

Lemma 6.1 (First Representation) *For any graph $G = (V, E)$ and any $\Lambda \subset V$,*

$$\widehat{\chi}_V(\varrho; G) = \inf_{\substack{p\in\mathcal{P}(V):\\ \mathrm{supp}(p)\subset\Lambda}} [I_E(p) + \varrho J_V(p)]\,. \tag{6.1}$$

In particular,

$$\widehat{\chi}_\Lambda(\varrho; G) \geq \widehat{\chi}_V(\varrho; G) = \chi_G(\varrho). \tag{6.2}$$

Proof For the proof of (6.1), see [9, Lemma 2.17]. Moreover, (6.2) follows from (6.1). □

We next consider the constant $\chi_G^{(0)}$ in (2.17) for infinite rooted graphs $G = (V, E, \mathcal{O})$. Note that, by (6.1), $\widehat{\chi}_{B_r}(\varrho; G)$ is non-increasing in r. With (6.2) this implies

$$\chi_G^{(0)}(\varrho) = \lim_{r\to\infty} \widehat{\chi}_{B_r}(\varrho; G) \geq \chi_G(\varrho). \tag{6.3}$$

Lemma 6.2 (Second Representation) *For any rooted $G \in \mathfrak{G}_\infty$, $\chi_G(\varrho) = \chi_G^{(0)}(\varrho)$.*

Proof Write $G = (V, E, \mathcal{O})$. By (1.9), Lemma 6.1 and (6.3), it suffices to show that, for any $p \in \mathcal{P}(V)$ and $r \in \mathbb{N}$, there is a $p_r \in \mathcal{P}(V)$ with support in B_r such that

$$\liminf_{r\to\infty} \{I_E(p_r) + \varrho J_V(p_r)\} \leq I_E(p) + \varrho J_V(p). \tag{6.4}$$

Simply take

$$p_r(x) = \frac{p(x)\mathbb{1}_{B_r}(x)}{p(B_r)}, \qquad x \in V, \tag{6.5}$$

i.e., the normalised restriction of p to B_r. Then we easily see that

$$\begin{aligned} J_V(p_r) - J_V(p) &= -\tfrac{1}{p(B_r)} \textstyle\sum_{x\in B_r} p(x)\log p(x) + \log p(B_r) + \sum_{x\in V} p(x)\log p(x) \\ &\leq \tfrac{J_V(p)}{p(B_r)}(1 - p(B_r)) \to 0, \qquad r \to \infty, \end{aligned} \tag{6.6}$$

where we use $\log p(B_r) \leq 0$ and $p(x)\log p(x) \leq 0$ for every x. As for the I-term,

$$\begin{aligned} I_E(p_r) = &\tfrac{1}{p(B_r)} \textstyle\sum_{\{x,y\}\in E:\, x,y\in B_r} \left(\sqrt{p(x)} - \sqrt{p(y)}\right)^2 \\ &+ \tfrac{1}{2} \textstyle\sum_{\{x,y\}\in E:\, x\in B_r,\, y\in B_r^{\mathrm{c}}} \tfrac{p(x)}{p(B_r)} \leq \tfrac{I_E(p)}{p(B_r)} + \tfrac{d_{\max}}{2} \tfrac{p(B_{r-1}^{\mathrm{c}})}{p(B_r)}, \end{aligned} \tag{6.7}$$

and therefore

$$I_E(p_r) - I_E(p) \leq \frac{I_E(p)}{p(B_r)}(1 - p(B_r)) + \frac{d_{\max}}{2}\frac{p(B_{r-1}^{\mathrm{c}})}{p(B_r)} \to 0, \qquad r \to \infty. \tag{6.8}$$

□

Proof (Of Proposition 2.3) The claim follows from Lemmas 6.1–6.2 and (6.3). □

6.2 Glueing Graphs

Here we analyse the constant χ of a graph obtained by connecting disjoint graphs. First we show that glueing two graphs together with one additional edge does not decrease the quantity χ:

Lemma 6.3 (Glue Two) *Let $G_i = (V_i, E_i)$, $i = 1, 2$, be two disjoint connected simple graphs, and let $x_i \in V_i$, $i = 1, 2$. Denote by G the union graph of G_1, G_2 with one extra edge between x_1 and x_2, i.e., $G = (V, E)$ with $V := V_1 \cup V_2$, $E := E_1 \cup E_2 \cup \{(x_1, x_2)\}$. Then*

$$\chi_G \geq \min\left\{\chi_{G_1}, \chi_{G_2}\right\}. \tag{6.9}$$

Proof Given $p \in \mathcal{P}(V)$, let $a_i = p(V_i)$, $i = 1, 2$, and define $p_i \in \mathcal{P}(V_i)$ by putting

$$p_i(x) := \begin{cases} \frac{1}{a_i} p(x) 1\!\!1_{V_i}(x) & \text{if } a_i > 0, \\ 1\!\!1_{x_i}(x) & \text{otherwise.} \end{cases} \tag{6.10}$$

Straightforward manipulations show that

$$I_E(p) = \sum_{i=1}^{2} a_i I_{E_i}(p_i) + \left(\sqrt{p(x_1)} - \sqrt{p(x_2)}\right)^2, \quad J_V(p) = \sum_{i=1}^{2} \left[a_i J_{V_i}(p_i) - a_i \log a_i\right], \tag{6.11}$$

and so

$$I_E(p) + \varrho J_V(p) \geq \sum_{i=1}^{2} a_i \Big[I_{E_i}(p_i) + \varrho J_{V_i}(p_i)\Big] \geq \min\{\chi_{G_1}, \chi_{G_2}\}. \tag{6.12}$$

The proof is completed by taking the infimum over $p \in \mathcal{P}(V)$. □

Below it will be useful to define, for $x \in V$,

$$\chi_G^{(x,b)} = \inf_{\substack{p \in \mathcal{P}(V), \\ p(x) = b}} [I_E(p) + \varrho J_V(p)], \tag{6.13}$$

i.e., a version of χ_G with "boundary condition" b at x. It is clear that $\chi_G^{(x,b)} \geq \chi_G$.

Next we glue several graphs together and derive representations and estimates for the corresponding χ. For $k \in \mathbb{N}$, let $G_i = (V_i, E_i)$, $1 \leq i \leq k$, be a collection of disjoint graphs. Let x be a point not belonging to $\bigcup_{i=1}^k V_i$. For a fixed choice $y_i \in V_i$, $1 \leq i \leq k$, we denote by $\overline{G}_k = (\overline{V}_k, \overline{E}_k)$ the graph obtained by adding an edge from each $y_1, \ldots, y_k$ to x, i.e., $\overline{V}_k = V_1 \cup \cdots \cup V_k \cup \{x\}$ and $\overline{E}_k = E_1 \cup \cdots \cup E_k \cup \{(y_1, \mathcal{O}), \ldots, (y_k, x)\}$.

Lemma 6.4 (Glue Many Plus Vertex) *For any $\varrho > 0$, any $k \in \mathbb{N}$, and any $G_i = (V_i, E_i)$, $y_i \in V_i$, $1 \le i \le k$,*

$$\chi_{\overline{G}_k} = \inf_{\substack{0\le c_i\le a_i\le 1,\\ a_1+\cdots+a_k\le 1}} \Big\{ \sum_{i=1}^k a_i \Big(\chi_{G_i}^{(y_i,c_i/a_i)} - \varrho \log a_i\Big) + \sum_{i=1}^k \Big(\sqrt{c_i} - \sqrt{1-\sum_{i=1}^k a_i}\Big)^2 - \varrho\Big(1-\sum_{i=1}^k a_i\Big)\log\Big(1-\sum_{i=1}^k a_i\Big)\Big\}. \tag{6.14}$$

Proof The claim follows from straightforward manipulations with (1.8). □

Lemma 6.4 leads to the following comparison lemma. For $j \in \mathbb{N}$, let

$$(G_i^j, y_i^j) = \begin{cases} (G_i, y_i) & \text{if } i < j,\\ (G_{i+1}, y_{i+1}) & \text{if } i \ge j,\end{cases} \tag{6.15}$$

i.e., $(G_i^j)_{i\in\mathbb{N}}$ is the sequence $(G_i)_{i\in\mathbb{N}}$ with the j-th graph omitted. Let $\overline{G}_k^j$ be the analogue of $\overline{G}_k$ obtained from G_i^j, $1 \le i \le k$, $i \ne j$, instead of G_i, $1 \le i \le k$.

Lemma 6.5 (Comparison) *For any $\varrho > 0$ and any $k \in \mathbb{N}$,*

$$\begin{aligned}\chi_{\overline{G}_{k+1}} &= \inf_{1\le j\le k+1}\ \inf_{0\le c\le u\le \frac{1}{k+1}}\ \inf_{\substack{0\le c_i\le a_i\le 1,\\ a_1+\cdots+a_k\le 1}}\\ &\Big\{(1-u)\Big[\sum_{i=1}^k a_i\big(\chi_{G_{\sigma_j(i)}}^{(y_{\sigma_j(i)},c_i/a_i)} - \varrho\log a_i\big)\\ &\quad + \sum_{i=1}^k \Big(\sqrt{c_i} - \sqrt{1-\sum_{i=1}^k a_i}\Big)^2 - \varrho\Big(1-\sum_{i=1}^k a_i\Big)\log\Big(1-\sum_{i=1}^k a_i\Big)\Big]\\ &\quad + u\chi_{G_j}^{(y_j,c/u)} + \Big(\sqrt{c} - \sqrt{(1-u)\Big(1-\sum_{i=1}^k a_i\Big)}\Big)^2\\ &\quad -\varrho\big[u\log u + (1-u)\log(1-u)\big]\Big\}.\end{aligned} \tag{6.16}$$

Moreover,

$$\begin{aligned}\chi_{\overline{G}_{k+1}} &\ge \inf_{1\le j\le k+1}\inf_{0\le u\le\frac{1}{k+1}}\\ &\Big\{(1-u)\chi_{\overline{G}_k^j} + \inf_{v\in[0,1]}\Big\{u\chi_{G_j}^{(y_j,v)} + \mathbb{1}_{\{u(1+v)\ge 1\}}\Big[\sqrt{vu} - \sqrt{1-u}\Big]^2\Big\}\\ &\quad -\varrho\big[u\log u + (1-u)\log(1-u)\big]\Big\}.\end{aligned} \tag{6.17}$$

Proof Note that

$$\begin{aligned}&\Big\{(c_i,a_i)_{i=1}^{k+1}\colon\ 0\le c_i\le a_i\le 1,\ \textstyle\sum_{i=1}^{k+1}a_i\le 1\Big\}\\&=\bigcup_{j=1}^{k+1}\left\{\begin{array}{l}\Big((1-u)(c_i,a_i)_{i=1}^{j-1},(c,u),(1-u)(c_i,a_i)_{i=j}^{k}\Big)\colon\\ 0\le c\le u\le\frac{1}{k+1},0\le c_i\le a_i\le 1,\sum_{i=1}^{k}a_i\le 1\end{array}\right\},\end{aligned}\tag{6.18}$$

from which (6.16) follows by straightforward manipulations on (6.14). To prove (6.17), note that the first term within the square brackets in the first two lines of (6.16) equals the term minimised in (6.14), and is therefore not smaller than $\chi_{\overline{G}_k^j}$. □

Lemma 6.6 (Propagation of Lower Bounds) *If $\varrho > 0$, $M \in \mathbb{R}$, $C > 0$ and $k \in \mathbb{N}$ satisfy $\varrho \ge C/\log(k+1)$ and*

$$\inf_{1\le j\le k+1}\chi_{\overline{G}_k^j}\ge M,\qquad \inf_{1\le j\le k+1}\inf_{v\in[0,1]}\chi_{G_j}^{(y_j,v)}\ge M-C,\tag{6.19}$$

then $\chi_{\overline{G}_{k+1}} \ge M$.

Proof Dropping some non-negative terms in (6.17), we obtain

$$\begin{aligned}\chi_{\overline{G}_{k+1}}-M&\ge\inf\nolimits_{0\le u\le 1/(k+1)}\Big\{u\Big(\chi_{G_j}^{(y_j,v)}-M\Big)-\varrho u\log u\Big\}\\&\ge\inf\nolimits_{0\le u\le 1/(k+1)}\{u\,(\varrho\log(k+1)-C)\}\ge 0\end{aligned}\tag{6.20}$$

by the assumption on ϱ. □

The above results will be applied in the next section to minimise χ over families of trees with minimum degrees.

6.3 Trees with Minimum Degrees

Fix $d \in \mathbb{N}$. Let $\mathring{\mathcal{T}}_d$ be an infinite tree rooted at $\mathcal{O}$ such that the degree of $\mathcal{O}$ equals $d-1$ and the degree of every other vertex in $\mathring{\mathcal{T}}_d$ is d. Let $\mathring{\mathscr{T}}_d^{(0)} = \{\mathring{\mathcal{T}}_d\}$ and, recursively, let $\mathring{\mathscr{T}}_d^{(n+1)}$ denote the set of all trees obtained from a tree in $\mathring{\mathscr{T}}_d^{(n)}$ and a disjoint copy of $\mathring{\mathcal{T}}_d$ by adding an edge between a vertex of the former and the root of the latter. Write $\mathring{\mathscr{T}}_d = \bigcup_{n\in\mathbb{N}_0}\mathring{\mathscr{T}}_d^{(n)}$. Assume that all trees in $\mathring{\mathscr{T}}_d$ are rooted at $\mathcal{O}$.

Recall that $\mathcal{T}_d$ is the infinite regular d-tree. Observe that $\mathcal{T}_d$ is obtained from $(\mathring{\mathcal{T}}_d, \mathcal{O})$ and a disjoint copy $(\mathring{\mathcal{T}}_d', \mathcal{O}')$ by adding one edge between $\mathcal{O}$ and $\mathcal{O}'$. Consider $\mathcal{T}_d$ to be rooted at $\mathcal{O}$. Let $\mathscr{T}_d^{(0)} = \{\mathcal{T}_d\}$ and, recursively, let $\mathscr{T}_d^{(n+1)}$ denote the set of all trees obtained from a tree in $\mathscr{T}_d^{(n)}$ and a disjoint copy of $\mathring{\mathcal{T}}_d$ by adding an edge between a vertex of the former and the root of the latter. Write $\mathscr{T}_d = \bigcup_{n\in\mathbb{N}_0}\mathscr{T}_d^{(n)}$, and still consider all trees in $\mathscr{T}_d$ to be rooted at $\mathcal{O}$. Note that

$\mathscr{T}_d^{(n)}$ contains precisely those trees of $\mathring{\mathscr{T}}_d^{(n+1)}$ that have $\mathcal{T}_d$ as a subgraph rooted at $\mathcal{O}$. In particular, $\mathscr{T}_d^{(n)} \subset \mathring{\mathscr{T}}_d^{(n+1)}$ and $\mathscr{T}_d \subset \mathring{\mathscr{T}}_d$.

Our objective is to prove the following.

Proposition 6.7 (Minimal Tree Is Optimal) *If* $\varrho \geq 1/\log(d+1)$, *then*

$$\chi_{\mathcal{T}_d}(\varrho) = \min_{T \in \mathscr{T}_d} \chi_T(\varrho).$$

For the proof of Proposition 6.7, we will need the following.

Lemma 6.8 (Minimal Half-Tree Is Optimal) *For all* $\varrho \in (0, \infty)$,

$$\chi_{\mathring{\mathcal{T}}_d}(\varrho) = \min_{T \in \mathring{\mathscr{T}}_d} \chi_T(\varrho).$$

Proof Fix $\varrho \in (0, \infty)$. It will be enough to show that

$$\chi_{\mathring{\mathcal{T}}_d} = \min_{T \in \mathring{\mathscr{T}}_d^{(n)}} \chi_T, \qquad n \in \mathbb{N}_0, \tag{6.21}$$

which we will achieve by induction in n. The case $n = 0$ is obvious. Assume that (6.21) holds for some $n \in \mathbb{N}_0$. Any tree $T \in \mathring{\mathscr{T}}_d^{(n+1)}$ can be obtained from a tree $\widetilde{T} \in \mathring{\mathscr{T}}_d^{(n)}$ and a disjoint copy $\mathring{\mathcal{T}}_d'$ of $\mathring{\mathcal{T}}_d$ by adding an edge between a point $\tilde{x}$ in the vertex set of $\widetilde{T}$ to the root of $\mathring{\mathcal{T}}_d'$. Applying Lemma 6.3 together with the induction hypothesis, we obtain

$$\chi_T \geq \min\left\{\chi_{\widetilde{T}}, \chi_{\mathring{\mathcal{T}}_d'}\right\} \geq \chi_{\mathring{\mathcal{T}}_d}, \tag{6.22}$$

which completes the induction step. □

Lemma 6.9 (A Priori Bounds) *For any* $d \in \mathbb{N}$ *and any* $\varrho \in (0, \infty)$,

$$\chi_{\mathring{\mathcal{T}}_d}(\varrho) \leq \chi_{\mathcal{T}_d}(\varrho) \leq \chi_{\mathring{\mathcal{T}}_d}(\varrho) + 1. \tag{6.23}$$

Proof The first inequality follows from Lemma 6.8. For the second inequality, note that $\mathcal{T}_d$ contains as subgraph a copy of $\mathring{\mathcal{T}}_d$, and restrict the minimum in (1.9) to $p \in \mathcal{P}(\mathring{\mathcal{T}}_d)$. □

Proof ***(Of Proposition 6.7)*** Fix $\varrho \geq 1/\log(d+1)$. It will be enough to show that

$$\chi_{\mathcal{T}_d} = \min_{T \in \mathscr{T}_d^{(n)}} \chi_T, \qquad n \in \mathbb{N}_0. \tag{6.24}$$

We will prove this by induction in n. The case $n = 0$ is trivial. Assume that, for some $n_0 \geq 0$, (6.24) holds for all $n \leq n_0$. Let $T \in \mathscr{T}_d^{(n_0+1)}$. Then there exists a vertex x of T with degree $k+1 \geq d+1$. Let $y_1, \ldots, y_{k+1}$ be set of neighbours of x in T. When we remove the edge between y_j and x, we obtain two connected trees; call G_j the one containing y_j, and $\overline{G}_k^j$ the other one. With this notation, T may be identified with $\overline{G}_{k+1}$.

Now, for each j, the rooted tree (G_j, y_j) is isomorphic (in the obvious sense) to a tree in $\mathring{\mathscr{T}}_d^{(\ell_j)}$, where $\ell_j \in \mathbb{N}_0$ satisfy $\ell_1 + \cdots + \ell_{k+1} \leq n_0$, while $\overline{G}_k^j$ belongs to $\mathscr{T}_d^{(n_j)}$ for some $n_j \leq n_0$. Therefore, by the induction hypothesis,

$$\chi_{\overline{G}_k^j} \geq \chi_{\mathcal{T}_d}, \tag{6.25}$$

while, by (6.13), Lemma 6.8 and Lemma 6.9,

$$\inf_{v \in [0,1]} \chi_{G_j}^{(y_j, v)} \geq \chi_{G_j} \geq \chi_{\mathring{\mathcal{T}}_d} \geq \chi_{\mathcal{T}_d} - 1. \tag{6.26}$$

Thus, by Lemma 6.3 applied with $M = \chi_{\mathcal{T}_d}$ and $C = 1$,

$$\chi_T = \chi_{\bar{G}_{k+1}} \geq \chi_{\mathcal{T}_d}, \tag{6.27}$$

which completes the induction step. □

Proof (Of Theorem 1.2) First note that, since $\mathcal{T}_{d_{\min}}$ has degrees in $\mathrm{supp}(D_g)$, $\widetilde{\chi}(\varrho) \leq \chi_{\mathcal{T}_{d_{\min}}}(\varrho)$. For the opposite inequality, we proceed as follows. Fix an infinite tree T with degrees in $\mathrm{supp}(D_g)$, and root it at a vertex $\mathcal{Y}$. For $r \in \mathbb{N}$, let $\widetilde{T}_r$ be the tree obtained from $B_r = B_r^T(\mathcal{Y})$ by attaching to each vertex $x \in B_r$ with $|x| = r$ a number $d_{\min} - 1$ of disjoint copies of $(\mathring{\mathcal{T}}_{d_{\min}}, \mathcal{O})$, i.e., adding edges between x and the corresponding roots. Then $\widetilde{T}_r \in \mathscr{T}_{d_{\min}}$ and, since B_r has more out-going edges in T than in $\widetilde{T}_r$, we may check using (6.1) that

$$\widehat{\chi}_{B_r}(\varrho; T) \geq \widehat{\chi}_{B_r}(\varrho; \widetilde{T}_r) \geq \chi_{\widetilde{T}_r}(\varrho) \geq \chi_{\mathcal{T}_{d_{\min}}}(\varrho). \tag{6.28}$$

Taking $r \to \infty$ and applying Proposition 2.3, we obtain $\chi_T(\varrho) \geq \chi_{\mathcal{T}_{d_{\min}}}(\varrho)$. Since T is arbitrary, the proof is complete. □

Acknowledgments The authors thank Remco van der Hofstad for helpful discussions and an anonymous referee for useful suggestions. FdH was supported by the Netherlands Organisation for Scientific Research through NWO Gravitation Grant NETWORKS-024.002.003 and by the Alexander von Humboldt Foundation. WK and RSdS were supported by the German Research Foundation through DFG SPP 1590 *Probabilistic Structures in Evolution*. RSdS thanks the Pró-Reitoria de Pesquisa da Universidade Federal de Minas Gerais.

References

1. Astrauskas, A.: From extreme values of i.i.d. random fields to extreme eigenvalues of finite-volume Anderson Hamiltonian. Probab. Surv. **13**, 156–244 (2016)
2. Avena, L., Gün, O., Hesse, M.: The parabolic Anderson model on the hypercube. Stoch. Proc. Appl. **130**(6), 3369–3393 (2020)
3. Biskup, M., König, W.: Eigenvalue order statistics for random Schrödinger operators with doubly-exponential tails. Commun. Math. Phys. **341**(1), 179–218 (2016)
4. Biskup, M., König, W., dos Santos, R.S.: Mass concentration and aging in the parabolic Anderson model with doubly-exponential tails. Probab. Theory Relat. Fields **171**(1–2), 251–331 (2018)
5. Federico, L., van der Hofstad, R.: Critical window for connectivity in the configuration model. Comb. Prob. Comp. **26**, 660–680 (2017)
6. Fleischmann, K., Molchanov, S.A.: Exact asymptotics in a mean field model with random potential. Probab. Theory Relat. Fields **86**(2), 239–251 (1990)
7. Gärtner, J., den Hollander, F.: Correlation structure of intermittency in the parabolic Anderson model. Probab. Theory Relat. Fields **114**, 1–54 (1999)
8. Gärtner, J., Molchanov, S.A.: Parabolic problems for the Anderson model I. Intermittency and related problems. Commun. Math. Phys. **132**, 613–655 (1990)
9. Gärtner, J., Molchanov, S.A.: Parabolic problems for the Anderson model II. Second-order asymptotics and structure of high peaks. Probab. Theory Relat. Fields **111**, 17–55 (1998)
10. Gärtner, J., König, W., Molchanov, S.: Geometric characterization of intermittency in the parabolic Anderson model. Ann. Probab. **35**(2), 439–499 (2007)
11. Grimmett, G.: *Percolation*, 2nd edn. (Springer, Berlin, 1999)
12. König, W.: *The Parabolic Anderson Model*. Pathways in Mathematics. Birkhäuser, Basel (2016)
13. Lyons, R., Peres, Y.: *Probability on Trees and Networks*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, New York (2016)
14. Muirhead, S., Pymar, R.: Localisation in the Bouchaud-Anderson model. Stoch. Proc. Appl. **126**(11), 3402–3462 (2016)
15. van der Hofstad, R.: *Random Graphs and Complex Networks*. Cambridge Series in Statistical and Probabilistic Mathematics, vol. 1. Cambridge University Press, Cambridge (2017)
16. van der Hofstad, R.: *Random Graphs and Complex Networks*, vol. 2. Pdf-file is available at https://www.win.tue.nl/~rhofstad/
17. van der Hofstad, R., König, W., Mörters, P.: The universality classes in the parabolic Anderson model. Commun. Math. Phys. **267**(2), 307–353 (2006)

Reflecting Random Walks in Curvilinear Wedges

Mikhail V. Menshikov, Aleksandar Mijatović, and Andrew R. Wade

Dedicated to our colleague Vladas Sidoravicius (1963–2019)

Abstract We study a random walk (Markov chain) in an unbounded planar domain bounded by two curves of the form $x_2 = a^+ x_1^{\beta^+}$ and $x_2 = -a^- x_1^{\beta^-}$, with $x_1 \geq 0$. In the interior of the domain, the random walk has zero drift and a given increment covariance matrix. From the vicinity of the upper and lower sections of the boundary, the walk drifts back into the interior at a given angle α^+ or α^- to the relevant inwards-pointing normal vector. Here we focus on the case where α^+ and α^- are equal but opposite, which includes the case of normal reflection. For $0 \leq \beta^+, \beta^- < 1$, we identify the phase transition between recurrence and transience, depending on the model parameters, and quantify recurrence via moments of passage times.

Keywords Reflected random walk · Generalized parabolic domain · Recurrence · Transience · Passage-time moments · Normal reflection · Oblique reflection

AMS Subject Classification 60J05 (Primary); 60J10, 60G50 (Secondary)

M. V. Menshikov (✉) · A. R. Wade
Durham University, Durham, UK
e-mail: mikhail.menshikov@durham.ac.uk; andrew.wade@durham.ac.uk

A. Mijatović
University of Warwick and the Alan Turing Institute, Coventry, UK
e-mail: a.mijatovic@warwick.ac.uk

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_26

1 Introduction and Main Results

1.1 Description of the Model

We describe our model and then state our main results: see Sect. 1.4 for a discussion of related literature. Write $x \in \mathbb{R}^2$ in Cartesian coordinates as $x = (x_1, x_2)$. For parameters $a^+, a^- > 0$ and $\beta^+, \beta^- \geq 0$, define, for $z \geq 0$, functions $d^+(z) := a^+ z^{\beta^+}$ and $d^-(z) := a^- z^{\beta^-}$. Set

$$\mathscr{D} := \left\{ x \in \mathbb{R}^2 : x_1 \geq 0, \ -d^-(x_1) \leq x_2 \leq d^+(x_1) \right\}.$$

Write $\| \cdot \|$ for the Euclidean norm on $\mathbb{R}^2$. For $x \in \mathbb{R}^2$ and $A \subseteq \mathbb{R}^2$, write $d(x, A) := \inf_{y \in A} \|x - y\|$ for the distance from x to A. Suppose that there exist $B \in (0, \infty)$ and a subset $\mathscr{D}_B$ of $\mathscr{D}$ for which every $x \in \mathscr{D}_B$ has $d(x, \mathbb{R}^2 \setminus \mathscr{D}) \leq B$. Let $\mathscr{D}_I := \mathscr{D} \setminus \mathscr{D}_B$; we call $\mathscr{D}_B$ the *boundary* and $\mathscr{D}_I$ the *interior*. Set $\mathscr{D}_B^{\pm} := \{x \in \mathscr{D}_B : \pm x_2 > 0\}$ for the parts of $\mathscr{D}_B$ in the upper and lower half-plane, respectively.

Let $\xi := (\xi_0, \xi_1, \ldots)$ be a discrete-time, time-homogeneous Markov chain on state-space $S \subseteq \mathscr{D}$. Set $S_I := S \cap \mathscr{D}_I$, $S_B := S \cap \mathscr{D}_B$, and $S_B^{\pm} := S \cap \mathscr{D}_B^{\pm}$. Write $\mathbb{P}_x$ and $\mathbb{E}_x$ for conditional probabilities and expectations given $\xi_0 = x \in S$, and suppose that $\mathbb{P}_x(\xi_n \in S \text{ for all } n \geq 0) = 1$ for all $x \in S$. Set $\Delta := \xi_1 - \xi_0$. Then $\mathbb{P}(\xi_{n+1} \in A \mid \xi_n = x) = \mathbb{P}_x(x + \Delta \in A)$ for all $x \in S$, all measurable $A \subseteq \mathscr{D}$, and all $n \in \mathbb{Z}_+$. In what follows, we will always treat vectors in $\mathbb{R}^2$ as column vectors.

We will assume that ξ has uniformly bounded $p > 2$ moments for its increments, that in S_I it has zero drift and a fixed increment covariance matrix, and that it *reflects* in S_B, meaning it has drift away from $\partial \mathscr{D}$ at a certain angle relative to the inwards-pointing normal vector. In fact we permit perturbations of this situation that are appropriately small as the distance from the origin increases. See Fig. 1 for an illustration.

To describe the assumptions formally, for $x_1 > 0$ let $n^+(x_1)$ denote the inwards-pointing unit normal vector to $\partial \mathscr{D}$ at $(x_1, d^+(x_1))$, and let $n^-(x_1)$ be the corresponding normal at $(x_1, -d^-(x_1))$; then $n^+(x_1)$ is a scalar multiple of $(a^+ \beta^+ x_1^{\beta^+ - 1}, -1)$, and $n^-(x_1)$ is a scalar multiple of $(a^- \beta^- x_1^{\beta^- - 1}, 1)$. Let $n^+(x_1, \alpha)$ denote the unit vector obtained by rotating $n^+(x_1)$ by angle α anticlockwise. Similarly, let $n^-(x_1, \alpha)$ denote the unit vector obtained by rotating $n^-(x_1)$ by angle α clockwise. (The orientation is such that, in each case, reflection at angle $\alpha < 0$ is pointing on the side of the normal towards 0.)

We write $\| \cdot \|_{\text{op}}$ for the matrix (operator) norm defined by $\|M\|_{\text{op}} := \sup_u \|Mu\|$, where the supremum is over all unit vectors $u \in \mathbb{R}^2$. We take $\xi_0 = x_0 \in S$ fixed, and impose the following assumptions for our main results.

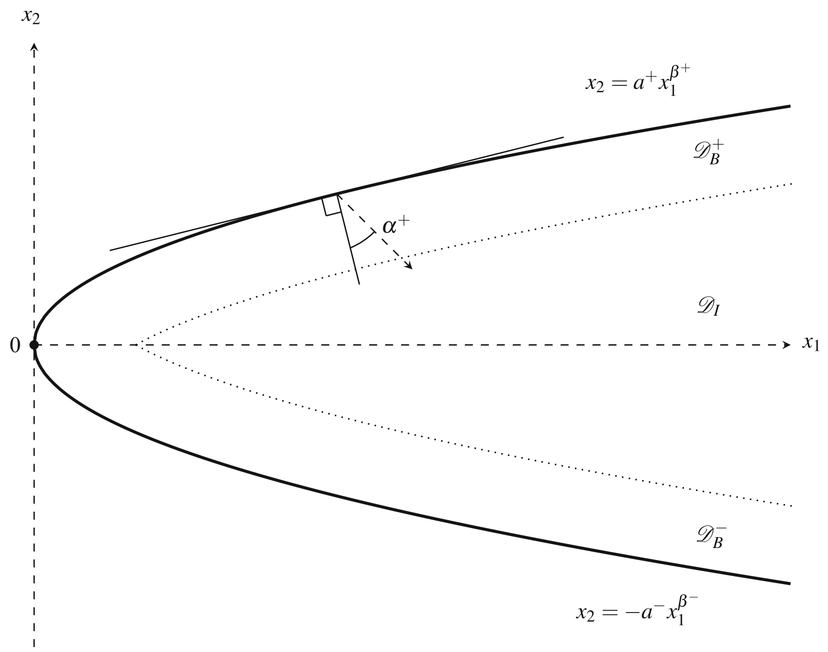


Fig. 1 An illustration of the model parameters, in the case where $\beta^+ = \beta^- \in (0, 1)$

(N) Suppose that $\mathbb{P}_x(\limsup_{n\to\infty} \|\xi_n\| = \infty) = 1$ for all $x \in S$.
(M$_p$) There exists $p > 2$ such that

$$\sup_{x\in S} \mathbb{E}_x(\|\Delta\|^p) < \infty. \tag{1}$$

(D) We have that $\sup_{x\in S_I:\|x\|\geq r} \|\mathbb{E}_x \Delta\| = o(r^{-1})$ as $r \to \infty$.
(R) There exist angles $\alpha^\pm \in (-\pi/2, \pi/2)$ and functions $\mu^\pm : S_B^\pm \to \mathbb{R}$ with $\liminf_{\|x\|\to\infty} \mu^\pm(x) > 0$, such that, as $r \to \infty$,

$$\sup_{x\in S_B^+:\|x\|\geq r} \|\mathbb{E}_x \Delta - \mu^+(x) n^+(x_1, \alpha^+)\| = O(r^{-1}); \tag{2}$$

$$\sup_{x\in S_B^-:\|x\|\geq r} \|\mathbb{E}_x \Delta - \mu^-(x) n^-(x_1, \alpha^-)\| = O(r^{-1}). \tag{3}$$

(C) There exists a positive-definite, symmetric 2×2 matrix Σ for which

$$\lim_{r\to\infty} \sup_{x\in S_I:\|x\|\geq r} \left\| \mathbb{E}_x(\Delta\Delta^\top) - \Sigma \right\|_{\mathrm{op}} = 0.$$

We write the entries of Σ in (C) as

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{pmatrix}.$$

Here ρ is the asymptotic increment covariance, and, since Σ is positive definite, $\sigma_1 > 0, \sigma_2 > 0$, and $\rho^2 < \sigma_1^2 \sigma_2^2$.

To identify the critically recurrent cases, we need slightly sharper control of the error terms in the drift assumption (D) and covariance assumption (C). In particular, we will in some cases impose the following stronger versions of these assumptions:

(D_+) There exists $\varepsilon > 0$ such that $\sup_{x \in S_I : \|x\| \geq r} \| \mathbb{E}_x \Delta \| = O(r^{-1-\varepsilon})$ as $r \to \infty$.

(C_+) There exists $\varepsilon > 0$ and a positive definite symmetric 2×2 matrix Σ for which

$$\sup_{x \in S_I : \|x\| \geq r} \left\| \mathbb{E}_x (\Delta \Delta^\top) - \Sigma \right\|_{\text{op}} = O(r^{-\varepsilon}), \text{ as } r \to \infty.$$

Without loss of generality, we may use the same constant $\varepsilon > 0$ for both (D_+) and (C_+).

The non-confinement condition (N) ensures our questions of recurrence and transience (see below) are non-trivial, and is implied by standard irreducibility or ellipticity conditions: see [26] and the following example.

Example 1 Let $S = \mathbb{Z}^2 \cap \mathcal{D}$, and take $\mathcal{D}_B$ to be the set of $x \in \mathcal{D}$ for which x is within unit ℓ_∞-distance of some $y \in \mathbb{Z}^2 \setminus \mathcal{D}$. Then S_B contains those points of S that have a neighbour outside of $\mathcal{D}$, and S_I consists of those points of S whose neighbours are all in $\mathcal{D}$. If ξ is irreducible on S, then (N) holds (see e.g. Corollary 2.1.10 of [26]). If $\beta^+ > 0$, then, for all $\|x\|$ sufficiently large, every point of $x \in S_B^+$ has its neighbours to the right and below in S, so if $\alpha^+ = 0$, for instance, we can achieve the asymptotic drift required by (2) using only nearest-neighbour jumps if we wish; similarly in S_B^-.

Under the non-confinement condition (N), the first question of interest is whether $\liminf_{n\to\infty} \|\xi_n\|$ is finite or infinite. We say that ξ is *recurrent* if there exists $r_0 \in \mathbb{R}_+$ for which $\liminf_{n\to\infty} \|\xi_n\| \leq r_0$, a.s., and that ξ is *transient* if $\lim_{n\to\infty} \|\xi_n\| = \infty$, a.s. The first main aim of this paper is to classify the process into one or other of these cases (which are not a priori exhaustive) depending on the parameters. Further, in the recurrent cases it is of interest to quantify the recurrence by studying the tails (or moments) of return times to compact sets. This is the second main aim of this paper.

In the present paper we focus on the case where $\alpha^+ + \alpha^- = 0$, which we call 'opposed reflection'. This case is the most subtle from the point of view of recurrence/transience, and, as we will see, exhibits a rich phase diagram depending on the model parameters. We emphasize that the model in the case $\alpha^+ + \alpha^- = 0$ is *near-critical* in that both recurrence and transience are possible, depending on the parameters, and moreover (i) in the recurrent cases, return-times to bounded sets have heavy tails being, in particular, non-integrable, and so stationary distributions will not exist, and (ii) in the transient cases, escape to infinity will be only diffusive. There is a sense in which the model studied here can be viewed as a perturbation of zero-drift random walks, in the manner of the seminal work of Lamperti [19]: see e.g. [26] for a discussion of near-critical phenomena. We leave for future work the case $\alpha^+ + \alpha^- \neq 0$, in which very different behaviour will occur: if $\beta^\pm < 1$, then the case $\alpha^+ + \alpha^- > 0$ gives super-diffusive (but sub-ballistic) transience, while the case $\alpha^+ + \alpha^- < 0$ leads to positive recurrence.

Opposed reflection includes the special case where $\alpha^+ = \alpha^- = 0$, which is 'normal reflection'. Since the results are in the latter case more easily digested, and since it is an important case in its own right, we present the case of normal reflection first, in Sect. 1.2. The general case of opposed reflection we present in Sect. 1.3. In Sect. 1.4 we review some of the extensive related literature on reflecting processes. Then Sect. 1.5 gives an outline of the remainder of the paper, which consists of the proofs of the results in Sects. 1.2–1.3.

1.2 Normal Reflection

First we consider the case of *normal* (i.e., orthogonal) reflection.

Theorem 1 *Suppose that* (N), (M$_p$), (D), (R), *and* (C) *hold with* $\alpha^+ = \alpha^- = 0$.

(a) Suppose that $\beta^+, \beta^- \in [0, 1)$. *Let* $\beta := \max(\beta^+, \beta^-)$. *Then the following hold.*

  (i) If $\beta < \sigma_1^2/\sigma_2^2$, *then* ξ *is recurrent.*
  (ii) If $\sigma_1^2/\sigma_2^2 < \beta < 1$, *then* ξ *is transient.*
  (iii) If, in addition, (D$_+$) *and* (C$_+$) *hold, then the case* $\beta = \sigma_1^2/\sigma_2^2$ *is recurrent.*

(b) Suppose that (D$_+$) *and* (C$_+$) *hold, and* $\beta^+, \beta^- > 1$. *Then* ξ *is recurrent.*

Remark 1

(i) Omitted from Theorem 1 is the case when at least one of $\beta^\pm$ is equal to 1, or their values fall each each side of 1. Here we anticipate behaviour similar to [5].
(ii) If $\sigma_1^2/\sigma_2^2 < 1$, then Theorem 1 shows a striking *non-monotonicity* property: there exist regions $\mathscr{D}_1 \subset \mathscr{D}_2 \subset \mathscr{D}_3$ such that the reflecting random walk is recurrent on $\mathscr{D}_1$ and $\mathscr{D}_3$, but transient on $\mathscr{D}_2$. This phenomenon does not occur in the classical case when Σ is the identity: see [28] for a derivation of monotonicity in the case of normally reflecting Brownian motion in unbounded domains in $\mathbb{R}^d$, $d \geq 2$.
(iii) Note that the correlation ρ and the values of a^+, a^- play no part in Theorem 1; ρ will, however, play a role in the more general Theorem 3 below.

Let $\tau_r := \min\{n \in \mathbb{Z}_+ : \|\xi_n\| \le r\}$. Define

$$s_0 := s_0(\Sigma, \beta) := \frac{1}{2}\left(1 - \frac{\sigma_2^2 \beta}{\sigma_1^2}\right). \tag{4}$$

Our next result concerns the moments of τ_r. Since most of our assumptions are asymptotic, we only make statements about r sufficiently large; with appropriate irreducibility assumptions, this restriction could be removed.

Theorem 2 *Suppose that* (N), (M$_p$), (D), (R), *and* (C) *hold with* $\alpha^+ = \alpha^- = 0$.

(a) Suppose that $\beta^+, \beta^- \in [0, 1)$. *Let* $\beta := \max(\beta^+, \beta^-)$. *Then the following hold.*

- *(i) If* $\beta < \sigma_1^2/\sigma_2^2$, *then* $\mathbb{E}_x(\tau_r^s) < \infty$ *for all* $s < s_0$ *and all* r *sufficiently large, but* $\mathbb{E}_x(\tau_r^s) = \infty$ *for all* $s > s_0$ *and all* x *with* $\|x\| > r$ *for* r *sufficiently large.*
- *(ii) If* $\beta \ge \sigma_1^2/\sigma_2^2$, *then* $\mathbb{E}_x(\tau_r^s) = \infty$ *for all* $s > 0$ *and all* x *with* $\|x\| > r$ *for* r *sufficiently large.*

(b) Suppose that $\beta^+, \beta^- > 1$. *Then* $\mathbb{E}_x(\tau_r^s) = \infty$ *for all* $s > 0$ *and all* x *with* $\|x\| > r$ *for* r *sufficiently large.*

Remark 2

- (i) Note that if $\beta < \sigma_1^2/\sigma_2^2$, then $s_0 > 0$, while $s_0 < 1/2$ for all $\beta > 0$, in which case the return time to a bounded set has a heavier tail than that for one-dimensional simple symmetric random walk.
- (ii) The transience result in Theorem 1(a)(ii) is essentially stronger than the claim in Theorem 2(a)(ii) for $\beta < \sigma_1^2/\sigma_2^2$, so the borderline (recurrent) case $\beta = \sigma_1^2/\sigma_2^2$ is the main content of the latter.
- (iii) Part (b) shows that the case $\beta^\pm > 1$ is critical: no moments of return times exist, as in the case of, say, simple symmetric random walk in $\mathbb{Z}^2$ [26, p. 77].

1.3 Opposed Reflection

We now consider the more general case where $\alpha^+ + \alpha^- = 0$, i.e., the two reflection angles are equal but opposite, relative to their respective normal vectors. For $\alpha^+ = -\alpha^- \neq 0$, this is a particular example of *oblique* reflection. The phase transition in β now depends on ρ and α in addition to σ_1^2 and σ_2^2. Define

$$\beta_{\mathrm{c}} := \beta_{\mathrm{c}}(\Sigma, \alpha) := \frac{\sigma_1^2}{\sigma_2^2} + \left(\frac{\sigma_2^2 - \sigma_1^2}{\sigma_2^2}\right)\sin^2\alpha + \frac{\rho}{\sigma_2^2}\sin 2\alpha. \tag{5}$$

The next result gives the key properties of the critical threshold function β_c which are needed for interpreting our main result.

Proposition 1 *For a fixed, positive-definite Σ such that $|\sigma_1^2 - \sigma_2^2| + |\rho| > 0$, the function $\alpha \mapsto \beta_c(\Sigma, \alpha)$ over the interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$ is strictly positive for $|\alpha| \le \pi/2$, with two stationary points, one in $(-\frac{\pi}{2}, 0)$ and the other in $(0, \frac{\pi}{2})$, at which the function takes its maximum/minimum values of*

$$\frac{1}{2} + \frac{\sigma_1^2}{2\sigma_2^2} \pm \frac{1}{2\sigma_2^2}\sqrt{\left(\sigma_1^2 - \sigma_2^2\right)^2 + 4\rho^2}. \tag{6}$$

The exception is the case where $\sigma_1^2 - \sigma_2^2 = \rho = 0$, when $\beta_c = 1$ is constant.

Here is the recurrence classification in this setting.

Theorem 3 *Suppose that (N), (M_p), (D), (R), and (C) hold with $\alpha^+ = -\alpha^- = \alpha$ for $|\alpha| < \pi/2$.*

(a) Suppose that $\beta^+, \beta^- \in [0, 1)$. Let $\beta := \max(\beta^+, \beta^-)$. Then the following hold.

  (i) If $\beta < \beta_c$, then ξ is recurrent.
  (ii) If $\beta > \beta_c$, then ξ is transient.
  (iii) If, in addition, (D_+) and (C_+) hold, then the case $\beta = \beta_c$ is recurrent.

(b) Suppose that (D_+) and (C_+) hold, and $\beta^+, \beta^- > 1$. Then ξ is recurrent.

Remark 3

(i) The threshold (5) is invariant under the map $(\alpha, \rho) \mapsto (-\alpha, -\rho)$.
(ii) For fixed Σ with $|\sigma_1^2 - \sigma_2^2| + |\rho| > 0$, Proposition 1 shows that β_c is non-constant and has exactly one maximum and exactly one minimum in $(-\frac{\pi}{2}, \frac{\pi}{2})$. Since $\beta_c(\Sigma, \pm\frac{\pi}{2}) = 1$, it follows from uniqueness of the minimum that the minimum is strictly less than 1, and so Theorem 3 shows that there is always an open interval of α for which there is transience.
(iii) Since $\beta_c > 0$ always, recurrence is certain for small enough β.
(iv) In the case where $\sigma_1^2 = \sigma_2^2$ and $\rho = 0$, then $\beta_c = 1$, so recurrence is certain for all $\beta^+, \beta^- < 1$ and all α.
(v) If $\alpha = 0$, then $\beta_c = \sigma_1^2/\sigma_2^2$, so Theorem 3 generalizes Theorem 1.

Next we turn to passage-time moments. We generalize (4) and define

$$s_0 := s_0(\Sigma, \alpha, \beta) := \frac{1}{2}\left(1 - \frac{\beta}{\beta_c}\right), \tag{7}$$

with β_c given by (5). The next result includes Theorem 2 as the special case $\alpha = 0$.

Theorem 4 *Suppose that (N), (M_p), (D), (R), and (C) hold with $\alpha^+ = -\alpha^- = \alpha$ for $|\alpha| < \pi/2$.*

(a) *Suppose that* $\beta^+, \beta^- \in [0, 1)$. *Let* $\beta := \max(\beta^+, \beta^-)$. *Then the following hold.*

(i) *If* $\beta < \beta_c$, *then* $s_0 \in (0, 1/2]$, *and* $\mathbb{E}_x(\tau_r^s) < \infty$ *for all* $s < s_0$ *and all* r *sufficiently large, but* $\mathbb{E}_x(\tau_r^s) = \infty$ *for all* $s > s_0$ *and all* x *with* $\|x\| > r$ *for* r *sufficiently large.*

(ii) *If* $\beta \geq \beta_c$, *then* $\mathbb{E}_x(\tau_r^s) = \infty$ *for all* $s > 0$ *and all* x *with* $\|x\| > r$ *for* r *sufficiently large.*

(b) *Suppose that* $\beta^+, \beta^- > 1$. *Then* $\mathbb{E}_x(\tau_r^s) = \infty$ *for all* $s > 0$ *and all* x *with* $\|x\| > r$ *for* r *sufficiently large.*

1.4 Related Literature

The stability properties of reflecting random walks or diffusions in unbounded domains in $\mathbb{R}^d$ have been studied for many years. A pre-eminent place in the development of the theory is occupied by processes in the quadrant $\mathbb{R}_+^2$ or quarter-lattice $\mathbb{Z}_+^2$, due to applications arising in queueing theory and other areas. Typically, the process is assumed to be maximally homogeneous in the sense that the transition mechanism is fixed in the interior and on each of the two half-lines making up the boundary. Distinct are the cases where the motion in the interior of the domain has *non-zero* or *zero drift*.

It was in 1961, in part motivated by queueing models, that Kingman [18] proposed a general approach to the non-zero drift problem on $\mathbb{Z}_+^2$ via Lyapunov functions and Foster's Markov chain classification criteria [14]. A formal statement of the classification was given in the early 1970s by Malyshev, who developed both an analytic approach [22] as well as the Lyapunov function one [23] (the latter, Malyshev reports, prompted by a question of Kolmogorov). Generically, the classification depends on the drift vector in the interior and the two boundary reflection angles. The Lyapunov function approach was further developed, so that the bounded jumps condition in [23] could be relaxed to finiteness of second moments [10, 27, 29] and, ultimately, of first moments [13, 30, 33]. The analytic approach was also subsequently developed [11], and although it seems to be not as robust as the Lyapunov function approach (the analysis in [22] was restricted to nearest-neighbour jumps), when it is applicable it can yield very precise information: see e.g. [15] for a recent application in the continuum setting. Intrinsically more complicated results are available for the non-zero drift case in $\mathbb{Z}_+^3$ [24] and $\mathbb{Z}_+^4$ [17].

The recurrence classification for the case of *zero-drift* reflecting random walk in $\mathbb{Z}_+^2$ was given in the early 1990s in [6, 12]; see also [13]. In this case, generically, the classification depends on the increment covariance matrix in the interior as well as the two boundary reflection angles. Subsequently, using a semimartingale approach extending work of Lamperti [19], passage-time moments were studied in [5], with refinements provided in [2, 3].

Parallel continuum developments concern reflecting Brownian motion in wedges in $\mathbb{R}^2$. In the zero-drift case with general (oblique) reflections, in the 1980s Varadhan and Williams [31] had showed that the process was well-defined, and then Williams [32] gave the recurrence classification, thus preceding the random walk results of [6, 12], and, in the recurrent cases, asymptotics of stationary measures (cf. [4] for the discrete setting). Passage-time moments were later studied in [7, 25], by providing a continuum version of the results of [5], and in [2], using discrete approximation [1]. The non-zero drift case was studied by Hobson and Rogers [16], who gave an analogue of Malyshev's theorem in the continuum setting.

For domains like our $\mathscr{D}$, Pinsky [28] established recurrence in the case of reflecting Brownian motion with normal reflections and standard covariance matrix in the interior. The case of general covariance matrix and oblique reflection does not appear to have been considered, and neither has the analysis of passage-time moments. The somewhat related problem of the asymptotics of the *first exit time* τ_e of planar Brownian motion from domains like our $\mathscr{D}$ has been considered [8, 9, 20]: in the case where $\beta^+ = \beta^- = \beta \in (0, 1)$, then $\log \mathbb{P}(\tau_e > t)$ is bounded above and below by constants times $-t^{(1-\beta)/(1+\beta)}$: see [20] and (for the case $\beta = 1/2$) [9].

1.5 Overview of the Proofs

The basic strategy is to construct suitable Lyapunov functions $f : \mathbb{R}^2 \to \mathbb{R}$ that satisfy appropriate semimartingale (i.e., drift) conditions on $\mathbb{E}_x[f(\xi_1) - f(\xi_0)]$ for x outside a bounded set. In fact, since the Lyapunov functions that we use are most suitable for the case where the interior increment covariance matrix is $\Sigma = I$, the identity, we first apply a linear transformation T of $\mathbb{R}^2$ and work with $T\xi$. The linear transformation is described in Sect. 2. Of course, one could combine these two steps and work directly with the Lyapunov function given by the composition $f \circ T$ for the appropriate f. However, for reasons of intuitive understanding and computational convenience, we prefer to separate the two steps.

Let $\beta^\pm < 1$. Then for $\alpha^+ = \alpha^- = 0$, the reflection angles are both pointing essentially vertically, with an asymptotically small component in the positive x_1 direction. After the linear transformation T, the reflection angles are no longer almost vertical, but instead are almost opposed at some oblique angle, where the deviation from direct opposition is again asymptotically small, and in the positive x_1 direction. For this reason, the case $\alpha^+ = -\alpha^- = \alpha \neq 0$ is not conceptually different from the simpler case where $\alpha = 0$, because after the linear transformation, both cases are oblique. In the case $\alpha \neq 0$, however, the details are more involved as both α and the value of the correlation ρ enter into the analysis of the Lyapunov functions, which is presented in Sect. 3, and is the main technical work of the paper. For $\beta^\pm > 1$, intuition is provided by the case of reflection in the half-plane (see e.g. [32] for the Brownian case).

Once the Lyapunov function estimates are in place, the proofs of the main theorems are given in Sect. 4, using some semimartingale results which are variations on those from [26]. The appendix contains the proof of Proposition 1 on the properties of the threshold function β_c defined at (5).

2 Linear Transformation

The inwards pointing normal vectors to $\partial\mathcal{D}$ at $(x_1, d^\pm(x_1))$ are

$$n^\pm(x_1) = \frac{1}{r^\pm(x_1)}\begin{pmatrix} a^\pm\beta^\pm x_1^{\beta^\pm-1} \\ \mp 1 \end{pmatrix}, \text{ where } r^\pm(x_1) := \sqrt{1+(a^\pm)^2(\beta^\pm)^2 x_1^{2\beta^\pm-2}}.$$

Define

$$n_\perp^\pm(x_1) := \frac{1}{r^\pm(x_1)}\begin{pmatrix} \pm 1 \\ a^\pm\beta^\pm x_1^{\beta^\pm-1} \end{pmatrix}.$$

Recall that $n^\pm(x_1, \alpha^\pm)$ is the unit vector at angle $\alpha^\pm$ to $n^\pm(x_1)$, with positive angles measured anticlockwise (for n^+) or clockwise (for n^-). Then (see Fig. 2 for the case of n^+) we have $n^\pm(x_1, \alpha^\pm) = n^\pm(x_1)\cos\alpha^\pm + n_\perp^\pm(x_1)\sin\alpha^\pm$, so

$$n^\pm(x_1, \alpha^\pm) = \frac{1}{r^\pm(x_1)}\begin{pmatrix} \sin\alpha^\pm + a^\pm\beta^\pm x_1^{\beta^\pm-1}\cos\alpha^\pm \\ \mp\cos\alpha^\pm \pm a^\pm\beta^\pm x_1^{\beta^\pm-1}\sin\alpha^\pm \end{pmatrix}.$$

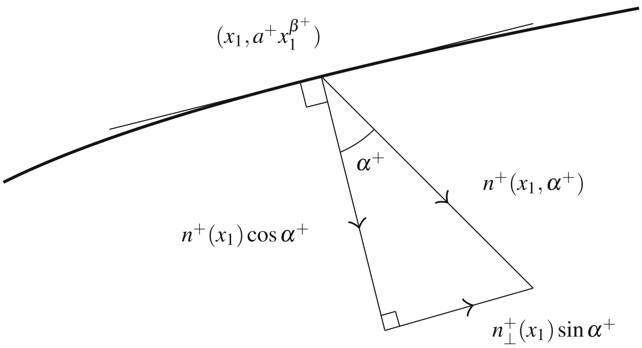


Fig. 2 Diagram describing oblique reflection at angle $\alpha^+ > 0$

In particular, if $\alpha^+ = -\alpha^- = \alpha$,

$$n^\pm(x_1, \alpha^\pm) = \frac{1}{r^\pm(x_1)} \begin{pmatrix} \pm \sin\alpha + a^\pm \beta^\pm x_1^{\beta^\pm - 1} \cos\alpha \\ \mp \cos\alpha + a^\pm \beta^\pm x_1^{\beta^\pm - 1} \sin\alpha \end{pmatrix} =: \begin{pmatrix} n_1^\pm(x_1, \alpha^\pm) \\ n_2^\pm(x_1, \alpha^\pm) \end{pmatrix}. \tag{8}$$

Recall that $\Delta = \xi_1 - \xi_0$. Write $\Delta = (\Delta_1, \Delta_2)$ in components.

Lemma 1 *Suppose that* (R) *holds, with* $\alpha^+ = -\alpha^- = \alpha$ *and* $\beta^+, \beta^- \geq 0$. *If* $\beta^\pm < 1$, *then, for* $x \in S_B^\pm$, *as* $\|x\| \to \infty$,

$$\begin{aligned} \mathbb{E}_x \Delta_1 &= \pm\mu^\pm(x) \sin\alpha + a^\pm \beta^\pm \mu^\pm(x) x_1^{\beta^\pm - 1} \cos\alpha \\ &\quad + O(\|x\|^{2\beta^\pm - 2}) + O(\|x\|^{-1}); \end{aligned} \tag{9}$$

$$\begin{aligned} \mathbb{E}_x \Delta_2 &= \mp\mu^\pm(x) \cos\alpha + a^\pm \beta^\pm \mu^\pm(x) x_1^{\beta^\pm - 1} \sin\alpha \\ &\quad + O(\|x\|^{2\beta^\pm - 2}) + O(\|x\|^{-1}). \end{aligned} \tag{10}$$

If $\beta^\pm > 1$, *then, for* $x \in S_B^\pm$, *as* $\|x\| \to \infty$,

$$\mathbb{E}_x \Delta_1 = \mu^\pm(x) \cos\alpha \pm \frac{\mu^\pm(x) \sin\alpha}{a^\pm \beta^\pm} x_1^{1-\beta^\pm} + O(x_1^{2-2\beta^\pm}) + O(\|x\|^{-1}); \tag{11}$$

$$\mathbb{E}_x \Delta_2 = \mu^\pm(x) \sin\alpha \mp \frac{\mu^\pm(x) \cos\alpha}{a^\pm \beta^\pm} x_1^{1-\beta^\pm} + O(x_1^{2-2\beta^\pm}) + O(\|x\|^{-1}). \tag{12}$$

Proof Suppose that $x \in S_B^\pm$. By (2), we have that $\| \mathbb{E}_x \Delta - \mu^\pm(x) n^\pm(x_1, \alpha^\pm) \| = O(\|x\|^{-1})$. First suppose that $0 \leq \beta^\pm < 1$. Then, $1/r^\pm(x_1) = 1 + O(x_1^{2\beta^\pm - 2})$, and hence, by (8),

$$\begin{aligned} n_1^\pm(x_1, \alpha^\pm) &= \pm \sin\alpha + a^\pm \beta^\pm x_1^{\beta^\pm - 1} \cos\alpha + O(x_1^{2\beta^\pm - 2}); \\ n_2^\pm(x_1, \alpha^\pm) &= \mp \cos\alpha + a^\pm \beta^\pm x_1^{\beta^\pm - 1} \sin\alpha + O(x_1^{2\beta^\pm - 2}). \end{aligned}$$

Then, since $\|x\| = x_1 + o(x_1)$ as $\|x\| \to \infty$ with $x \in \mathscr{D}$, we obtain (9) and (10).

On the other hand, if $\beta^\pm > 1$, then

$$\frac{1}{r^\pm(x_1)} = \frac{x_1^{1-\beta^\pm}}{a^\pm \beta^\pm} + O(x_1^{3-3\beta^\pm}),$$

and hence, by (8),

$$n_1^{\pm}(x_1, \alpha^{\pm}) = \cos\alpha \pm \frac{\sin\alpha}{a^{\pm}\beta^{\pm}} x_1^{1-\beta^{\pm}} + O(x_1^{2-2\beta^{\pm}});$$

$$n_2^{\pm}(x_1, \alpha^{\pm}) = \sin\alpha \mp \frac{\cos\alpha}{a^{\pm}\beta^{\pm}} x_1^{1-\beta^{\pm}} + O(x_1^{2-2\beta^{\pm}}).$$

The expressions (11) and (12) follow. □

It is convenient to introduce a linear transformation of $\mathbb{R}^2$ under which the asymptotic increment covariance matrix Σ appearing in (C) is transformed to the identity. Define

$$T := \begin{pmatrix} \frac{\sigma_2}{s} & -\frac{\rho}{s\sigma_2} \\ 0 & \frac{1}{\sigma_2} \end{pmatrix}, \text{ where } s := \sqrt{\det \Sigma} = \sqrt{\sigma_1^2\sigma_2^2 - \rho^2};$$

recall that $\sigma_2, s > 0$, since Σ is positive definite. The choice of T is such that $T\Sigma T^{\top} = I$ (the identity), and $x \mapsto Tx$ leaves the horizontal direction unchanged. Explicitly,

$$T\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{\sigma_2}{s}x_1 - \frac{\rho}{s\sigma_2}x_2 \\ \frac{1}{\sigma_2}x_2 \end{pmatrix}. \tag{13}$$

Note that T is positive definite, and so $\|Tx\|$ is bounded above and below by positive constants times $\|x\|$. Also, if $x \in \mathscr{D}$ and $\beta^+, \beta^- < 1$, the fact that $|x_2| = o(x_1)$ means that Tx has the properties (i) $(Tx)_1 > 0$ for all x_1 sufficiently large, and (ii) $|(Tx)_2| = o(|(Tx)_1|)$ as $x_1 \to \infty$. See Fig. 3 for a picture.

The next result describes the increment moment properties of the process under the transformation T. For convenience, we set $\tilde{\Delta} := T\Delta$ for the transformed increment, with components $\tilde{\Delta}_i = (T\Delta)_i$.

Lemma 2 *Suppose that* (D), (R), *and* (C) *hold, with* $\alpha^+ = -\alpha^- = \alpha$, *and* $\beta^+, \beta^- \geq 0$. *Then, if* $\|x\| \to \infty$ *with* $x \in S_I$,

$$\| \mathbb{E}_x \tilde{\Delta} \| = o(\|x\|^{-1}), \text{ and } \left\| \mathbb{E}_x(\tilde{\Delta}\tilde{\Delta}^{\top}) - I \right\|_{\text{op}} = o(1). \tag{14}$$

If, in addition, (D_+) *and* (C_+) *hold with* $\varepsilon > 0$, *then, if* $\|x\| \to \infty$ *with* $x \in S_I$,

$$\| \mathbb{E}_x \tilde{\Delta} \| = O(\|x\|^{-1-\varepsilon}), \text{ and } \left\| \mathbb{E}_x(\tilde{\Delta}\tilde{\Delta}^{\top}) - I \right\|_{\text{op}} = O(\|x\|^{-\varepsilon}). \tag{15}$$

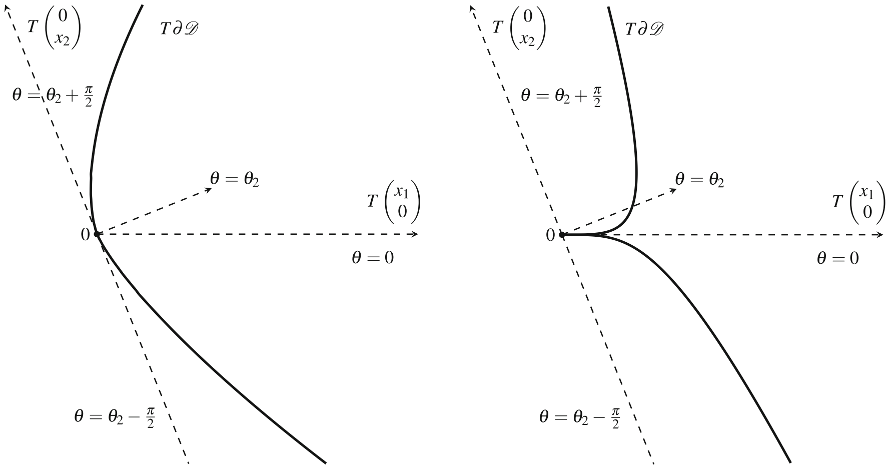


Fig. 3 An illustration of the transformation T with $\rho > 0$ acting on a domain $\mathscr{D}$ with $\beta^+ = \beta^- = \beta$ for $\beta \in (0, 1)$ (*left*) and $\beta > 1$ (*right*). The angle θ_2 is given by $\theta_2 = \arctan(\rho/s)$, measured anticlockwise from the positive horizontal axis

If $\beta^\pm < 1$, *then, as* $\|x\| \to \infty$ *with* $x \in S_B^\pm$,

$$\mathbb{E}_x \tilde{\Delta}_1 = \pm\frac{\sigma_2 \mu^\pm(x)}{s} \sin\alpha \pm \frac{\rho\mu^\pm(x)}{s\sigma_2} \cos\alpha + \frac{\sigma_2 a^\pm \beta^\pm \mu^\pm(x)}{s} x_1^{\beta^\pm - 1} \cos\alpha - \frac{\rho a^\pm \beta^\pm \mu^\pm(x)}{s\sigma_2} x_1^{\beta^\pm - 1} \sin\alpha + O(\|x\|^{2\beta^\pm - 2}) + O(\|x\|^{-1}); \tag{16}$$

$$\mathbb{E}_x \tilde{\Delta}_2 = \mp\frac{\mu^\pm(x)}{\sigma_2} \cos\alpha + \frac{a^\pm \beta^\pm \mu^\pm(x)}{\sigma_2} x_1^{\beta^\pm - 1} \sin\alpha + O(\|x\|^{2\beta^\pm - 2}) + O(\|x\|^{-1}). \tag{17}$$

If $\beta^\pm > 1$, *then, as* $\|x\| \to \infty$ *with* $x \in S_B^\pm$,

$$\mathbb{E}_x \tilde{\Delta}_1 = \frac{\sigma_2 \mu^\pm(x)}{s} \cos\alpha - \frac{\rho\mu^\pm(x)}{s\sigma_2} \sin\alpha \pm \frac{\sigma_2 \mu^\pm(x)}{a^\pm \beta^\pm s} x_1^{1-\beta^\pm} \sin\alpha \pm \frac{\rho\mu^\pm(x)}{a^\pm \beta^\pm s\sigma_2} x_1^{1-\beta^\pm} \cos\alpha + O(x_1^{2-2\beta^\pm}) + O(\|x\|^{-1}); \tag{18}$$

$$\mathbb{E}_x \tilde{\Delta}_2 = \frac{\mu^\pm(x)}{\sigma_2} \sin\alpha \mp \frac{\mu^\pm(x)}{a^\pm \beta^\pm \sigma_2} x_1^{1-\beta^\pm} \cos\alpha + O(x_1^{2-2\beta^\pm}) + O(\|x\|^{-1}). \tag{19}$$

Proof By linearity,

$$\mathbb{E}_x \tilde{\Delta} = T \,\mathbb{E}_x \Delta, \tag{20}$$

which, by (D) or (D_+), is, respectively, $o(\|x\|^{-1})$ or $O(\|x\|^{-1-\varepsilon})$ for $x \in S_I$. Also, since $T\Sigma T^\top = I$, we have

$$\mathbb{E}_x(\tilde{\Delta}\tilde{\Delta}^\top) - I = T\,\mathbb{E}_x(\Delta\Delta^\top)T^\top - I = T\left(\mathbb{E}_x(\Delta\Delta^\top) - \Sigma\right)T^\top.$$

For $x \in S_I$, the middle matrix in the last product here has norm $o(1)$ or $O(\|x\|^{-\varepsilon})$, by (C) or (C_+). Thus we obtain (14) and (15). For $x \in S_B^{\pm}$, the claimed results follow on using (20), (13), and the expressions for $\mathbb{E}_x \Delta$ in Lemma 1. □

3 Lyapunov Functions

For the rest of the paper, we suppose that $\alpha^+ = -\alpha^- = \alpha$ for some $|\alpha| < \pi/2$. Our proofs will make use of some carefully chosen functions of the process. Most of these functions are most conveniently expressed in polar coordinates.

We write $x = (r, \theta)$ in polar coordinates, with angles measured relative to the positive horizontal axis: $r := r(x) := \|x\|$ and $\theta := \theta(x) \in (-\pi, \pi]$ is the angle between the ray through 0 and x and the ray in the Cartesian direction $(1, 0)$, with the convention that anticlockwise angles are positive. Then $x_1 = r\cos\theta$ and $x_2 = r\sin\theta$.

For $w \in \mathbb{R}$, $\theta_0 \in (-\pi/2, \pi/2)$, and $\gamma \in \mathbb{R}$, define

$$h_w(x) := h_w(r, \theta) := r^w \cos(w\theta - \theta_0), \text{ and } f_w^\gamma(x) := (h_w(Tx))^\gamma, \tag{21}$$

where T is the linear transformation described at (13). The functions h_w were used in analysis of processes in wedges in e.g. [5, 21, 29, 31]. Since the h_w are harmonic for the Laplacian (see below for a proof), Lemma 2 suggests that $h_w(T\xi_n)$ will be approximately a martingale in S_I, and the choice of the geometrical parameter θ_0 gives us the flexibility to try to arrange things so that the level curves of h_w are incident to the boundary at appropriate angles relative to the reflection vectors. The level curves of h_w cross the horizontal axis at angle θ_0: see Fig. 4, and (33) below. In the case $\beta^{\pm} < 1$, the interest is near the horizontal axis, and we take θ_0 to be such that the level curves cut $\partial\mathscr{D}$ at the reflection angles (asymptotically), so that $h_w(T\xi_n)$ will be approximately a martingale also in S_B. Then adjusting w and γ will enable us to obtain a supermartingale with the properties suitable to apply some Foster–Lyapunov theorems. This intuition is solidified in Lemma 4 below, where we show that the parameters w, θ_0, and γ can be chosen so that $f_w^\gamma(\xi_n)$ satisfies an appropriate supermartingale condition outside a bounded set. For the case $\beta^{\pm} < 1$, since we only need to consider $\theta \approx 0$, we could replace these harmonic functions in polar coordinates by suitable polynomial approximations in Cartesian components,

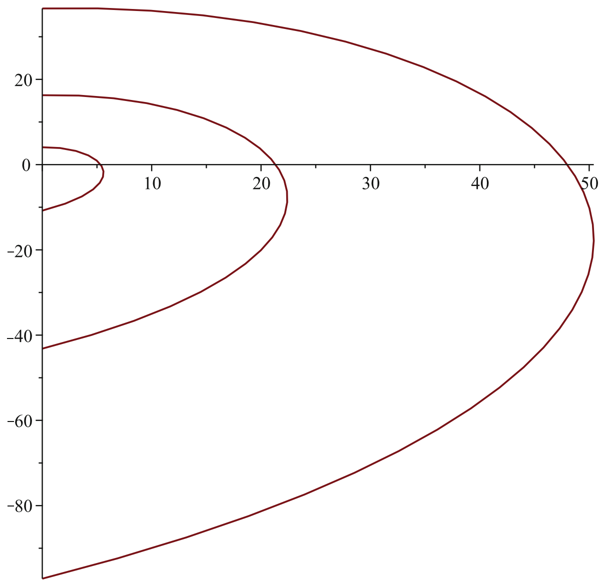


Fig. 4 Level curves of the function $h_w(x)$ with $\theta_0 = \pi/6$ and $w = 1/4$. The level curves cut the horizontal axis at angle θ_0 to the vertical

but since we also want to consider $\beta^{\pm} > 1$, it is convenient to use the functions in the form given. When $\beta^{\pm} > 1$, the recurrence classification is particularly delicate, so we must use another function (see (57) below), although the functions at (21) will still be used to study passage time moments in that case.

If $\beta^+, \beta^- < 1$, then $\theta(x) \to 0$ as $\|x\| \to \infty$ with $x \in \mathscr{D}$, which means that, for any $|\theta_0| < \pi/2$, $h_w(x) \geq \delta \|x\|^w$ for some $\delta > 0$ and all $x \in S$ with $\|x\|$ sufficiently large. On the other hand, for $\beta^+, \beta^- > 1$, we will restrict to the case with $w > 0$ sufficiently small such that $\cos(w\theta - \theta_0)$ is bounded away from zero, uniformly in $\theta \in [-\pi/2, \pi/2]$, so that we again have the estimate $h_w(x) \geq \delta \|x\|^w$ for some $\delta > 0$ and all $x \in \mathscr{D}$, but where now $\mathscr{D}$ is close to the whole half-plane (see Remark 4). In the calculations that follow, we will often use the fact that $h_w(x)$ is bounded above and below by a constant times $\|x\|^w$ as $\|x\| \to \infty$ with $x \in \mathscr{D}$.

We use the notation $D_i := \frac{\mathrm{d}}{\mathrm{d}x_i}$ for differentials, and for $f : \mathbb{R}^2 \to \mathbb{R}$ write Df for the vector with components $(Df)_i = D_i f$. We use repeatedly

$$D_1 r = \cos\theta, \ \ D_2 r = \sin\theta, \ \ D_1\theta = -\frac{\sin\theta}{r}, \ \ D_2\theta = \frac{\cos\theta}{r}. \tag{22}$$

Define

$$\theta_1 := \theta_1(\Sigma, \alpha) := \arctan\left(\frac{\sigma_2^2}{s}\tan\alpha + \frac{\rho}{s}\right) \in (-\pi/2, \pi/2). \tag{23}$$

For $\beta^{\pm} > 1$, we will also need

$$\theta_2 := \theta_2(\Sigma) := \arctan\left(\frac{\rho}{s}\right) \in (-\pi/2, \pi/2), \tag{24}$$

and $\theta_3 := \theta_3(\Sigma, \alpha) \in (-\pi, \pi)$ for which

$$\sin\theta_3 = \frac{s\sin\alpha}{\sigma_2 d}, \text{ and } \cos\theta_3 = \frac{\sigma_2^2\cos\alpha - \rho\sin\alpha}{\sigma_2 d}, \tag{25}$$

where

$$d := d(\Sigma, \alpha) := \sqrt{\sigma_2^2\cos^2\alpha - 2\rho\sin\alpha\cos\alpha + \sigma_1^2\sin^2\alpha}. \tag{26}$$

The geometric interpretation of θ_1, θ_2, and θ_3 is as follows.

- The angle between $(0, \pm 1)$ and $T(0, \pm 1)$ has magnitude θ_2. Thus, if $\beta^{\pm} < 1$, then θ_2 is, as $x_1 \to \infty$, the limiting angle of the transformed inwards pointing normal at x_1 relative to the vertical. On the other hand, if $\beta^{\pm} > 1$, then θ_2 is, as $x_1 \to \infty$, the limiting angle, relative to the horizontal, of the inwards pointing normal to $T\partial\mathcal{D}$. See Fig. 3.
- The angle between $(0, -1)$ and $T(\sin\alpha, -\cos\alpha)$ is θ_1. Thus, if $\beta^{\pm} < 1$, then θ_1 is, as $x_1 \to \infty$, the limiting angle between the vertical and the transformed reflection vector. Since the normal in the transformed domain remains asymptotically vertical, θ_1 is in this case the limiting reflection angle, relative to the normal, after the transformation.
- The angle between $(1, 0)$ and $T(\cos\alpha, \sin\alpha)$ is θ_3. Thus, if $\beta^{\pm} > 1$, then θ_3 is, as $x_1 \to \infty$, the limiting angle between the horizontal and the transformed reflection vector. Since the transformed normal is, asymptotically, at angle θ_2 relative to the horizontal, the limiting reflection angle, relative to the normal, after the transformation is in this case $\theta_3 - \theta_2$.

We need two simple facts.

Lemma 3 *We have (i)* $\inf_{\alpha\in[-\frac{\pi}{2},\frac{\pi}{2}]} d(\Sigma, \alpha) > 0$*, and (ii)* $|\theta_3 - \theta_2| < \pi/2$.

Proof For (i), from (26) we may write

$$d^2 = \sigma_2^2 + \left(\sigma_1^2 - \sigma_2^2\right)\sin^2\alpha - \rho\sin 2\alpha. \tag{27}$$

If $\sigma_1^2 \neq \sigma_2^2$, then, by Lemma 11, the extrema over $\alpha \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ of (27) are

$$\sigma_2^2 + \frac{\sigma_1^2 - \sigma_2^2}{2}\left(1 \pm \sqrt{1 + \frac{4\rho^2}{(\sigma_1^2 - \sigma_2^2)^2}}\right).$$

Hence

$$d^2 \geq \frac{\sigma_1^2 + \sigma_2^2}{2} - \frac{1}{2}\sqrt{(\sigma_1^2 - \sigma_2^2)^2 + 4\rho^2},$$

which is strictly positive since $\rho^2 < \sigma_1^2\sigma_2^2$. If $\sigma_1^2 = \sigma_2^2$, then $d^2 \geq \sigma_2^2 - |\rho|$, and $|\rho| < |\sigma_1\sigma_2| = \sigma_2^2$, so d is also strictly positive in that case.

For (ii), we use the fact that $\cos(\theta_3 - \theta_2) = \cos\theta_3\cos\theta_2 + \sin\theta_3\sin\theta_2$, where, by (24), $\sin\theta_2 = \frac{\rho}{\sigma_1\sigma_2}$ and $\cos\theta_2 = \frac{s}{\sigma_1\sigma_2}$, and (25), to get $\cos(\theta_3 - \theta_2) = \frac{s}{\sigma_1 d}\cos\alpha > 0$. Since $|\theta_3 - \theta_2| < 3\pi/2$, it follows that $|\theta_3 - \theta_2| < \pi/2$, as claimed. □

We estimate the expected increments of our Lyapunov functions in two stages: the main term comes from a Taylor expansion valid when the jump of the walk is not too big compared to its current distance from the origin, while we bound the (smaller) contribution from big jumps using the moments assumption (M_p). For the first stage, let $B_b(x) := \{z \in \mathbb{R}^2 : \|x - z\| \leq b\}$ denote the (closed) Euclidean ball centred at x with radius $b \geq 0$. We use the multivariable Taylor theorem in the following form. Suppose that $f : \mathbb{R}^2 \to \mathbb{R}$ is thrice continuously differentiable in $B_b(x)$. Recall that $Df(x)$ is the vector function whose components are $D_i f(x)$. Then, for $y \in B_b(x)$,

$$f(x + y) = f(x) + \langle Df(x), y\rangle + y_1^2\frac{D_1^2 f(x)}{2} + y_2^2\frac{D_2^2 f(x)}{2} + y_1 y_2 D_1 D_2 f(x) + R(x, y), \tag{28}$$

where, for all $y \in B_b(x)$, $|R(x, y)| \leq C\|y\|^3 R(x)$ for an absolute constant $C < \infty$ and

$$R(x) := \max_{i,j,k} \sup_{z \in B_b(x)} \left|D_i D_j D_k f(z)\right|.$$

For dealing with the large jumps, we observe the useful fact that if $p > 2$ is a constant for which (1) holds, then for some constant $C < \infty$, all $\delta \in (0, 1)$, and all $q \in [0, p]$,

$$\mathbb{E}_x\left[\|\Delta\|^q \mathbf{1}\{\|\Delta\| \geq \|x\|^\delta\}\right] \leq C\|x\|^{-\delta(p-q)}, \tag{29}$$

for all $\|x\|$ sufficiently large. To see (29), write $\|\Delta\|^q = \|\Delta\|^p\|\Delta\|^{q-p}$ and use the fact that $\|\Delta\| \geq \|x\|^\delta$ to bound the second factor.

Here is our first main Lyapunov function estimate.

Lemma 4 *Suppose that* (M_p), (D), (R), *and* (C) *hold, with* $p > 2$, $\alpha^+ = -\alpha^- = \alpha$ *for* $|\alpha| < \pi/2$, *and* $\beta^+, \beta^- \geq 0$. *Let* $w, \gamma \in \mathbb{R}$ *be such that* $2 - p < \gamma w < p$. *Take*

$\theta_0 \in (-\pi/2, \pi/2)$. Then as $\|x\| \to \infty$ with $x \in S_I$,

$$\mathbb{E}[f_w^\gamma(\xi_{n+1}) - f_w^\gamma(\xi_n) \mid \xi_n = x] = \frac{\gamma(\gamma-1)}{2} w^2 (h_w(Tx))^{\gamma-2} \|Tx\|^{2w-2} + o(\|x\|^{\gamma w - 2}). \tag{30}$$

We separate the boundary behaviour into two cases.

(i) If $0 \le \beta^\pm < 1$, take $\theta_0 = \theta_1$ given by (23). Then, as $\|x\| \to \infty$ with $x \in S_B^\pm$,

$$\mathbb{E}[f_w^\gamma(\xi_{n+1}) - f_w^\gamma(\xi_n) \mid \xi_n = x] \\ = \gamma w \|Tx\|^{w-1} (h_w(Tx))^{\gamma-1} \frac{a^\pm \mu^\pm(x) \sigma_2 \cos\theta_1}{s \cos\alpha} \left(\beta^\pm - (1-w)\beta_c\right) x_1^{\beta^\pm - 1} \\ + o(\|x\|^{w\gamma + \beta^\pm - 2}), \tag{31}$$

where β_c is given by (5).

(ii) If $\beta^\pm > 1$, suppose that $w \in (0, 1/2)$ and $\theta_0 = \theta_0(\Sigma, \alpha, w) = \theta_3 - (1-w)\theta_2$, where θ_2 and θ_3 are given by (24) and (25), such that $\sup_{\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]} |w\theta - \theta_0| < \pi/2$. Then, with $d = d(\Sigma, \alpha)$ as defined at (26), as $\|x\| \to \infty$ with $x \in S_B^\pm$,

$$\mathbb{E}[f_w^\gamma(\xi_{n+1}) - f_w^\gamma(\xi_n) \mid \xi_n = x] \\ = \gamma w \|Tx\|^{w-1} (h_w(Tx))^{\gamma-1} \frac{d\mu^\pm(x)}{s} \left(\cos((1-w)(\pi/2)) + o(1)\right). \tag{32}$$

Remark 4 We can choose $w > 0$ small enough so that $|\theta_3 - (1-w)\theta_2| < \pi/2$, by Lemma 3(ii), and so if $\theta_0 = \theta_3 - (1-w)\theta_2$, we can always choose $w > 0$ small enough so that $\sup_{\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]} |w\theta - \theta_0| < \pi/2$, as required for the $\beta^\pm > 1$ part of Lemma 4.

Proof (of Lemma 4) Differentiating (21) and using (22) we see that

$$D_1 h_w(x) = w r^{w-1} \cos\left((w-1)\theta - \theta_0\right), \text{ and} \\ D_2 h_w(x) = -w r^{w-1} \sin\left((w-1)\theta - \theta_0\right). \tag{33}$$

Moreover,

$$D_1^2 h_w(x) = w(w-1) r^{w-2} \cos\left((w-2)\theta - \theta_0\right) = -D_2^2 h_w(x),$$

verifying that h_w is harmonic. Also, for any i, j, k, $|D_i D_j D_k h_w(x)| = O(r^{w-3})$. Writing $h_w^\gamma(x) := (h_w(x))^\gamma$, we also have that $D_i h_w^\gamma(x) = \gamma h_w^{\gamma-1}(x) D_i h_w(x)$, that

$$D_i D_j h_w^\gamma(x) = \gamma h_w^{\gamma-1}(x) D_i D_j h_w(x) + \gamma(\gamma-1) h_w^{\gamma-2}(x) (D_i h_w(x))(D_j h_w(x)),$$

and $|D_i D_j D_k h_w^\gamma(x)| = O(r^{\gamma w-3})$. We apply Taylor's formula (28) in the ball $B_{r/2}(x)$ together with the harmonic property of h_w, to obtain, for $y \in B_{r/2}(x)$,

$$\begin{aligned} h_w^\gamma(x+y) &= h_w^\gamma(x) + \gamma \langle Dh_w(x), y\rangle h_w^{\gamma-1}(x) + \frac{\gamma(\gamma-1)}{2}\langle Dh_w(x), y\rangle^2 h_w^{\gamma-2}(x) \\ &\quad + \gamma\left(\frac{(y_1^2-y_2^2)D_1^2 h_w(x)}{2} + y_1 y_2 D_1 D_2 h_w(x)\right) h_w^{\gamma-1}(x) \\ &\quad + R(x,y), \end{aligned} \tag{34}$$

where $|R(x,y)| \le C\|y\|^3 \|x\|^{\gamma w-3}$, using the fact that $h_w(x)$ is bounded above and below by a constant times $\|x\|^w$.

Let $E_x := \{\|\Delta\| < \|x\|^\delta\}$, where we fix a constant δ satisfying

$$\frac{\max\{2, \gamma w, 2-\gamma w\}}{p} < \delta < 1; \tag{35}$$

such a choice of δ is possible since $p > 2$ and $2 - p < \gamma w < p$. If $\xi_0 = x$ and E_x occurs, then $Tx + \tilde{\Delta} \in B_{r/2}(Tx)$ for all $\|x\|$ sufficiently large. Thus, conditioning on $\xi_0 = x$, on the event E_x we may use the expansion in (34) for $h_w^\gamma(Tx + \tilde{\Delta})$, which, after taking expectations, yields

$$\begin{aligned} \mathbb{E}_x\left[(f_w^\gamma(\xi_1) - f_w^\gamma(\xi_0))\mathbf{1}_{E_x}\right] &= \gamma\,(h_w(Tx))^{\gamma-1}\,\mathbb{E}_x\left[\langle Dh_w(Tx), \tilde{\Delta}\rangle \mathbf{1}_{E_x}\right] \\ &\quad + \gamma\,(h_w(Tx))^{\gamma-1}\left[\frac{D_1^2 h_w(Tx)\,\mathbb{E}_x\left[(\tilde{\Delta}_1^2 - \tilde{\Delta}_2^2)\mathbf{1}_{E_x}\right]}{2} + D_1 D_2 h_w(Tx)\,\mathbb{E}_x\left[\tilde{\Delta}_1\tilde{\Delta}_2\mathbf{1}_{E_x}\right]\right] \\ &\quad + \frac{\gamma(\gamma-1)}{2}\,(h_w(Tx))^{\gamma-2}\,\mathbb{E}_x\left[\langle Dh_w(Tx), \tilde{\Delta}\rangle^2 \mathbf{1}_{E_x}\right] + \mathbb{E}_x\left[R(Tx, \tilde{\Delta})\mathbf{1}_{E_x}\right]. \end{aligned} \tag{36}$$

Let $p' = p \wedge 3$, so that (1) also holds for $p' \in (2, 3]$. Then, writing $\|\tilde{\Delta}\|^3 = \|\tilde{\Delta}\|^{p'}\|\tilde{\Delta}\|^{3-p'}$,

$$\mathbb{E}_x\left[|R(Tx, \tilde{\Delta})|\mathbf{1}_{E_x}\right] \le C\|x\|^{\gamma w - 3 + (3-p')\delta}\,\mathbb{E}_x\left[\|\tilde{\Delta}\|^{p'}\right] = o(\|x\|^{\gamma w-2}),$$

since $(3 - p')\delta < 1$. If $x \in S_I$, then (14) shows $|\,\mathbb{E}_x\langle Dh_w(Tx), \tilde{\Delta}\rangle| = o(\|x\|^{w-2})$, so

$$\mathbb{E}_x\left|\langle Dh_w(Tx), \tilde{\Delta}\rangle \mathbf{1}_{E_x}\right| \le C\|x\|^{w-1}\,\mathbb{E}_x(\|\Delta\|\mathbf{1}_{E_x^c}) + o(\|x\|^{w-2}).$$

Note that, by (35), $\delta > \frac{2}{p} > \frac{1}{p-1}$. Then, using the $q = 1$ case of (29), we get

$$\mathbb{E}_x\left|\langle Dh_w(Tx), \tilde{\Delta}\rangle \mathbf{1}_{E_x}\right| = o(\|x\|^{w-2}). \tag{37}$$

A similar argument using the $q = 2$ case of (29) gives

$$\mathbb{E}_x \left[\langle D h_w(Tx), \tilde{\Delta} \rangle^2 \mathbf{1}_{E_x^c} \right] \le C \|x\|^{2w-2-\delta(p-2)} = o(\|x\|^{2w-2}).$$

If $x \in S_I$, then (14) shows that $\mathbb{E}_x(\tilde{\Delta}_1^2 - \tilde{\Delta}_2^2)$ and $\mathbb{E}_x(\tilde{\Delta}_1 \tilde{\Delta}_2)$ are both $o(1)$, and hence, by the $q = 2$ case of (29) once more, we see that $\mathbb{E}_x[|\tilde{\Delta}_1^2 - \tilde{\Delta}_2^2| \mathbf{1}_{E_x}]$ and $\mathbb{E}_x[|\tilde{\Delta}_1 \tilde{\Delta}_2| \mathbf{1}_{E_x}]$ are both $o(1)$. Moreover, (14) also shows that

$$\begin{aligned}
\mathbb{E}_x \langle D h_w(Tx), \tilde{\Delta} \rangle^2 &= \mathbb{E}_x \left((D h_w(Tx))^\top \tilde{\Delta} \tilde{\Delta}^\top D h_w(Tx) \right) \\
&= (D h_w(Tx))^\top D h_w(Tx) + o(\|x\|^{2w-2}) \\
&= (D_1 h_w(Tx))^2 + (D_2 h_w(Tx))^2 + o(\|x\|^{2w-2}).
\end{aligned}$$

Putting all these estimates into (36) we get, for $x \in S_I$,

$$\begin{aligned}
\mathbb{E}_x \left[(f_w^\gamma(\xi_1) - f_w^\gamma(\xi_0)) \mathbf{1}_{E_x} \right] = \frac{\gamma(\gamma-1)}{2} (h_w(Tx))^{\gamma-2} \left((D_1 h_w(Tx))^2 + (D_2 h_w(Tx))^2 \right) \\
+ o(\|x\|^{\gamma w - 2}). \qquad (38)
\end{aligned}$$

On the other hand, given $\xi_0 = x$, if $\gamma w \ge 0$, by the triangle inequality,

$$\begin{aligned}
\left| f_w^\gamma(\xi_1) - f_w^\gamma(x) \right| &\le \|T\xi_1\|^{\gamma w} + \|Tx\|^{\gamma w} \le 2\big(\|T\xi_1\| + \|Tx\|\big)^{\gamma w} \\
&\le 2\big(2\|Tx\| + \|\tilde{\Delta}\|\big)^{\gamma w}. \qquad (39)
\end{aligned}$$

It follows from (39) that $|f_w^\gamma(\xi_1) - f_w^\gamma(x)| \mathbf{1}_{E_x^c} \le C \|\Delta\|^{\gamma w/\delta}$, for some constant $C < \infty$ and all $\|x\|$ sufficiently large. Hence

$$\mathbb{E}_x \left| (f_w^\gamma(\xi_1) - f_w^\gamma(\xi_0)) \mathbf{1}_{E_x^c} \right| \le C\, \mathbb{E}_x \left[\|\Delta\|^{\gamma w/\delta} \mathbf{1}_{E_x^c} \right].$$

Since $\delta > \frac{\gamma w}{p}$, by (35), we may apply (29) with $q = \frac{\gamma w}{\delta}$ to get

$$\mathbb{E}_x \left| (f_w^\gamma(\xi_1) - f_w^\gamma(\xi_0)) \mathbf{1}_{E_x^c} \right| = O(\|x\|^{\gamma w - \delta p}) = o(\|x\|^{\gamma w - 2}), \qquad (40)$$

since $\delta > \frac{2}{p}$. If $w\gamma < 0$, then we use the fact that f_w^γ is uniformly bounded to get

$$\mathbb{E}_x \left| (f_w^\gamma(\xi_1) - f_w^\gamma(\xi_0)) \mathbf{1}_{E_x^c} \right| \le C \mathbb{P}_x(E_x^c) = O(\|x\|^{-\delta p}),$$

by the $q = 0$ case of (29). Thus (40) holds in this case too, since $\gamma w > 2 - \delta p$ by choice of δ at (35). Then (30) follows from combining (38) and (40) with (33).

Next suppose that $x \in S_B$. Truncating (34), we see that for all $y \in B_{r/2}(x)$,

$$h_w^\gamma(x+y) = h_w^\gamma(x) + \gamma \langle D h_w(x), y \rangle h_w^{\gamma-1}(x) + R(x,y), \qquad (41)$$

where now $|R(x,y)| \le C\|y\|^2\|x\|^{\gamma w-2}$. It follows from (41) and (M_p) that

$$\mathbb{E}_x\left[(f_w^\gamma(\xi_1) - f_w^\gamma(\xi_0))\mathbf{1}_{E_x}\right] = \gamma h_w^{\gamma-1}(Tx)\,\mathbb{E}_x\left[\langle Dh_w(Tx), \tilde{\Delta}\rangle \mathbf{1}_{E_x}\right] + O(\|x\|^{\gamma w-2}).$$

By the $q = 1$ case of (29), since $\delta > \frac{1}{p-1}$, we see that $\mathbb{E}_x[\langle Dh_w(Tx), \tilde{\Delta}\rangle \mathbf{1}_{E_x^{\mathrm{c}}}] = o(\|x\|^{w-2})$, while the estimate (40) still applies, so that

$$\mathbb{E}_x\left[f_w^\gamma(\xi_1) - f_w^\gamma(\xi_0)\right] = \gamma h_w^{\gamma-1}(Tx)\,\mathbb{E}_x\langle Dh_w(Tx), \tilde{\Delta}\rangle + O(\|x\|^{\gamma w-2}). \tag{42}$$

From (33) we have

$$Dh_w(Tx) = w\|Tx\|^{w-1}\begin{pmatrix}\cos((1-w)\theta(Tx)+\theta_0)\\ \sin((1-w)\theta(Tx)+\theta_0)\end{pmatrix}. \tag{43}$$

First suppose that $\beta^\pm < 1$. Then, by (13), for $x \in S_B^\pm$, $x_2 = \pm a^\pm x_1^{\beta^\pm} + O(1)$ and

$$\sin\theta(Tx) = \pm\frac{sa^\pm}{\sigma_2^2}x_1^{\beta^\pm-1} + O(x_1^{2\beta^\pm-2}) + O(x_1^{-1}).$$

Since $\arcsin z = z + O(z^3)$ as $z \to 0$, it follows that

$$\theta(Tx) = \pm\frac{sa^\pm}{\sigma_2^2}x_1^{\beta^\pm-1} + O(x_1^{2\beta^\pm-2}) + O(x_1^{-1}).$$

Hence

$$\cos((1-w)\theta(Tx)+\theta_0) = \cos\theta_0 \mp (1-w)\frac{sa^\pm}{\sigma_2^2}x_1^{\beta^\pm-1}\sin\theta_0 + O(x_1^{2\beta^\pm-2}) + O(x_1^{-1});$$

$$\sin((1-w)\theta(Tx)+\theta_0) = \sin\theta_0 \pm (1-w)\frac{sa^\pm}{\sigma_2^2}x_1^{\beta^\pm-1}\cos\theta_0 + O(x_1^{2\beta^\pm-2}) + O(x_1^{-1}).$$

Then (43) with (16) and (17) shows that

$$\begin{aligned}&\mathbb{E}_x\langle Dh_w(Tx), \tilde{\Delta}\rangle\\ &\quad = w\|Tx\|^{w-1}\frac{\mu^\pm(x)\cos\theta_0\cos\alpha}{s\sigma_2}\left(\pm A_1 + (a^\pm A_2 + o(1))x_1^{\beta^\pm-1}\right),\end{aligned} \tag{44}$$

where, for $|\theta_0| < \pi/2$, $A_1 = \sigma_2^2\tan\alpha + \rho - s\tan\theta_0$, and

$$\begin{aligned}A_2 = {}&\sigma_2^2\beta^\pm - \rho\beta^\pm\tan\alpha - (1-w)s\tan\theta_0\tan\alpha - (1-w)\frac{s\rho}{\sigma_2^2}\tan\theta_0\\ &+ s\beta^\pm\tan\theta_0\tan\alpha - (1-w)\frac{s^2}{\sigma_2^2}.\end{aligned}$$

Now take $\theta_0 = \theta_1$ as given by (23), so that $s \tan \theta_0 = \sigma_2^2 \tan \alpha + \rho$. Then $A_1 = 0$, eliminating the leading order term in (44). Moreover, with this choice of θ_0 we get, after some further cancellation and simplification, that

$$A_2 = \frac{\sigma_2^2 \left(\beta^{\pm} - (1-w)\beta_{\mathrm{c}}\right)}{\cos^2 \alpha},$$

with β_{c} as given by (5). Thus with (44) and (42) we verify (31).

Finally suppose that $\beta^{\pm} > 1$, and restrict to the case $w \in (0, 1/2)$. Let $\theta_2 \in (-\pi/2, \pi/2)$ be as given by (24). Then if $x = (0, x_2)$, we have $\theta(Tx) = \theta_2 - \frac{\pi}{2}$ if $x_2 < 0$ and $\theta(Tx) = \theta_2 + \frac{\pi}{2}$ if $x_2 > 0$ (see Fig. 3). It follows from (13) that

$$\theta(Tx) = \theta_2 \pm \frac{\pi}{2} + O(x_1^{1-\beta^{\pm}}), \text{ for } x \in S_B^{\pm},$$

as $\|x\| \to \infty$ (and $x_1 \to \infty$). Now (43) with (18) and (19) shows that

$$\begin{aligned} \mathbb{E}_x \langle Dh_w(Tx), \tilde{\Delta} \rangle = w \|Tx\|^{w-1} \frac{\mu^{\pm}(x)}{s\sigma_2} \Big(& \sigma_2^2 \cos \alpha \cos\left((1-w)\theta(Tx) + \theta_0\right) \\ & - \rho \sin \alpha \cos\left((1-w)\theta(Tx) + \theta_0\right) \\ & + s \sin \alpha \sin\left((1-w)\theta(Tx) + \theta_0\right) + O(x_1^{1-\beta^{\pm}}) \Big). \end{aligned} \tag{45}$$

Set $\phi := (1-w)\frac{\pi}{2}$. Choose $\theta_0 = \theta_3 - (1-w)\theta_2$, where $\theta_3 \in (-\pi, \pi)$ satisfies (25). Then we have that, for $x \in S_B^{\pm}$,

$$\begin{aligned} \cos\left((1-w)\theta(Tx) + \theta_0\right) &= \cos\left(\theta_3 \pm \phi\right) + O(x_1^{1-\beta^{\pm}}) \\ &= \cos \phi \cos \theta_3 \mp \sin \phi \sin \theta_3 + O(x_1^{1-\beta^{\pm}}). \end{aligned} \tag{46}$$

Similarly, for $x \in S_B^{\pm}$,

$$\sin\left((1-w)\theta(Tx) + \theta_0\right) = \cos \phi \sin \theta_3 \pm \sin \phi \cos \theta_3 + O(x_1^{1-\beta^{\pm}}). \tag{47}$$

Using (46) and (47) in (45), we obtain

$$\mathbb{E}_x \langle Dh_w(Tx), \tilde{\Delta} \rangle = w \|Tx\|^{w-1} \frac{\mu^{\pm}(x)}{s\sigma_2} \left(A_3 \cos \phi \mp A_4 \sin \phi + o(1)\right),$$

where

$$\begin{aligned} A_3 &= \left(\sigma_2^2 \cos \alpha - \rho \sin \alpha\right) \cos \theta_3 + s \sin \alpha \sin \theta_3 \\ &= \sigma_2 d \cos^2 \theta_3 + \sigma_2 d \sin^2 \theta_3 = \sigma_2 d, \end{aligned}$$

by (25), and, similarly,

$$A_4 = \left(\sigma_2^2 \cos\alpha - \rho \sin\alpha\right) \sin\theta_3 - s \sin\alpha \cos\theta_3 = 0.$$

Then with (42) we obtain (32). □

In the case where $\beta^+, \beta^- < 1$ with $\beta^+ \neq \beta^-$, we will in some circumstances need to modify the function f_w^γ so that it can be made insensitive to the behaviour near the boundary with the smaller of β^+, β^-. To this end, define for $w, \gamma, \nu, \lambda \in \mathbb{R}$,

$$F_w^{\gamma,\nu}(x) := f_w^\gamma(x) + \lambda x_2 \|Tx\|^{2\nu}. \tag{48}$$

We state a result for the case $\beta^- < \beta^+$; an analogous result holds if $\beta^+ < \beta^-$.

Lemma 5 *Suppose that* (M_p)*,* (D)*,* (R)*, and* (C) *hold, with* $p > 2$*,* $\alpha^+ = -\alpha^- = \alpha$ *for* $|\alpha| < \pi/2$*, and* $0 \leq \beta^- < \beta^+ < 1$*. Let* $w, \gamma \in \mathbb{R}$ *be such that* $2-p < \gamma w < p$*. Take* $\theta_0 = \theta_1 \in (-\pi/2, \pi/2)$ *given by* (23)*. Suppose that*

$$\gamma w + \beta^- - 2 < 2\nu < \gamma w + \beta^+ - 2.$$

Then as $\|x\| \to \infty$ *with* $x \in S_I$*,*

$$\begin{aligned}
&\mathbb{E}[F_w^{\gamma,\nu}(\xi_{n+1}) - F_w^{\gamma,\nu}(\xi_n) \mid \xi_n = x] \\
&\quad = \frac{1}{2}\gamma(\gamma-1)(w^2 + o(1))(h_w(Tx))^{\gamma-2}\|Tx\|^{2w-2}.
\end{aligned} \tag{49}$$

As $\|x\| \to \infty$ *with* $x \in S_B^+$*,*

$$\begin{aligned}
&\mathbb{E}[F_w^{\gamma,\nu}(\xi_{n+1}) - F_w^{\gamma,\nu}(\xi_n) \mid \xi_n = x] \\
&\quad = \gamma w \|Tx\|^{w-1} (h_w(Tx))^{\gamma-1} \frac{a^+\mu^+(x)\sigma_2 \cos\theta_1}{s\cos\alpha} \left(\beta^+ - (1-w)\beta_c\right) x_1^{\beta^+-1} \\
&\qquad + o(\|x\|^{w\gamma+\beta^+-2}).
\end{aligned} \tag{50}$$

As $\|x\| \to \infty$ *with* $x \in S_B^-$*,*

$$\mathbb{E}[F_w^{\gamma,\nu}(\xi_{n+1}) - F_w^{\gamma,\nu}(\xi_n) \mid \xi_n = x] = \lambda \|Tx\|^{2\nu} \left(\mu^-(x)\cos\alpha + o(1)\right). \tag{51}$$

Proof Suppose that $0 \leq \beta^- < \beta^+ < 1$. As in the proof of Lemma 4, let $E_x = \{\|\Delta\| < \|x\|^\delta\}$, where $\delta \in (0,1)$ satisfies (35). Set $v_\nu(x) := x_2\|Tx\|^{2\nu}$. Then, using Taylor's formula in one variable, for $x, y \in \mathbb{R}^2$ with $y \in B_{r/2}(x)$,

$$\|x+y\|^{2\nu} = \|x\|^{2\nu}\left(1 + \frac{2\langle x, y\rangle + \|y\|^2}{\|x\|^2}\right)^\nu = \|x\|^{2\nu} + 2\nu\langle x, y\rangle\|x\|^{2\nu-2} + R(x,y),$$

where $|R(x,y)| \leq C\|y\|^2\|x\|^{2\nu-2}$. Thus, for $x \in S$ with $y \in B_{r/2}(x)$ and $x+y \in S$,

$$\begin{aligned} v_\nu(x+y) - v_\nu(x) &= (x_2+y_2)\|Tx+Ty\|^{2\nu} - x_2\|Tx\|^{2\nu} \\ &= y_2\|Tx\|^{2\nu} + 2\nu x_2\langle Tx, Ty\rangle\|Tx\|^{2\nu-2} + 2\nu y_2\langle Tx, Ty\rangle\|Tx\|^{2\nu-2} \\ &\quad + R(x,y), \end{aligned} \tag{52}$$

where now $|R(x,y)| \leq C\|y\|^2\|x\|^{2\nu+\beta^+-2}$, using the fact that both $|x_2|$ and $|y_2|$ are $O(\|x\|^{\beta^+})$. Taking $x = \xi_0$ and $y = \Delta$ so $Ty = \tilde{\Delta}$, we obtain

$$\begin{aligned} \mathbb{E}_x\left[(v_\nu(\xi_1) - v_\nu(\xi_0))\mathbf{1}_{E_x}\right] &= \|Tx\|^{2\nu}\,\mathbb{E}_x\left[\Delta_2\mathbf{1}_{E_x}\right] + 2\nu x_2\|Tx\|^{2\nu-2}\,\mathbb{E}_x\left[\langle Tx, \tilde{\Delta}\rangle\mathbf{1}_{E_x}\right] \\ &\quad + 2\nu\|Tx\|^{2\nu-2}\,\mathbb{E}\left[\Delta_2\langle Tx, \tilde{\Delta}\rangle\mathbf{1}_{E_x}\right] \\ &\quad + \mathbb{E}\left[R(x,\Delta)\mathbf{1}_{E_x}\right]. \end{aligned} \tag{53}$$

Suppose that $x \in S_I$. Similarly to (37), we have $\mathbb{E}_x[\langle Tx, \tilde{\Delta}\rangle\mathbf{1}_{E_x}] = o(1)$, and, by similar arguments using (29), $\mathbb{E}[\Delta_2\mathbf{1}_{E_x}] = o(\|x\|^{-1})$, $\mathbb{E}_x|\Delta_2\langle Tx, \tilde{\Delta}\rangle\mathbf{1}_{E_x^c}| = o(\|x\|)$, and $\mathbb{E}_x|R(x,\Delta)\mathbf{1}_{E_x}| = o(\|x\|^{2\nu-1})$, since $\beta^+ < 1$. Also, by (13),

$$\begin{aligned} \mathbb{E}_x(\Delta_2\langle Tx, \tilde{\Delta}\rangle) &= \sigma_2\,\mathbb{E}_x(\tilde{\Delta}_2\langle Tx, \tilde{\Delta}\rangle) \\ &= \sigma_2(Tx)_1\,\mathbb{E}_x(\tilde{\Delta}_1\tilde{\Delta}_2) + \sigma_2(Tx)_2\,\mathbb{E}_x(\tilde{\Delta}_2^2). \end{aligned}$$

Here, by (14), $\mathbb{E}_x(\tilde{\Delta}_1\tilde{\Delta}_2) = o(1)$ and $\mathbb{E}_x(\tilde{\Delta}_2^2) = O(1)$, while $\sigma_2(Tx)_2 = x_2 = O(\|x\|^{\beta^+})$. Thus $\mathbb{E}_x(\Delta_2\langle Tx, \tilde{\Delta}\rangle) = o(\|x\|)$. Hence also

$$\mathbb{E}_x\left[\Delta_2\langle Tx, \tilde{\Delta}\rangle\mathbf{1}_{E_x}\right] = o(\|x\|).$$

Thus from (53) we get that, for $x \in S_I$,

$$\mathbb{E}_x\left[(v_\nu(\xi_1) - v_\nu(\xi_0))\mathbf{1}_{E_x}\right] = o(\|x\|^{2\nu-1}). \tag{54}$$

On the other hand, since $|v_\nu(x+y) - v_\nu(x)| \leq C(\|x\| + \|y\|)^{2\nu+\beta^+}$ we get

$$\mathbb{E}_x\left[|v_\nu(\xi_1) - v_\nu(\xi_0)|\mathbf{1}_{E_x^c}\right] \leq C\,\mathbb{E}_x\left[\|\Delta\|^{(2\nu+\beta^+)/\delta}\mathbf{1}_{E_x^c}\right].$$

Here $2\nu + \beta^+ < 2\nu + 1 < \gamma w < \delta p$, by choice of ν and (35), so we may apply (29) with $q = (2\nu + \beta^+)/\delta$ to get

$$\mathbb{E}_x\left[|v_\nu(\xi_1) - v_\nu(\xi_0)|\mathbf{1}_{E_x^c}\right] = O(\|x\|^{2\nu+\beta^+-\delta p}) = o(\|x\|^{2\nu-1}), \tag{55}$$

since $\delta p > 2$, by (35). Combining (54), (55) and (30), we obtain (49), provided that $2\nu - 1 < \gamma w - 2$, which is the case since $2\nu < \gamma w + \beta^+ - 2$ and $\beta^+ < 1$.

Now suppose that $x \in S_B^{\pm}$. We truncate (52) to see that, for $x \in S$ with $y \in B_{r/2}(x)$ and $x + y \in S$,

$$v_\nu(x+y) - v_\nu(x) = y_2 \|Tx\|^{2\nu} + R(x, y),$$

where now $|R(x, y)| \leq C\|y\|\|x\|^{2\nu+\beta^{\pm}-1}$, using the fact that for $x \in S_B^{\pm}$, $|x_2| = O(\|x\|^{\beta^{\pm}})$. It follows that, for $x \in S_B^{\pm}$,

$$\mathbb{E}_x\left[(v_\nu(\xi_1) - v_\nu(\xi_0))\mathbf{1}_{E_x}\right] = \|Tx\|^{2\nu}\,\mathbb{E}_x\left[\Delta_2 \mathbf{1}_{E_x}\right] + O(\|x\|^{2\nu+\beta^{\pm}-1}).$$

By (29) and (35) we have that $\mathbb{E}[|\Delta_2|\mathbf{1}_{E_x^c}] = O(\|x\|^{-\delta(p-1)}) = o(\|x\|^{-1})$, while if $x \in S_B^{\pm}$, then, by (10), $\mathbb{E}_x\,\Delta_2 = \mp\mu^{\pm}(x)\cos\alpha + O(\|x\|^{\beta^{\pm}-1})$. On the other hand, the estimate (55) still applies, so we get, for $x \in S_B^{\pm}$,

$$\mathbb{E}_x[v_\nu(\xi_1) - v_\nu(\xi_0)] = \mp\|Tx\|^{2\nu}\mu^{\pm}(x)\cos\alpha + O(\|x\|^{2\nu+\beta^{\pm}-1}). \tag{56}$$

If we choose ν such that $2\nu < \gamma w + \beta^+ - 2$, then we combine (56) and (31) to get (50), since the term from (31) dominates. If we choose ν such that $2\nu > \gamma w + \beta^- - 2$, then the term from (56) dominates that from (31), and we get (51).

□

In the critically recurrent cases, where $\max(\beta^+, \beta^-) = \beta_c \in (0, 1)$ or $\beta^+, \beta^- > 1$, in which no passage-time moments exist, the functions of polynomial growth based on h_w as defined at (21) are not sufficient to prove recurrence. Instead we need functions which grow more slowly. For $\eta \in \mathbb{R}$ let

$$h(x) := h(r, \theta) := \log r + \eta\theta, \text{ and } \ell(x) := \log h(Tx), \tag{57}$$

where we understand $\log y$ to mean $\max(1, \log y)$. The function h is again harmonic (see below) and was used in the context of reflecting Brownian motion in a wedge in [31]. Set

$$\eta_0 := \eta_0(\Sigma, \alpha) := \frac{\sigma_2^2 \tan\alpha + \rho}{s}, \text{ and } \eta_1 := \eta_1(\Sigma, \alpha) := \frac{\sigma_1^2 \tan\alpha - \rho}{s}. \tag{58}$$

Lemma 6 *Suppose that* (M_p), (D_+), (R), *and* (C_+) *hold, with* $p > 2$, $\varepsilon > 0$, $\alpha^+ = -\alpha^- = \alpha$ *for* $|\alpha| < \pi/2$, *and* $\beta^+, \beta^- \geq 0$. *For any* $\eta \in \mathbb{R}$, *as* $\|x\| \to \infty$ *with* $x \in S_I$,

$$\mathbb{E}[\ell(\xi_{n+1}) - \ell(\xi_n) \mid \xi_n = x] = -\frac{1 + \eta^2 + o(1)}{2\|Tx\|^2(\log\|Tx\|)^2}. \tag{59}$$

If $0 \le \beta^{\pm} < 1$, *take* $\eta = \eta_0$ *as defined at* (58). *Then, as* $\|x\| \to \infty$ *with* $x \in S_B^{\pm}$,

$$\mathbb{E}[\ell(\xi_{n+1}) - \ell(\xi_n) \mid \xi_n = x] \\ = \frac{\sigma_2^2 a^{\pm}\mu^{\pm}(x)}{s^2 \cos\alpha} \frac{1}{\|Tx\|^2 \log \|Tx\|} \left((\beta^{\pm} - \beta_c) x_1^{\beta^{\pm}} + O(\|x\|^{2\beta^{\pm}-1}) + O(1) \right). \tag{60}$$

If $\beta^{\pm} > 1$, *take* $\eta = \eta_1$ *as defined at* (58). *Then as* $\|x\| \to \infty$ *with* $x \in S_B^{\pm}$,

$$\mathbb{E}[\ell(\xi_{n+1}) - \ell(\xi_n) \mid \xi_n = x] \\ = \frac{\mu^{\pm}(x)}{s^2 \cos\alpha} \frac{x_1}{\|Tx\|^2 \log \|Tx\|} \left(\sigma_1^2 \sin^2\alpha + \sigma_2^2 \cos^2\alpha - \frac{\sigma_1^2}{\beta^{\pm}} - \rho \sin 2\alpha + o(1) \right). \tag{61}$$

Proof Given $\eta \in \mathbb{R}$, for $r_0 = r_0(\eta) = \exp(\mathrm{e} + |\eta|\pi)$, we have from (58) that both h and $\log h$ are infinitely differentiable in the domain $\mathscr{R}_{r_0} := \{x \in \mathbb{R}^2 : x_1 > 0,\ r(x) > r_0\}$. Differentiating (58) and using (22) we obtain, for $x \in \mathscr{R}_{r_0}$,

$$D_1 h(x) = \frac{1}{r}(\cos\theta - \eta\sin\theta), \text{ and } D_2 h(x) = \frac{1}{r}(\sin\theta + \eta\cos\theta). \tag{62}$$

We verify that h is harmonic in $\mathscr{R}_{r_0}$, since

$$D_1^2 h(x) = \frac{\eta \sin 2\theta}{r^2} - \frac{\cos 2\theta}{r^2} = -D_2^2 h(x).$$

Also, for any i, j, k, $|D_i D_j D_k h(x)| = O(r^{-3})$. Moreover, $D_i \log h(x) = (h(x))^{-1} D_i h(x)$,

$$D_i D_j \log h(x) = \frac{D_i D_j h(x)}{h(x)} - \frac{(D_i h(x))(D_j h(x))}{(h(x))^2},$$

and $|D_i D_j D_k \log h(x)| = O(r^{-3}(\log r)^{-1})$. Recall that $Dh(x)$ is the vector function whose components are $D_i h(x)$. Then Taylor's formula (28) together with the harmonic property of h shows that for $x \in \mathscr{R}_{2r_0}$ and $y \in B_{r/2}(x)$,

$$\log h(x+y) = \log h(x) + \frac{\langle Dh(x), y\rangle}{h(x)} + \frac{(y_1^2 - y_2^2) D_1^2 h(x)}{2h(x)} + \frac{y_1 y_2 D_1 D_2 h(x)}{h(x)} \\ - \frac{\langle Dh(x), y\rangle^2}{2(h(x))^2} + R(x, y), \tag{63}$$

where $|R(x, y)| \leq C\|y\|^3\|x\|^{-3}(\log\|x\|)^{-1}$ for some constant $C < \infty$, all $y \in B_{r/2}(x)$, and all $\|x\|$ sufficiently large. As in the proof of Lemma 4, let $E_x = \{\|\Delta\| < \|x\|^\delta\}$ for $\delta \in (\frac{2}{p}, 1)$. Then applying the expansion in (63) to $\log h(Tx + \tilde{\Delta})$, conditioning on $\xi_0 = x$, and taking expectations, we obtain, for $\|x\|$ sufficiently large,

$$\mathbb{E}_x\left[(\ell(\xi_1) - \ell(\xi_0))\mathbf{1}_{E_x}\right] = \frac{\mathbb{E}_x\left[\langle Dh(Tx), \tilde{\Delta}\rangle \mathbf{1}_{E_x}\right]}{h(Tx)} + \frac{D_1^2 h(Tx)\,\mathbb{E}_x\left[(\tilde{\Delta}_1^2 - \tilde{\Delta}_2^2)\mathbf{1}_{E_x}\right]}{2h(Tx)}$$
$$+ \frac{D_1 D_2 h(Tx)\,\mathbb{E}_x\left[\tilde{\Delta}_1\tilde{\Delta}_2\mathbf{1}_{E_x}\right]}{h(Tx)} - \frac{\mathbb{E}_x\left[\langle Dh(Tx), \tilde{\Delta}\rangle^2 \mathbf{1}_{E_x}\right]}{2(h(Tx))^2} + \mathbb{E}_x\left[R(Tx, \tilde{\Delta})\mathbf{1}_{E_x}\right]. \tag{64}$$

Let $p' \in (2, 3]$ be such that (1) holds. Then

$$\mathbb{E}_x\left|R(Tx, \tilde{\Delta})\mathbf{1}_{E_x}\right| \leq C\|x\|^{-3+(3-p')\delta}\,\mathbb{E}_x(\|\Delta\|^{p'}) = O(\|x\|^{-2-\varepsilon'}),$$

for some $\varepsilon' > 0$.

Suppose that $x \in S_I$. By (15), $\mathbb{E}_x(\tilde{\Delta}_1\tilde{\Delta}_2) = O(\|x\|^{-\varepsilon})$ and, by (29), $\mathbb{E}_x|\tilde{\Delta}_1\tilde{\Delta}_2\mathbf{1}_{E_x^c}| \leq C\,\mathbb{E}[\|\Delta\|^2\mathbf{1}_{E_x^c}] = O(\|x\|^{-\varepsilon'})$, for some $\varepsilon' > 0$. Thus $\mathbb{E}_x(\tilde{\Delta}_1\tilde{\Delta}_2\mathbf{1}_{E_x}) = O(\|x\|^{-\varepsilon'})$. A similar argument gives the same bound for $\mathbb{E}_x[(\tilde{\Delta}_1^2 - \tilde{\Delta}_2^2)\mathbf{1}_{E_x}]$. Also, from (15) and (62), $\mathbb{E}_x(\langle Dh(Tx), \tilde{\Delta}\rangle) = O(\|x\|^{-2-\varepsilon})$ and, by (29), $\mathbb{E}_x|\langle Dh(Tx), \tilde{\Delta}\rangle\mathbf{1}_{E_x^c}| = O(\|x\|^{-2-\varepsilon'})$ for some $\varepsilon' > 0$. Hence $\mathbb{E}_x[\langle Dh(Tx), \tilde{\Delta}\rangle\mathbf{1}_{E_x}] = O(\|x\|^{-2-\varepsilon'})$. Finally, by (15) and (62),

$$\begin{aligned}\mathbb{E}_x\langle Dh(Tx), \tilde{\Delta}\rangle^2 &= \mathbb{E}_x\left((Dh(Tx))^\top\tilde{\Delta}\tilde{\Delta}^\top Dh(Tx)\right)\\ &= (Dh(Tx))^\top Dh(Tx) + O(\|x\|^{-2-\varepsilon})\\ &= (D_1h(Tx))^2 + (D_2h(Tx))^2 + O(\|x\|^{-2-\varepsilon}),\end{aligned}$$

while, by (29), $\mathbb{E}_x|\langle Dh(Tx), \tilde{\Delta}\rangle^2\mathbf{1}_{E_x^c}| = O(\|x\|^{-2-\varepsilon'})$. Putting all these estimates into (64) gives

$$\mathbb{E}_x\left[(\ell(\xi_1) - \ell(\xi_0))\mathbf{1}_{E_x}\right] = -\frac{(D_1h(Tx))^2 + (D_2h(Tx))^2}{2(h(Tx))^2} + O(\|x\|^{-2-\varepsilon'}),$$

for some $\varepsilon' > 0$. On the other hand, for all $\|x\|$ sufficiently large, $|\ell(x+y) - \ell(x)| \leq C\log\log\|x\| + C\log\log\|y\|$. For any $p > 2$ and $\delta \in (\frac{2}{p}, 1)$, we may (and do) choose $q > 0$ sufficiently small such that $\delta(p-q) > 2$, and then, by (29),

$$\begin{aligned}\mathbb{E}_x\left[(\ell(\xi_1) - \ell(\xi_0))\mathbf{1}_{E_x^c}\right] &\leq C\,\mathbb{E}_x\left[\|\Delta\|^q\mathbf{1}_{E_x^c}\right]\\ &= O(\|x\|^{-\delta(p-q)}) = O(\|x\|^{-2-\varepsilon'}),\end{aligned} \tag{65}$$

for some $\varepsilon' > 0$. Thus we conclude that

$$\mathbb{E}_x\left[\ell(\xi_1) - \ell(\xi_0)\right] = -\frac{(D_1 h(Tx))^2 + (D_2 h(Tx))^2}{2(h(Tx))^2} + O(\|x\|^{-2-\varepsilon'}),$$

for some $\varepsilon' > 0$. Then (59) follows from (62).

Next suppose that $x \in S_B$. Truncating (63), we have for $x \in \mathscr{R}_{2r_0}$ and $y \in B_{r/2}(x)$,

$$\log h(x+y) = \log h(x) + \frac{\langle Dh(x), y\rangle}{h(x)} + R(x, y),$$

where now $|R(x, y)| \leq C\|y\|^2\|x\|^{-2}(\log\|x\|)^{-1}$ for $\|x\|$ sufficiently large. Hence

$$\mathbb{E}_x\left[(\ell(\xi_1) - \ell(\xi_0))\mathbf{1}_{E_x}\right] = \frac{\mathbb{E}_x\left[\langle Dh(Tx), \tilde{\Delta}\rangle\mathbf{1}_{E_x}\right] + O(\|x\|^{-2})}{h(Tx)}.$$

Then by (65) and the fact that $\mathbb{E}_x\,|\langle Dh(Tx), \tilde{\Delta}\rangle\mathbf{1}_{E_x^c}| = O(\|x\|^{-2-\varepsilon'})$ (as above),

$$\mathbb{E}_x\left[\ell(\xi_1) - \ell(\xi_0)\right] = \frac{\mathbb{E}_x\left[\langle Dh(Tx), \tilde{\Delta}\rangle\right] + O(\|x\|^{-2})}{h(Tx)}. \tag{66}$$

From (62) we have

$$Dh(x) = \frac{1}{\|x\|^2}\begin{pmatrix} x_1 - \eta x_2 \\ x_2 + \eta x_1 \end{pmatrix}, \text{ and hence } Dh(Tx) = \frac{1}{\|Tx\|^2}\begin{pmatrix} \frac{\sigma_2}{s}x_1 - \frac{\rho}{s\sigma_2}x_2 - \frac{\eta}{\sigma_2}x_2 \\ \frac{1}{\sigma_2}x_2 + \frac{\eta\sigma_2}{s}x_1 - \frac{\eta\rho}{s\sigma_2}x_2 \end{pmatrix},$$

using (13). If $\beta^\pm < 1$ and $x \in S_B^\pm$, we have from (16) and (17) that

$$\begin{aligned}
&\mathbb{E}_x\langle Dh(Tx), \tilde{\Delta}\rangle \\
&\quad = \frac{\mu^\pm(x)}{s^2}\frac{1}{\|Tx\|^2}\Big\{a^\pm\Big[\big(s\eta(\beta^\pm - 1) - \rho(1+\beta^\pm)\big)\sin\alpha + \left(\sigma_2^2\beta^\pm - \sigma_1^2\right)\cos\alpha\Big]x_1^{\beta^\pm} \\
&\qquad \pm\Big[\sigma_2^2\sin\alpha + (\rho - s\eta)\cos\alpha\Big]x_1 + O(x_1^{2\beta^\pm - 1}) + O(1)\Big\}.
\end{aligned}$$

Taking $\eta = \eta_0$ as given by (58), the $\pm x_1$ term vanishes; after simplification, we get

$$\mathbb{E}_x\langle Dh(Tx), \tilde{\Delta}\rangle = \frac{\sigma_2^2 a^\pm \mu^\pm(x)}{\|Tx\|^2 s^2\cos\alpha}\left(\left(\beta^\pm - \beta_c\right)x_1^{\beta^\pm} + O(x_1^{2\beta^\pm - 1}) + O(1)\right). \tag{67}$$

Using (67) in (66) gives (60).

On the other hand, if $\beta^\pm > 1$ and $x \in S_B^\pm$, we have from (18) and (19) that

$$\begin{aligned}
&\mathbb{E}_x \langle Dh(Tx), \tilde{\Delta} \rangle \\
&= \frac{\mu^\pm(x)}{s^2} \frac{1}{\|Tx\|^2} \left\{ \frac{1}{\beta^\pm} \left[\left(s\eta(\beta^\pm - 1) - \rho(1+\beta^\pm)\right) \sin\alpha + \left(\sigma_2^2 \beta^\pm - \sigma_1^2\right) \cos\alpha \right] x_1 \right. \\
&\quad \left. \pm a^\pm \left[\sigma_1^2 \sin\alpha - (\rho + s\eta)\cos\alpha \right] x_1^{\beta^\pm} + O(x_1^{2-\beta^\pm}) + O(1) \right\}.
\end{aligned}$$

Taking $\eta = \eta_1$ as given by (58), the $\pm x_1^{\beta^\pm}$ term vanishes, and we get

$$\mathbb{E}_x \langle Dh(Tx), \tilde{\Delta} \rangle = \frac{\mu^\pm(x)}{s^2 \cos\alpha} \frac{x_1}{\|Tx\|^2} \left(\sigma_1^2 \sin^2\alpha + \sigma_2^2 \cos^2\alpha - \frac{\sigma_1^2}{\beta^\pm} - \rho \sin 2\alpha + o(1) \right),$$

as $\|x\| \to \infty$ (and $x_1 \to \infty$). Then using the last display in (66) gives (61). □

The function ℓ is not by itself enough to prove recurrence in the critical cases, because the estimates in Lemma 6 do not guarantee that ℓ satisfies a supermartingale condition for all parameter values of interest. To proceed, we modify the function slightly to improve its properties near the boundary. In the case where $\max(\beta^+, \beta^-) = \beta_c \in (0, 1)$, the following function will be used to prove recurrence,

$$g_\gamma(x) := g_\gamma(r, \theta) := \ell(x) + \frac{\theta^2}{(1+r)^\gamma},$$

where the parameter η in ℓ is chosen as $\eta = \eta_0$ as given by (58).

Lemma 7 *Suppose that* (M_p), (D_+), (R), *and* (C_+) *hold, with* $p > 2$, $\varepsilon > 0$, $\alpha^+ = -\alpha^- = \alpha$ *for* $|\alpha| < \pi/2$, *and* $\beta^+, \beta^- \in (0, 1)$ *with* $\beta^+, \beta^- \le \beta_c$. *Let* $\eta = \eta_0$, *and suppose*

$$0 < \gamma < \min(\beta^+, \beta^-, 1 - \beta^+, 1 - \beta^-, p - 2).$$

Then as $\|x\| \to \infty$ *with* $x \in S_I$,

$$\mathbb{E}[g_\gamma(\xi_{n+1}) - g_\gamma(\xi_n) \mid \xi_n = x] = -\frac{1 + \eta^2 + o(1)}{2\|Tx\|^2 (\log \|Tx\|)^2}. \tag{68}$$

Moreover, as $\|x\| \to \infty$ *with* $x \in S_B^\pm$,

$$\mathbb{E}[g_\gamma(\xi_{n+1}) - g_\gamma(\xi_n) \mid \xi_n = x] \le -2a^\pm \mu^\pm(x)(\cos\alpha + o(1))\|x\|^{\beta^\pm - 2 - \gamma}. \tag{69}$$

Proof Set $u_\gamma(x) := u_\gamma(r,\theta) := \theta^2(1+r)^{-\gamma}$, and note that, by (22), for $x_1 > 0$,

$$D_1 u_\gamma(x) = -\frac{2\theta \sin\theta}{r(1+r)^\gamma} - \frac{\gamma\theta^2\cos\theta}{(1+r)^{1+\gamma}}, \quad D_2 u_\gamma(x) = \frac{2\theta\cos\theta}{r(1+r)^\gamma} - \frac{\gamma\theta^2\sin\theta}{(1+r)^{1+\gamma}},$$

and $|D_i D_j u_\gamma(x)| = O(r^{-2-\gamma})$ for any i, j. So, by Taylor's formula (28), for all $y \in B_{r/2}(x)$,

$$u_\gamma(x+y) = u_\gamma(x) + \langle Du_\gamma(x), y\rangle + R(x,y),$$

where $|R(x,y)| \le C\|y\|^2\|x\|^{-2-\gamma}$ for all $\|x\|$ sufficiently large. Once more define the event $E_x = \{\|\Delta\| < \|x\|^\delta\}$, where now $\delta \in (\frac{2+\gamma}{p}, 1)$. Then

$$\mathbb{E}_x\left[(u_\gamma(\xi_1) - u_\gamma(\xi_0))\mathbf{1}_{E_x}\right] = \mathbb{E}_x\left[\langle Du_\gamma(x), \Delta\rangle \mathbf{1}_{E_x}\right] + O(\|x\|^{-2-\gamma}).$$

Moreover, $\mathbb{E}_x |\langle Du_\gamma(x), \Delta\rangle \mathbf{1}_{E_x^c}| \le C\|x\|^{-1-\gamma}\,\mathbb{E}_x(\|\Delta\|\mathbf{1}_{E_x^c}) = O(\|x\|^{-2-\gamma})$, by (29) and the fact that $\delta > \frac{2}{p} > \frac{1}{p-1}$. Also, since u_γ is uniformly bounded,

$$\mathbb{E}_x\left[|u_\gamma(\xi_1) - u_\gamma(\xi_0)|\mathbf{1}_{E_x^c}\right] \le C\mathbb{P}_x(E_x^c) = O(\|x\|^{-p\delta}),$$

by (29). Since $p\delta > 2+\gamma$, it follows that

$$\mathbb{E}_x\left[u_\gamma(\xi_1) - u_\gamma(\xi_0)\right] = \mathbb{E}_x\langle Du_\gamma(x), \Delta\rangle + O(\|x\|^{-2-\gamma}). \tag{70}$$

For $x \in S_I$, it follows from (70) and (D_+) that $\mathbb{E}_x[u_\gamma(\xi_1) - u_\gamma(\xi_0)] = O(\|x\|^{-2-\gamma})$, and combining this with (59) we get (68).

Let $\beta = \max(\beta^+, \beta^-) < 1$. For $x \in S$, $|\theta(x)| = O(r^{\beta-1})$ as $\|x\| \to \infty$, so (70) gives

$$\mathbb{E}_x[u_\gamma(\xi_1) - u_\gamma(\xi_0)] = \frac{2\theta\cos\theta\,\mathbb{E}_x\,\Delta_2}{\|x\|(1+\|x\|)^\gamma} + O(\|x\|^{2\beta-3-\gamma}) + O(\|x\|^{-2-\gamma}).$$

If $x \in S_B^\pm$ then $\theta = \pm a^\pm(1+o(1))x_1^{\beta^\pm-1}$ and, by (10), $\mathbb{E}_x\,\Delta_2 = \mp\mu^\pm(x)\cos\alpha + o(1)$, so

$$\mathbb{E}_x[u_\gamma(\xi_1) - u_\gamma(\xi_0)] = -2a^\pm\mu^\pm(x)(\cos\alpha + o(1))\|x\|^{\beta^\pm-2-\gamma}. \tag{71}$$

For $\eta = \eta_0$ and $\beta^+, \beta^- \le \beta_c$, we have from (60) that

$$\mathbb{E}_x[\ell(\xi_1) - \ell(\xi_0)] \le \frac{1}{\|Tx\|^2\log\|Tx\|}\left(O(\|x\|^{2\beta^\pm-1}) + O(1)\right).$$

Combining this with (71), we obtain (69), provided that we choose γ such that $\beta^{\pm} - 2 - \gamma > 2\beta^{\pm} - 3$ and $\beta^{\pm} - 2 - \gamma > -2$, that is, $\gamma < 1 - \beta^{\pm}$ and $\gamma < \beta^{\pm}$. □

In the case where $\beta^+, \beta^- > 1$, we will use the function

$$w_\gamma(x) := \ell(x) - \frac{x_1}{(1 + \|x\|^2)^\gamma},$$

where the parameter η in ℓ is now chosen as $\eta = \eta_1$ as defined at (58). A similar function was used in [6].

Lemma 8 *Suppose that* (M_p)*,* (D_+)*,* (R)*, and* (C_+) *hold, with* $p > 2$*,* $\varepsilon > 0$*,* $\alpha^+ = -\alpha^- = \alpha$ *for* $|\alpha| < \pi/2$*, and* $\beta^+, \beta^- > 1$ *Let* $\eta = \eta_1$*, and suppose that*

$$\frac{1}{2} < \gamma < \min\left(1 - \frac{1}{2\beta^+}, 1 - \frac{1}{2\beta^-}, \frac{p-1}{2}\right).$$

Then as $\|x\| \to \infty$ *with* $x \in S_I$*,*

$$\mathbb{E}[w_\gamma(\xi_{n+1}) - w_\gamma(\xi_n) \mid \xi_n = x] = -\frac{1 + \eta^2 + o(1)}{2\|Tx\|^2(\log\|Tx\|)^2}. \tag{72}$$

Moreover, as $\|x\| \to \infty$ *with* $x \in S_B^{\pm}$*,*

$$\mathbb{E}[w_\gamma(\xi_{n+1}) - w_\gamma(\xi_n) \mid \xi_n = x] = -\frac{\mu^{\pm}(x)\cos\alpha + o(1)}{\|x\|^{2\gamma}}. \tag{73}$$

Proof Let $q_\gamma(x) := x_1(1 + \|x\|^2)^{-\gamma}$. Then

$$D_1 q_\gamma(x) = \frac{1}{(1 + \|x\|^2)^\gamma} - \frac{2\gamma x_1^2}{(1 + \|x\|^2)^{1+\gamma}}, \quad D_2 q_\gamma(x) = -\frac{2\gamma x_1 x_2}{(1 + \|x\|^2)^{1+\gamma}},$$

and $|D_i D_j q_\gamma(x)| = O(\|x\|^{-1-2\gamma})$ for any i, j. Thus by Taylor's formula, for $y \in B_{r/2}(x)$,

$$q_\gamma(x + y) - q_\gamma(x) = \langle Dq_\gamma(x), y\rangle + R(x, y),$$

where $|R(x, y)| \le C\|y\|^2\|x\|^{-1-2\gamma}$ for $\|x\|$ sufficiently large. Once more let $E_x = \{\|\Delta\| < \|x\|^\delta\}$, where now we take $\delta \in (\frac{1+2\gamma}{p}, 1)$. Then

$$\mathbb{E}_x\left[(q_\gamma(\xi_1) - q_\gamma(\xi_0))\mathbf{1}_{E_x}\right] = \mathbb{E}_x\left[\langle Dq_\gamma(x), \Delta\rangle\mathbf{1}_{E_x}\right] + O(\|x\|^{-1-2\gamma}).$$

Moreover, we get from (29) that $\mathbb{E}_x |\langle Dq_\gamma(x), \Delta \rangle \mathbf{1}_{E_x^c}| = O(\|x\|^{-2\gamma-\delta(p-1)})$, where $\delta(p-1) > 2\gamma > 1$, and, since q_γ is uniformly bounded for $\gamma > 1/2$,

$$\mathbb{E}_x \left[(q_\gamma(\xi_1) - q_\gamma(\xi_0)) \mathbf{1}_{E_x^c} \right] = O(\|x\|^{-p\delta}),$$

where $p\delta > 1 + 2\gamma$. Thus

$$\mathbb{E}_x \left[q_\gamma(\xi_1) - q_\gamma(\xi_0) \right] = \mathbb{E}_x \langle Dq_\gamma(x), \Delta \rangle + O(\|x\|^{-1-2\gamma}). \tag{74}$$

If $x \in S_I$, then (D_+) gives $\mathbb{E}_x \langle Dq_\gamma(x), \Delta \rangle = O(\|x\|^{-1-2\gamma})$ and with (59) we get (72), since $\gamma > 1/2$. On the other hand, suppose that $x \in S_B^\pm$ and $\beta^\pm > 1$. Then $\|x\| \geq c x_1^{\beta^\pm}$ for some $c > 0$, so $x_1 = O(\|x\|^{1/\beta^\pm})$. So, by (74),

$$\mathbb{E}_x[q_\gamma(\xi_1) - q_\gamma(\xi_0)] = \frac{\mathbb{E}_x \Delta_1}{(1 + \|x\|^2)^\gamma} + O\left(\|x\|^{\frac{1}{\beta^\pm} - 1 - 2\gamma} \right).$$

Moreover, by (11), $\mathbb{E}_x \Delta_1 = \mu^\pm(x) \cos \alpha + o(1)$. Combined with (61), this yields (73), provided that $2\gamma \leq 2 - (1/\beta^\pm)$, again using the fact that $x_1 = O(\|x\|^{1/\beta^\pm})$. This completes the proof. □

4 Proofs of Main Results

We obtain our recurrence classification and quantification of passage-times via Foster–Lyapunov criteria (cf. [14]). As we do not assume any irreducibility, the most convenient form of the criteria are those for discrete-time adapted processes presented in [26]. However, the recurrence criteria in [26, §3.5] are formulated for processes on $\mathbb{R}_+$, and, strictly, do not apply directly here. Thus we present appropriate generalizations here, as they may also be useful elsewhere. The following recurrence result is based on Theorem 3.5.8 of [26].

Lemma 9 *Let $X_0, X_1, \ldots$ be a stochastic process on $\mathbb{R}^d$ adapted to a filtration $\mathscr{F}_0, \mathscr{F}_1, \ldots$. Let $f : \mathbb{R}^d \to \mathbb{R}_+$ be such that $f(x) \to \infty$ as $\|x\| \to \infty$, and $\mathbb{E} f(X_0) < \infty$. Suppose that there exist $r_0 \in \mathbb{R}_+$ and $C < \infty$ for which, for all $n \in \mathbb{Z}_+$,*

$$\mathbb{E}[f(X_{n+1}) - f(X_n) \mid \mathscr{F}_n] \leq 0, \text{ on } \{\|X_n\| \geq r_0\};$$
$$\mathbb{E}[f(X_{n+1}) - f(X_n) \mid \mathscr{F}_n] \leq C, \text{ on } \{\|X_n\| < r_0\}.$$

Then if $\mathbb{P}(\limsup_{n\to\infty} \|X_n\| = \infty) = 1$, we have $\mathbb{P}(\liminf_{n\to\infty} \|X_n\| \leq r_0) = 1$.

Proof By hypothesis, $\mathbb{E} f(X_n) < \infty$ for all n. Fix $n \in \mathbb{Z}_+$ and let $\lambda_n := \min\{m \geq n : \|X_m\| \leq r_0\}$ and, for some $r > r_0$, set $\sigma_n := \min\{m \geq n : \|X_m\| \geq r\}$. Since $\limsup_{n\to\infty} \|X_n\| = \infty$ a.s., we have that $\sigma_n < \infty$, a.s. Then $f(X_{m\wedge\lambda_n\wedge\sigma_n})$, $m \geq n$, is a non-negative supermartingale with $\lim_{m\to\infty} f(X_{m\wedge\lambda_n\wedge\sigma_n}) = f(X_{\lambda_n\wedge\sigma_n})$, a.s. By Fatou's lemma and the fact that f is non-negative,

$$\mathbb{E} f(X_n) \geq \mathbb{E} f(X_{\lambda_n\wedge\sigma_n}) \geq \mathbb{P}(\sigma_n < \lambda_n) \inf_{y:\|y\|\geq r} f(y).$$

So

$$\mathbb{P}\left(\inf_{m\geq n} \|X_m\| \leq r_0\right) \geq \mathbb{P}(\lambda_n < \infty) \geq \mathbb{P}(\lambda_n < \sigma_n) \geq 1 - \frac{\mathbb{E} f(X_n)}{\inf_{y:\|y\|\geq r} f(y)}.$$

Since $r > r_0$ was arbitrary, and $\inf_{y:\|y\|\geq r} f(y) \to \infty$ as $r \to \infty$, it follows that, for fixed $n \in \mathbb{Z}_+$, $\mathbb{P}(\inf_{m\geq n} \|X_m\| \leq r_0) = 1$. Since this holds for all $n \in \mathbb{Z}_+$, the result follows. □

The corresponding transience result is based on Theorem 3.5.6 of [26].

Lemma 10 *Let $X_0, X_1, \ldots$ be a stochastic process on $\mathbb{R}^d$ adapted to a filtration $\mathcal{F}_0, \mathcal{F}_1, \ldots$. Let $f : \mathbb{R}^d \to \mathbb{R}_+$ be such that $\sup_x f(x) < \infty$, $f(x) \to 0$ as $\|x\| \to \infty$, and $\inf_{x:\|x\|\leq r} f(x) > 0$ for all $r \in \mathbb{R}_+$. Suppose that there exists $r_0 \in \mathbb{R}_+$ for which, for all $n \in \mathbb{Z}_+$,*

$$\mathbb{E}[f(X_{n+1}) - f(X_n) \mid \mathcal{F}_n] \leq 0, \text{ on } \{\|X_n\| \geq r_0\}.$$

Then if $\mathbb{P}(\limsup_{n\to\infty} \|X_n\| = \infty) = 1$, we have that $\mathbb{P}(\lim_{n\to\infty} \|X_n\| = \infty) = 1$.

Proof Since f is bounded, $\mathbb{E} f(X_n) < \infty$ for all n. Fix $n \in \mathbb{Z}_+$ and $r_1 \geq r_0$. For $r \in \mathbb{Z}_+$ let $\sigma_r := \min\{n \in \mathbb{Z}_+ : \|X_n\| \geq r\}$. Since $\mathbb{P}(\limsup_{n\to\infty} \|X_n\| = \infty) = 1$, we have $\sigma_r < \infty$, a.s. Let $\lambda_r := \min\{n \geq \sigma_r : \|X_n\| \leq r_1\}$. Then $f(X_{n\wedge\lambda_r})$, $n \geq \sigma_r$, is a non-negative supermartingale, which converges, on $\{\lambda_r < \infty\}$, to $f(X_{\lambda_r})$. By optional stopping (e.g. Theorem 2.3.11 of [26]), a.s.,

$$\sup_{x:\|x\|\geq r} f(x) \geq f(X_{\sigma_r}) \geq \mathbb{E}[f(X_{\lambda_r}) \mid \mathcal{F}_{\sigma_r}] \geq \mathbb{P}(\lambda_r < \infty \mid \mathcal{F}_{\sigma_r}) \inf_{x:\|x\|\leq r_1} f(x).$$

So

$$\mathbb{P}(\lambda_r < \infty) \leq \frac{\sup_{x:\|x\|\geq r} f(x)}{\inf_{x:\|x\|\leq r_1} f(x)},$$

which tends to 0 as $r \to \infty$, by our hypotheses on f. Thus,

$$\mathbb{P}\left(\liminf_{n\to\infty} \|X_n\| \leq r_1\right) = \mathbb{P}\left(\cap_{r\in\mathbb{Z}_+} \{\lambda_r < \infty\}\right) = \lim_{r\to\infty} \mathbb{P}(\lambda_r < \infty) = 0.$$

Since $r_1 \geq r_0$ was arbitrary, we get the result. □

Now we can complete the proof of Theorem 3, which includes Theorem 1 as the special case $\alpha = 0$.

Proof (of Theorem 3) Let $\beta = \max(\beta^+, \beta^-)$, and recall the definition of β_c from (5) and that of s_0 from (7). Suppose first that $0 \leq \beta < 1 \wedge \beta_c$. Then $s_0 > 0$ and we may (and do) choose $w \in (0, 2s_0)$. Also, take $\gamma \in (0, 1)$; note $0 < \gamma w < 1$. Consider the function f_w^γ with $\theta_0 = \theta_1$ given by (23). Then from (30), we see that there exist $c > 0$ and $r_0 < \infty$ such that, for all $x \in S_I$,

$$\mathbb{E}[f_w^\gamma(\xi_{n+1}) - f_w^\gamma(\xi_n) \mid \xi_n = x] \leq -c\|x\|^{\gamma w - 2}, \text{ for all } \|x\| \geq r_0. \tag{75}$$

By choice of w, we have $\beta - (1 - w)\beta_c < 0$, so (31) shows that, for all $x \in S_B^\pm$,

$$\mathbb{E}[f_w^\gamma(\xi_{n+1}) - f_w^\gamma(\xi_n) \mid \xi_n = x] \leq -c\|x\|^{\gamma w - 2 + \beta^\pm},$$

for some $c > 0$ and all $\|x\|$ sufficiently large. In particular, this means that (75) holds throughout S. On the other hand, it follows from (39) and (M_p) that there is a constant $C < \infty$ such that

$$\mathbb{E}[f_w^\gamma(\xi_{n+1}) - f_w^\gamma(\xi_n) \mid \xi_n = x] \leq C, \text{ for all } \|x\| \leq r_0. \tag{76}$$

Since $w, \gamma > 0$, we have that $f_w^\gamma(x) \to \infty$ as $\|x\| \to \infty$. Then by Lemma 9 with the conditions (75) and (76) and assumption (N), we establish recurrence.

Next suppose that $\beta_c < \beta < 1$. If $\beta^+ = \beta^- = \beta$, we use the function f_w^γ, again with $\theta_0 = \theta_1$ given by (23). We may (and do) choose $\gamma \in (0, 1)$ and $w < 0$ with $w > -2|s_0|$ and $\gamma w > w > 2 - p$. By choice of w, we have $\beta - (1 - w)\beta_c > 0$. We have from (30) and (31) that (75) holds in this case also, but now $f_w^\gamma(x) \to 0$ as $\|x\| \to \infty$, since $\gamma w < 0$. Lemma 10 then gives transience when $\beta^+ = \beta^-$.

Suppose now that $\beta_c < \beta < 1$ with $\beta^+ \neq \beta^-$. Without loss of generality, suppose that $\beta = \beta^+ > \beta^-$. We now use the function $F_w^{\gamma,\nu}$ defined at (48), where, as above, we take $\gamma \in (0, 1)$ and $w \in (-2|s_0|, 0)$, and we choose the constants λ, ν with $\lambda < 0$ and $\gamma w + \beta^- - 2 < 2\nu < \gamma w + \beta^+ - 2$. Note that $2\nu < \gamma w - 1$, so $F_w^{\gamma,\nu}(x) = f_w^\gamma(x)(1 + o(1))$. With $\theta_0 = \theta_1$ given by (23), and this choice of ν, Lemma 5 applies. The choice of γ ensures that the right-hand side of (49) is eventually negative, and the choice of w ensures the same for (50). Since $\lambda < 0$, the right-hand side of (51) is also eventually negative. Combining these three estimates shows, for all $x \in S$ with $\|x\|$ large enough,

$$\mathbb{E}[F_w^{\gamma,\nu}(\xi_{n+1}) - F_w^{\gamma,\nu}(\xi_n) \mid \xi_n = x] \leq 0.$$

Since $F_w^{\gamma,\nu}(x) \to 0$ as $\|x\| \to \infty$, Lemma 10 gives transience.

Of the cases where $\beta^+, \beta^- < 1$, it remains to consider the borderline case where $\beta = \beta_c \in (0, 1)$. Here Lemma 7 together with Lemma 9 proves recurrence. Finally, if $\beta^+, \beta^- > 1$, we apply Lemma 8 together with Lemma 9 to obtain recurrence. Note that both of these critical cases require (D_+) and (C_+). □

Next we turn to moments of passage times: we prove Theorem 4, which includes Theorem 2 as the special case $\alpha = 0$. Here the criteria we apply are from [26, §2.7], which are heavily based on those from [5].

Proof (of Theorem 4) Again let $\beta = \max(\beta^+, \beta^-)$. First we prove the existence of moments part of (a)(i). Suppose that $0 \le \beta < 1 \wedge \beta_c$, so s_0 as defined at (7) satisfies $s_0 > 0$. We use the function f_w^γ, with $\gamma \in (0, 1)$ and $w \in (0, 2s_0)$ as in the first part of the proof of Theorem 3. We saw in that proof that for these choices of γ, w we have that (75) holds for all $x \in S$. Rewriting this slightly, using the fact that $f_w^\gamma(x)$ is bounded above and below by constants times $\|x\|^{\gamma w}$ for all $\|x\|$ sufficiently large, we get that there are constants $c > 0$ and $r_0 < \infty$ for which

$$\mathbb{E}[f_w^\gamma(\xi_{n+1}) - f_w^\gamma(\xi_n) \mid \xi_n = x] \le -c(f_w^\gamma(x))^{1-\frac{2}{\gamma w}}, \text{ for all } x \in S \text{ with } \|x\| \ge r_0. \tag{77}$$

Then we may apply Corollary 2.7.3 of [26] to get $\mathbb{E}_x(\tau_r^s) < \infty$ for any $r \ge r_0$ and any $s < \gamma w/2$. Taking $\gamma < 1$ and $w < 2s_0$ arbitrarily close to their upper bounds, we get $\mathbb{E}_x(\tau_r^s) < \infty$ for all $s < s_0$.

Next suppose that $0 \le \beta \le \beta_c$. Let $s > s_0$. First consider the case where $\beta^+ = \beta^-$. Then we consider f_w^γ with $\gamma > 1$, $w > 2s_0$ (so $w > 0$), and $0 < w\gamma < 2$. Then, since $\beta - (1-w)\beta_c = \beta_c - \beta + (w - 2s_0)\beta_c > 0$, we have from (30) and (31) that

$$\mathbb{E}[f_w^\gamma(\xi_{n+1}) - f_w^\gamma(\xi_n) \mid \xi_n = x] \ge 0, \tag{78}$$

for all $x \in S$ with $\|x\|$ sufficiently large. Now set $Y_n := f_w^{1/w}(\xi_n)$, and note that Y_n is bounded above and below by constants times $\|\xi_n\|$, and $Y_n^{\gamma w} = f_w^\gamma(\xi_n)$. Write $\mathscr{F}_n = \sigma(\xi_0, \xi_1, \dots, \xi_n)$. Then we have shown in (78) that

$$\mathbb{E}[Y_{n+1}^{\gamma w} - Y_n^{\gamma w} \mid \mathscr{F}_n] \ge 0, \text{ on } \{Y_n > r_1\}, \tag{79}$$

for some r_1 sufficiently large. Also, from the $\gamma = 1/w$ case of (30) and (31),

$$\mathbb{E}[Y_{n+1} - Y_n \mid \mathscr{F}_n] \ge -\frac{B}{Y_n}, \text{ on } \{Y_n > r_2\}, \tag{80}$$

for some $B < \infty$ and r_2 sufficiently large. (The right-hand side of (31) is still eventually positive, while the right-hand-side of (30) will be eventually negative if $\gamma < 1$.) Again let $E_x = \{\|\varDelta\| < \|x\|^\delta\}$ for $\delta \in (0, 1)$. Then from the $\gamma = 1/w$ case of (41),

$$\left| f_w^{1/w}(\xi_1) - f_w^{1/w}(\xi_0) \right|^2 \mathbf{1}_{E_x} \le C\|\varDelta\|^2,$$

while from the $\gamma = 1/w$ case of (39) we have

$$\left| f_w^{1/w}(\xi_1) - f_w^{1/w}(\xi_0) \right|^2 \mathbf{1}_{E_x^c} \le C \|\Delta\|^{2/\delta}.$$

Taking $\delta \in (2/p, 1)$, it follows from (M_p) that for some $C < \infty$, a.s.,

$$\mathbb{E}[(Y_{n+1} - Y_n)^2 \mid \mathscr{F}_n] \le C. \tag{81}$$

The three conditions (79)–(81) show that we may apply Theorem 2.7.4 of [26] to get $\mathbb{E}_x(\tau_r^s) = \infty$ for all $s > \gamma w/2$, all r sufficiently large, and all $x \in S$ with $\|x\| > r$. Hence, taking $\gamma > 1$ and $w > 2s_0$ arbitrarily close to their lower bounds, we get $\mathbb{E}_x(\tau_r^s) = \infty$ for all $s > s_0$ and appropriate r, x. This proves the non-existence of moments part of (a)(i) in the case $\beta^+ = \beta^-$.

Next suppose that $0 \le \beta^+, \beta^- \le \beta_c$ with $\beta^+ \ne \beta^-$. Without loss of generality, suppose that $0 \le \beta^- < \beta^+ = \beta \le \beta_c$. Then $0 \le s_0 < 1/2$. We consider the function $F_w^{\gamma,\nu}$ given by (48) with $\theta_0 = \theta_1$ given by (23), $\lambda > 0$, $w \in (2s_0, 1)$, and $\gamma > 1$ such that $\gamma w < 1$. Also, take ν for which $\gamma w + \beta^- - 2 < 2\nu < \gamma w + \beta^+ - 2$. Then by choice of γ and w, we have that the right-hand sides of (49) and (50) are both eventually positive. Since $\lambda > 0$, the right-hand side of (51) is also eventually positive. Thus

$$\mathbb{E}[F_w^{\gamma,\nu}(\xi_{n+1}) - F_w^{\gamma,\nu}(\xi_n) \mid \xi_n = x] \ge 0,$$

for all $x \in S$ with $\|x\|$ sufficiently large. Take $Y_n := (F_w^{\gamma,\nu}(\xi_n))^{1/(\gamma w)}$. Then we have shown that, for this Y_n, the condition (79) holds. Moreover, since $\gamma w < 1$ we have from convexity that (80) also holds. Again let $E_x = \{\|\Delta\| < \|x\|^\delta\}$. From (41) and (52),

$$\left| F_w^{\gamma,\nu}(x + y) - F_w^{\gamma,\nu}(x) \right| \le C \|y\| \|x\|^{\gamma w - 1},$$

for all $y \in B_{r/2}(x)$. Then, by another Taylor's theorem calculation,

$$\left| \left(F_w^{\gamma,\nu}(x + y) \right)^{1/(\gamma w)} - \left(F_w^{\gamma,\nu}(x) \right)^{1/(\gamma w)} \right| \le C \|y\|,$$

for all $y \in B_{r/2}(x)$. It follows that $\mathbb{E}_x[(Y_1 - Y_0)^2 \mathbf{1}_{E_x}] \le C$. Moreover, by a similar argument to (40), $|Y_1 - Y_0|^2 \le C \|\Delta\|^{2\gamma w/\delta}$ on E_x^c, so taking $\delta \in (2/p, 1)$ and using the fact that $\gamma w < 1$, we get $\mathbb{E}_x[(Y_1 - Y_0)^2 \mathbf{1}_{E_x^c}] \le C$ as well. Thus we also verify (81) in this case. Then we may again apply Theorem 2.7.4 of [26] to get $\mathbb{E}_x(\tau_r^s) = \infty$ for all $s > \gamma w/2$, and hence all $s > s_0$. This completes the proof of (a)(i).

For part (a)(ii), suppose first that $\beta^+ = \beta^- = \beta$, and that $\beta_c \le \beta < 1$. We apply the function f_w^γ with $w > 0$ and $\gamma > 1$. Then we have from (30) and (31) that (78) holds. Repeating the argument below (78) shows that $\mathbb{E}_x(\tau_r^s) = \infty$ for all

$s > \gamma w/2$, and hence all $s > 0$. The case where $\beta^+ \neq \beta^-$ is similar, using an appropriate $F_w^{\gamma,\nu}$. This proves (a)(ii).

It remains to consider the case where $\beta^+, \beta^- > 1$. Now we apply f_w^γ with $\gamma > 1$ and $w \in (0, 1/2)$ small enough, noting Remark 4. In this case (30) with (32) and Lemma 3 show that (78) holds, and repeating the argument below (78) shows that $\mathbb{E}_x(\tau_r^s) = \infty$ for all $s > 0$. This proves part (b). □

Appendix: Properties of the Threshold Function

For a constant $b \neq 0$, consider the function

$$\phi(\alpha) = \sin^2 \alpha + b \sin 2\alpha.$$

Set $\alpha_0 := \frac{1}{2} \arctan(-2b)$, which has $0 < |\alpha_0| < \pi/4$.

Lemma 11 *There are two stationary points of ϕ in $[-\frac{\pi}{2}, \frac{\pi}{2}]$. One of these is a local minimum at α_0, with*

$$\phi(\alpha_0) = \frac{1}{2}\left(1 - \sqrt{1 + 4b^2}\right) < 0.$$

The other is a local maximum, at $\alpha_1 = \alpha_0 + \frac{\pi}{2}$ if $b > 0$, or at $\alpha_1 = \alpha_0 - \frac{\pi}{2}$ if $b < 0$, with

$$\phi(\alpha_1) = \frac{1}{2}\left(1 + \sqrt{1 + 4b^2}\right) > 1.$$

Proof We compute $\phi'(\alpha) = \sin 2\alpha + 2b \cos 2\alpha$ and $\phi''(\alpha) = 2\cos 2\alpha - 4b \sin 2\alpha$. Then $\phi'(\alpha) = 0$ if and only if $\tan 2\alpha = -2b$. Thus the stationary values of ϕ are $\alpha_0 + k\frac{\pi}{2}$, $k \in \mathbb{Z}$. Exactly two of these values fall in $[-\frac{\pi}{2}, \frac{\pi}{2}]$, namely α_0 and α_1 as defined in the statement of the lemma. Also

$$\phi''(\alpha_0) = 2\cos 2\alpha_0 - 4b \sin 2\alpha_0 = \left(2 + 8b^2\right) \cos 2\alpha_0 > 0,$$

so α_0 is a local minimum. Similarly, if $|\delta| = \pi/2$, then $\sin 2\delta = 0$ and $\cos 2\delta = -1$, so

$$\phi''(\alpha_0 + \delta) = -\cos 2\alpha_0 + 4b \sin 2\alpha_0 = -\phi''(\alpha_0),$$

and hence the stationary point at α_1 is a local maximum. Finally, to evaluate the values of ϕ at the stationary points, note that

$$\cos 2\alpha_0 = \frac{1}{\sqrt{1 + 4b^2}}, \text{ and } \sin 2\alpha_0 = \frac{-2b}{\sqrt{1 + 4b^2}},$$

and use the fact that $2\sin^2\alpha_0 = 1 - \cos 2\alpha_0$ to get $\phi(\alpha_0)$, and that $2\cos^2\alpha_0 = \cos 2\alpha_0 + 1$ to get $\phi(\alpha_1) = \cos^2\alpha_0 - b\sin 2\alpha_0 = 1 - \phi(\alpha_0)$. □

Proof (of Proposition 1) By Lemma 11 (and considering separately the case $\sigma_1^2 = \sigma_2^2$) we see that the extrema of $\beta_c(\Sigma, \alpha)$ over $\alpha \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ are

$$\frac{\sigma_1^2 + \sigma_2^2}{2\sigma_2^2} \pm \frac{1}{2\sigma_2^2}\sqrt{\left(\sigma_2^2 - \sigma_1^2\right)^2 + 4\rho^2},$$

as claimed at (6). It remains to show that the minimum is strictly positive, which is a consequence of the fact that

$$\sigma_1^2 + \sigma_2^2 - \sqrt{\left(\sigma_1^2 + \sigma_2^2\right)^2 - 4\left(\sigma_1^2\sigma_2^2 - \rho^2\right)} > 0,$$

since $\rho^2 < \sigma_1^2\sigma_2^2$ (as Σ is positive definite). □

Acknowledgments A.M. is supported by The Alan Turing Institute under the EPSRC grant EP/N510129/1 and by the EPSRC grant EP/P003818/1 and the Turing Fellowship funded by the Programme on Data-Centric Engineering of Lloyd's Register Foundation. Some of the work reported in this paper was undertaken during visits by M.M. and A.W. to The Alan Turing Institute, whose hospitality is gratefully acknowledged. The authors also thank two referees for helpful comments and suggestions.

References

1. Aspandiiarov, S.: On the convergence of 2-dimensional Markov chains in quadrants with boundary reflection. Stochastics Stochastics Rep. **53**, 275–303 (1995)
2. Aspandiiarov, S., Iasnogorodski, R.: Tails of passage-times and an application to stochastic processes with boundary reflection in wedges. Stochastic Process. Appl. **66**, 115–145 (1997)
3. Aspandiiarov, S., Iasnogorodski, R.: General criteria of integrability of functions of passage-times for nonnegative stochastic processes and their applications. Theory Probab. Appl. **43**, 343–369 (1999). Translated from Teor. Veroyatnost. i Primenen. **43**, 509–539 (1998) (in Russian)
4. Aspandiiarov, S., Iasnogorodski, R.: Asymptotic behaviour of stationary distributions for countable Markov chains, with some applications. Bernoulli **5**, 535–569 (1999)
5. Aspandiiarov, S., Iasnogorodski, R., Menshikov, M.: Passage-time moments for nonnegative stochastic processes and an application to reflected random walks in a quadrant. Ann. Probab. **24**, 932–960 (1996)
6. Asymont, I.M., Fayolle, G., Menshikov, M.V.: Random walks in a quarter plane with zero drifts: transience and recurrence. J. Appl. Probab. **32**, 941–955 (1995)
7. Balaji, S., Ramasubramanian, S.: Passage time moments for multidimensional diffusions. J. Appl. Probab. **37**, 246–251 (2000)
8. Bañuelos, R., Carroll, T.: Sharp integrability for Brownian motion in parabola-shaped regions. J. Funct. Anal. **218**, 219–253 (2005)
9. Bañuelos, R., DeBlassie, R.D., Smits, R.: The first exit time of a planar Brownian motion from the interior of a parabola. Ann. Probab. **29**, 882–901 (2001)

10. Fayolle, G.: On random walks arising in queueing systems: ergodicity and transience via quadratic forms as Lyapunov functions – Part I. Queueing Syst. **5**, 167–184 (1989)
11. Fayolle, G., Iasnogorodski, R., Malyshev, V.: Random Walks in the Quarter-Plane. 2nd edn. Springer, Berlin (2017)
12. Fayolle, G., Malyshev, V.A., Menshikov, M.V.: Random walks in a quarter plane with zero drifts. I. Ergodicity and null recurrence. Ann. Inst. Henri Poincaré **28**, 179–194 (1992)
13. Fayolle, G., Malyshev, V.A., Menshikov, M.V.: Topics in the Constructive Theory of Countable Markov Chains. Cambridge University Press, Cambridge (1995)
14. Foster, F.G.: On the stochastic matrices associated with certain queuing processes. Ann. Math. Stat. **24**, 355–360 (1953)
15. Franceschi, S., Raschel, K.: Integral expression for the stationary distribution of reflected Brownian motion in a wedge. Bernoulli **25**, 3673–3713 (2019)
16. Hobson, D.G., Rogers, L.C.G.: Recurrence and transience of reflecting Brownian motion in the quadrant. Math. Proc. Cambridge Philos. Soc. **113**, 387–399 (1993)
17. Ignatyuk, I.A., Malyshev, V.A.: Classification of random walks in $\mathbb{Z}_+^4$. Selecta Math. **12**, 129–194 (1993)
18. Kingman, J.F.C.: The ergodic behaviour of random walks. Biometrika **48**, 391–396 (1961)
19. Lamperti, J.: Criteria for stochastic processes II: passage-time moments. J. Math. Anal. Appl. **7**, 127–145 (1963)
20. Li, W.V.: The first exit time of a Brownian motion from unbounded convex domains. Ann. Probab. **31**, 1078–1096 (2003)
21. MacPhee, I.M., Menshikov, M.V., Wade, A.R.: Moments of exit times from wedges for non-homogeneous random walks with asymptotically zero drifts. J. Theoret. Probab. **26**, 1–30 (2013)
22. Malyshev, V.A.: Random Walks, The Wiener-Hopf Equation in a Quarter Plane, Galois Automorphisms. Moscow State University, Moscow (1970) (in Russian)
23. Malyshev, V.A.: Classification of two-dimensional positive random walks and almost linear semimartingales. Soviet Math. Dokl. **13**, 136–139 (1972). Translated from Dokl. Akad. Nauk SSSR **202**, 526–528 (1972) (in Russian)
24. Menshikov, M.V.: Ergodicity and transience conditions for random walks in the positive octant of space. Soviet Math. Dokl. **15**, 1118–1121 (1974). Translated from Dokl. Akad. Nauk SSSR **217**, 755–758 (1974) (in Russian)
25. Menshikov, M., Williams, R.J.: Passage-time moments for continuous non-negative stochastic processes and applications. Adv. Appl. Probab. **28**, 747–762 (1996)
26. Menshikov, M., Popov, S., Wade, A.: Non-homogeneous random walks. Cambridge University Press, Cambridge (2017)
27. Mikhailov, V.A.: Methods of Random Multiple Access, Thesis. Dolgoprudny, Moscow, 1979 (in Russian)
28. Pinsky, R.G.: Transience/recurrence for normally reflected Brownian motion in unbounded domains. Ann. Probab. **37**, 676–686 (2009)
29. Rosenkrantz, W.A.: Ergodicity conditions for two-dimensional Markov chains on the positive quadrant. Probab. Theory Relat. Fields **83**, 309–319 (1989)
30. Vaninskii, K.L., Lazareva, B.V.: Ergodicity and nonrecurrence conditions of a homogeneous Markov chain in the positive quadrant. Probl. Inf. Transm. **24**, 82–86 (1988). Translated from Problemy Peredachi Informatsii **24**, 105–110 (1988) (in Russian)
31. Varadhan, S.R.S., Williams, R.: Brownian motion in a wedge with oblique reflection. Comm. Pure Appl. Math. **38**, 405–443 (1985)
32. Williams, R.J.: Recurrence classification and invariant measures for reflected Brownian motion in a wedge. Ann. Probab. **13**, 758–778 (1985)
33. Zachary, S.: On two-dimensional Markov chains in the positive quadrant with partial spatial homogeneity. Markov Process. Related Fields **1**, 267–280 (1995)

Noise Stability of Weighted Majority

Yuval Peres

Abstract Benjamini et al. (Inst Hautes Études Sci Publ Math 90:5–43, 2001) showed that weighted majority functions of n independent unbiased bits are uniformly stable under noise: when each bit is flipped with probability ϵ, the probability p_ϵ that the weighted majority changes is at most $C\epsilon^{1/4}$. They asked what is the best possible exponent that could replace $1/4$. We prove that the answer is $1/2$. The upper bound obtained for p_ϵ is within a factor of $\sqrt{\pi/2} + o(1)$ from the known lower bound when $\epsilon \to 0$ and $n\epsilon \to \infty$.

Keywords Noise sensitivity · Boolean functions · Weighted majority

MSC 60C05

1 Introduction

In their study of noise sensitivity and stability of Boolean functions, Benjamini et al. [2] showed that weighted majority functions of n independent unbiased ± 1-valued variables are uniformly stable under noise: When each variable is flipped with probability ϵ, the weighted majority changes with probability at most $C\epsilon^{1/4}$. They asked what is the best possible exponent that could replace $1/4$. In this note we prove that the answer is $1/2$. The upper bound obtained for p_ϵ is within a factor of $\sqrt{\pi/2} + o(1)$ from the known lower bound when $\epsilon \to 0$ and $n\epsilon \to \infty$.

Remark The result presented here was obtained in 1999 (as mentioned in [2]) and was uploaded on the arxiv in 2004, but was not previously published. It was featured as a highlight in Section 5.5 of the book [9].

Y. Peres (✉)
Kent State University, Kent, OH, USA

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_27

Denote $\operatorname{sgn}(u) = u/|u|$ for $u \neq 0$ and $\operatorname{sgn}(0) = 0$, and let $N_\epsilon : \mathbb{R}^n \to \mathbb{R}^n$ be the noise operator that flips each variable in its input independently with probability ϵ. Formally, given a random vector $X = (X_1, \ldots, X_n)$, the random vector $N_\epsilon(X)$ is defined as $(\sigma_1 X_1, \ldots, \sigma_n X_n)$ where the i.i.d. random variables σ_i are independent of X and take the values $1, -1$ with probabilities $1-\epsilon$, ϵ respectively.

Theorem 1 *Let $X = (X_1, \ldots, X_n)$ be a random vector uniformly distributed over $\{-1, 1\}^n$. Given nonzero weights $w_1, \ldots, w_n \in \mathbb{R}$ and a threshold $t \in \mathbb{R}$, consider the weighted majority function $f : \mathbb{R}^n \to \{-1, 0, 1\}$ defined by*

$$f(x) = \operatorname{sgn}\Big(\sum_{i=1}^n w_i x_i - t\Big) \tag{1}$$

Then for $\epsilon \le 1/2$,

$$p_\epsilon(n, w, t) = \mathbb{P}\Big(f(X) \neq f(N_\epsilon(X))\Big) \le 2\epsilon^{1/2}. \tag{2}$$

Moreover, $p_\epsilon^* = \limsup_{n\to\infty} \sup_{w,t} p_\epsilon(n, w, t)$ satisfies

$$\limsup_{\epsilon\to 0} \frac{p_\epsilon^*}{\sqrt{\epsilon}} \le \sqrt{2/\pi}\,. \tag{3}$$

In the statement of the theorem we opted for a simple formulation: Our proof yields the following sharper, but more involved estimate:

$$p_\epsilon(n, w, t) \le \frac{2}{m} \cdot \mathbb{E}\,|B_m - \frac{m}{2}| + [1 - (1-\epsilon)^n]\binom{n}{\lfloor n/2 \rfloor} 2^{-n}\,, \tag{4}$$

where $m = \lfloor \epsilon^{-1} \rfloor$ and B_m is a Binomial$(m, 1/2)$ variable.

It easy to see, and classical [4, 11], that for simple majority (when all weights are equal) we have

$$\lim_{n\to\infty} \mathbb{P}(\operatorname{sgn} \sum_{i=1}^n X_i \neq \operatorname{sgn} \sum_{i=1}^n (N_\epsilon X)_i) = \frac{1}{\pi} \arccos(1 - 2\epsilon) = \frac{2}{\pi}\sqrt{\epsilon} + O(\epsilon^{3/2})\,. \tag{5}$$

For the reader's convenience we include a brief argument:

Since $\operatorname{Cov}\Big(\sum_{i=1}^n X_i,\ \sum_{i=1}^n (N_\epsilon X)_i\Big) = n(1 - 2\epsilon)$, the central limit theorem implies that as $n \to \infty$,

$$\frac{1}{\sqrt{n}}\Big(\sum_{i=1}^n X_i, \sum_{i=1}^n (N_\epsilon X)_i\Big) \Rightarrow (Z_1, Z_1^*) \quad \text{in law,}$$

where Z_1, Z_1^* are standard normals with covariance $1-2\epsilon$. We can write $Z_1^* = Z_1 \cos\alpha - Z_2 \sin\alpha$ where Z_1, Z_2 are i.i.d. standard normals and $\alpha \in (0,\pi)$ satisfies $\cos\alpha = 1-2\epsilon$. Rotating the random vector (Z_1, Z_2) by the angle α yields a vector with first coordinate Z_1^*. Since (Z_1, Z_2) has a rotationally-symmetric law, the rotation changes the sign of the first coordinate with probability α/π. This verifies the left-hand side of (5); the right-hand side follows from Taylor expansion of cosine.

Thus the estimate (2) is sharp (up to the value of the constant). Moreover, the ratio between the upper bound in (3) and the value for simple majority in (5) tends to $\sqrt{\pi/2} < 1.26$ as $\epsilon \to 0$. We remark that the stability result in theorem 1 is stronger than an assertion about stability of half-spaces, $\{x : \sum_i w_i x_i > \theta\}$, because we consider the weighted majority as taking three values, rather than two.

2 Proof of Theorem 1

Using symmetry of X_i, we may assume that $w_i > 0$ for $i = 1, \ldots, n$. Let $\langle w, X\rangle = \sum_{i=1}^n w_i X_i$. We first consider the threshold $t = 0$. Later, we will extend the argument to thresholds $t \neq 0$.

We will need the following well-known fact from [3]:

$$\mathbb{P}\Big(\langle w, X\rangle = 0\Big) \le \binom{n}{\lfloor n/2\rfloor} 2^{-n}. \tag{6}$$

Indeed, the collection $\mathscr{D}(w)$ of sets $D \subset \{1, \ldots n\}$ such that $\sum_{i\in D} w_i = \sum_{k\notin D} w_k$ forms an anti-chain with respect to inclusion, so Sperner's theorem (see [1, Ch. 11]) implies that the cardinality of $\mathscr{D}(w)$ is at most $\binom{n}{\lfloor n/2\rfloor}$. Finally, observe that a vector $x \in \{-1,1\}^n$ satisfies $\langle w, x\rangle = 0$ iff $\{i : x_i = 1\}$ is in $\mathscr{D}(w)$.

Let $m = \lfloor \epsilon^{-1} \rfloor$ and let τ be a random variable taking the values $0, 1, \ldots, m$, with $\mathbb{P}(\tau = j) = \epsilon$ for $j = 1, \ldots, m$ and $\mathbb{P}(\tau = 0) = 1 - m\epsilon$. We use a sequence $\tau_1, \tau_2, \ldots, \tau_n$ of i.i.d. random variables with the same law as τ, to partition $[n] = \{1, \ldots, n\}$ into $m+1$ random sets

$$A_j = \Big\{ i \in [n] : \tau_i = j \Big\} \quad \text{for } 0 \le j \le m. \tag{7}$$

Denote $S_j = \sum_{i\in A_j} w_i X_i$ and let $Y_1 = \sum_{i\notin A_1} w_i X_i = \langle w, X\rangle - S_1$. Observe that $Y_1 - S_1$ has the same law, given X, as $\langle w, N_\epsilon(X)\rangle$. Therefore,

$$\begin{aligned} p_\epsilon(n, w, 0) &= \mathbb{P}\Big(\operatorname{sgn}\langle w, X\rangle \neq \operatorname{sgn}\langle w, N_\epsilon(X)\rangle\Big) \\ &= \mathbb{P}\Big(\operatorname{sgn}(Y_1 + S_1) \neq \operatorname{sgn}(Y_1 - S_1)\Big). \end{aligned} \tag{8}$$

Denote $\xi_j = \operatorname{sgn}(S_j)$. A key step in the proof is the pointwise identity

$$\begin{aligned}&\mathbf{1}_{\{\operatorname{sgn}(Y_1+S_1)\neq \operatorname{sgn}(Y_1-S_1)\}} \\ &\quad = 2\cdot \mathbf{1}_{\{S_1\neq 0\}}\,\mathbb{E}\Big(\frac{1}{2}-\mathbf{1}_{\{\operatorname{sgn}(S_1+Y_1)=-\xi_1\}}\Big|\, Y_1, |S_1|\Big)\,.\end{aligned}\tag{9}$$

To verify this, we consider three cases:

(i) Clearly both sides vanish if $S_1 = 0$.

(ii) Suppose that $0 < |S_1| < |Y_1|$ and therefore $\operatorname{sgn}(Y_1 + S_1) = \operatorname{sgn}(Y_1)$. The conditional distribution of S_1 given Y_1 and $|S_1|$ is uniform over $\{-|S_1|, |S_1|\}$, whence the conditional probability that $\operatorname{sgn}(S_1 + Y_1) = -\xi_1$ is $1/2$. Thus both sides of (9) also vanish in this case.

(iii) Finally, suppose that $S_1 \neq 0$ and $|S_1| \geq |Y_1|$. In this case $\operatorname{sgn}(S_1 + Y_1) \neq -\xi_1$, so both sides of (9) equal 1.

Taking expectations in (9) and using (8), we deduce that

$$\begin{aligned}p_\epsilon(n,w,0) &= 2\,\mathbb{E}\Big[\mathbf{1}_{\{S_1\neq 0\}}\Big(\frac{1}{2}-\mathbf{1}_{\{\operatorname{sgn}\langle w,X\rangle=-\xi_1\}}\Big)\Big]\\ &= \frac{2}{m}\mathbb{E}\sum_{j\in\Lambda}\Big(\frac{1}{2}-\mathbf{1}_{\{\operatorname{sgn}\langle w,X\rangle=-\xi_j\}}\Big)\,,\end{aligned}\tag{10}$$

where $\Lambda = \{j \in [1, m] :\ S_j \neq 0\}$.

The random variable $B_\Lambda = \#\{j \in \Lambda : \xi_j = 1\}$ has a Binomial$(\#\Lambda, \frac{1}{2})$ distribution given Λ, and satisfies the pointwise inequality

$$\sum_{j\in\Lambda}\Big(\frac{1}{2}-\mathbf{1}_{\{\operatorname{sgn}\langle w,X\rangle=-\xi_j\}}\Big) \leq \Big|B_\Lambda-\frac{\#\Lambda}{2}\Big| + \frac{1}{2}\mathbf{1}_{\{\langle w,X\rangle=0\}}\sum_{j=1}^{m}\mathbf{1}_{\{A_j\neq\emptyset\}}\,.$$

To see this, consider the three possibilities for $\operatorname{sgn}\langle w, X\rangle$. Taking expectations and using (10), we get

$$p_\epsilon(n,w,0) \leq \frac{2}{m}\mathbb{E}\Big|B_\Lambda-\frac{\#\Lambda}{2}\Big| + \mathbb{P}(A_1\neq\emptyset)\mathbb{P}(\langle w,X\rangle=0).\tag{11}$$

Let B_ℓ denote a Binomial$(\ell, \frac{1}{2})$ random variable. Since for any martingale $\{M_\ell\}_{\ell\geq 1}$ the absolute values $|M_\ell|$ form a submartingale, the expression $\mathbb{E}|B_\ell - \frac{\ell}{2}|$ is increasing in ℓ. By averaging over Λ, we see that $\mathbb{E}|B_\Lambda - \frac{\#\Lambda}{2}| \leq \mathbb{E}|B_m - \frac{m}{2}|$. In conjunction with (11) and (6), this implies

$$p_\epsilon(n,w,0) \leq \frac{2}{m}\mathbb{E}|B_m-\frac{m}{2}| + [1-(1-\epsilon)^n]\binom{n}{\lfloor n/2\rfloor}2^{-n}.\tag{12}$$

Next, suppose that $f(x) = \operatorname{sgn}\Big(\sum_{i=1}^n w_i x_i - t\Big)$, where $t \neq 0$ is a given threshold. Let X_{n+1} be a ± 1 valued symmetric random variable, independent of $X = (X_1, \ldots, X_n)$, and define $w_{n+1} = t$. Then

$$p_\epsilon(n, w, t) = \mathbb{P}\Big(f(X) \neq f(N_\epsilon(X))\Big) \tag{13}$$

$$= \mathbb{P}\Big(\operatorname{sgn}\sum_{i=1}^{n+1} w_i X_i \neq \operatorname{sgn}\big[\sum_{i=1}^{n} w_i (N_\epsilon X)_i + w_{n+1} X_{N+1}\big]\Big),$$

and the argument used above to establish the bound (12) for $p_\epsilon(n, w, 0)$, yields the same bound for $p_\epsilon(n, w, t)$. This proves (4).

To derive (2), we may assume that $\epsilon \leq 1/4$. Use Cauchy-Schwarz to write $\mathbb{E}|B_m - \frac{m}{2}| \leq \sqrt{\operatorname{Var}(B_m)} = \sqrt{m/4}$ and apply the elementary inequalities

$$\binom{n}{\lfloor n/2 \rfloor} 2^{-n} \leq \sqrt{3/4}\, n^{-1/2},$$

(see, e.g., [10, Section 2.3]) and $[1 - (1-\epsilon)^n] \leq \min\{n\epsilon, 1\} \leq \sqrt{n\epsilon}$, to obtain

$$p_\epsilon(n, w, t) \leq m^{-1/2} + \sqrt{n\epsilon} \cdot \sqrt{3/4}\, n^{-1/2}. \tag{14}$$

Since $m = \lfloor \epsilon^{-1} \rfloor \geq 4/(5\epsilon)$ for $\epsilon \leq 1/4$, we conclude that

$$p_\epsilon(n, w, t) \leq \left(\sqrt{5/4} + \sqrt{3/4}\right) \epsilon^{1/2} < 2\epsilon^{1/2},$$

and this proves (2).

Finally, the central limit theorem implies that

$$\lim_{m \to \infty} \frac{\mathbb{E}|2B_m - m|}{\sqrt{m}} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |u| e^{-u^2/2}\, du = \sqrt{2/\pi}.$$

This proves (3). □

Remark

1. Theorem 1 states that among linear threshold functions, majority is the most sensitive (up to a constant factor). Conversely, [8] showed that among balanced functions where each variable has low influence, majority is asymptotically the most stable.
2. The randomization idea which is crucial to our proof of Theorem 1 was inspired by an argument of Matthews [7] to bound cover times for Markov chains. See also [12] for related random walk estimates.

3. After I presented the proof of Theorem 1 to R. O'Donnell in 2001, he found (jointly with A. Klivans and R. Servedio) some variants and applications of the argument to learning theory, see [6] and the book [9].

4. The proof of Theorem 1 extends verbatim to the case where X_i are independent symmetric real-valued random variables with $\mathbb{P}(X_i = 0) = 0$ for all i. However, this extension reduces to Theorem 1 by conditioning on $|X_i|$.

5. In [2, Remark 3.6], the authors asked whether simple majority is the most noise sensitive of the weighted majority functions. Several people, including Noam Berger (2003, personal communication), Sivakanth Gopi and Daniel Kane found counterexamples for a small number of variables; see, e.g., [5]. However, an asymptotic version of this question is still open.

6. Is simple majority the most noise sensitive of the weighted majority functions, asymptotically when $\epsilon \to 0$ and $n\epsilon \to \infty$?
In particular, is it possible to replace the right-hand side of (3) by $2/\pi$?

Acknowledgments I am grateful to I. Benjamini, G. Kalai and O. Schramm for suggesting the problem, and to E. Mossel, R. Peled, O. Schramm, R. Siegmund-Schultze and H. V. Weizsäcker for useful discussions.

References

1. Alon, N., Spencer, J.: The Probabilistic Method. Wiley, New York (1992)
2. Benjamini, I., Kalai, G., Schramm, O.: Noise sensitivity of Boolean functions and applications to percolation. Inst. Hautes Études Sci. Publ. Math. **90**, 5–43 (2001)
3. Erdős, P.: On a lemma of Littlewood and Offord. Bull. Am. Math. Soc. **51**, 898–902 (1945)
4. Guilbaud, G.: Theories of the general interest, and the logical problem of aggregation. In: Lazarsfeld, P.F., Henry, N.W. (eds.) Readings in Mathematical Social Science, pp. 262–307. MIT Press, Cambridge (1966)
5. Jain, V.: A counterexample to the "Majority is Least Stable" conjecture (2017). Available via arxiv. http://arxiv.org/abs/1703.07657 cited 21 July 2020
6. Klivans, A., O'Donnell, R., Servedio, R.: Learning intersections and thresholds of halfspaces. J. Comput. Syst. Sci. **68**, 808–840 (2004)
7. Matthews, P.: Covering problems for Brownian motion on spheres. Ann. Probab. **16**, 189–199 (1988)
8. Mossel, E., O'Donnell, R., Oleszkiewicz, K.: Noise stability of functions with low influences: invariance and optimality. Ann. Math. **171**(1), 295–341 (2010)
9. O'Donnell, R.: Analysis of Boolean Functions. Cambridge University Press, Cambridge (2014)
10. Pitman, J.: Probability. Springer Texts in Statistics. Springer, New York (1993)
11. Sheppard, W.: On the application of the theory of error to cases of normal distribution and normal correlations. Philos. Trans. R. Soc. Lond. **192**, 101–168 (1899)
12. Siegmund-Schultze, R., von Weizsäcker, H.: Level crossing probabilities I: one-dimensional random walks and symmetrization. Adv. Math. **208**(2), 672–679 (2007)

Scaling Limits of Linear Random Fields on $\mathbb{Z}^2$ with General Dependence Axis

Vytautė Pilipauskaitė and Donatas Surgailis

To the memory of Vladas and the time we enjoyed together in Vilnius, Rio and Shanghai

Abstract We discuss anisotropic scaling limits of long-range dependent linear random fields X on $\mathbb{Z}^2$ with arbitrary dependence axis (direction in the plane along which the moving-average coefficients decay at a smallest rate). The scaling limits V^X_γ are random fields on $\mathbb{R}^2_+$ defined as the limits (in the sense of finite-dimensional distributions) of partial sums of X taken over rectangles with sides increasing along horizontal and vertical directions at rates λ and λ^γ respectively as $\lambda \to \infty$ for arbitrary fixed $\gamma > 0$. The scaling limits generally depend on γ and constitute a one-dimensional family $\{V^X_\gamma, \gamma > 0\}$ of random fields. The scaling transition occurs at some $\gamma^X_0 > 0$ if V^X_γ are different and do not depend on γ for $\gamma > \gamma^X_0$ and $\gamma < \gamma^X_0$. We prove that the fact of 'oblique' dependence axis (or incongruous scaling) dramatically changes the scaling transition in the above model so that $\gamma^X_0 = 1$ independently of other parameters, contrasting the results in Pilipauskaitė and Surgailis (2017) on the scaling transition under congruous scaling.

Keywords Random field · Long-range dependence · Dependence axis · Anisotropic scaling limits · Scaling transition · Fractional Brownian sheet

MSC 2010 Primary 60G60; Secondary 60G15, 60618, 60G22

V. Pilipauskaitė
Department of Mathematics, Aarhus University, Aarhus C, Denmark

D. Surgailis (✉)
Faculty of Mathematics and Informatics, Vilnius University, Vilnius, Lithuania
e-mail: donatas.surgailis@mif.vu.lt

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_28

1 Introduction

Pilipauskaitė and Surgailis [12–14], Damarackas and Paulauskas [2, 3], Puplinskaitė and Surgailis [16, 17], and Surgailis [19, 21] discussed scaling limits

$$\left\{A_{\lambda,\gamma}^{-1} S_{\lambda,\gamma}^{X}(\mathbf{x}), \mathbf{x} \in \mathbb{R}_{+}^{\nu}\right\} \stackrel{\mathrm{fdd}}{\to} V_{\gamma}^{X}, \quad \lambda \to \infty, \tag{1}$$

for some classes of stationary random fields (RFs) $X = \{X(\mathbf{t}), \mathbf{t} \in \mathbb{Z}^{\nu}\}$, where $A_{\lambda,\gamma} \to \infty$ is a normalization and

$$S_{\lambda,\gamma}^{X}(\mathbf{x}) := \sum_{\mathbf{t} \in K_{\lambda,\gamma}(\mathbf{x})} X(\mathbf{t}), \quad \mathbf{x} \in \mathbb{R}_{+}^{\nu}, \tag{2}$$

are partial sums of RF X taken over rectangles $K_{\lambda,\gamma}(\mathbf{x}) := \{\mathbf{t} = (t_1, \ldots, t_\nu)^\top \in \mathbb{Z}^\nu : 0 < t_i \le \lambda^{\gamma_i} x_i,\ i = 1, \ldots, \nu\}$ for *arbitrary* fixed collection $\gamma = (\gamma_1, \ldots, \gamma_\nu)^\top \in \mathbb{R}_{+}^{\nu}$ of exponents, with sides $[1, \ldots, \lfloor \lambda^{\gamma_i} x_i \rfloor]$ increasing at generally different rates $O(\lambda^{\gamma_i})$, $i = 1, \ldots, \nu$. Following [13, 19] the family $\{V_{\gamma}^{X}, \gamma \in \mathbb{R}_{+}^{\nu}\}$ of all scaling limits in (1) will be called the *scaling diagram of RF X*. Recall that a stationary RF X with $\mathrm{Var}(X(\mathbf{0})) < \infty$ is said long-range dependent (LRD) if $\sum_{\mathbf{t} \in \mathbb{Z}^\nu} |\mathrm{Cov}(X(\mathbf{0}), X(\mathbf{t}))| = \infty$, see [10, 17]. [14, 16, 17] observed that for a large class of LRD RFs X in dimension $\nu = 2$, the scaling diagram essentially consists of three points. More precisely (assuming $\gamma_1 = 1, \gamma_2 = \gamma$ w.l.g.), there exists a (nonrandom) $\gamma_0^X > 0$ such that $V_\gamma^X \equiv V_{(1,\gamma)}^X$ do not depend on γ for $\gamma > \gamma_0^X$ and $\gamma < \gamma_0^X$, viz.,

$$V_{\gamma}^{X} = \begin{cases} V_{+}^{X}, & \gamma > \gamma_0^X, \\ V_{-}^{X}, & \gamma < \gamma_0^X, \\ V_{0}^{X}, & \gamma = \gamma_0^X, \end{cases} \tag{3}$$

and $V_{+}^{X} \stackrel{\mathrm{fdd}}{\neq} a V_{-}^{X}$, $\forall a > 0$. We note that $V_{\pm}^{X}$, V_0^X and γ_0^X generally depend on the distribution (the model parameters) of X. The above fact was termed the *scaling transition* [16, 17], V_0^X called the *well-balanced* and $V_{\pm}^X$ the *unbalanced* scaling limits of X. In the sequel, we shall also refer to $\gamma_0^X > 0$ in (3) as the *scaling transition* or the *critical point*. Intuitively, $V_{\pm}^X$ arise when the scaling in the vertical direction is much stronger than in the horizontal one or vice versa. The well-balanced limit V_0^X corresponds to the balanced scaling in the two directions and is generally different from both $V_{\pm}^X$. The existence of the scaling transition was established for a wide class of planar linear and nonlinear RF models including those appearing in telecommunications and econometrics. See the review paper [20] for further discussion and recent developments.

Particularly, [14] discussed anisotropic scaling of linear LRD RFs X on $\mathbb{Z}^2$ written as a moving-average

$$X(\mathbf{t}) = \sum_{\mathbf{s}\in\mathbb{Z}^2} a(\mathbf{t}-\mathbf{s})\varepsilon(\mathbf{s}), \quad \mathbf{t} \in \mathbb{Z}^2, \tag{4}$$

of standardized i.i.d. sequence $\{\varepsilon(\mathbf{t}), \mathbf{t} \in \mathbb{Z}^2\}$ with deterministic coefficients

$$\begin{aligned} a(\mathbf{t}) = &\frac{1}{|t_1|^{q_1} + |t_2|^{q_2}} \\ &\times \left(L_{\mathrm{sign}(t_2)} \left(\frac{t_1}{(|t_1|^{q_1} + |t_2|^{q_2})^{1/q_1}} \right) + o(1) \right), \quad |\mathbf{t}| := |t_1| + |t_2| \to \infty, \end{aligned} \tag{5}$$

where $q_i > 0$, $i = 1, 2$, satisfy

$$1 < Q := \frac{1}{q_1} + \frac{1}{q_2} < 2 \tag{6}$$

and $L_\pm$ are continuous functions on $[-1, 1]$ with $L_+(\pm 1) = L_-(\pm 1) =: L_0(\pm 1)$. The latter condition on $L_\pm$ and natural definition of L_0 guarantee continuity of the angular factor in (5) on the set $\{\mathbf{t} \in \mathbb{R}^2 : |t_1|^{q_1} + |t_2|^{q_2} = 1\}$. ([14, 21] use a slightly different form of moving-average coefficients a, and assume $L_+ = L_-$ but their results are valid for a in (5). See also Sect. 4 below.) Since $a(t, 0) = O(|t|^{-q_1})$, $a(0, t) = O(|t|^{-q_2})$, $t \to \infty$, for $q_1 \neq q_2$ decay at different rate in the horizontal and vertical directions, the ratio $\gamma^0 := q_1/q_2$ can be regarded as 'intrinsic (internal) scale ratio' and the exponent $\gamma > 0$ as 'external scale ratio', characterizing the anisotropy of the RF X in (4) and the scaling procedure in (1)–(2), respectively. Indeed, the scaling transition for the above X occurs at the point $\gamma_0^X = \gamma^0$ where these ratios coincide [14]. Let us remark that isotropic scaling of linear and nonlinear RFs on $\mathbb{Z}^\nu$ and $\mathbb{R}^\nu$ was discussed in [5, 6, 10, 11] and other works, while the scaling limits of linear random processes with one-dimensional 'time' (case $\nu = 1$) were identified in [4]. We also refer to the monographs [1, 7, 8] on various probabilistic and statistical aspects of long-range dependence.

A direction in the plane (a line passing through the origin) along which the moving-average coefficients a, decay at the smallest rate may be called the *dependence axis* of RF X in (4). The rigorous definition of dependence axis is given in Sect. 4. Due to the form in (5) the dependence axis agrees with the horizontal axis if $q_1 < q_2$ and with the vertical axis if $q_1 > q_2$, see Proposition 4. Since the scaling in (1)–(2) is parallel to the coordinate axes, we may say that for RF X in (4)–(5), *the scaling is congruous with the dependence axis of* X and the results of [14, 21] (as well as of [16, 17]) refer to this rather specific situation. The situation when the dependence axis does not agree with any of the two coordinate axes (the case of *incongruous scaling*) seems to be more common and then one may naturally ask about the scaling transition and the scaling transition point γ_0^X under incongruous scaling.

The present paper discusses the above problem for linear RF

$$X(\mathbf{t}) = \sum_{\mathbf{s}\in\mathbb{Z}^2} b(\mathbf{t}-\mathbf{s})\varepsilon(\mathbf{s}), \quad \mathbf{t}\in\mathbb{Z}^2, \tag{7}$$

with the moving-average coefficients

$$b(\mathbf{t}) = \frac{1}{|\mathbf{b}_1\cdot\mathbf{t}|^{q_1} + |\mathbf{b}_2\cdot\mathbf{t}|^{q_2}} \times \left(L_{\mathrm{sign}(\mathbf{b}_2\cdot\mathbf{t})}\left(\frac{\mathbf{b}_1\cdot\mathbf{t}}{(|\mathbf{b}_1\cdot\mathbf{t}|^{q_1} + |\mathbf{b}_2\cdot\mathbf{t}|^{q_2})^{1/q_1}}\right) + o(1)\right), \quad |\mathbf{t}|\to\infty, \tag{8}$$

where $\mathbf{b}_i\cdot\mathbf{t} := b_{i1}t_1 + b_{i2}t_2$ is the scalar product, $\mathbf{b}_i = (b_{i1}, b_{i2})^\top$, $i = 1, 2$, are real vectors, $B = (b_{ij})_{i,j=1,2}$ is a nondegenerate matrix and $q_i > 0$, $i = 1, 2$, $Q \in (1, 2)$, $L_\pm$ are the same as in (5). The above assumptions imply

$$\sum_{\mathbf{t}\in\mathbb{Z}^2} |b(\mathbf{t})| = \infty \quad \text{and} \quad \sum_{\mathbf{t}\in\mathbb{Z}^2} b(\mathbf{t})^2 < \infty, \tag{9}$$

in other words, X in (7) is a well-defined RF with LRD. The dependence axis of X with coefficients b in (8) is given by $\mathbf{t}\in\mathbb{R}^2$ for which

$$\mathbf{b}_2\cdot\mathbf{t} = 0 \;\; (q_1 < q_2) \quad \text{or} \quad \mathbf{b}_1\cdot\mathbf{t} = 0 \;\; (q_1 > q_2), \tag{10}$$

see Proposition 4 below, and generally does not agree with the coordinate axes, which results in incongruous scaling in (1). We prove that the last fact completely changes the scaling transition. Namely, under incongruous scaling *the scaling transition point* γ_0^X *in* (3) *is always 1*: $\gamma_0^X = 1$ for any $q_1 > 0$, $q_2 > 0$ satisfying (6), and the unbalanced limits $V_\pm^X$ are generally different from the corresponding limits in the congruous scaling case. The main results of this paper are illustrated in Table 1. Throughout the paper we use the notation

$$\tilde{Q}_i := Q - \frac{1}{2q_i}, \quad \tilde{H}_i := 1 - \frac{q_i}{2}(2-Q), \quad H_i := \frac{1}{2} + q_i(Q-1), \quad i = 1, 2, \tag{11}$$

and $B_{\mathscr{H}_1,\mathscr{H}_2} = \{B_{\mathscr{H}_1,\mathscr{H}_2}(\mathbf{x}), \mathbf{x}\in\mathbb{R}_+^2\}$ for fractional Brownian sheet (FBS) with Hurst parameters $0 < \mathscr{H}_i \le 1$, $i = 1, 2$, defined as a Gaussian RF with zero mean and covariance $\mathrm{E}B_{\mathscr{H}_1,\mathscr{H}_2}(\mathbf{x})B_{\mathscr{H}_1,\mathscr{H}_2}(\mathbf{y}) = (1/4)\prod_{i=1}^2 (x_i^{2\mathscr{H}_i} + y_i^{2\mathscr{H}_i} - |x_i - y_i|^{2\mathscr{H}_i})$, $\mathbf{x}, \mathbf{y}\in\mathbb{R}_+^2$.

If $Q < 1$ and $\sum_{\mathbf{t}\in\mathbb{Z}^2} b(\mathbf{t}) \ne 0$ then the corresponding RF X in (7) is short-range dependent and its all scaling limits V_γ^X agree with the Brownian Sheet $B_{1/2,1/2}$ for

Table 1 Unbalanced scaling limits $V^X_\pm$ (without asymptotic constants) under congruous and incongruous scaling

Parameter region	Congruous scaling			Incongruous scaling		
	Critical γ^X_0	V^X_+	V^X_-	Critical γ^X_0	V^X_+	V^X_-
$\tilde{Q}_1 \wedge \tilde{Q}_2 > 1$	q_1/q_2	$B_{1,\tilde{H}_2}$	$B_{\tilde{H}_1,1}$	1	$B_{1,\tilde{H}_1\wedge\tilde{H}_2}$	$B_{\tilde{H}_1\wedge\tilde{H}_2,1}$
$\tilde{Q}_1 < 1 < \tilde{Q}_2$	q_1/q_2	$B_{H_1,1/2}$	$B_{\tilde{H}_1,1}$	1	$B_{H_1\wedge H_2,1/2}$	$B_{1/2,H_1\wedge H_2}$
$\tilde{Q}_2 < 1 < \tilde{Q}_1$	q_1/q_2	$B_{1,\tilde{H}_2}$	$B_{1/2,H_2}$	1	$B_{H_1\wedge H_2,1/2}$	$B_{1/2,H_1\wedge H_2}$
$\tilde{Q}_1 \vee \tilde{Q}_2 < 1$	q_1/q_2	$B_{H_1,1/2}$	$B_{1/2,H_2}$	1	$B_{H_1\wedge H_2,1/2}$	$B_{1/2,H_1\wedge H_2}$

any $\gamma > 0$, see Theorem 5, in other words, X does not exhibit a scaling transition. On the other hand, we expect that the results in Table 1 can be extended to *negatively dependent* linear RFs with coefficients as in (8) satisfying $Q < 1$ (which guarantees their summability) and the zero-sum condition $\sum_{\mathbf{t}\in\mathbb{Z}^2} b(\mathbf{t}) = 0$. The existence of the scaling transition for negatively dependent RFs with coefficients as in (5) (i.e., under congruous scaling) was established in [21]. Let us note that the case of negative dependence is more delicate, due to the possible occurrence of edge effects, see [10, 21]. Further interesting open problems concern incongruous scaling of *nonlinear* or *subordinated* RFs on $\mathbb{Z}^2$ (see [14]) and possible extensions to $\mathbb{Z}^3$ and higher dimensions. We mention that the scaling diagram of linear LRD RF on $\mathbb{Z}^3$ under congruous scaling is quite complicated, see [19]; the incongruous scaling may lead to a much more simple result akin to Table 1.

Section 2 contains the main results (Theorems 1–4), together with rigorous assumptions and the definitions of the limit RFs. The proofs of these facts are given in Sect. 3. Section 4 (Appendix) contains the definition and the existence of the dependence axis for moving-average coefficients b as in (8) (Proposition 4). We also prove in Sect. 4 that the dependence axis is preserved under convolution, implying that the covariance function of the linear RF X also decays along this axis at the smallest rate.

Notation In what follows, C denote generic positive constants which may be different at different locations. We write $\xrightarrow{\text{fdd}}$, $\overset{\text{fdd}}{=}$, and $\overset{\text{fdd}}{\neq}$ for the weak convergence, equality, and inequality of finite-dimensional distributions, respectively. $\mathbf{1} := (1,1)^\top$, $\mathbf{0} := (0,0)^\top$, $\mathbb{R}^2_0 := \mathbb{R}^2 \setminus \{\mathbf{0}\}$, $\mathbb{R}^2_+ := \{\mathbf{x} = (x_1,x_2)^\top \in \mathbb{R}^2 : x_i > 0, i = 1,2\}$, $\mathbb{R}_+ := (0,\infty)$ and $(\mathbf{0},\mathbf{x}] := (0,x_1] \times (0,x_2]$, $\mathbf{x} = (x_1,x_2)^\top \in \mathbb{R}^2_+$. Also, $\lfloor x \rfloor := \max\{k \in \mathbb{Z} : k \le x\}$, $\lceil x \rceil := \min\{k \in \mathbb{Z} : k \ge x\}$, $x \in \mathbb{R}$, and $\lfloor \mathbf{x} \rfloor := (\lfloor x_1 \rfloor, \lfloor x_2 \rfloor)^\top$, $\lceil \mathbf{x} \rceil := (\lceil x_1 \rceil, \lceil x_2 \rceil)^\top$, $|\mathbf{x}| := |x_1| + |x_2|$, $\mathbf{x} = (x_1,x_2)^\top \in \mathbb{R}^2$. We also write $f(\mathbf{x}) = f(x_1,x_2)$, $\mathbf{x} = (x_1,x_2)^\top \in \mathbb{R}^2$.

2 Main Results

For $\gamma > 0$, we study the limit distribution in (1) of partial sums

$$S_{\lambda,\gamma}(\mathbf{x}) = \sum_{\mathbf{t}\in(0,\lambda x_1]\times(0,\lambda^\gamma x_2]\cap\mathbb{Z}^2} X(\mathbf{t}), \quad \mathbf{x} \in \mathbb{R}^2_+, \tag{12}$$

over rectangles of a linear RF X in (7) satisfying the following assumptions.

Assumption A Innovations $\varepsilon(\mathbf{t})$, $\mathbf{t} \in \mathbb{Z}^2$, in (7) are i.i.d. r.v.s with $\mathrm{E}\varepsilon(\mathbf{0}) = 0$, $\mathrm{E}|\varepsilon(\mathbf{0})|^2 = 1$.

Assumption B Coefficients $b(\mathbf{t})$, $\mathbf{t} \in \mathbb{Z}^2$, in (7) satisfy

$$b(\mathbf{t}) = \rho(B\mathbf{t})^{-1}(L(B\mathbf{t}) + o(1)), \quad |\mathbf{t}| \to \infty, \tag{13}$$

where $B = (b_{ij})_{i,j=1,2}$ is a real nondegenerate matrix, and

$$\rho(\mathbf{u}) := |u_1|^{q_1} + |u_2|^{q_2}, \quad \mathbf{u} \in \mathbb{R}^2, \tag{14}$$

with $q_i > 0$, $i = 1, 2$, satisfying (6), and

$$L(\mathbf{u}) := L_+(u_1/\rho(\mathbf{u})^{1/q_1})\mathbf{1}(u_2 \ge 0) + L_-(u_1/\rho(\mathbf{u})^{1/q_1})\mathbf{1}(u_2 < 0), \quad \mathbf{u} \in \mathbb{R}^2_0, \tag{15}$$

where $L_\pm(x)$, $x \in [-1, 1]$, are continuous functions such that $L_+(1) = L_-(1)$, $L_+(-1) = L_-(-1)$.

We note that the boundedness and continuity assumptions of the 'angular functions' $L_\pm$ in (15) do not seem necessary for our results and possibly can be relaxed. Note $q_1 < q_2$ for $1 < Q < 2$ implies $H_1 \wedge H_2 = H_1$, $\tilde{H}_1 \wedge \tilde{H}_2 = \tilde{H}_2$. Then from Proposition 4 we see that $\{\mathbf{t} \in \mathbb{R}^2 : \mathbf{b}_2 \cdot \mathbf{t} = 0\}$ with $\mathbf{b}_2 = (b_{21}, b_{22})^\top$ is the dependence axis of X, which agrees with the coordinate axes if and only if $b_{21} = 0$ or $b_{22} = 0$ leading to the two cases, namely $b_{21} = 0$ (congruous scaling) and $b_{21}b_{22} \neq 0$ (incongruous scaling).

The limit Gaussian RFs in our theorems are defined as stochastic integrals w.r.t. (real-valued) Gaussian white noise $W = \{W(\mathrm{d}\mathbf{u}), \mathbf{u} \in \mathbb{R}^2\}$ with zero mean and variance $\mathrm{E}W(\mathrm{d}\mathbf{u})^2 = \mathrm{d}\mathbf{u}$ (= the Lebesgue measure on $\mathbb{R}^2$). Let

$$a_\infty(\mathbf{u}) := \rho(\mathbf{u})^{-1}L(\mathbf{u}), \quad \mathbf{u} \in \mathbb{R}^2_0, \tag{16}$$

and

$$\tilde{V}_{\tilde{D}}(\mathbf{x}) := |\det(B)|^{-\frac{1}{2}} \int_{\mathbb{R}^2} \left\{ \int_{(\mathbf{0},\mathbf{x}]} a_\infty(\tilde{D}\mathbf{t} - \mathbf{u})\mathrm{d}\mathbf{t} \right\} W(\mathrm{d}\mathbf{u}), \quad \mathbf{x} \in \mathbb{R}^2_+, \tag{17}$$

where $\tilde{D}$ is any 2×2 matrix in (18) below:

$$\tilde{B}_{00} := \begin{bmatrix} b_{11} & 0 \\ 0 & b_{22} \end{bmatrix}, \quad \tilde{B}_{01} := \begin{bmatrix} b_{11} & 0 \\ b_{21} & 0 \end{bmatrix}, \quad \tilde{B}_{02} := \begin{bmatrix} 0 & b_{12} \\ 0 & b_{22} \end{bmatrix}, \quad \tilde{B}_{20} := \begin{bmatrix} 0 & 0 \\ b_{21} & b_{22} \end{bmatrix},$$
$$\tilde{B}_{11} := \begin{bmatrix} b_{11} & 0 \\ 0 & 0 \end{bmatrix}, \quad \tilde{B}_{21} := \begin{bmatrix} 0 & 0 \\ b_{21} & 0 \end{bmatrix}, \quad \tilde{B}_{22} := \begin{bmatrix} 0 & 0 \\ 0 & b_{22} \end{bmatrix}. \tag{18}$$

Also let

$$V_D(\mathbf{x}) := |\det(B)|^{-1} \int_{\mathbb{R}^2} \left\{ \int_{\mathbb{R}^2} a_\infty(\mathbf{t}) \mathbf{1}(D\mathbf{t} + \mathbf{u} \in (\mathbf{0}, \mathbf{x}]) \mathrm{d}\mathbf{t} \right\} W(\mathrm{d}\mathbf{u}), \quad \mathbf{x} \in \mathbb{R}^2_+, \tag{19}$$

where D is any 2×2 matrix in (20) below:

$$B_{01} := \frac{1}{\det(B)} \begin{bmatrix} b_{22} & 0 \\ -b_{21} & 0 \end{bmatrix}, \quad B_{11} := \frac{1}{\det(B)} \begin{bmatrix} b_{22} & 0 \\ 0 & 0 \end{bmatrix},$$
$$B_{10} := \frac{1}{\det(B)} \begin{bmatrix} b_{22} & -b_{12} \\ 0 & 0 \end{bmatrix}, \quad B_{20} := \frac{1}{\det(B)} \begin{bmatrix} 0 & 0 \\ -b_{21} & b_{11} \end{bmatrix},$$
$$B_{21} := \frac{1}{\det(B)} \begin{bmatrix} 0 & 0 \\ -b_{21} & 0 \end{bmatrix}, \quad B_{22} := \frac{1}{\det(B)} \begin{bmatrix} 0 & 0 \\ 0 & b_{11} \end{bmatrix}. \tag{20}$$

Recall that in (18), (20) b_{ij} are entries of the matrix B in (13). To shorten notation, write $\tilde{V}_{ij} := \tilde{V}_{\tilde{B}_{ij}}$, $V_{ij} := V_{B_{ij}}$ and also define $\tilde{V}_0 := \tilde{V}_B$, $V_0 := V_{B^{-1}}$ satisfying $\tilde{V}_0 \overset{\mathrm{fdd}}{=} V_0$. The existence of all these RFs in the corresponding regions of parameters q_1, q_2 is established in Proposition 1, which also identifies some of these RFs with FBS having one of its parameters equal to 1 or $\frac{1}{2}$. Recall that stochastic integral $\int_{\mathbb{R}^2} h(\mathbf{u}) W(\mathrm{d}\mathbf{u}) =: I(h)$ w.r.t. Gaussian white noise W is well defined for any $h \in L^2(\mathbb{R}^2)$ and has a Gaussian distribution with zero mean and variance $\mathrm{E}|I(h)|^2 = \|h\|^2 = \int_{\mathbb{R}^2} |h(\mathbf{u})|^2 \mathrm{d}\mathbf{u}$, c.f. [18]. Recall the definitions of $\tilde{Q}_i$, $\tilde{H}_i$, H_i, $i = 1, 2$, in (11). In Proposition 1 and Theorems 1–4 below, Assumptions A and B hold without further notice. Let $\tilde{\sigma}^2_{ij} := \mathrm{E}|\tilde{V}_{ij}(\mathbf{1})|^2$, $\sigma^2_{ij} := \mathrm{E}|V_{ij}(\mathbf{1})|^2$.

Proposition 1 *The following RFs in (17)–(20) are well defined:*

(i) $\tilde{V}_{00}$, $\tilde{V}_{11}$, $\tilde{V}_{20}$, $\tilde{V}_{21}$, *and* $\tilde{V}_{22}$ *provided* $q_1 < q_2$ *and* $\tilde{Q}_1 > 1$ *hold; moreover,*

$$\tilde{V}_{22} \overset{\mathrm{fdd}}{=} \tilde{\sigma}_{22} B_{1,\tilde{H}_2}, \quad \tilde{V}_{21} \overset{\mathrm{fdd}}{=} \tilde{\sigma}_{21} B_{\tilde{H}_2,1}, \quad \tilde{V}_{11} \overset{\mathrm{fdd}}{=} \tilde{\sigma}_{11} B_{\tilde{H}_1,1}. \tag{21}$$

(ii) $\tilde{V}_{00}$, $\tilde{V}_{11}$, V_{01}, V_{11}, *and* V_{21} *provided* $q_1 < q_2$ *and* $\tilde{Q}_1 < 1 < \tilde{Q}_2$ *hold; moreover,*

$$V_{11} \stackrel{\mathrm{fdd}}{=} \sigma_{11} B_{H_1, \frac{1}{2}}, \quad V_{21} \stackrel{\mathrm{fdd}}{=} \sigma_{21} B_{\frac{1}{2}, H_1}. \tag{22}$$

(iii) $\tilde{V}_{00}$, V_{01}, V_{11}, V_{21}, *and* V_{22} *provided* $q_1 < q_2$ *and* $\tilde{Q}_2 < 1$ *hold; moreover,*

$$V_{22} \stackrel{\mathrm{fdd}}{=} \sigma_{22} B_{\frac{1}{2}, H_2}. \tag{23}$$

(iv) $\tilde{V}_{02}$, $\tilde{V}_0$, *and* $\tilde{V}_{01}$ *provided* $q_1 = q_2 =: q \in (1, \frac{3}{2})$*; moreover,*

$$\tilde{V}_{01} \stackrel{\mathrm{fdd}}{=} \tilde{\sigma}_{01} B_{\tilde{H}, 1}, \quad \tilde{V}_{02} \stackrel{\mathrm{fdd}}{=} \tilde{\sigma}_{02} B_{1, \tilde{H}}, \quad \tilde{H} := 2 - q \in \left(\frac{1}{2}, 1\right). \tag{24}$$

(v) V_{10}, V_0, *and* V_{20} *provided* $q_1 = q_2 =: q \in (\frac{3}{2}, 2)$*; moreover,*

$$V_{10} \stackrel{\mathrm{fdd}}{=} \sigma_{10} B_{H, \frac{1}{2}}, \quad V_{20} \stackrel{\mathrm{fdd}}{=} \sigma_{20} B_{\frac{1}{2}, H}, \quad H := \frac{5}{2} - q \in \left(\frac{1}{2}, 1\right). \tag{25}$$

Remark 1 While some of the RFs in (17)–(20) are well identified by (21)–(25) as FBS with special values of the Hurst parameters, the remaining ones, namely $\tilde{V}_{00}$, $\tilde{V}_{20}$, $\tilde{V}_0$, V_{01}, and V_0 which arise as well-balanced limits in Theorems 1–4, are less explicit and depend on a_∞ in (16) and the matrix B in a more complicated way. Using the isometry $\mathrm{E} I(h_1) I(h_2) = \int_{\mathbb{R}^2} h_1(\mathbf{u}) h_2(\mathbf{u}) \mathrm{d}\mathbf{u}$ for any $h_1, h_2 \in L^2(\mathbb{R})$ the covariance function of these RFs write as integral of the product of the corresponding kernels; for instance, $\mathrm{E}[\tilde{V}_{00}(\mathbf{x}) \tilde{V}_{00}(\mathbf{y})] = |\det(B)|^{-1} \int_{\mathbb{R}^2} \{\int_{(\mathbf{0}, \mathbf{x}]} a_\infty(\tilde{B}_{00}\mathbf{t} - \mathbf{u}) \mathrm{d}\mathbf{t}\} \times \{\int_{(\mathbf{0}, \mathbf{y}]} a_\infty(\tilde{B}_{00}\mathbf{s} - \mathbf{u}) \mathrm{d}\mathbf{s}\} \mathrm{d}\mathbf{u}$ which is a complex function of $\mathbf{x}, \mathbf{y}$ even for $q_1 = q_2$ and $L_\pm = 1$. We note that all RFs in (17)–(20) have rectangular stationary increments (see [14] for definition) and satisfy the self-similarity property in (29) below; however, these properties are very general and are satisfied by many Gaussian RFs. See also Dobrushin [5] for (spectral) characterization of stationary self-similar Gaussian RFs.

As noted above, our main results (Theorems 1–4) describe the anisotropic scaling limits and the scaling transition of the linear RF X in (7), viz.,

$$\left\{ A_{\lambda,\gamma}^{-1} S_{\lambda,\gamma}^X(\mathbf{x}), \mathbf{x} \in \mathbb{R}_+^2 \right\} \stackrel{\mathrm{fdd}}{\to} V_\gamma^X = \begin{cases} V_+^X, & \gamma > \gamma_0^X, \\ V_-^X, & \gamma < \gamma_0^X, \\ V_0^X, & \gamma = \gamma_0^X, \end{cases} \quad \lambda \to \infty, \tag{26}$$

where $S^X_{\lambda,\gamma}$ is the partial-sum RF in (12). As explained above, in Theorems 1–3 we distinguish between two cases:

$$\begin{aligned} &\text{Case 1:} \quad b_{21}b_{22} \neq 0 \ \ \text{(incongruous scaling)}, \\ &\text{Case 2:} \quad b_{21} = 0 \ \ \text{(congruous scaling)}. \end{aligned} \tag{27}$$

Theorem 1 *Let $q_1 < q_2$ and $\tilde{Q}_1 > 1$. Then the convergence in (26) holds for all $\gamma > 0$ with*

Case 1: $\gamma^X_0 = 1$ and $V^X_+ = \tilde{V}_{22}$, $V^X_- = \tilde{V}_{21}$, $V^X_0 = \tilde{V}_{20}$,
Case 2: $\gamma^X_0 = \frac{q_1}{q_2}$ and $V^X_+ = \tilde{V}_{22}$, $V^X_- = \tilde{V}_{11}$, $V^X_0 = \tilde{V}_{00}$.

Theorem 2 *Let $q_1 < q_2$ and $\tilde{Q}_1 < 1 < \tilde{Q}_2$. Then the convergence in (26) holds for all $\gamma > 0$ with*

Case 1: $\gamma^X_0 = 1$ and $V^X_+ = V_{11}$, $V^X_- = V_{21}$, $V^X_0 = V_{01}$,
Case 2: $\gamma^X_0 = \frac{q_1}{q_2}$ and $V^X_+ = V_{11}$, $V^X_- = \tilde{V}_{11}$, $V^X_0 = \tilde{V}_{00}$.

Theorem 3 *Let $q_1 < q_2$ and $\tilde{Q}_2 < 1$. Then the convergence in (26) holds for all $\gamma > 0$ with*

Case 1: γ^X_0 and V^X_+, V^X_-, V^X_0 the same as in Case 1 of Theorem 2,
Case 2: $\gamma^X_0 = \frac{q_1}{q_2}$ and $V^X_+ = V_{11}$, $V^X_- = V_{22}$, $V^X_0 = \tilde{V}_{00}$.

Theorem 4 discusses the case $q_1 = q_2$ when the dependence axis is undefined.

Theorem 4 *Let $q_1 = q_2 =: q$ and $\tilde{Q}_1 = \tilde{Q}_2 =: \tilde{Q}$. Then the convergence in (26) holds for all $\gamma > 0$ with $\gamma^X_0 = 1$ and*

(i) $V^X_+ = \tilde{V}_{02}$, $V^X_- = \tilde{V}_{01}$, $V^X_0 = \tilde{V}_0$ if $q \in (1, \frac{3}{2})$ or $\tilde{Q} > 1$.
(ii) $V^X_+ = V_{10}$, $V^X_- = V_{20}$, $V^X_0 = V_0$ if $q \in (\frac{3}{2}, 2)$ or $\tilde{Q} < 1$.

Remark 2 In the above theorems the convergence in (26) holds under normalization

$$A_{\lambda,\gamma} = \lambda^{H(\gamma)}, \tag{28}$$

where $H(\gamma) > 0$ is defined in the proof of these theorems below. Under congruous scaling $b_{21} = 0$ the exponent $H(\gamma)$ in (28) is the same as in the case $B = I$ (= the identity matrix) studied in [14]. As shown in [17], V^X_γ in (26) satisfies the following self-similarity property:

$$\left\{V^X_\gamma(\lambda^\Gamma \mathbf{x}), \mathbf{x} \in \mathbb{R}^2_+\right\} \overset{\text{fdd}}{=} \left\{\lambda^{H(\gamma)} V^X_\gamma(\mathbf{x}), \mathbf{x} \in \mathbb{R}^2_+\right\} \quad \forall \lambda > 0, \tag{29}$$

where $\lambda^\Gamma = \mathrm{diag}(\lambda, \lambda^\gamma)$ and $H(\gamma)$ is the same as in (28). Note that an FBS $B_{\mathscr{H}_1,\mathscr{H}_2}$ with $\mathscr{H}_i \in (0, 1]$, $i = 1, 2$, satisfies (29) with

$$H(\gamma) = \mathscr{H}_1 + \gamma \mathscr{H}_2. \tag{30}$$

Thus, in Theorems 1–4 in the case of unbalanced (FBS) limit $A_{\lambda,\gamma}$ in (28) can also be identified from (30) and the expressions for H_i, $\tilde{H}_i$, $i = 1, 2$, in (11).

In the above theorems, a crucial LRD condition (guaranteeing (9)) is $Q \in (1, 2)$. For $Q < 1$ the linear RF X in (4) is short-range dependent (SRD), viz., $\sum_{\mathbf{t}\in\mathbb{Z}^2} |\operatorname{Cov}(X(\mathbf{0}), X(\mathbf{t}))| < \infty$, and the scaling transition does not occur, see [14, Thm. 3.4]. These facts extend also to RF X in (7) with 'rotated' coefficients b of (8), see Theorem 5 below. Its proof is omitted because it mimics that of [14, Thm. 3.4].

Theorem 5

(i) Let X be a linear RF in (7) with coefficients b such that

$$\sum_{\mathbf{t}\in\mathbb{Z}^2} |b(\mathbf{t})| < \infty \quad \textit{and} \quad \sum_{\mathbf{t}\in\mathbb{Z}^2} b(\mathbf{t}) \neq 0. \tag{31}$$

Then for any $\gamma > 0$

$$\left\{\lambda^{-(1+\gamma)/2} S^X_{\lambda,\gamma}(\mathbf{x}), \mathbf{x} \in \mathbb{R}^2_+\right\} \overset{\text{fdd}}{\to} \sigma B_{1/2,1/2}, \tag{32}$$

where $\sigma^2 = (\sum_{\mathbf{t}\in\mathbb{Z}^2} b(\mathbf{t}))^2 > 0$.

(ii) If b satisfy Assumption B with $Q < 1$ instead of $Q \in (1, 2)$, then $\sum_{\mathbf{t}\in\mathbb{Z}^2} |b(\mathbf{t})| < \infty$. If, in addition, $\sum_{\mathbf{t}\in\mathbb{Z}^2} b(\mathbf{t}) \neq 0$, then the corresponding linear RF X in (7) satisfies the CLT in (32).

3 Proofs of Proposition 1 and Theorems 1–4

For $\gamma > 0$, the limit distribution of $S^X_{\lambda,\gamma}$ is obtained using a general criterion for the weak convergence of linear forms in i.i.d. r.v.s towards a stochastic integral w.r.t. the white noise. Consider a linear form

$$S(g) := \sum_{\mathbf{s}\in\mathbb{Z}^2} g(\mathbf{s})\varepsilon(\mathbf{s}) \tag{33}$$

with real coefficients $\sum_{\mathbf{s}\in\mathbb{Z}^2} g(\mathbf{s})^2 < \infty$ and innovations satisfying Assumption A. The following proposition extends [8, Prop. 14.3.2], [21, Prop. 5.1], [19, Prop. 3.1].

Proposition 2 *For $\lambda > 0$, let $S(g_\lambda)$ be as in (33) with $g_\lambda : \mathbb{Z}^2 \to \mathbb{R}$ satisfying $\sum_{\mathbf{s}\in\mathbb{Z}^2} g_\lambda(\mathbf{s})^2 < \infty$. Assume that for some 2×2 non-degenerate matrix A and $\Lambda = \operatorname{diag}(l_1, l_2)$ with $l_i = l_i(\lambda)$, $i = 1, 2$, such that $l_1 \wedge l_2 \to \infty$, $\lambda \to \infty$, the functions*

$$\tilde{g}_\lambda(\mathbf{u}) := |\det(A\Lambda)|^{1/2} g_\lambda(\lceil A\Lambda\mathbf{u}\rceil), \quad \mathbf{u} \in \mathbb{R}^2, \quad \lambda > 0, \tag{34}$$

tend to a limit h in $L^2(\mathbb{R}^2)$, i.e.

$$\|\tilde{g}_\lambda - h\|^2 = \int_{\mathbb{R}^2} |\tilde{g}_\lambda(\mathbf{u}) - h(\mathbf{u})|^2 \mathrm{d}\mathbf{u} \to 0, \quad \lambda \to \infty. \tag{35}$$

Then

$$S(g_\lambda) \xrightarrow{\mathrm{d}} I(h) = \int_{\mathbb{R}^2} h(\mathbf{u}) W(\mathrm{d}\mathbf{u}), \quad \lambda \to \infty. \tag{36}$$

Proof Denote by $S(\mathbb{R}^2)$ the set of simple functions $f : \mathbb{R}^2 \to \mathbb{R}$, which are finite linear combinations of indicator functions of disjoint squares $\Box^K_{\mathbf{k}} := \prod_{i=1}^2 (k_i/K, (k_i+1)/K]$, $\mathbf{k} \in \mathbb{Z}^2$, $K \in \mathbb{N}$. The set $S(\mathbb{R}^2)$ is dense in $L^2(\mathbb{R}^2)$: given $f \in L^2(\mathbb{R}^2)$, for every $\epsilon > 0$ there exists $f_\epsilon \in S(\mathbb{R}^2)$ such that $\|f - f_\epsilon\| < \epsilon$. (36) follows once we show that for every $\epsilon > 0$ there exists $h_\epsilon \in S(\mathbb{R}^2)$ such that as $\lambda \to \infty$, the following relations (i)–(iii) hold: (i) $\mathrm{E}|S(g_\lambda) - S(h_{\epsilon,\lambda})|^2 < \epsilon$, (ii) $S(h_{\epsilon,\lambda}) \xrightarrow{\mathrm{d}} I(h_\epsilon)$, (iii) $\mathrm{E}|I(h_\epsilon) - I(h)|^2 < \epsilon$, where

$$h_{\epsilon,\lambda}(\mathbf{s}) := |\det(A\Lambda)|^{-1/2} h_\epsilon((A\Lambda)^{-1}\mathbf{s}), \quad \mathbf{s} \in \mathbb{Z}^2, \quad \lambda > 0. \tag{37}$$

As for (i), note that

$$\begin{aligned}
\mathrm{E}|S(g_\lambda) - S(h_{\epsilon,\lambda})|^2 &= \int_{\mathbb{R}^2} |g_\lambda(\lceil \mathbf{s} \rceil) - h_{\epsilon,\lambda}(\lceil \mathbf{s} \rceil)|^2 \mathrm{d}\mathbf{s} \\
&= |\det(A\Lambda)| \int_{\mathbb{R}^2} |g_\lambda(\lceil A\Lambda\mathbf{u} \rceil) - h_{\epsilon,\lambda}(\lceil A\Lambda\mathbf{u} \rceil)|^2 \mathrm{d}\mathbf{u} \\
&= \|\tilde{g}_\lambda - \tilde{h}_{\epsilon,\lambda}\|^2,
\end{aligned}$$

where $\tilde{h}_{\epsilon,\lambda}$ is derived from $h_{\epsilon,\lambda}$ in the same way as $\tilde{g}_\lambda$ is derived from g_λ in (34). To prove (i) we need to find suitable $h_\epsilon \in S(\mathbb{R}^2)$ and thus $h_{\epsilon,\lambda}$ in (37). By (35), there exists $\lambda_0 > 0$ such that $\|\tilde{g}_\lambda - h\| < \epsilon/4$, $\forall \lambda \geq \lambda_0$. Given $\tilde{g}_{\lambda_0} \in L^2(\mathbb{R}^2)$, there exists $h_\epsilon \in S(\mathbb{R}^2)$ such that $\|\tilde{g}_{\lambda_0} - h_\epsilon\| < \epsilon/4$. Note that

$$\|h_\epsilon - \tilde{h}_{\epsilon,\lambda}\|^2 = \int_{\mathbb{R}^2} |h_\epsilon(\mathbf{u}) - h_\epsilon((A\Lambda)^{-1}\lceil A\Lambda\mathbf{u} \rceil)|^2 \mathrm{d}\mathbf{u} \to 0, \quad \lambda \to \infty,$$

follows from $|(A\Lambda)^{-1}\lceil A\Lambda\mathbf{u} \rceil - \mathbf{u}| = |(A\Lambda)^{-1}(\lceil A\Lambda\mathbf{u} \rceil - A\Lambda\mathbf{u})| \leq C \min(l_1, l_2)^{-1} = o(1)$ uniformly in $\mathbf{u} \in \mathbb{R}^2$ and the fact that h_ϵ is bounded and has a compact support. Thus, there exists $\lambda_1 > 0$ such that $\|h_\epsilon - \tilde{h}_{\epsilon,\lambda}\| < \epsilon/4$, $\forall \lambda \geq \lambda_1$. Hence,

$$\|\tilde{g}_\lambda - \tilde{h}_{\epsilon,\lambda}\| \leq \|\tilde{g}_\lambda - h\| + \|h - \tilde{g}_{\lambda_0}\| + \|\tilde{g}_{\lambda_0} - h_\epsilon\| + \|h_\epsilon - \tilde{h}_{\epsilon,\lambda}\| < \epsilon, \quad \forall \lambda \geq \lambda_0 \vee \lambda_1,$$

completing the proof of (i). The above reasoning implies also (iii) since $(\mathrm{E}|I(h_\epsilon) - I(h)|^2)^{1/2} = \|h_\epsilon - h\| \le \|h_\epsilon - \tilde g_{\lambda_0}\| + \|\tilde g_{\lambda_0} - h\| < \epsilon/2$.

It remains to prove (ii). The step function h_ϵ in the above proof of (i) can be written as $h_\epsilon(\mathbf{u}) = \sum_{\mathbf{k}\in\mathbb{Z}^2} h_\epsilon^{\square^K_{\mathbf{k}}} \mathbf{1}(\mathbf{u} \in \square^K_{\mathbf{k}})$, $\mathbf{u} \in \mathbb{R}^2$, $(\exists K \in \mathbb{N})$, where $h_\epsilon^{\square^K_{\mathbf{k}}} = 0$ except for a finite number of $\mathbf{k} \in \mathbb{Z}^2$. Then, by (37),

$$h_{\epsilon,\lambda}(\mathbf{s}) = |\det(A\Lambda)|^{-1/2} \sum_{\mathbf{k}\in\mathbb{Z}^2} h_\epsilon^{\square^K_{\mathbf{k}}} \mathbf{1}(\mathbf{s} \in A\Lambda\square^K_{\mathbf{k}}), \quad \mathbf{s} \in \mathbb{Z}^2,$$

and $S(h_{\epsilon,\lambda}) = \sum_{\mathbf{k}\in\mathbb{Z}^2} h_\epsilon^{\square^K_{\mathbf{k}}} W_\lambda(\square^K_{\mathbf{k}})$, where

$$W_\lambda(\square^K_{\mathbf{k}}) := |\det(A\Lambda)|^{-1/2} \sum_{\mathbf{s}\in A\Lambda\square^K_{\mathbf{k}}} \varepsilon(\mathbf{s}), \quad \mathbf{k} \in \mathbb{Z}^2.$$

Since the r.v.s $\varepsilon(\mathbf{s})$, $\mathbf{s} \in \mathbb{Z}^2$, are i.i.d. with $\mathrm{E}\varepsilon(\mathbf{0}) = 0$, $\mathrm{E}|\varepsilon(\mathbf{0})|^2 = 1$ and the parallelograms $A\Lambda\square^K_{\mathbf{k}}$, $\mathbf{k} \in \mathbb{Z}^2$, are disjoint, the r.v.s $W_\lambda(\square^K_{\mathbf{k}})$, $\mathbf{k} \in \mathbb{Z}^2$, are independent and satisfy $\mathrm{E}W_\lambda(\square^K_{\mathbf{k}}) = 0$, $\mathrm{E}|W_\lambda(\square^K_{\mathbf{k}})|^2 = |\det(A\Lambda)|^{-1} \sum_{\mathbf{s}\in\mathbb{Z}^2} \mathbf{1}(\mathbf{s} \in A\Lambda\square^K_{\mathbf{k}}) \to \int_{\square^K_{\mathbf{k}}} \mathrm{d}\mathbf{u}$, $\lambda \to \infty$. Hence, by the classical CLT, for every $J \in \mathbb{N}$,

$$\Big\{W_\lambda(\square^K_{\mathbf{k}}), \mathbf{k} \in \{-J,\dots,J\}^2\Big\} \xrightarrow{\mathrm{d}} \Big\{W(\square^K_{\mathbf{k}}), \mathbf{k} \in \{-J,\dots,J\}^2\Big\}, \quad \lambda \to \infty,$$

implying the convergence $S(h_{\epsilon,\lambda}) \xrightarrow{\mathrm{d}} \sum_{\mathbf{k}\in\mathbb{Z}^2} h_\epsilon^{\square^K_{\mathbf{k}}} W(\square^K_{\mathbf{k}}) = I(h_\epsilon)$, $\lambda \to \infty$, or part (ii), and completing the proof of the proposition. □

We shall also need some properties of the generalized homogeneous function ρ in (14) for $q_i > 0$, $i = 1, 2$, with $Q := \frac{1}{q_1} + \frac{1}{q_2}$. Note the elementary inequality

$$C_1\rho(\mathbf{u})^{1/q_1} \le (|u_1|^2 + |u_2|^{2q_2/q_1})^{1/2} \le C_2\rho(\mathbf{u})^{1/q_1}, \quad \mathbf{u} \in \mathbb{R}^2, \tag{38}$$

with $C_i > 0$, $i = 1, 2$, independent of $\mathbf{u}$, see [19, (2.16)]. From (38) and [14, Prop. 5.1] we obtain for any $\delta > 0$,

$$\int_{\mathbb{R}^2} \rho(\mathbf{u})^{-1}\mathbf{1}(\rho(\mathbf{u}) < \delta)\mathrm{d}\mathbf{u} < \infty \iff Q > 1, \tag{39}$$
$$\int_{\mathbb{R}^2} \rho(\mathbf{u})^{-2}\mathbf{1}(\rho(\mathbf{u}) \ge \delta)\mathrm{d}\mathbf{u} < \infty \iff Q < 2.$$

Moreover, with $q = \max\{q_1, q_2, 1\}$,

$$\rho(\mathbf{u} + \mathbf{v})^{1/q} \le \rho(\mathbf{u})^{1/q} + \rho(\mathbf{v})^{1/q}, \quad \mathbf{u}, \mathbf{v} \in \mathbb{R}^2, \tag{40}$$

see [14, (7.1)], and, for $1 < Q < 2$,

$$(\rho^{-1} \star \rho^{-1})(\mathbf{u}) := \int_{\mathbb{R}^2} \rho(\mathbf{v})^{-1} \rho(\mathbf{v}+\mathbf{u})^{-1} \mathrm{d}\mathbf{v} = \tilde{\rho}(\mathbf{u})^{-1} \tilde{L}(|u_1|/\tilde{\rho}(\mathbf{u})^{1/\tilde{q}_1}), \quad \mathbf{u} \in \mathbb{R}^2_0, \tag{41}$$

where with $\tilde{q}_i := q_i(2-Q)$, $i = 1, 2$,

$$\tilde{\rho}(\mathbf{u}) := |u_1|^{\tilde{q}_1} + |u_2|^{\tilde{q}_2}, \quad \mathbf{u} \in \mathbb{R}^2, \tag{42}$$

and $\tilde{L}(z) := (\rho^{-1} \star \rho^{-1})(z, (1-z^{\tilde{q}_1})^{1/\tilde{q}_2})$, $z \in [0, 1]$, is a continuous function. Note $\tilde{q}_2 < 1$ (respectively, $\tilde{q}_1 < 1$) is equivalent to $\tilde{Q}_1 > 1$ (respectively, $\tilde{Q}_2 > 1$). The proof of (41) is similar to that of [14, (5.6)] (see also Proposition 5 below).

Proof of Proposition 1 Since $\tilde{V}_{ij}(\mathbf{1}) = I(\tilde{h}_{ij})$, $V_{ij}(\mathbf{1}) = I(h_{ij})$, it suffices to prove

$$\|\tilde{h}_{ij}\| < \infty, \quad \|h_{ij}\| < \infty \tag{43}$$

for suitable i, j in the corresponding regions of parameters q_1, q_2. Using the boundedness of $L_\pm$ in (16) and (38) we can replace $|a_\infty|$ by ρ^{-1} in the subsequent proofs of (43). Hence and from (41) it follows that

$$\begin{aligned}\|\tilde{h}_{ij}\|^2 &\le C \int_{(0,1]^2 \times (0,1]^2} (\rho^{-1} \star \rho^{-1})(\tilde{B}_{ij}(\mathbf{t}-\mathbf{s}))\mathrm{d}\mathbf{t}\mathrm{d}\mathbf{s} \\ &\le C \int_{(0,1]^2 \times (0,1]^2} \tilde{\rho}(\tilde{B}_{ij}(\mathbf{t}-\mathbf{s}))^{-1}\mathrm{d}\mathbf{t}\mathrm{d}\mathbf{s}.\end{aligned} \tag{44}$$

Existence of $\tilde{V}_{00}$ Relation $\|\tilde{h}_{00}\|^2 \le C \int_{[-1,1]^2} \tilde{\rho}(b_{11}t_1, b_{22}t_2)^{-1}\mathrm{d}\mathbf{t} < \infty$ follows from (39) since $\frac{1}{\tilde{q}_1} + \frac{1}{\tilde{q}_2} = \frac{Q}{2-Q} > 1$ for $1 < Q < 2$. This proves the existence of $\tilde{V}_{00}$ in all cases (i)–(iii) of Proposition 1.

Existence of $\tilde{V}_{20}$, $\tilde{V}_{21}$, $\tilde{V}_{22}$ From (44) we get $\|\tilde{h}_{20}\|^2 \le C \int_{[-1,1]^2} |b_{21}t_1 + b_{22}t_2|^{-\tilde{q}_2}\mathrm{d}\mathbf{t} < \infty$ since $\tilde{q}_2 < 1$. The proof of $\|\tilde{h}_{21}\| < \infty$, $\|\tilde{h}_{22}\| < \infty$ is completely analogous. This proves the existence of $\tilde{V}_{20}$, $\tilde{V}_{21}$ and $\tilde{V}_{22}$ for $\tilde{Q}_1 > 1$.

Existence of $\tilde{V}_{11}$ Similarly as above, from (44) we get $\|\tilde{h}_{11}\|^2 \le C \int_{[-1,1]} |b_{11}t|^{-\tilde{q}_1} \mathrm{d}t < \infty$ since $\tilde{q}_1 > 1$. This proves the existence of $\tilde{V}_{11}$ for $\tilde{Q}_2 > 1$.

Existence of V_{01}, V_{11}, V_{21} We have

$$\begin{aligned}\|h_{01}\|^2 &\le C \int_{\mathbb{R}^2} \Big(\int_{\mathbb{R}} |t_1|^{-q_1(1-\frac{1}{q_2})} \mathbf{1}(b_{22}t_1 + u_1 \in (0, 1], -b_{21}t_1 + u_2 \in (0, 1])\,\mathrm{d}t_1\Big)^2 \mathrm{d}\mathbf{u} \\ &\le C \int_{\mathbb{R}^2} |t_1|^{-q_1(1-\frac{1}{q_2})} |t_2|^{-q_1(1-\frac{1}{q_2})} \mathbf{1}(|b_{22}(t_1-t_2)| \le 1, |b_{21}(t_1-t_2)| \le 1)\mathrm{d}t_1\mathrm{d}t_2 \\ &\le C \int_{\mathbb{R}^2} |t_1|^{-q_1(1-\frac{1}{q_2})} |t_2|^{-q_1(1-\frac{1}{q_2})} \mathbf{1}(|t_1-t_2| \le 1)\mathrm{d}t_1\mathrm{d}t_2 < \infty\end{aligned}$$

since $\frac{1}{2} < q_1(1 - \frac{1}{q_2}) < 1$ or $\tilde{Q}_1 < 1 < Q$. The proof of $\|h_{11}\| < \infty$ and $\|h_{21}\| < \infty$ is completely analogous.

Existence of V_{22} Similarly as above, $\|h_{22}\|^2 \le C \int_{\mathbb{R}^2} |t_1|^{-q_2(1-\frac{1}{q_1})} |t_2|^{-q_2(1-\frac{1}{q_1})} \mathbf{1}(|t_1 - t_2| \le 1) \mathrm{d}t_1 \mathrm{d}t_2 < \infty$ since $\frac{1}{2} < q_2(1 - \frac{1}{q_1}) < 1$ or $\tilde{Q}_2 < 1 < Q$.

Existence of $\tilde{V}_{02}, \tilde{V}_{01}$ From (44) we get $\|\tilde{h}_{02}\|^2 \le C \int_{[-1,1]^2} \tilde{\rho}(b_{12}t_2, b_{22}t_2)^{-1} \mathrm{d}\mathbf{t} < \infty$ since $\tilde{q} = 2(q-1) < 1$ for $q \in (1, \frac{3}{2})$. The proof of $\|\tilde{h}_{01}\| < \infty$ is completely analogous.

Existence of V_{20}, V_{10} We have

$$\begin{aligned}
\|h_{20}\|^2 &\le C \int_{\mathbb{R}} \Big(\int_{\mathbb{R}^2} \rho(\mathbf{t})^{-1} \mathbf{1}(b_{22}t_1 - b_{12}t_2 + u \in (0,1]) \mathrm{d}\mathbf{t} \Big)^2 \mathrm{d}u \\
&\le C \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} \rho(1 + \frac{b_{12}}{b_{22}} t_2, t_2)^{-1} \mathrm{d}t_2 \int_{\mathbb{R}} |t_1|^{1-q} \mathbf{1}(b_{22}t_1 + u \in (0,1]) \mathrm{d}t_1 \Big)^2 \mathrm{d}u < \infty
\end{aligned}$$

for $q \in (\frac{3}{2}, 2)$. The proof of $\|h_{10}\| < \infty$ is completely analogous.

Existence of $\tilde{V}_0, V_0$ Relation $\|\tilde{h}_0\|^2 \le C \int_{[-1,1]^2} \tilde{\rho}(B\mathbf{t})^{-1} \mathrm{d}\mathbf{t} < \infty$ follows from (39) since $\frac{2}{\tilde{q}} = \frac{1}{q-1} > 1$ for $q \in (1,2)$. We have $\|h_0\| = \|\tilde{h}_0\| < \infty$.

It remains to show the relations (21)–(25), which follow from the variance expressions: for any $\mathbf{x} \in \mathbb{R}^2_+$, we have that

$$\begin{aligned}
&\mathrm{E}|\tilde{V}_{22}(\mathbf{x})|^2 = \tilde{\sigma}^2_{22} x_1^2 x_2^{2\tilde{H}_2}, \quad \mathrm{E}|\tilde{V}_{21}(\mathbf{x})|^2 = \tilde{\sigma}^2_{21} x_1^{2\tilde{H}_2} x_2^2, \quad \mathrm{E}|\tilde{V}_{11}(\mathbf{x})|^2 = \tilde{\sigma}^2_{11} x_1^{2\tilde{H}_1} x_2^2, \\
&\mathrm{E}|V_{11}(\mathbf{x})|^2 = \sigma^2_{11} x_1^{2H_1} x_2, \quad \mathrm{E}|V_{21}(\mathbf{x})|^2 = \sigma^2_{21} x_1 x_2^{2H_1}, \quad \mathrm{E}|V_{22}(\mathbf{x})|^2 = \sigma^2_{22} x_1 x_2^{2H_2}, \\
&\mathrm{E}|\tilde{V}_{01}(\mathbf{x})|^2 = \tilde{\sigma}^2_{01} x_1^{2\tilde{H}} x_2^2, \quad \mathrm{E}|\tilde{V}_{02}(\mathbf{x})|^2 = \tilde{\sigma}^2_{02} x_1^2 x_2^{2\tilde{H}}, \\
&\mathrm{E}|V_{10}(\mathbf{x})|^2 = \sigma^2_{10} x_1^{2H} x_2, \quad \mathrm{E}|V_{20}(\mathbf{x})|^2 = \sigma^2_{20} x_1 x_2^{2H}.
\end{aligned} \tag{45}$$

Relations (45) follow by a change of variables in the corresponding integrals, using the invariance property: $\lambda a_\infty(\lambda^{\frac{1}{q_1}} t_1, \lambda^{\frac{1}{q_2}} t_2) = a_\infty(\mathbf{t})$, $\mathbf{t} \in \mathbb{R}^2_0$, for all $\lambda > 0$. E.g., after a change of variables $t_2 \to x_2 t_2$, $u_2 \to x_2 u_2$, $u_1 \to x_2^{\frac{q_2}{q_1}} u_1$, the first expectation in (45) writes as $\mathrm{E}|\tilde{V}_{22}(\mathbf{x})|^2 = x_1^2 |\det(B)|^{-1} \int_{\mathbb{R}^2} |\int_0^{x_2} a_\infty(u_1, b_{22}t_2 + u_2) \mathrm{d}t_2|^2 \mathrm{d}\mathbf{u} = x_1^2 x_2^{2\tilde{H}_2} \tilde{\sigma}^2_{22}$, where $\tilde{\sigma}^2_{22} := |\det(B)|^{-1} \int_{\mathbb{R}^2} |\int_0^1 a_\infty(u_1, b_{22}t_2 + u_2) \mathrm{d}t_2|^2 \mathrm{d}\mathbf{u} < \infty$. Proposition 1 is proved. □

Similarly as in [14, 19] and other papers, in Theorems 1–4 we restrict the proof of (26) to one-dimensional convergence at $\mathbf{x} \in \mathbb{R}^2_+$. Towards this end, we use

Proposition 2 and rewrite every $\lambda^{-H(\gamma)} S^X_{\lambda,\gamma}(\mathbf{x}) = S(g_\lambda)$ as a linear form in (33) with

$$g_\lambda(\mathbf{u}) := \lambda^{-H(\gamma)} \int_{(0,\lambda x_1]\times(0,\lambda^\gamma x_2]} b(\lceil \mathbf{t} \rceil - \mathbf{u})\mathrm{d}\mathbf{t}, \quad \mathbf{u} \in \mathbb{Z}^2. \tag{46}$$

In what follows, w.l.g., we set $|\det(B)| = 1$.

Proof of Theorem 1

Case 1 and $\gamma > 1$ **or** $V^X_+ = \tilde{V}_{22}$ Set $H(\gamma) = 1 + \gamma(\frac{3}{2} + \frac{q_2}{2q_1} - q_2) = 1 + \gamma \tilde{H}_2$ in agreement with (28)–(30). Rewrite

$$g_\lambda(\mathbf{s}) = \lambda^{1+\gamma-H(\gamma)} \int_{\mathbb{R}^2} b(\lceil \Lambda'\mathbf{t} \rceil - \mathbf{s})\mathbf{1}(\Lambda'\mathbf{t} \in (\mathbf{0}, \lfloor \Lambda'\mathbf{x} \rfloor])\mathrm{d}\mathbf{t}, \quad \mathbf{s} \in \mathbb{Z}^2,$$

with $\Lambda' = \mathrm{diag}(\lambda, \lambda^\gamma)$. Use Proposition 2 with $A = B^{-1}$ and $\Lambda = \mathrm{diag}(l_1, l_2)$, where $l_1 = \lambda^{\gamma\frac{q_2}{q_1}}$, $l_2 = \lambda^\gamma$. According to the definition in (34),

$$\tilde{g}_\lambda(\mathbf{u}) = \int_{\mathbb{R}^2} \lambda^{\gamma q_2} b(\lceil \Lambda'\mathbf{t} \rceil - \lceil B^{-1}\Lambda\mathbf{u} \rceil)\mathbf{1}(\Lambda'\mathbf{t} \in (\mathbf{0}, \lfloor \Lambda'\mathbf{x} \rfloor])\mathrm{d}\mathbf{t}, \quad \mathbf{u} \in \mathbb{R}^2, \tag{47}$$

for which we need to show the L^2-convergence in (35) with $h(\mathbf{u})$ replaced by

$$\tilde{h}_{22}(\mathbf{u}) := \int_{(\mathbf{0},\mathbf{x}]} a_\infty(\tilde{B}_{22}\mathbf{t} - \mathbf{u})\mathrm{d}\mathbf{t}, \quad \mathbf{u} \in \mathbb{R}^2, \tag{48}$$

where the integrand does not depend on t_1. Note since $q_1 < q_2$ and $\gamma > 1$ that

$$\Lambda^{-1} B(\lceil \Lambda'\mathbf{t} \rceil - \lceil B^{-1}\Lambda\mathbf{u} \rceil) \to \tilde{B}_{22}\mathbf{t} - \mathbf{u}, \tag{49}$$

point-wise for any $\mathbf{t}, \mathbf{u} \in \mathbb{R}^2$ and therefore, by continuity of ρ^{-1},

$$\lambda^{\gamma q_2} \rho(B(\lceil \Lambda'\mathbf{t} \rceil - \lceil B^{-1}\Lambda\mathbf{u} \rceil))^{-1} \to \rho(\tilde{B}_{22}\mathbf{t} - \mathbf{u})^{-1} \tag{50}$$

for any $\mathbf{t}, \mathbf{u} \in \mathbb{R}^2$ such that $\tilde{B}_{22}\mathbf{t} - \mathbf{u} \neq \mathbf{0}$. Later use (13), (49), (50) and continuity of $L_\pm$ to get

$$\lambda^{\gamma q_2} b(\lceil \Lambda'\mathbf{t} \rceil - \lceil B^{-1}\Lambda\mathbf{u} \rceil) \to a_\infty(\tilde{B}_{22}\mathbf{t} - \mathbf{u}) \tag{51}$$

for $\mathbf{t}, \mathbf{u} \in \mathbb{R}^2$ such that $\tilde{B}_{22}\mathbf{t} - \mathbf{u} \neq \mathbf{0}$. Therefore, $\tilde{g}_\lambda(\mathbf{u}) \to \tilde{h}_{22}(\mathbf{u})$ for all $\mathbf{u} \in \mathbb{R}^2$. This point-wise convergence can be extended to that in $L^2(\mathbb{R}^2)$ by applying Pratt's lemma, c.f. [15], [14, proof of Theorem 3.2], to the domination $|\tilde{g}_\lambda(\mathbf{u})| \le C\tilde{G}_\lambda(\mathbf{u})$, where

$$\tilde{G}_\lambda(\mathbf{u}) := \int_{(\mathbf{0},\mathbf{x}]} \rho(\Lambda^{-1}B\Lambda'\mathbf{t}-\mathbf{u})^{-1}\mathrm{d}\mathbf{t} \to \int_{(\mathbf{0},\mathbf{x}]} \rho(\tilde{B}_{22}\mathbf{t}-\mathbf{u})^{-1}\mathrm{d}\mathbf{t} =: \tilde{G}(\mathbf{u}), \tag{52}$$

for all $\mathbf{u} \in \mathbb{R}^2$. To get this domination we use $|b(\mathbf{s})| \le C \max\{\rho(B\mathbf{s}), 1\}^{-1}$, $\mathbf{s} \in \mathbb{Z}^2$, and by (40) and $\rho(\Lambda^{-1}\mathbf{t}) = \lambda^{-\gamma q_2}\rho(\mathbf{t})$, we further see that

$$\begin{aligned}\rho(\Lambda^{-1}B\Lambda'\mathbf{t} - \mathbf{u}) \le C\{&\rho(\Lambda^{-1}B(\lceil\Lambda'\mathbf{t}\rceil - \lceil B^{-1}\Lambda\mathbf{u}\rceil)) \\ &+\rho(\Lambda^{-1}B(\Lambda'\mathbf{t} - B^{-1}\Lambda\mathbf{u} - \lceil\Lambda'\mathbf{t}\rceil + \lceil B^{-1}\Lambda\mathbf{u}\rceil))\} \\ \le C\lambda^{-\gamma q_2} &\max\{\rho(B(\lceil\Lambda'\mathbf{t}\rceil - \lceil B^{-1}\Lambda\mathbf{u}\rceil)), 1\},\end{aligned}$$

where C does not depend on $\mathbf{t}, \mathbf{u} \in \mathbb{R}^2$. Then in view of the domination $|\tilde{g}_\lambda(\mathbf{u})| \le C\tilde{G}_\lambda(\mathbf{u})$, $\mathbf{u} \in \mathbb{R}^2$, and (52), the L^2-convergence in (35) follows by Pratt's lemma from the convergence of norms $\|\tilde{G}_\lambda\|^2 = \int_{(\mathbf{0},\mathbf{x}]\times(\mathbf{0},\mathbf{x}]}(\rho^{-1} \star \rho^{-1})(\Lambda^{-1}B\Lambda'(\mathbf{t} - \mathbf{s}))\mathrm{d}\mathbf{t}\mathrm{d}\mathbf{s} \to \|\tilde{G}\|^2$. Indeed, after such a change of variables in the last integral that $b_{22}(t_2 - s_2) = b_{22}(t_2' - s_2') - \lambda^{1-\gamma}b_{21}(t_1 - s_1)$ we see that

$$\begin{aligned}\|\tilde{G}_\lambda\|^2 = \int_{\mathbb{R}^4}&(\rho^{-1} \star \rho^{-1})(\lambda^{1-\gamma\frac{q_2}{q_1}}\frac{\det(B)}{b_{22}}(t_1 - s_1) \\ &+\lambda^{\gamma(1-\frac{q_2}{q_1})}b_{12}(t_2' - s_2'), b_{22}(t_2' - s_2')) \\ &\times\mathbf{1}(t_1 \in (0, x_1],\ -\lambda^{1-\gamma}\frac{b_{21}}{b_{22}}t_1 + t_2' \in (0, x_2]) \\ &\times\mathbf{1}(s_1 \in (0, x_1],\ -\lambda^{1-\gamma}\frac{b_{21}}{b_{22}}s_1 + s_2' \in (0, x_2])\mathrm{d}t_1\mathrm{d}t_2'\mathrm{d}s_1\mathrm{d}s_2' \\ \to \int_{(\mathbf{0},\mathbf{x}]\times(\mathbf{0},\mathbf{x}]}&(\rho^{-1} \star \rho^{-1})(0, b_{22}(t_2' - s_2'))\mathrm{d}t_1\mathrm{d}t_2'\mathrm{d}s_1\mathrm{d}s_2' = \|\tilde{G}\|^2 \qquad (53)\end{aligned}$$

by the dominated convergence theorem using the continuity of $(\rho^{-1} \star \rho^{-1})(\mathbf{t})$ and $(\rho^{-1} \star \rho^{-1})(\mathbf{t}) \le C|t_2|^{-\tilde{q}_2}$ for $t_2 \ne 0$ with $\tilde{q}_2 = q_2(2 - Q) < 1$, see (41), (42).

Case 1 and $\gamma = 1$ or $V_0^X = \tilde{V}_{20}$ Set $H(\gamma) = \frac{5}{2} + \frac{q_2}{2q_1} - q_2 = 1 + \tilde{H}_2$. The proof is similar to that in Case 1, $\gamma > 1$ above. We use Proposition 2 with $A = B^{-1}$, $\Lambda' = \mathrm{diag}(\lambda, \lambda)$ and $\Lambda = \mathrm{diag}(\lambda^{\frac{q_2}{q_1}}, \lambda)$. Accordingly, we need to prove $\|\tilde{g}_\lambda - \tilde{h}_{20}\| \to 0$, where

$$\tilde{h}_{20}(\mathbf{u}) := \int_{(\mathbf{0},\mathbf{x}]} a_\infty(\tilde{B}_{20}\mathbf{t} - \mathbf{u})\mathrm{d}\mathbf{t}, \quad \mathbf{u} \in \mathbb{R}^2,$$

and $\tilde{g}_\lambda$ is defined as in (47) with $\gamma = 1$. Note that now (49) must be replaced by

$$\Lambda^{-1}B(\lceil\Lambda'\mathbf{t}\rceil - \lceil B^{-1}\Lambda\mathbf{u}\rceil) \to \tilde{B}_{20}\mathbf{t} - \mathbf{u},$$

leading to

$$\lambda^{q_2}b(\lceil\Lambda'\mathbf{t}\rceil - \lceil B^{-1}\Lambda\mathbf{u}\rceil) \to a_\infty(\tilde{B}_{20}\mathbf{t} - \mathbf{u})$$

for $\tilde{B}_{20}\mathbf{t} - \mathbf{u} \neq \mathbf{0}$. Therefore, $\tilde{g}_\lambda$ converges to $\tilde{h}_{20}$ point-wise. To prove the L^2-convergence use Pratt's lemma as in the case $\gamma > 1$ above, with $\tilde{G}_\lambda$, $\tilde{G}$ defined as in (52) with $\gamma = 1$ and $\tilde{B}_{22}$ replaced by $\tilde{B}_{20}$. Then $\tilde{G}_\lambda(\mathbf{u}) \to \tilde{G}(\mathbf{u})$ for all $\mathbf{u} \in \mathbb{R}^2$ as in (52) and

$$\begin{aligned}\|\tilde{G}_\lambda\|^2 &= \int_{(\mathbf{0},\mathbf{x}]\times(\mathbf{0},\mathbf{x}]} (\rho^{-1} \star \rho^{-1})(\Lambda^{-1} B \Lambda'(\mathbf{t}-\mathbf{s}))\mathrm{d}\mathbf{t}\mathrm{d}\mathbf{s} \\ &= \int_{(\mathbf{0},\mathbf{x}]\times(\mathbf{0},\mathbf{x}]} (\rho^{-1} \star \rho^{-1})(\lambda^{1-\frac{q_2}{q_1}}(b_{11}(t_1-s_1)+b_{12}(t_2-s_2)), \\ &\qquad b_{21}(t_1-s_1)+b_{22}(t_2-s_2))\mathrm{d}\mathbf{t}\mathrm{d}\mathbf{s} \\ &\to \int_{(\mathbf{0},\mathbf{x}]\times(\mathbf{0},\mathbf{x}]} (\rho^{-1} \star \rho^{-1})(\tilde{B}_{20}(\mathbf{t}-\mathbf{s}))\mathrm{d}\mathbf{t}\mathrm{d}\mathbf{s} = \|\tilde{G}\|^2\end{aligned}$$

follows similarly to (53).

Case 1 and $\gamma < 1$ **or** $V_-^X = \tilde{V}_{21}$ Set $H(\gamma) = \gamma + \frac{3}{2} + \frac{q_2}{2q_1} - q_2 = \gamma + \tilde{H}_2$. The proof proceeds similarly as above with $A = B^{-1}$, $\Lambda' = \mathrm{diag}(\lambda, \lambda^\gamma)$ and $\Lambda = \mathrm{diag}(\lambda^{\frac{q_2}{q_1}}, \lambda)$. Then

$$\begin{aligned}\tilde{g}_\lambda(\mathbf{u}) &= \int_{\mathbb{R}^2} \lambda^{q_2} b(\lceil \Lambda'\mathbf{t}\rceil - \lceil B^{-1}\Lambda\mathbf{u}\rceil)\mathbf{1}(\Lambda'\mathbf{t} \in (\mathbf{0}, \lfloor \Lambda'\mathbf{x}\rfloor])\mathrm{d}\mathbf{t} \\ &\to \int_{(\mathbf{0},\mathbf{x}]} a_\infty(\tilde{B}_{21}\mathbf{t}-\mathbf{u})\mathrm{d}\mathbf{t} =: \tilde{h}_{21}(\mathbf{u}), \quad \mathbf{u} \in \mathbb{R}^2,\end{aligned}$$

in view of $\Lambda^{-1}B(\lceil \Lambda'\mathbf{t}\rceil - \lceil B^{-1}\Lambda\mathbf{u}\rceil) \to \tilde{B}_{21}\mathbf{t} - \mathbf{u}$ and $\lambda^{q_2} b(\lceil \Lambda'\mathbf{t}\rceil - \lceil B^{-1}\Lambda\mathbf{u}\rceil) \to a_\infty(\tilde{B}_{21}\mathbf{t}-\mathbf{u})$ for $\tilde{B}_{21}\mathbf{t}-\mathbf{u} \neq \mathbf{0}$. The proof of $\|\tilde{g}_\lambda - \tilde{h}_{21}\| \to 0$ using Pratt's lemma also follows similarly as above, with $\tilde{G}_\lambda(\mathbf{u}) := \int_{(\mathbf{0},\mathbf{x}]} \rho(\Lambda^{-1}B\Lambda'\mathbf{t}-\mathbf{u})^{-1}\mathrm{d}\mathbf{t} \to \int_{(\mathbf{0},\mathbf{x}]} \rho(\tilde{B}_{21}\mathbf{t}-\mathbf{u})^{-1}\mathrm{d}\mathbf{t} =: \tilde{G}(\mathbf{u})$ for all $\mathbf{u} \in \mathbb{R}^2$ and

$$\begin{aligned}\|\tilde{G}_\lambda\|^2 &= \int_{\mathbb{R}^4} (\rho^{-1} \star \rho^{-1})(\lambda^{1-\frac{q_2}{q_1}} b_{11}(t_1'-s_1') - \lambda^{\gamma-\frac{q_2}{q_1}}\frac{\det(B)}{b_{21}}(t_2-s_2), b_{21}(t_1'-s_1')) \\ &\qquad \times \mathbf{1}(t_1' - \lambda^{\gamma-1}\frac{b_{22}}{b_{21}}t_2 \in (0,x_1],\ t_2 \in (0,x_2]) \\ &\qquad \times \mathbf{1}(s_1' - \lambda^{\gamma-1}\frac{b_{22}}{b_{21}}t_2 \in (0,x_1],\ s_2 \in (0,x_2])\mathrm{d}t_1'\mathrm{d}t_2\mathrm{d}s_1'\mathrm{d}s_2 \\ &\to \int_{(\mathbf{0},\mathbf{x}]\times(\mathbf{0},\mathbf{x}]} (\rho^{-1} \star \rho^{-1})(0, b_{21}(t_1'-s_1'))\mathrm{d}t_1'\mathrm{d}t_2\mathrm{d}s_1'\mathrm{d}s_2 = \|\tilde{G}\|^2\end{aligned}$$

as in (53).

Case 2 and $\gamma > \frac{q_1}{q_2}$ **or** $V_+^X = \tilde{V}_{22}$ Set $H(\gamma) = 1+\gamma\tilde{H}_2$. The proof of $\|\tilde{g}_\lambda - \tilde{h}_{22}\| \to 0$ is completely analogous to that in Case 1, $\gamma > 1$, with the same $\tilde{g}_\lambda$, Λ', Λ, $\tilde{h}_{22}$ as in (47), (48) using the fact that

$$\Lambda^{-1}B\Lambda' = \begin{bmatrix} \lambda^{1-\gamma\frac{q_2}{q_1}} b_{11} & \lambda^{\gamma(1-\frac{q_2}{q_1})} b_{12} \\ 0 & b_{22} \end{bmatrix} \to \tilde{B}_{22}.$$

Case 2 and $\gamma = \frac{q_1}{q_2}$ **or** $V_0^X = \tilde{V}_{00}$ Set $H(\gamma) = \frac{3}{2} + \frac{3q_1}{2q_2} - q_1 = \tilde{H}_1 + \frac{q_1}{q_2} = 1 + \frac{q_1}{q_2}\tilde{H}_2$, $A = B^{-1}$, $\Lambda = \Lambda' = \mathrm{diag}(\lambda, \lambda^{\frac{q_1}{q_2}})$. The proof of $\|\tilde{g}_\lambda - \tilde{h}_{00}\| \to 0$ with $\tilde{h}_{00}(\mathbf{u}) := \int_{(\mathbf{0},\mathbf{x}]} a_\infty(\tilde{B}_{00}\mathbf{t} - \mathbf{u})\mathrm{d}\mathbf{t}$, $\mathbf{u} \in \mathbb{R}^2$, follows similar lines as in the other cases. The point-wise convergence $\tilde{g}_\lambda \to \tilde{h}_{00}$ uses $\Lambda^{-1}B(\lceil\Lambda'\mathbf{t}\rceil - \lceil B^{-1}\Lambda\mathbf{u}\rceil) \to \tilde{B}_{00}\mathbf{t} - \mathbf{u}$ and $\lambda^{q_1} b(\lceil\Lambda'\mathbf{t}\rceil - \lceil B^{-1}\Lambda\mathbf{u}\rceil) \to a_\infty(\tilde{B}_{00}\mathbf{t} - \mathbf{u})$ for $\tilde{B}_{00}\mathbf{t} - \mathbf{u} \neq \mathbf{0}$. The L^2-convergence can be verified using Pratt's lemma with the dominating function $\tilde{G}_\lambda(\mathbf{u}) := \int_{(\mathbf{0},\mathbf{x}]} \rho(\Lambda^{-1}B\Lambda'\mathbf{t} - \mathbf{u})^{-1}\mathrm{d}\mathbf{t} \to \int_{(\mathbf{0},\mathbf{x}]} \rho(\tilde{B}_{00}\mathbf{t} - \mathbf{u})^{-1}\mathrm{d}\mathbf{t} =: \tilde{G}(\mathbf{u})$, $\mathbf{u} \in \mathbb{R}^2$, satisfying

$$\begin{aligned}
\|\tilde{G}_\lambda\|^2 &= \int_{\mathbb{R}^4} (\rho^{-1} \star \rho^{-1})(b_{11}(t_1' - s_1'), b_{22}(t_2 - s_2)) \\
&\qquad \times \mathbf{1}\Big(t_1' - \lambda^{\frac{q_1}{q_2}-1}\frac{b_{12}}{b_{11}} t_2 \in (0, x_1],\ t_2 \in (0, x_2]\Big) \\
&\qquad \times \mathbf{1}\Big(s_1' - \lambda^{\frac{q_1}{q_2}-1}\frac{b_{12}}{b_{11}} s_2 \in (0, x_1],\ s_2 \in (0, x_2]\Big)\mathrm{d}t_1\mathrm{d}t_2\mathrm{d}s_1\mathrm{d}s_2 \\
&\to \int_{(\mathbf{0},\mathbf{x}]\times(\mathbf{0},\mathbf{x}]} (\rho^{-1} \star \rho^{-1})(b_{11}(t_1' - s_1'), b_{22}(t_2 - s_2))\mathrm{d}t_1'\mathrm{d}t_2\mathrm{d}s_1'\mathrm{d}s_2 = \|\tilde{G}\|^2,
\end{aligned}$$

which follows from the dominated convergence theorem using $(\rho^{-1} \star \rho^{-1})(\mathbf{t}) \le C\tilde{\rho}(\mathbf{t})^{-1}$, $\mathbf{t} \in \mathbb{R}_0^2$, and the (local) integrability of the function $\tilde{\rho}^{-1}$ with $\frac{1}{\tilde{q}_1} + \frac{1}{\tilde{q}_2} = \frac{Q}{2-Q} > 1$, see (41), (42) and (39).

We note that the above proof applies for all $q_1 < q_2$ satisfying $1 < Q < 2$, hence also in Cases 2, $V_0^X = \tilde{V}_{00}$ of Theorems 2 and 3.

Case 2 and $\gamma < \frac{q_1}{q_2}$ **or** $V_-^X = \tilde{V}_{11}$ Set $H(\gamma) = \gamma + \frac{3}{2} + \frac{q_1}{2q_2} - q_1 = \gamma + \tilde{H}_1$, $A = B^{-1}$, $\Lambda = \mathrm{diag}(\lambda, \lambda^{\frac{q_1}{q_2}})$, $\Lambda' = \mathrm{diag}(\lambda, \lambda^\gamma)$. Then $\Lambda^{-1}B(\lceil\Lambda'\mathbf{t}\rceil - \lceil B^{-1}\Lambda\mathbf{u}\rceil) \to \tilde{B}_{11}\mathbf{t} - \mathbf{u}$ and $\lambda^{q_1} b(\lceil\Lambda'\mathbf{t}\rceil - \lceil B^{-1}\Lambda\mathbf{u}\rceil) \to a_\infty(\tilde{B}_{11}\mathbf{t} - \mathbf{u})$ for $\tilde{B}_{11}\mathbf{t} - \mathbf{u} \neq \mathbf{0}$. This leads to $\tilde{g}_\lambda(\mathbf{u}) \to h_{11}(\mathbf{u}) := \int_{(\mathbf{0},\mathbf{x}]} a_\infty(\tilde{B}_{11}\mathbf{t} - \mathbf{u})\mathrm{d}\mathbf{t}$ for all $\mathbf{u} \in \mathbb{R}^2$. The required convergence $\|\tilde{g}_\lambda - \tilde{h}_{11}\| \to 0$ follows similarly as in the other cases using $\tilde{G}_\lambda(\mathbf{u}) := \int_{(\mathbf{0},\mathbf{x}]} \rho(\Lambda^{-1} B \Lambda' \mathbf{t} - \mathbf{u})^{-1} \mathrm{d}\mathbf{t} \to \int_{(\mathbf{0},\mathbf{x}]} \rho(\tilde{B}_{11}\mathbf{t} - \mathbf{u})^{-1} \mathrm{d}\mathbf{t} =: \tilde{G}(\mathbf{u})$ for all $\mathbf{u} \in \mathbb{R}^2$ and

$$
\begin{aligned}
\|\tilde{G}_\lambda\|^2 = & \int_{\mathbb{R}^4} (\rho^{-1} \star \rho^{-1})(b_{11}(t_1' - s_1'), \lambda^{\gamma - \frac{q_1}{q_2}} b_{22}(t_2 - s_2)) \\
& \times \mathbf{1}(t_1' - \lambda^{\gamma-1} \frac{b_{12}}{b_{11}} t_2 \in (0, x_1],\ t_2 \in (0, x_2]) \\
& \times \mathbf{1}(s_1' - \lambda^{\gamma-1} \frac{b_{12}}{b_{11}} s_2 \in (0, x_1],\ s_2 \in (0, x_2]) \mathrm{d}t_1' \mathrm{d}t_2 \mathrm{d}s_1' \mathrm{d}s_2 \\
\to & \int_{(\mathbf{0},\mathbf{x}] \times (\mathbf{0},\mathbf{x}]} (\rho^{-1} \star \rho^{-1})(b_{11}(t_1' - s_1'), 0) \mathrm{d}t_1' \mathrm{d}t_2 \mathrm{d}s_1' \mathrm{d}s_2 = \|\tilde{G}\|^2
\end{aligned}
$$

which follows from the dominated convergence theorem using $(\rho^{-1} \star \rho^{-1})(\mathbf{t}) \le C|t_1|^{-\tilde{q}_1}$ for all $t_1 \neq 0$ with $\tilde{q}_1 < 1$.

Note that the above proof applies for all $q_1 < q_2$ satisfying $\tilde{q}_1 < 1$ or $\tilde{Q}_2 > 1$ hence also in Case 2, $V_-^X = \tilde{V}_{11}$ of Theorem 2. Theorem 1 is proved. □

Proof of Theorem 2

Case 1 and $\gamma > 1$ or $V_+^X = V_{11}$ Set $H(\gamma) = \frac{3}{2} + \frac{q_1}{q_2} - q_1 + \frac{\gamma}{2} = H_1 + \frac{\gamma}{2}$. Rewrite $g_\lambda(\mathbf{s})$ as

$$
\begin{aligned}
g_\lambda(\mathbf{s}) &= \lambda^{-H(\gamma)} \int_{\mathbb{R}^2} b(\lceil \mathbf{t} \rceil) \mathbf{1}(\lceil \mathbf{t} \rceil + \mathbf{s} \in (0, \lfloor \lambda x_1 \rfloor] \times (0, \lfloor \lambda^\gamma x_2 \rfloor]) \mathrm{d}\mathbf{t} \\
&= \lambda^{1 + \frac{q_1}{q_2} - H(\gamma)} \int_{\mathbb{R}^2} b(\lceil B^{-1} \Lambda' \mathbf{t} \rceil) \\
&\qquad \times \mathbf{1}(\lceil B^{-1} \Lambda' \mathbf{t} \rceil + \mathbf{s} \in (0, \lfloor \lambda x_1 \rfloor] \times (0, \lfloor \lambda^\gamma x_2 \rfloor]) \mathrm{d}\mathbf{t}, \quad \mathbf{s} \in \mathbb{Z}^2, \qquad (54)
\end{aligned}
$$

where $\Lambda' := \mathrm{diag}(\lambda, \lambda^{\frac{q_1}{q_2}})$. We use Proposition 2 with

$$
A := \begin{bmatrix} 1 & 0 \\ -\frac{b_{21}}{b_{22}} & 1 \end{bmatrix}, \quad \Lambda := \begin{bmatrix} \lambda & 0 \\ 0 & \lambda^\gamma \end{bmatrix}. \qquad (55)
$$

Note $\Lambda^{-1} B^{-1} \Lambda' \to B_{11}$, $\Lambda^{-1} A \Lambda \to I$ since $\gamma > 1$. Then, according to the definition (34), $\tilde{g}_\lambda(\mathbf{u}) = \int_{\mathbb{R}^2} \tilde{b}_\lambda(\mathbf{u}, \mathbf{t}) \mathrm{d}\mathbf{t}$, where

$$
\begin{aligned}
\tilde{b}_\lambda(\mathbf{u}, \mathbf{t}) &:= \lambda^{q_1} b(\lceil B^{-1} \Lambda' \mathbf{t} \rceil) \mathbf{1}(\lceil B^{-1} \Lambda' \mathbf{t} \rceil + \lceil A \Lambda \mathbf{u} \rceil \in (\mathbf{0}, \lfloor \Lambda \mathbf{x} \rfloor]) \\
&\to a_\infty(\mathbf{t}) \mathbf{1}(B_{11}\mathbf{t} + \mathbf{u} \in (\mathbf{0}, \mathbf{x}]) \qquad (56)
\end{aligned}
$$

for all $\mathbf{u}, \mathbf{t} \in \mathbb{R}^2$ such that $\mathbf{t} \neq \mathbf{0}$, $b_{22}t_1 + u_1 \not\in \{0, x_1\}$, $u_2 \not\in \{0, x_2\}$. It suffices to prove that the following point-wise convergence also holds

$$\tilde{g}_\lambda(\mathbf{u}) \to h_{11}(\mathbf{u}) := \int_{\mathbb{R}^2} a_\infty(\mathbf{t})\mathbf{1}(B_{11}\mathbf{t} + \mathbf{u} \in (\mathbf{0}, \mathbf{x}])\mathrm{d}\mathbf{t} \quad \text{in } L^2(\mathbb{R}^2). \tag{57}$$

To show (57), decompose $\tilde{g}_\lambda(\mathbf{u}) = \tilde{g}_{\lambda,0}(\mathbf{u}) + \tilde{g}_{\lambda,1}(\mathbf{u})$ with $\tilde{g}_{\lambda,j}(\mathbf{u}) = \int_{\mathbb{R}^2} \tilde{b}_{\lambda,j}(\mathbf{u}, \mathbf{t})\mathrm{d}\mathbf{t}$ given by

$$\tilde{b}_{\lambda,0}(\mathbf{u}, \mathbf{t}) := \tilde{b}_\lambda(\mathbf{u}, \mathbf{t})\mathbf{1}(|t_2| \ge \lambda^{1-\frac{q_1}{q_2}}),$$
$$\tilde{b}_{\lambda,1}(\mathbf{u}, \mathbf{t}) := \tilde{b}_\lambda(\mathbf{u}, \mathbf{t})\mathbf{1}(|t_2| < \lambda^{1-\frac{q_1}{q_2}}), \quad \mathbf{u}, \mathbf{t} \in \mathbb{R}^2.$$

Then (57) follows from

$$\|\tilde{g}_{\lambda,1} - h_{11}\| \to 0 \quad \text{and} \quad \|\tilde{g}_{\lambda,0}\| \to 0. \tag{58}$$

The first relation in (58) follows from (56) and the dominated convergence theorem, as follows. To justify the domination, combine $|b(\mathbf{s})| \le C \max\{\rho(B\mathbf{s}), 1\}^{-1}$, $\mathbf{s} \in \mathbb{Z}^2$, and $\rho(\mathbf{t}) \le C\lambda^{-q_1} \max\{\rho(B\lceil B^{-1}\Lambda'\mathbf{t}\rceil), 1\}$, $\mathbf{t} \in \mathbb{R}^2$, to get $\lambda^{q_1}|b(\lceil B^{-1}\Lambda'\mathbf{t}\rceil)| \le C\rho(\mathbf{t})^{-1}$, $\mathbf{t} \in \mathbb{R}^2$, $\lambda > 1$. Also note that

$$\begin{aligned}
&\mathbf{1}(|t_2| < \lambda^{1-\frac{q_1}{q_2}}, \lceil B^{-1}\Lambda'\mathbf{t}\rceil + \lceil A\Lambda\mathbf{u}\rceil \in (\mathbf{0}, \lfloor \Lambda\mathbf{x}\rfloor]) \\
&\quad \le \mathbf{1}(|t_2| < \lambda^{1-\frac{q_1}{q_2}}, B^{-1}\Lambda'\mathbf{t} + A\Lambda\mathbf{u} \in (-\mathbf{2}, \Lambda\mathbf{x}]) \\
&\quad \le \mathbf{1}(|t_2| < \lambda^{1-\frac{q_1}{q_2}}, b_{22}t_1 - \lambda^{\frac{q_1}{q_2}-1}b_{12}t_2 + u_1 \in (-\frac{2}{\lambda}, x_1], \\
&\qquad\quad -\lambda^{1-\gamma}\frac{b_{21}}{b_{22}}(b_{22}t_1 - \lambda^{\frac{q_1}{q_2}-1}b_{12}t_2 + u_1) + \lambda^{\frac{q_1}{q_2}-\gamma}\frac{\det(B)}{b_{22}}t_2 + u_2 \in (-\frac{2}{\lambda^\gamma}, x_2]) \\
&\quad \le \mathbf{1}(|b_{22}t_1 + u_1| \le C_1, |u_2| \le C_2)
\end{aligned}$$

for some $C_1, C_2 > 0$ independent of $\mathbf{t}, \mathbf{u} \in \mathbb{R}^2$ and $\lambda > 0$. Thus, $|\tilde{g}_{\lambda,1}(\mathbf{u})| \le \bar{g}(\mathbf{u})$, where the dominating function $\bar{g}(\mathbf{u}) := C\mathbf{1}(|u_2| \le C_2)\int_{\mathbb{R}^2} \rho(\mathbf{t})^{-1}\mathbf{1}(|b_{22}t_1 + u_1| \le C_1)\mathrm{d}\mathbf{t} \le C\mathbf{1}(|u_2| \le C_2)\int_{\mathbb{R}} |t_1|^{\frac{q_1}{q_2}-q_1}\mathbf{1}(|b_{22}t_1 + u_1| \le C_1)\mathrm{d}t_1$, $\mathbf{u} \in \mathbb{R}^2$, satisfies $\|\bar{g}\| < \infty$ since $\tilde{Q}_1 < 1$, proving the first relation in (58). The second relation in (58) follows by Minkowski's inequality:

$$\begin{aligned}
\|\tilde{g}_{\lambda,0}\| \le C\int_{\mathbb{R}^2}\Big(\int_{\mathbb{R}^2} &\rho(t_1 - \frac{1}{b_{22}}u_1, t_2)^{-2}\mathbf{1}(|t_2| \ge \lambda^{1-\frac{q_1}{q_2}}) \\
&\times \mathbf{1}(|b_{22}t_1 - \lambda^{\frac{q_1}{q_2}-1}b_{12}t_2| \le C_1, |\lambda^{\frac{q_1}{q_2}-\gamma}\frac{\det(B)}{b_{22}}t_2 + u_2| \le C_2)\mathrm{d}\mathbf{u}\Big)^{\frac{1}{2}}\mathrm{d}\mathbf{t}
\end{aligned}$$

$$= C \int_{\mathbb{R}^2} \bigg(\int_{\mathbb{R}} \rho(u_1, t_2)^{-2} \mathbf{1}(|t_2| \ge \lambda^{1-\frac{q_1}{q_2}}) \times \mathbf{1}(|b_{22}t_1 - \lambda^{\frac{q_1}{q_2}-1} b_{12}t_2| \le C_1) \mathrm{d}u_1 \bigg)^{\frac{1}{2}} \mathrm{d}\mathbf{t}$$

$$= C \int_{\mathbb{R}^2} |t_2|^{q_2(\frac{1}{2q_1}-1)} \mathbf{1}(|t_2| \ge \lambda^{1-\frac{q_1}{q_2}}, |b_{22}t_1 - \lambda^{\frac{q_1}{q_2}-1} b_{12}t_2| \le C_1) \mathrm{d}\mathbf{t}$$

$$= C \int_{\mathbb{R}} |t_2|^{q_2(\frac{1}{2q_1}-1)} \mathbf{1}(|t_2| \ge \lambda^{1-\frac{q_1}{q_2}}) \mathrm{d}t_2 = o(1)$$

since $\frac{1}{2q_1} < \tilde{Q}_1 < 1$. This proves (58) and (57).

We note that the above argument applies to the proof of the limit $V_+^X = V_{11}$ in both Cases 1 and 2 of Theorems 2 and 3 as well, including Case 2 and $\frac{q_1}{q_2} < \gamma \le 1$, with the difference that in the latter case we make the change of variable $\mathbf{t} \to B^{-1}\Lambda'\mathbf{t}$ in (54) with $\Lambda' := \mathrm{diag}(\lambda^{\gamma}, \lambda^{\gamma\frac{q_1}{q_2}})$.

Case 1 and $\gamma = 1$ or $V_+^X = V_{01}$ Set $H(\gamma) = H_1 + \frac{1}{2}$. The proof proceeds as in Case 1, $\gamma > 1$ above, by writing $\tilde{g}_\lambda(\mathbf{u}) = \int_{\mathbb{R}^2} \tilde{b}_\lambda(\mathbf{u}, \mathbf{t})\mathrm{d}\mathbf{t}$ with A as in (55), $\Lambda := \mathrm{diag}(\lambda, \lambda)$, $\Lambda' := \mathrm{diag}(\lambda, \lambda^{\frac{q_1}{q_2}})$ in

$$\tilde{b}_\lambda(\mathbf{u}, \mathbf{t}) := \lambda^{q_1} b(\lceil B^{-1}\Lambda'\mathbf{t}\rceil)\mathbf{1}(\lceil B^{-1}\Lambda'\mathbf{t}\rceil + \lceil \Lambda\mathbf{u}\rceil \in (\mathbf{0}, \lfloor \Lambda\mathbf{x}\rfloor]) \to a_\infty(\mathbf{t})\mathbf{1}(B_{01}\mathbf{t} + \mathbf{u} \in (\mathbf{0}, \mathbf{x}])$$

for all $\mathbf{u}, \mathbf{t} \in \mathbb{R}^2$ such that $\mathbf{t} \ne \mathbf{0}$, $b_{22}t_1 + u_1 \notin \{0, x_1\}$, $-b_{21}t_1 + u_2 \notin \{0, x_2\}$, c.f. (56). The details of the convergence $\tilde{g}_\lambda(\mathbf{u}) \to h_{01}(\mathbf{u}) := \int_{\mathbb{R}^2} a_\infty(\mathbf{t})\mathbf{1}(B_{01}\mathbf{t} + \mathbf{u} \in (\mathbf{0}, \mathbf{x}])\mathrm{d}\mathbf{t}$, $\mathbf{u} \in \mathbb{R}^2$, in $L^2(\mathbb{R}^2)$ are similar as above and omitted.

We note that the above argument applies to the proof of $V_0^X = V_{01}$ in Case 1 of Theorem 3.

Case 1 and $\gamma < 1$ or $V_-^X = V_{21}$ Set $H(\gamma) = \frac{1}{2} + \gamma(\frac{3}{2} + \frac{q_1}{q_2} - q_1) = \frac{1}{2} + \gamma H_1$. Use Proposition 2 with

$$A := \begin{bmatrix} 1 & -\frac{b_{22}}{b_{21}} \\ 0 & 1 \end{bmatrix}, \quad \Lambda := \begin{bmatrix} \lambda & 0 \\ 0 & \lambda^{\gamma} \end{bmatrix}.$$

Similarly to (56), $\tilde{g}_\lambda(\mathbf{u}) = \int_{\mathbb{R}^2} \tilde{b}_\lambda(\mathbf{u}, \mathbf{t})\mathrm{d}\mathbf{t}$, where

$$\tilde{b}_\lambda(\mathbf{u}, \mathbf{t}) := \lambda^{\gamma q_1} b(\lceil B^{-1}\Lambda'\mathbf{t}\rceil)\mathbf{1}(\lceil B^{-1}\Lambda'\mathbf{t}\rceil + \lceil \Lambda\mathbf{u}\rceil \in (\mathbf{0}, \lfloor \Lambda\mathbf{x}\rfloor]) \to a_\infty(\mathbf{t})\mathbf{1}(B_{21}\mathbf{t} + \mathbf{u} \in (\mathbf{0}, \mathbf{x}])$$

for all $\mathbf{u}, \mathbf{t} \in \mathbb{R}^2$ such that $\mathbf{t} \neq \mathbf{0}$, $u_1 \notin \{0, x_1\}$, $-b_{21}t_1 + u_2 \notin \{0, x_2\}$. It suffices to prove $\tilde{g}_\lambda(\mathbf{u}) \to h_{21}(\mathbf{u}) := \int_{\mathbb{R}^2} a_\infty(\mathbf{t})\mathbf{1}(B_{21}\mathbf{t} + \mathbf{u} \in (\mathbf{0}, \mathbf{x}])\mathrm{d}\mathbf{t}$ in $L^2(\mathbb{R}^2)$ or

$$\|\tilde{g}_{\lambda,1} - h_{21}\| \to 0 \quad \text{and} \quad \|\tilde{g}_{\lambda,0}\| \to 0, \tag{59}$$

where $\tilde{g}_{\lambda,j}(\mathbf{u}) = \int_{\mathbb{R}^2} \tilde{b}_{\lambda,j}(\mathbf{u}, \mathbf{t})\mathrm{d}\mathbf{t}$ and

$$\tilde{b}_{\lambda,0}(\mathbf{u}, \mathbf{t}) := \tilde{b}_\lambda(\mathbf{u}, \mathbf{t})\mathbf{1}(|t_2| \geq \lambda^{\gamma(1-\frac{q_1}{q_2})}),$$
$$\tilde{b}_{\lambda,1}(\mathbf{u}, \mathbf{t}) := \tilde{b}_\lambda(\mathbf{u}, \mathbf{t})\mathbf{1}(|t_2| < \lambda^{\gamma(1-\frac{q_1}{q_2})}), \quad \mathbf{u}, \mathbf{t} \in \mathbb{R}^2.$$

The proof of (59) is similar to that of (58) and omitted. This proves Case 1. The above proofs also included Case 2 of Theorem 2. Theorem 2 is proved. □

Proof of Theorem 3

Case 2 and $\gamma < \frac{q_1}{q_2}$ or $V_-^X = V_{22}$ Set $H(\gamma) = \frac{1}{2} + \gamma(\frac{3}{2} + \frac{q_2}{q_1} - q_2) = \frac{1}{2} + \gamma H_2$. The proof is similar to that in Case 1 of Theorem 2 using

$$A := \begin{bmatrix} 1 & -\frac{b_{12}}{b_{11}} \\ 0 & 1 \end{bmatrix}, \quad \Lambda := \begin{bmatrix} \lambda & 0 \\ 0 & \lambda^\gamma \end{bmatrix}, \quad \Lambda' := \begin{bmatrix} \lambda^{\gamma\frac{q_2}{q_1}} & 0 \\ 0 & \lambda^\gamma \end{bmatrix}.$$

Accordingly, $\tilde{g}_\lambda(\mathbf{u}) = \int_{\mathbb{R}^2} \tilde{b}_\lambda(\mathbf{u}, \mathbf{t})\mathrm{d}\mathbf{t}$, where

$$\tilde{b}_\lambda(\mathbf{u}, \mathbf{t}) := \lambda^{\gamma q_2} b(\lceil B^{-1}\Lambda'\mathbf{t}\rceil)\mathbf{1}(\lceil B^{-1}\Lambda'\mathbf{t}\rceil + \lceil \Lambda\mathbf{u}\rceil \in (\mathbf{0}, \lfloor \Lambda\mathbf{x}\rfloor])$$
$$\to a_\infty(\mathbf{t})\mathbf{1}(B_{22}\mathbf{t} + \mathbf{u} \in (\mathbf{0}, \mathbf{x}])$$

for all $\mathbf{u}, \mathbf{t} \in \mathbb{R}^2$ such that $\mathbf{t} \neq \mathbf{0}$, $u_1 \notin \{0, x_1\}$, $b_{11}t_2 + u_2 \notin \{0, x_2\}$. It suffices to prove $\tilde{g}_\lambda(\mathbf{u}) \to h_{22}(\mathbf{u}) := \int_{\mathbb{R}^2} a_\infty(\mathbf{t})\mathbf{1}(B_{22}\mathbf{t} + \mathbf{u} \in (\mathbf{0}, \mathbf{x}])\mathrm{d}\mathbf{t}$ in $L^2(\mathbb{R}^2)$ or

$$\|\tilde{g}_{\lambda,1} - h_{22}\| \to 0 \quad \text{and} \quad \|\tilde{g}_{\lambda,0}\| \to 0, \tag{60}$$

where $\tilde{g}_{\lambda,j}(\mathbf{u}) = \int_{\mathbb{R}^2} \tilde{b}_{\lambda,j}(\mathbf{u}, \mathbf{t})\mathrm{d}\mathbf{t}$ and

$$\tilde{b}_{\lambda,0}(\mathbf{u}, \mathbf{t}) := \tilde{b}_\lambda(\mathbf{u}, \mathbf{t})\mathbf{1}(|t_1| \geq \lambda^{\gamma\frac{q_2}{q_1}-1}),$$
$$\tilde{b}_{\lambda,1}(\mathbf{u}, \mathbf{t}) := \tilde{b}_\lambda(\mathbf{u}, \mathbf{t})\mathbf{1}(|t_1| < \lambda^{\gamma\frac{q_2}{q_1}-1}), \quad \mathbf{u}, \mathbf{t} \in \mathbb{R}^2.$$

The proof of (60) using $\tilde{Q}_2 < 1$ is similar to that of (58) and omitted. The remaining cases of Theorem 3 follow from Theorem 2, thereby completing the proof of Theorem 3. □

Proof of Theorem 4

Case (i) and $\gamma > 1$ **or** $V_+^X = \tilde{V}_{02}$ Set $H(\gamma) = 1 + \gamma\tilde{H}$. We follow the proof of Theorem 1, Case 1, $\gamma > 1$. Let $\Lambda := \mathrm{diag}(\lambda^\gamma, \lambda^\gamma)$, $\Lambda' := \mathrm{diag}(\lambda, \lambda^\gamma)$. Then (49) and (51) hold with $q_2 = q$ and $\tilde{B}_{22}$ replaced by $\tilde{B}_{02}$. Then the convergence $\tilde{g}_\lambda(\mathbf{u}) \to \tilde{h}_{02}(\mathbf{u}) := \int_{(\mathbf{0},\mathbf{x}]} a_\infty(\tilde{B}_{02}\mathbf{t} - \mathbf{u})\mathrm{d}\mathbf{t}$ in $L^2(\mathbb{R}^2)$ follows similarly by Pratt's lemma with (53) replaced by

$$\begin{aligned}\|\tilde{G}_\lambda\|^2 &= \int_{\mathbb{R}^4} (\rho^{-1} \star \rho^{-1})(\lambda^{1-\gamma}\frac{\det(B)}{b_{22}}(t_1 - s_1) + b_{12}(t_2' - s_2'), b_{22}(t_2' - s_2')) \\ &\qquad \times \mathbf{1}(t_1 \in (0, x_1],\ -\lambda^{1-\gamma}\frac{b_{21}}{b_{22}}t_1 + t_2' \in (0, x_2]) \\ &\qquad \times \mathbf{1}(s_1 \in (0, x_1],\ -\lambda^{1-\gamma}\frac{b_{21}}{b_{22}}s_1 + s_2' \in (0, x_2])\mathrm{d}t_1\mathrm{d}t_2'\mathrm{d}s_1\mathrm{d}s_2' \\ &\to \int_{(\mathbf{0},\mathbf{x}]\times(\mathbf{0},\mathbf{x}]} (\rho^{-1} \star \rho^{-1})(\tilde{B}_{22}(\mathbf{t} - \mathbf{s}))\mathrm{d}\mathbf{t}\mathrm{d}\mathbf{s} = \|\tilde{G}\|^2.\end{aligned}$$

The proof in Case (i), $\gamma = 1$ and $\gamma < 1$ is similar to that of Theorem 1, Case 1 and is omitted.

Case (ii) and $\gamma > 1$ **or** $V_+^X = V_{10}$ Set $H(\gamma) = H + \frac{\gamma}{2}$. We follow the proof of Theorem 2, Case 1, $\gamma > 1$, with $\Lambda' := \mathrm{diag}(\lambda, \lambda)$ and A, Λ as in (55). Then $\Lambda^{-1}B^{-1}\Lambda' \to B_{10}$ and the result follows from $\|\tilde{g}_\lambda - h_{10}\| \to 0$, where $h_{10}(\mathbf{u}) := \int_{\mathbb{R}^2} a_\infty(\mathbf{t})\mathbf{1}(B_{10}\mathbf{t} + \mathbf{u} \in (\mathbf{0}, \mathbf{x}])\mathrm{d}\mathbf{t}$. Following the proof of (57), we decompose $\tilde{g}_\lambda(\mathbf{u}) = \tilde{g}_{\lambda,0}(\mathbf{u}) + \tilde{g}_{\lambda,1}(\mathbf{u})$ with $\tilde{g}_{\lambda,j}(\mathbf{u}) = \int_{\mathbb{R}^2} \tilde{b}_{\lambda,j}(\mathbf{u}, \mathbf{t})\mathrm{d}\mathbf{t}$ given by

$$\tilde{b}_{\lambda,0}(\mathbf{u}, \mathbf{t}) := \tilde{b}_\lambda(\mathbf{u}, \mathbf{t})\mathbf{1}(|t_2| \ge \tilde{C}), \quad \tilde{b}_{\lambda,1}(\mathbf{u}, \mathbf{t}) := \tilde{b}_\lambda(\mathbf{u}, \mathbf{t})\mathbf{1}(|t_2| < \tilde{C}),$$

and $\tilde{b}_\lambda(\mathbf{u}, \mathbf{t})$, $\mathbf{u}, \mathbf{t} \in \mathbb{R}^2$, as in (56), where $\tilde{C} > 0$ is a sufficiently large constant. It suffices to prove

$$\lim_{\tilde{C}\to\infty} \limsup_{\lambda\to\infty} \|\tilde{g}_{\lambda,1} - h_{10}\| = 0 \quad \text{and} \quad \lim_{\tilde{C}\to\infty} \limsup_{\lambda\to\infty} \|\tilde{g}_{\lambda,0}\| = 0. \tag{61}$$

The proof of (61) mimics that of (59) and we omit the details. The remaining statements in Theorem 4, Case (ii) also follow similarly to the proof of Theorem 2, Case 1. Theorem 4 is proved. □

4 Appendix

4.1 Generalized Homogeneous Functions

Let $q_i > 0$, $i = 1, 2$, with $Q := \frac{1}{q_1} + \frac{1}{q_2}$.

Definition 1 A measurable function $h : \mathbb{R}^2 \to \mathbb{R}$ is said to be:

(i) *generalized homogeneous* if for all $\lambda > 0$,

$$\lambda h(\lambda^{1/q_1} t_1, \lambda^{1/q_2} t_2) = h(\mathbf{t}), \quad \mathbf{t} \in \mathbb{R}^2_0. \tag{62}$$

(ii) *generalized invariant* if $\lambda \mapsto h(\lambda^{1/q_1} t_1, \lambda^{1/q_2} t_2)$ is a constant function on $\mathbb{R}_+$ for any $\mathbf{t} \in \mathbb{R}^2_0$.

Proposition 3 *Any generalized homogeneous function h can be represented as*

$$h(\mathbf{t}) = L(\mathbf{t})/\rho(\mathbf{t}), \quad \mathbf{t} \in \mathbb{R}^2_0, \tag{63}$$

where $\rho(\mathbf{t}) = |t_1|^{q_1} + |t_2|^{q_2}$, $\mathbf{t} \in \mathbb{R}^2$, *and L is a generalized invariant function. Moreover, L can be written as*

$$L(\mathbf{t}) = L_{\mathrm{sign}(t_2)}(t_1/\rho(\mathbf{t})^{1/q_1}), \quad t_2 \neq 0, \tag{64}$$

where $L_\pm(z) := h(z, \pm(1 - |z|^{q_1})^{1/q_2})$, $z \in [-1, 1]$.

Proof Equation (63) follows from (62), by taking $\lambda = 1/\rho(\mathbf{t})$. Then $L(\mathbf{t}) := h(t_1/\rho(\mathbf{t})^{1/q_1}, t_2/\rho(\mathbf{t})^{1/q_2})$ is a generalized invariant function. Whence, (64) follows since $t_2/\rho(\mathbf{t})^{1/q_2} = \mathrm{sign}(t_2)(1 - |t_1/\rho(\mathbf{t})^{1/q_1}|^{q_1})^{1/q_2}$. □

The notion of generalized homogeneous function was introduced in [9]. The last paper also obtained a representation of such functions different from (63). Note $\mathbf{t} \mapsto (t_1, \rho(\mathbf{t})^{1/q_1})$ is a 1-1 transformation of the upper half-plane $\{\mathbf{t} \in \mathbb{R}^2 : t_2 \geq 0\}$ onto itself. Following [5], the form in (63) will be called the *polar representation of h*. The two factors in (63), viz., ρ^{-1} and L are called the *radial* and *angular* functions, respectively. Note that h being strictly positive and continuous on $\mathbb{R}^2_0$ is equivalent to $L_\pm$ both being strictly positive and continuous on $[-1, 1]$ with $L_+(\pm 1) = L_-(\pm 1)$.

4.2 Dependence Axis

Definition 2 Let $g : \mathbb{R}^2 \to \mathbb{R}$ be a measurable function. We say that a line passing through the origin and given by $\{\mathbf{t} \in \mathbb{R}^2 : \mathbf{a} \cdot \mathbf{t} = 0\}$ with $\mathbf{a} \in \mathbb{R}^2_0$ is the *dependence axis of g* if for all $\mathbf{c} \in \mathbb{R}^2_0$ such that $a_1 c_2 \neq c_1 a_2$,

$$\liminf_{|\mathbf{t}| \to \infty,\, \mathbf{c} \cdot \mathbf{t} = 0} \frac{\log(1/|g(\mathbf{t})|)}{\log |\mathbf{t}|} > \limsup_{|\mathbf{t}| \to \infty,\, \mathbf{a} \cdot \mathbf{t} = 0} \frac{\log(1/|g(\mathbf{t})|)}{\log |\mathbf{t}|}. \tag{65}$$

We say that a line $\{\mathbf{t} \in \mathbb{R}^2 : \mathbf{a} \cdot \mathbf{t} = 0\}$ with $\mathbf{a} \in \mathbb{R}^2_0$ is the *dependence axis of* $g : \mathbb{Z}^2 \to \mathbb{R}$ if this line is the dependence axis of $g(\lfloor \mathbf{t} \rfloor)$, $\mathbf{t} \in \mathbb{R}^2$.

Proposition 4 *Let* $g : \mathbb{Z}^2 \to \mathbb{R}$ *satisfy*

$$g(\mathbf{t}) = \rho(B\mathbf{t})^{-1}(L(B\mathbf{t}) + o(1)), \quad |\mathbf{t}| \to \infty, \tag{66}$$

where $B = (b_{ij})_{i,j=1,2}$ *is a* 2×2 *nondegenerate matrix,* $\rho(\mathbf{t}) := |t_1|^{q_1} + |t_2|^{q_2}$, $\mathbf{t} \in \mathbb{R}^2$*, with* $q_i > 0$*,* $i = 1, 2$*, and* $L : \mathbb{R}^2_0 \to \mathbb{R}$ *satisfies Assumption B. In addition,*

(i) *let* $q_1 < q_2$ *and* $|L_+(1)| = |L_-(1)| > 0$*,* $|L_+(-1)| = |L_-(-1)| > 0$*. Then the dependence axis of* g *is* $\{\mathbf{t} \in \mathbb{R}^2 : \mathbf{b}_2 \cdot \mathbf{t} = 0\}$ *with* $\mathbf{b}_2 = (b_{21}, b_{22})^\top$*;*

(ii) *let* $q_1 > q_2$ *and* $|L_+(0)| > 0$*,* $|L_-(0)| > 0$*. Then the dependence axis of* g *is* $\{\mathbf{t} \in \mathbb{R}^2 : \mathbf{b}_1 \cdot \mathbf{t} = 0\}$ *with* $\mathbf{b}_1 = (b_{11}, b_{12})^\top$.

Proof It suffices to show part (i) only since (ii) is analogous. Below we prove that

$$\lim_{|\mathbf{t}|\to\infty,\, \mathbf{b}_2\cdot\mathbf{t}=0} |\mathbf{t}|^{q_1}|g(\lfloor\mathbf{t}\rfloor)| = \frac{|\mathbf{b}_2|^{q_1}}{|\det(B)|^{q_1}} \begin{cases} L_\pm(1), & \mathbf{b}_1\cdot\mathbf{t} \to +\infty, \\ L_\pm(-1), & \mathbf{b}_1\cdot\mathbf{t} \to -\infty, \end{cases} \tag{67}$$

$$\limsup_{|\mathbf{t}|\to\infty,\, \mathbf{c}\cdot\mathbf{t}=0} |\mathbf{t}|^{q_2}|g(\lfloor\mathbf{t}\rfloor)| < \infty, \quad \forall \mathbf{c} \in \mathbb{R}^2_0,\ b_{21}c_2 \neq c_1 b_{22}. \tag{68}$$

Note (67) implies $\lim_{|\mathbf{t}|\to\infty,\, \mathbf{b}_2\cdot\mathbf{t}=0} \frac{\log(1/|g(\lfloor\mathbf{t}\rfloor)|)}{\log|\mathbf{t}|} = q_1$ while (68) implies $\liminf_{|\mathbf{t}|\to\infty,\, \mathbf{c}\cdot\mathbf{t}=0} \frac{\log(1/|g(\lfloor\mathbf{t}\rfloor)|)}{\log|\mathbf{t}|} \geq q_2$, hence the statement of the proposition.

Let us prove (67). We have

$$\frac{\mathbf{b}_1\cdot\mathbf{t}}{\rho(B\lfloor\mathbf{t}\rfloor)^{1/q_1}} = \frac{\mathrm{sign}(\mathbf{b}_1\cdot\mathbf{t})}{(|\mathbf{b}_1\cdot\lfloor\mathbf{t}\rfloor/|\mathbf{b}_1\cdot\mathbf{t}||^{q_1} + |\mathbf{b}_2\cdot\lfloor\mathbf{t}\rfloor|^{q_2}/|\mathbf{b}_1\cdot\mathbf{t}|^{q_1})^{1/q_1}} \to \pm 1, \quad \mathbf{b}_1\cdot\mathbf{t} \to \pm\infty$$

since $|\mathbf{b}_2\cdot\lfloor\mathbf{t}\rfloor| = O(1)$ on $\mathbf{b}_2\cdot\mathbf{t}=0$. In a similar way, $\lim_{|\mathbf{t}|\to\infty,\, \mathbf{b}_2\cdot\mathbf{t}=0} |\mathbf{t}|^{q_1}\rho(B\lfloor\mathbf{t}\rfloor)^{-1} = (\frac{|b_{21}|+|b_{22}|}{|\det(B)|})^{q_1}$. Whence, (67) follows by the asymptotic form of g and the assumption of the continuity of $L_\pm$.

Consider (68). In view of (66) and the boundedness of $L_\pm$ it suffices to show (68) for $\rho(B\mathbf{t})^{-1}$ in place of $g(\mathbf{t})$, $\mathbf{t} \in \mathbb{Z}^2$. Then $|\mathbf{t}|^{q_2}\rho(B\lfloor\mathbf{t}\rfloor)^{-1} = (\frac{|\mathbf{b}_1\cdot\lfloor\mathbf{t}\rfloor|^{q_1}}{|\mathbf{t}|^{q_2}} + \frac{|\mathbf{b}_2\cdot\lfloor\mathbf{t}\rfloor|^{q_2}}{|\mathbf{t}|^{q_2}})^{-1}$, where $\frac{|\mathbf{b}_1\cdot\lfloor\mathbf{t}\rfloor|^{q_1}}{|\mathbf{t}|^{q_2}} \to 0$ and $\frac{|\mathbf{b}_2\cdot\lfloor\mathbf{t}\rfloor|}{|\mathbf{t}|} \to \frac{|b_{21}c_2 - b_{22}c_1|}{|c_1|+|c_2|} > 0$, proving (68). □

Below, we show that the dependence axis is preserved under 'discrete' convolution $[g_1 \star g_2](\mathbf{t}) := \sum_{\mathbf{u}\in\mathbb{Z}^2} g_1(\mathbf{u})g_2(\mathbf{u}+\mathbf{t})$, $\mathbf{t} \in \mathbb{Z}^2$, of two functions $g_i : \mathbb{Z}^2 \to \mathbb{R}$, $i = 1, 2$.

Proposition 5 *For* $i = 1, 2$*, let* $g_i : \mathbb{Z}^2 \to \mathbb{R}$ *satisfy*

$$g_i(\mathbf{t}) = \rho(B\mathbf{t})^{-1}(L_i(B\mathbf{t}) + o(1)), \quad |\mathbf{t}| \to \infty, \tag{69}$$

where B is a 2×2 nondegenerate matrix, ρ with $Q = q_1^{-1} + q_2^{-1} \in (1, 2)$ and L_i are functions as in Assumption B. For $i = 1, 2$, let $a_{\infty,i}(\mathbf{t}) := \rho(\mathbf{t})^{-1} L_i(\mathbf{t})$, $\mathbf{t} \in \mathbb{R}^2_0$. Then

$$[g_1 \star g_2](\mathbf{t}) = \tilde{\rho}(B\mathbf{t})^{-1}(\tilde{L}(B\mathbf{t}) + o(1)), \quad |\mathbf{t}| \to \infty, \tag{70}$$

where $\tilde{\rho}(\mathbf{t}) := |t_1|^{\tilde{q}_1} + |t_2|^{\tilde{q}_2}$, $\mathbf{t} \in \mathbb{R}^2$, with $\tilde{q}_i := q_i(2 - Q)$, $i = 1, 2$, and

$$\tilde{L}(\mathbf{t}) := |\det(B)|^{-1}(a_{\infty,1} \star a_{\infty,2})(t_1/\tilde{\rho}(\mathbf{t})^{1/\tilde{q}_1}, t_2/\tilde{\rho}(\mathbf{t})^{1/\tilde{q}_2}), \quad \mathbf{t} \in \mathbb{R}^2_0, \tag{71}$$

is a generalized invariant function in the sense of Definition 1 (ii) (with q_i replaced by $\tilde{q}_i$, $i = 1, 2$). Moreover, if $L_1 = L_2 \geq 0$ then $\tilde{L}$ is strictly positive.

Proof We follow the proof in [14, Prop. 5.1 (iii)]. For $\mathbf{t} \in \mathbb{Z}^2$, split every $g_i(\mathbf{t})$ as a sum of $g_i^1(\mathbf{t}) := g_i(\mathbf{t}) - g_i^0(\mathbf{t})$ and $g_i^0(\mathbf{t}) := (1 \vee \rho(B\mathbf{t}))^{-1} L_i(B\mathbf{t})$ using the convention $g_i^0(\mathbf{0}) = L_i(\mathbf{0}) := 0$. Then $[g_1 \star g_2](\mathbf{t}) = \sum_{k,j=0}^1 [g_1^k \star g_2^j](\mathbf{t})$ and (70) follows from

$$\lim_{|\mathbf{t}| \to \infty} \left| \tilde{\rho}(B\mathbf{t})[g_1^0 \star g_2^0](\mathbf{t}) - \tilde{L}(B\mathbf{t}) \right| = 0 \tag{72}$$

and

$$\tilde{\rho}(B\mathbf{t})[g_1^k \star g_2^j](\mathbf{t}) = o(1), \quad |\mathbf{t}| \to \infty, \quad (k, j) \neq (0, 0). \tag{73}$$

To prove (72) we write the 'discrete' convolution as integral $[g_1^0 \star g_2^0](\mathbf{t}) = \int_{\mathbb{R}^2} g_1^0(\lceil \mathbf{u} \rceil) \times g_2^0(\lceil \mathbf{u} \rceil + \mathbf{t}) \mathrm{d}\mathbf{u}$, where we change a variable: $\mathbf{u} \to B^{-1} R_{\tilde{\varrho}} \mathbf{u}$ with

$$\mathbf{t}' := B\mathbf{t}, \quad \tilde{\varrho} := \tilde{\rho}(\mathbf{t}'), \quad R_{\tilde{\varrho}} := \mathrm{diag}(\tilde{\varrho}^{1/\tilde{q}_1}, \tilde{\varrho}^{1/\tilde{q}_2}).$$

Then with $\tilde{Q} := \tilde{q}_1^{-1} + \tilde{q}_2^{-1}$ we have

$$\begin{aligned}
&\tilde{\rho}(B\mathbf{t})[g_1^0 \star g_2^0](\mathbf{t}) \\
&\quad = |\det(B)|^{-1} \tilde{\varrho}^{1+\tilde{Q}} \int_{\mathbb{R}^2} \frac{L_1(B\lceil B^{-1} R_{\tilde{\varrho}} \mathbf{u} \rceil)}{\rho(B\lceil B^{-1} R_{\tilde{\varrho}} \mathbf{u} \rceil) \vee 1} \times \frac{L_2(B\lceil B^{-1} R_{\tilde{\varrho}} \mathbf{u} \rceil + \mathbf{t}')}{\rho(B\lceil B^{-1} R_{\tilde{\varrho}} \mathbf{u} \rceil + \mathbf{t}') \vee 1} \mathrm{d}\mathbf{u} \\
&\quad = |\det(B)|^{-1} \int_{\mathbb{R}^2} g_{\tilde{\varrho},\mathbf{z}}(\mathbf{u}) \mathrm{d}\mathbf{u}, \quad \mathbf{z} = R_{\tilde{\varrho}}^{-1} \mathbf{t}',
\end{aligned}$$

where for all $\tilde{\rho} > 0$, $\mathbf{z} \in \mathbb{R}^2$ such that $\tilde{\rho}(\mathbf{z}) = 1$, $\mathbf{u} \in \mathbb{R}^2$,

$$g_{\tilde{\varrho},\mathbf{z}}(\mathbf{u}) := \frac{L_1(R_{\tilde{\varrho}}^{-1} B\lceil B^{-1} R_{\tilde{\varrho}} \mathbf{u} \rceil)}{\rho(R_{\tilde{\varrho}}^{-1} B\lceil B^{-1} R_{\tilde{\varrho}} \mathbf{u} \rceil) \vee \tilde{\varrho}^{-q_1/\tilde{q}_1}} \times \frac{L_2(R_{\tilde{\varrho}}^{-1} B\lceil B^{-1} R_{\tilde{\varrho}} \mathbf{u} \rceil + \mathbf{z})}{\rho(R_{\tilde{\varrho}}^{-1} B\lceil B^{-1} R_{\tilde{\varrho}} \mathbf{u} \rceil + \mathbf{z}) \vee \tilde{\varrho}^{-q_1/\tilde{q}_1}}$$

and we used generalized homogeneous and generalized invariance properties of ρ and L_i, $i = 1, 2$, and the facts that $q_1/\tilde{q}_1 = q_2/\tilde{q}_2$, $1 + \tilde{Q} = 2q_1/\tilde{q}_1 = 2/(2 - Q)$. Whence using continuity of ρ and L_i, $i = 1, 2$, it follows that $g_{\tilde{\varrho},\mathbf{z}}(\mathbf{u}) - a_{\infty,1}(\mathbf{u}) \times a_{\infty,2}(\mathbf{u} + \mathbf{z}) \to 0$ as $\tilde{\varrho} \to \infty$ or $|\mathbf{t}| \to \infty$ for all $\mathbf{u} \in \mathbb{R}^2_0$, $\mathbf{u} + \mathbf{z} \in \mathbb{R}^2_0$. Then similarly as in [14, (7.8)] we conclude that $\sup_{\mathbf{z}\in\mathbb{R}^2:\rho(\mathbf{z})=1} |\int_{\mathbb{R}^2} g_{\tilde{\varrho},\mathbf{z}}(\mathbf{u})\mathrm{d}\mathbf{u} - (a_{\infty,1} \star a_{\infty,2})(\mathbf{z})| \to 0$, $\tilde{\varrho} \to \infty$, and (72) holds. The remaining details including the proof of (73) are similar to those in [14]. Proposition 5 is proved. □

Corollary 1 *Let X be a linear RF on $\mathbb{Z}^2$ satisfying Assumptions A, B and having a covariance function $r_X(\mathbf{t}) := \mathrm{E}X(\mathbf{0})X(\mathbf{t}) = [b \star b](\mathbf{t})$, $\mathbf{t} \in \mathbb{Z}^2$. Then*

$$r_X(\mathbf{t}) = \tilde{\rho}(B\mathbf{t})^{-1}(\tilde{L}(B\mathbf{t}) + o(1)), \quad |\mathbf{t}| \to \infty, \tag{74}$$

where $\tilde{\rho}$, $\tilde{L}$ are as in (70), (71) (with $a_{\infty,1} = a_{\infty,2} = a_\infty$ of (16)). Particularly, if $q_1 \neq q_2$ and $L_\pm$ satisfy the conditions in Proposition 4, the dependence axes of the covariance function r_X in (74) and the moving-average coefficients b in (8) coincide.

Acknowledgments We are grateful to an anonymous referee for useful comments. The authors thank Shanghai New York University for hosting their visits in April–May, 2019 during which this work was initiated and partially completed. Vytautė Pilipauskaitė acknowledges the financial support from the project "Ambit fields: probabilistic properties and statistical inference" funded by Villum Fonden.

References

1. Beran, J., Feng, Y., Gosh, S., Kulik, R.: Long-Memory Processes: Probabilistic Properties and Statistical Methods. Springer, New York (2013)
2. Damarackas, J., Paulauskas, V.: Spectral covariance and limit theorems for random fields with infinite variance. J. Multiv. Anal. **153**, 156–175 (2017)
3. Damarackas, J., Paulauskas, V.: Some remarks on scaling transition in limit theorems for random fields (2019). Preprint. arXiv:1903.09399 [math.PR]
4. Davydov, Y.A.: The invariance principle for stationary processes. Theor. Probab. Appl. **15**, 487–498 (1970)
5. Dobrushin, R.L.: Gaussian and their subordinated self-similar random generalized fields. Ann. Probab. **7**, 1–28 (1979)
6. Dobrushin, R.L., Major, P.: Non-central limit theorems for non-linear functionals of Gaussian fields. Probab. Theory Relat. Fields **50**, 27–52 (1979)
7. Doukhan, P., Oppenheim, G., Taqqu, M.S. (eds.): Theory and Applications of Long-Range Dependence. Birkhäuser, Boston (2003)
8. Giraitis, L., Koul, H.L., Surgailis, D.: Large Sample Inference for Long Memory Processes. Imperial College Press, London (2012)
9. Hankey, A., Stanley, H.E.: Systematic application of generalized homogeneous functions to static scaling, dynamic scaling, and universality. Phys. Rev. B **6**, 3515–3542 (1972)
10. Lahiri, S.N., Robinson, P.M.: Central limit theorems for long range dependent spatial linear processes. Bernoulli **22**, 345–375 (2016)
11. Leonenko, N.N.: Random Fields with Singular Spectrum. Kluwer, Dordrecht (1999)

12. Pilipauskaitė, V., Surgailis, D.: Joint temporal and contemporaneous aggregation of random-coefficient AR(1) processes. Stoch. Process. Appl. **124**, 1011–1035 (2014)
13. Pilipauskaitė, V., Surgailis, D.: Anisotropic scaling of random grain model with application to network traffic. J. Appl. Probab. **53**, 857–879 (2016)
14. Pilipauskaitė, V., Surgailis, D.: Scaling transition for nonlinear random fields with long-range dependence. Stoch. Process. Appl. **127**, 2751–2779 (2017)
15. Pratt, J.W.: On interchanging limits and integrals. Ann. Math. Stat. **31**, 74–77 (1960)
16. Puplinskaitė, D., Surgailis, D.: Scaling transition for long-range dependent Gaussian random fields. Stoch. Process. Appl. **125**, 2256–2271 (2015)
17. Puplinskaitė, D., Surgailis, D.: Aggregation of autoregressive random fields and anisotropic long-range dependence. Bernoulli **22**, 2401–2441 (2016)
18. Samorodnitsky, G., Taqqu, M.S.: Stable Non-Gaussian Random Processes. Chapman and Hall, London (1994)
19. Surgailis, D.: Anisotropic scaling limits of long-range dependent linear random fields on $\mathbb{Z}^3$. J. Math. Anal. Appl. **472**, 328–351 (2019)
20. Surgailis, D.: Anisotropic scaling limits of long-range dependent random fields. Lith. Math. J. **59**, 595–615 (2019)
21. Surgailis, D.: Scaling transition and edge effects for negatively dependent linear random fields on $\mathbb{Z}^2$. Stoch. Process. Appl. **130**, 7518–7546 (2020)

Brownian Aspects of the KPZ Fixed Point

Leandro P. R. Pimentel

Abstract The Kardar-Parisi-Zhang (KPZ) fixed point is a Markov process that is conjectured to be at the core of the KPZ universality class. In this article we study two aspects the KPZ fixed point that share the same Brownian limiting behaviour: the local space regularity and the long time evolution. Most of the results that we will present here were obtained by either applying explicit formulas for the transition probabilities or applying the coupling method to discrete approximations. Instead we will use the variational description of the KPZ fixed point, allowing us the possibility of running the process starting from different initial data (basic coupling), to prove directly the aforementioned limiting behaviours.

Keywords Random growth · Kardar-Parisi-Zhang fixed point · Brownian motion

1 Introduction and Main Results

The universality class concept is an instrument of modern statistical mechanics that systemizes the idea that there are but a few important characteristics that determine the scaling behaviour of a stochastic model. In $d + 1$ stochastic growth models the object of interest is a height function $h(x, t)$ over the d-dimensional substrate $x \in \mathbb{R}^d$ at time $t \geq 0$, whose evolution is described by a random mechanism. For fairly general models one has a deterministic macroscopic shape for the height function and its fluctuations, under proper space and time scaling, are expected to be characterized by a universal distribution. A well known example is given by the random deposition growth model, where blocks are pilled in columns (indexed by $x \in \mathbb{Z}$) according to independent Poisson processes. The existence of a macroscopic shape follows from the law of large numbers and, due to the classical central limit theorem, the height function at $x \in \mathbb{R}$ has Gaussian fluctuations that are independent

L. P. R. Pimentel (✉)
Instituto de Matemática, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil
e-mail: leandro@im.ufrj.br

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_29

in space. In 1986 [18], Kardar, Parisi and Zhang (KPZ) proposed a stochastic partial differential equation (the KPZ equation) for a growth model where a non-linear local slope dependent rate is added to a diffusion equation with additive noise: $\partial_t h = \frac{1}{2}(\partial_x h)^2 + \partial_x^2 h + \xi$. In opposition to the previous random deposition growth model, they predicted that for $d = 1$ the solution of the KPZ equation at time nt has fluctuations of order $n^{1/3}$, and on a scale of $n^{2/3}$ that non-trivial spatial correlation is achieved (KPZ scaling exponents). Since then it is expected that $1 + 1$ interface growth models that exhibit a similar KPZ growth mechanism would satisfy

$$h(an^{2/3}x, nt) \sim bnt + cn^{1/3}\mathfrak{h}_t(x)\,,$$

for some constants $a, b, c \in \mathbb{R} \setminus \{0\}$ that might depend on the microscopic distributional details of the model, but where $\mathfrak{h}_t(x)$ is a universal space-time process called the KPZ fixed point [11]. Illustrations of natural phenomena within the KPZ universality class include turbulent liquid crystals, bacteria colony growth and paper wetting [27]. For a more complete introduction to the KPZ equation and universality class, and its relation with other discrete growth models in statistical physics, the author address to [9].

After [18], the study of KPZ fluctuations became a famous subject in the literature of physics and mathematics and, in the late nineties, a breakthrough was presented by Baik, Deift and Johansson [2, 16]. By applying an exact formula (in terms of a Toeplitz determinant) for the Hammersley last-passage percolation growth model with narrow wedge initial profile, and then by analysing asymptotics of the resulting expressions, they were able to prove convergence of shape fluctuations at $x = 0$ to the Tracy-Widom (GUE) distribution. In the past 20 years there has been a significant amount of improvements of the theory and the exact statistics for some special initial conditions, resulting in different types of limiting distributions, were computed using connections with integrable probability [1, 6, 17, 25]. Recently, a unifying approach was developed by Matetski et al. [19] in the totally asymmetric simple exclusion process (TASEP) context that conducted to a rigorous construction of the Markov process $(\mathfrak{h}_t\,,\ t \geq 0)$ and the explicit computation of its transition probabilities.

Alongside the rich structure of integrable probability, the study of the KPZ universality class was also developed by techniques based on the graphical representation of an interacting particle system due to Harris [15]. There are many advantages of this approach, also known as the coupling method, comprising the possibility of running the process starting from different initial data on the same probability space. In the seminal paper by Cator and Groeneboom [7], the authors applied the coupling method to derive the KPZ scaling exponents (1/3 and 2/3) for the Hammersley last-passage percolation growth model. This method was further developed in the TASEP context by Balázs et al. [4], and became a successful tool to analyse fluctuations of models [3, 5, 26] lying within the KPZ universality class, and local properties of different types of of Airy processes [8, 14, 23]. Related to that, there has been considerable developments in describing the KPZ fixed point for fixed $x \in \mathbb{R}$ and $t > 0$ in terms of a variational formula [11, 12, 14, 19].

The full variational space-time picture of the KPZ fixed point in terms of the directed landscape was constructed by Dauvergne et al. [13], which relies on the existence and uniqueness of a two-dimensional random scalar field, called the Airy sheet. In analogy with Harris graphical representation, the directed landscape allows one to run simultaneously the process starting from different initial data on the same probability space (basic coupling). Thereby, it seems natural to expect that particle systems techniques that were applied to discrete approximations of the KPZ fixed point [23, 24] can be developed in the continuous space-time context itself. In the course of this article we prove Brownian behaviour of the KPZ fixed point (Theorems 1–3) by using soft arguments based on geometrical aspects of the directed landscape.

1.1 The Airy Sheet and the Directed Landscape

The construction of the directed landscape is based on the existence and uniqueness of the so called Airy Sheet, which in turn is defined through a last-passage percolation model over the Gibbsian Airy line ensemble [10, 13, 25]. For a sequence of differentiable functions $\mathbf{F} = (\dots, \mathbf{F}_{-1}, \mathbf{F}_0, \mathbf{F}_1, \dots)$ with domain $\mathbb{R}$, and coordinates $x \le y$ and $n \le m$, define the last-passage percolation time

$$\mathbf{F}\left((x,m) \to (y,n)\right) := \sup_{\pi} \int_x^y \mathbf{F}'_{\pi(t)}(t)dt \,,$$

where the supremum is over nonincreasing functions $\pi : [x, y] \to \mathbb{Z}$ with $\pi(x) = m$ and $\pi(y) = n$. Notice that, for such paths, the integral is just the sum of the increments of $\mathbf{F}$ (over each line), so the same can be defined for continuous $\mathbf{F}$. An important example is given by setting $\mathbf{F} \equiv \mathbf{B}$ a sequence of independent standard two-sided Brownian motions (Brownian last-passage percolation). In the literature of last-passage percolation it is normally considered maximization over nondecreasing paths instead, but to accommodate the natural order of the Airy line ensemble from top to bottom (as below), Dauvergne et al. [13] defined it for nonincreasing paths.

The Airy line ensemble [10, 25] is a random sequence of ordered real functions $\mathbf{L}_1 > \mathbf{L}_2 > \dots$ with domain $\mathbb{R}$. The function $\mathbf{L}_n(x) + x^2$ is stationary for all $n \ge 1$, and the top line $\mathbf{L}_1(x) + x^2$ is known as the Airy$_2$ process and represents the limit fluctuations of some integrable last-passage percolation models, including the Brownian one.

Definition 1 The stationary Airy sheet is a random continuous function $\mathscr{A} : \mathbb{R}^2 \to \mathbb{R}$ such that:

- $\mathscr{A} \stackrel{dist.}{=} T_{(z,w)}\mathscr{A}$ for all $(z, w) \in \mathbb{R}^2$, where $T_{(z,w)}\mathfrak{f}(x, y) := \mathfrak{f}(x + z, y + w)$.
- $\mathscr{A}$ can be coupled with the Airy line ensemble so that

$$(\mathscr{A}(0,x)\,,\; x \in \mathbb{R}) \stackrel{dist.}{=} \left(\mathbf{L}_1(x) + x^2\,,\; x \in \mathbb{R}\right)\,,$$

and for all $(x, y, z) \in \mathbb{Q}^+ \times \mathbb{Q}^2$ almost surely there exists a random $K_{x,y,z}$ such that for all $k \geq K_{x,y,z}$ we have

$$\mathbf{L}\left((-\sqrt{k/2x}, k) \to (z, 1)\right) - \mathbf{L}\left((-\sqrt{k/2x}, k) \to (y, 1)\right) = \mathscr{L}(x, z) - \mathscr{L}(x, y)\,,$$

where

$$\mathscr{L}(x, y) := \mathscr{A}(x, y) - (x - y)^2\,. \tag{1}$$

In [13] the authors have a similar definition for the Airy sheet $\mathscr{L}(x, y)$ (see Definition 1.2[13] and notice that they used different notation to represent the Airy line ensemble and the Airy sheet), but it follows from their results (Remark 1.1 and Theorem 1.3 [13]) that the stationary Airy sheet exists and is unique in law. The Airy sheet satisfies a version of the 1:2:3 scaling with respect to metric composition. For each $\gamma > 0$ let S_γ denote the diffusive scaling transform, which we will apply to real functions of one or two variables:

$$S_\gamma \mathfrak{f}(x) := \gamma^{-1}\mathfrak{f}(\gamma^2 x) \text{ and } S_\gamma \mathfrak{f}(x, y) := \gamma^{-1}\mathfrak{f}(\gamma^2 x, \gamma^2 y)\,.$$

Define the Airy sheet $\mathscr{L}_s$ of scale $s > 0$ by

$$\mathscr{L}_s(x, y) := S_{s^{-1}}\mathscr{L}(x, y) = s\mathscr{L}(x/s^2, y/s^2)\,.$$

Then

$$\mathscr{L}_r(x, y) \stackrel{dist.}{=} \max_{z\in\mathbb{R}} \left\{\mathscr{L}_s^{(1)}(x, z) + \mathscr{L}_t^{(2)}(z, y)\right\}\,, \text{ with } r^3 = s^3 + t^3\,,$$

(as random functions) where $\mathscr{L}_s^{(1)}$ and $\mathscr{L}_t^{(2)}$ are two independent copies of the Airy sheet of scales $s, t > 0$, respectively. (For the Airy sheet (1) we have a true maximum!)

To introduce the directed landscape we consider an oriented four-dimensional parameter space defined as

$$\mathbb{R}^4_\uparrow := \left\{(x, s; y, t) \in \mathbb{R}^4 \,:\, s < t\right\}\,.$$

Coordinates s and t represents time while coordinates x and y represents space. In the next we follow Definition 10.1 [13] to introduce the directed landscape. By Theorem 10.9 [13], the directed landscape exists and is unique in law.

Definition 2 The directed landscape is a random continuous function $\mathscr{L} : \mathbb{R}^4_\uparrow \to \mathbb{R}$ that satisfies the following properties.

- Airy sheets marginals: for each $t \in \mathbb{R}$ and $s > 0$ we have

$$\mathscr{L}(\cdot, t; \cdot, t+s) \overset{dist.}{=} \mathscr{L}_s(\cdot, \cdot) . \tag{2}$$

- Independent increments: if $\{(t_i, t_i + s_i) : i = 1, \ldots, k\}$ is a collection of disjoint intervals then $\{\mathscr{L}(\cdot, t_i; \cdot, t_i + s_i) : i = 1, \ldots, k\}$ is a collection of independent random functions.
- Metric composition: almost surely

$$\mathscr{L}(x, r; y, t) = \max_{z \in \mathbb{R}} \{\mathscr{L}(x, r; z, s) + \mathscr{L}(z, s; y, t)\} , \ \forall (x, s; y, t) \in \mathbb{R}^4_\uparrow \text{ and } s \in (r, t) . \tag{3}$$

Dauvergne et al. [13] showed that the directed landscape describes the full space and time scaling limit of the fluctuations of the Brownian last-passage percolation model (Theorem 1.5 [13]). By setting $(x, s)_n := (s + 2x/n^{1/3}, -\lfloor sn \rfloor)$, they proved that there exists a coupling between the directed landscape and the Brownian last-passage percolation model such that

$$\mathbf{B}^{(n)}\left((x, s)_n \to (y, t)_n\right) = 2(t-s)\sqrt{n} + 2(y-x)n^{1/6} + n^{-1/6}\left(\mathscr{L} + o_n\right)(x, s; y, t) , \tag{4}$$

where $\mathbf{B}^{(n)}(\ldots, \mathbf{B}^{(n)}_{-1}, \mathbf{B}^{(n)}_0, \mathbf{B}^{(n)}_1, \ldots)$ is a sequence of Brownian motions and o_n is a random function asymptotically small in the sense that for each compact $K \subseteq \mathbb{R}^4_\uparrow$ there exists $a > 1$ such that $\mathbb{E}\left(a^{\sup_K o_n}\right) \to 1$ as $n \to \infty$.

The directed landscape induces an evolution which takes into account the metric composition (3). The state space UC is defined below and, as our initial data, we incorporate (generalized) functions that might take value $-\infty$.

Definition 3 We say that a function $\mathfrak{f} : \mathbb{R} \to [-\infty, \infty)$ is upper semicontinuous if

$$\limsup_{x \to y} \mathfrak{f}(x) \le \mathfrak{f}(y) .$$

Let UC denote the space of upper semicontinuous generalized functions $\mathfrak{f} : \mathbb{R} \to [-\infty, \infty)$ with $\mathfrak{f}(x) \le C_1|x| + C_2$ for all $x \in \mathbb{R}$, for some $C_1, C_2 < \infty$, and $\mathfrak{f}(x) > -\infty$ for some $x \in \mathbb{R}$.

A canonical example of a (generalized) upper semicontinuous function that will be consider here several time is

$$\mathfrak{d}_x(z) = \begin{cases} 0 & \text{for } z = x \\ -\infty & \text{for } z \neq x . \end{cases} \tag{5}$$

The state space UC can be endowed with the topology of local convergence turning it into a Polish space (Section 3.1 [19]).

Proposition 1 *Let* $\mathfrak{h} \in \mathsf{UC}$*. Then a.s. for all* $0 < s < t$ *and* $x \in \mathbb{R}$ *the random function* $z \in \mathbb{R} \mapsto \mathfrak{h}(z) + \mathscr{L}(z, s; x, t)$ *attains it maximum and the process*

$$\mathfrak{h}_{s,t}(x; \mathfrak{h}) := \max_{z\in\mathbb{R}} \{\mathfrak{h}(z) + \mathscr{L}(z, s; x, t)\} , \tag{6}$$

defines a Markov process acting on UC*, i.e.*

$$\mathfrak{h}_{r,t+s}(\cdot; \mathfrak{h}) = \mathfrak{h}_{t,t+s}(\cdot; \mathfrak{h}_{r,t}) .$$

From now on we denote $\mathfrak{h}_t \equiv \mathfrak{h}_{0,t}$*.*

Proposition 1 follows from Proposition 2, which will be proved in the next section. Notice that, by independence of increments (Definition 2), $\mathfrak{h}_{r,t}(\cdot; \mathfrak{h})$ and $\mathscr{L}(\cdot, t; \cdot, t + s)$ are independent. The directed landscape can be recovered in terms of $\mathfrak{h}_{s,t}$ by choosing a proper initial condition (5):

$$\mathscr{L}(x, s; y, t) = \mathfrak{h}_{s,t}(y; \mathfrak{d}_x) . \tag{7}$$

The KPZ fixed point satisfies the 1:2:3 scaling invariance:

$$S_{\gamma^{-1}}\mathfrak{h}_{\gamma^{-3}t}(\cdot; S_\gamma \mathfrak{h}) \stackrel{dist.}{=} \mathfrak{h}_t(\cdot; \mathfrak{h}) . \tag{8}$$

Furthermore, if we set

$$\mathfrak{b} \equiv \text{ two-sided Brownian motion with diffusion coefficient } 2 , \tag{9}$$

then

$$\varDelta\mathfrak{h}_t(\cdot; \mathfrak{b}^\mu) \stackrel{dist.}{=} \mathfrak{b}^\mu(\cdot) , \quad \text{for all } t \geq 0 , \tag{10}$$

where $\varDelta\mathfrak{f}(x) := \mathfrak{f}(x) - \mathfrak{f}(0)$ and $\mathfrak{b}^\mu(x) := \mu x + \mathfrak{b}(x)$. The time invariance (10) can be justified in two ways: one can use that the directed landscape is the limit fluctuations of the Brownian the last-passage percolation model (4), and that the Brownian motion is invariant for the Markov evolution induced by the Brownian last-passage percolation model [21]; or Theorem 4.5 in [19], where it is used that the KPZ fixed point is of the limit fluctuations of the TASEP, whose invariant measure are given by Bernoulli i.i.d. random variables.

The transition probabilities of $\mathfrak{h}_{s,t}$ were computed by Matetski et al. [19], and the connection with the directed landscape construction was established by Nica et al. [20]. We give a brief description of the transition probabilities as follows. The collection composed by cylindrical subsets

$$\mathrm{Cy}(\mathbf{x}, \mathbf{a}) := \left\{\mathfrak{f} \in \mathsf{UC} \, : \, \mathfrak{f}(x_1) \leq a_1, \ldots, \mathfrak{f}(x_m) \leq a_m\right\} \text{ for } \mathbf{x}, \mathbf{a} \in \mathbb{R}^m ,$$

is a generating sub-algebra for the Borel σ-algebra over UC. The KPZ fixed point $(\mathfrak{h}_t(\cdot)\,,\ t \geq 0)$ is the unique time homogenous Markov process taking values in UC with transition probabilities given by the extension from the cylindrical sub-algebra to the Borel sets of

$$\mathbb{P}\Big(\mathfrak{h}_t \in \mathrm{Cy}(\mathbf{x},\mathbf{a}) \mid \mathfrak{h}_0 = \mathfrak{h}\Big) = \det\Big(I - K^{\mathfrak{h}}_{t,\mathbf{x},\mathbf{a}}\Big)_{\mathbb{L}^2(\{x_1,\ldots,x_m\}\times\mathbb{R})}\,. \tag{11}$$

On the right hand side of (11) we have a Fredholm determinant of the operator $K^{\mathfrak{h}}_{t,\mathbf{x},\mathbf{a}}$, whose definition we address to [19] (I is the identity operator). From this formula one can recover several of the classical Airy processes by starting with special profiles for which the respective operators K are explicit (see Section 4.4 of [19]). For instance, the Airy$_2$ process $\mathscr{A}(\cdot) = \mathfrak{h}(\cdot;\mathfrak{d}_0)$ is defined by taking the initial profile (5) with $x = 0$.

1.2 Space Hölder Regularity and Brownian Behaviour

Using kernel estimates for discrete approximations of the integral operator in (11), Matetski et al. [19] proved that $\mathfrak{h}_t$ has Hölder $1/2-$ regularity in space (Theorem 4.13 [19] and Proposition 1.6 [13]), and also that $S_{\sqrt{\epsilon}}\Delta\mathfrak{h}_t$ converges to $\mathfrak{b}$, as $\epsilon \to 0^+$, in terms of finite dimensional distributions (Theorem 4.14 [19]). Functional convergence was proved by Pimentel [23] for several versions of Airy processes, which are obtained from the fundamental initial profiles $\mathfrak{h} \equiv \mathfrak{d}_0$, $\mathfrak{h} \equiv 0$ and $\mathfrak{h} \equiv \mathfrak{b}$, and stronger forms of local Brownian behaviour were proved by Corwin and Hammond [10] and Hammond [15]. Here we use geometrical properties related to (6) to control space regularity of $\mathfrak{h}_t$.

Let $\beta \in [0, 1]$ and define the Hölder semi-norm of a real function $\mathfrak{f} : \mathbb{R} \to \mathbb{R}$ as

$$\|\mathfrak{f}\|_{\beta,[-a,a]} := \sup\left\{\frac{|\mathfrak{f}(x) - \mathfrak{f}(y)|}{|x-y|^{\beta}}\ :\ x, y \in [-a, a] \text{ and } x \neq y\right\}.$$

Theorem 1 *Fix $a, t > 0$ and $\beta \in [0, 1/2)$. Then*

$$\mathbb{P}\left(\|\mathfrak{h}_t\|_{\beta,[-a,a]} < \infty\right) = 1\,. \tag{12}$$

Furthermore,

$$\lim_{\epsilon\to 0^+} S_{\sqrt{\epsilon}}\Delta\mathfrak{h}_t(\cdot) \overset{dist.}{=} \mathfrak{b}(\cdot)\,, \tag{13}$$

where the distribution of $\mathfrak{b}$ is given by (9).

For $\mathfrak{f} : \mathbb{R}^2 \to \mathbb{R}$, define the Hölder semi-norm as follows

$$\|\mathfrak{f}\|_{\beta,[-a,a]^2} := \sup \left\{ \frac{|\mathfrak{f}(\mathbf{x}) - \mathfrak{f}(\mathbf{y})|}{|\mathbf{x}-\mathbf{y}|_\infty^\beta} \ : \ \mathbf{x}, \mathbf{y} \in [-a,a]^2 \text{ and } \mathbf{x} \neq \mathbf{y} \right\} .$$

Denote

$$\Delta\mathfrak{f}(x,y) := \mathfrak{f}(x,y) - \mathfrak{f}(0,0) ,$$

and let $\mathscr{B}(x,y) := \mathfrak{b}_1(x) + \mathfrak{b}_2(y)$, where $\mathfrak{b}_1$ and $\mathfrak{b}_2$ are two independent copies of (9).

Theorem 2 *Consider the stationary Airy sheet and* $\beta \in [0, 1/2)$. *Then*

$$\mathbb{P}\left(\|\mathscr{A}\|_{\beta,[-a,a]^2} < \infty\right) = 1 .$$

Furthermore[1]

$$\lim_{\epsilon \to 0^+} S_{\sqrt{\epsilon}} \Delta\mathscr{A}(\cdot,\cdot) \stackrel{dist.}{=} \mathscr{B}(\cdot,\cdot) .$$

In view of (1) *and* (2)*, we also have that*

$$\lim_{t \to \infty} \Delta\mathscr{L}(\cdot, 0; \cdot, t) \stackrel{dist.}{=} \mathscr{B}(\cdot,\cdot) .$$

1.3 Brownian Long Time Behaviour

From (8), one can see that the long time behaviour of $\Delta\mathfrak{h}_t$ can be written in terms of the local space behaviour of $\Delta\mathfrak{h}_1$ (take $\gamma = t^{1/3}$), which allows one to obtain long time convergence (in terms of finite dimensional distributions) from the local convergence to Brownian motion, as soon as $S_\gamma \mathfrak{h}$ converges in distribution in UC as $\gamma \to \infty$ (Theorem 4.15 [19]). Based on the same geometrical tools to study the space regularity of the KPZ fixed point, we will prove long time convergence of the KPZ fixed.

Theorem 3 *Assume that there exist* $c > 0$ *and a real function* ψ *such that for all* $\gamma \geq c$ *and* $r \geq 1$

$$\mathbb{P}\left(S_\gamma \mathfrak{h}(z) \leq r|z| \,,\ \forall\, |z| \geq 1\right) \geq 1 - \psi(r) \ \ \text{and} \ \ \lim_{r \to \infty} \psi(r) = 0 . \tag{14}$$

[1] Convergence in terms of a sequence of random elements in the space of continuous scalar fields on a fixed compact subset of $\mathbb{R}^2$, endowed with the uniform metric.

Let $a, t, \eta > 0$ *and set* $r_t := \sqrt[4]{t^{2/3}a^{-1}}$*. Under* (14)*, where* $\mathfrak{b}$ (9) *and* $\mathfrak{h}$ *are sample independently, there exists a real function* ϕ*, which does not depend on* $a, t, \eta > 0$*, such that for all* $t \geq \max\{c^3, a^{3/2}\}$ *and* $\eta > 0$ *we have*

$$\mathbb{P}\left(\sup_{x\in[-a,a]} |\Delta\mathfrak{h}_t(x;\mathfrak{h}) - \Delta\mathfrak{h}_t(x;\mathfrak{b})| > \eta\sqrt{a}\right) \leq \phi\,(r_t) + \frac{1}{\eta r_t} \;\; \textit{and} \;\; \lim_{r\to\infty}\phi(r) = 0\,. \tag{15}$$

In particular, if $\lim_{t\to\infty} a_t t^{-2/3} = 0$ *then*

$$\lim_{t\to\infty}\mathbb{P}\left(\sup_{x\in[-a_t,a_t]} |\Delta\mathfrak{h}_t(x;\mathfrak{h}) - \Delta\mathfrak{h}_t(x;\mathfrak{b})| > \eta\sqrt{a_t}\right) = 0\,.$$

Since $S_{\sqrt{a_t}}\Delta\mathfrak{h}_t(\cdot;\mathfrak{b}) \stackrel{dist.}{=} \mathfrak{b}(\cdot)$*, we also have that*

$$\lim_{t\to\infty} S_{\sqrt{a_t}}\Delta\mathfrak{h}_t(\cdot;\mathfrak{h}) \stackrel{dist.}{=} \mathfrak{b}(\cdot)\,.$$

Remark 1 For deterministic $\mathfrak{h}(x) = x^{\zeta}$, for $\zeta \in [0,1]$, we have that $S_\gamma\mathfrak{h}(x) = \gamma^{2\zeta-1}x^{\zeta}$. If $\zeta \in [0,1/2]$, then $\mathfrak{h}$ does satisfy (14), while for $\zeta \in (1/2,1]$ it does not. We use assumption (14) to ensure that, for all large values of t,

$$\mathbb{P}\left(|Z_t(\pm a;\mathfrak{h})| > rt^{2/3}\right) \leq \phi_1(r) \to 0\,, \text{ as } r \to \infty\,, \tag{16}$$

where $Z_t(x;\mathfrak{h})$ is the rightmost $z \in \mathbb{R}$ to attain the maximum (6), and ϕ_1 is a real function that does not depend on $a > 0$ or $t > 0$ (Lemma 2). If one can prove (16), based on possible different assumptions, then (15) will follow as well.

Remark 2 Theorem 3 does not imply immediately that the only spatially ergodic (in terms of its increments) and time invariant process with zero drift is $\mathfrak{b}$. This would follow as soon as one can verify (14) or (16) for such a process.

2 Geometry, Comparison and Attractiveness

Given an upper semicontinuous function $\mathfrak{f}$ such that

$$\lim_{|z|\to\infty}\mathfrak{f}(z) = -\infty\,, \tag{17}$$

then the supremum of $\mathfrak{f}(z)$ over $z \in \mathbb{R}$ is indeed a maximum, i.e. $\exists\, Z \in \mathbb{R}$ such that $\mathfrak{f}(Z) \geq \mathfrak{f}(z)$ for all $z \in \mathbb{R}$. Additionally, the set

$$\arg\max_{z\in\mathbb{R}} \mathfrak{f}(z) := \left\{ Z \in \mathbb{R} \,:\, \mathfrak{f}(Z) = \max_{z\in\mathbb{R}} \mathfrak{f}(z) \right\}.$$

is compact. Since with probability one, for all $\mathfrak{h} \in \mathsf{UC}$, $\mathfrak{h}(z) + \mathscr{L}(z, s; x, t)$ satisfies (17), for all $s < t$ and $x \in \mathbb{R}$, (due to the parabolic drift (1)) we can use these aforementioned facts to study the Markov evolution (6).

We call a continuous path $\mathscr{P} : [r, t] \to \mathbb{R}$ a geodesic between the space-time points (x, r) and (y, t) if $\mathscr{P}(r) = x$, $\mathscr{P}(t) = y$ and for $s \in (r, t)$

$$\mathscr{L}(x, r; y, t) = \mathscr{L}\left(x, r; \mathscr{P}(s), s\right) + \mathscr{L}\left(\mathscr{P}(s), s; y, t\right) \tag{18}$$

Define $\mathscr{P}^{y,t}_{x,r}(r) = x$, $\mathscr{P}^{y,t}_{x,r}(t) = y$ and

$$\mathscr{P}^{y,t}_{x,r}(s) := \max \arg\max_{z\in\mathbb{R}} \left\{\mathscr{L}(x, r; z, s) + \mathscr{L}(z, s; y, t)\right\} \text{ for } s \in (r, t)\,.$$

By Lemma 13.3 [13], almost surely, $\mathscr{P}^{y,t}_{x,r}$ is a geodesic for every (x, r) and (y, t). We also identify the geodesic path (or function) $\mathscr{P}$ with its graph $\{(\mathscr{P}(s), s) : s \in [r, t]\}$ in order to handle intersection points between different paths. For each $\mathfrak{h} \in \mathsf{UC}$, $0 < t$ and $x \in \mathbb{R}$, let

$$Z_t(x; \mathfrak{h}) := \max \arg\max_{z\in\mathbb{R}} \left\{\mathfrak{h}(z) + \mathscr{L}(z, s; x, t)\right\}\,. \tag{19}$$

Proposition 2 *Almost surely $\mathfrak{h}_t$ and Z_t are a well defined real functions for which we have the following properties.*

(i) $\mathfrak{h}_t(x) = \mathfrak{h}\left(Z_t(x)\right) + \mathscr{L}\left(Z_t(x), 0; x, t\right)$.

(ii) For every $w \in \mathbb{R}$ and $u \in [0, t)$,

$$\mathfrak{h}_t(x) \geq \mathfrak{h}_u(w) + \mathscr{L}(w, u; x, t)\,.$$

(iii) For every $(w, u) \in \mathscr{P}^{x,t}_{Z_t(x),0}$,

$$\mathfrak{h}_t(x) = \mathfrak{h}_u(w) + \mathscr{L}(w, u; x, t) \text{ and } \mathfrak{h}_u(w) = \mathfrak{h}(Z_t(x)) + \mathscr{L}(0, Z_t(x); w, u)\,.$$

(iv) For fixed $t > 0$, $Z_t(x)$ is a nondecreasing function of $x \in \mathbb{R}$.

(v) $\mathfrak{h}_{s,t}$ satisfies a composition property: $\mathfrak{h}_{t+s}(\cdot; \mathfrak{h}) = \mathfrak{h}_{t,t+s}(\cdot; \mathfrak{h}_t)$, i.e.

$$\mathfrak{h}_{t+s}(x; \mathfrak{h}) = \max_{z\in\mathbb{R}} \left\{\mathfrak{h}_t(z; \mathfrak{h}) + \mathscr{L}(z, t; x, t + s)\right\}\,,\ \forall\, x \in \mathbb{R}\,.$$

Proof By compactness, $Z_t(x) \in \arg\max_{z\in\mathbb{R}} \{\mathfrak{h}(z) + \mathcal{L}(z, 0; x, t)\}$, which implies (i). Now we use (6) and (3) to get (ii): for any $z, w \in \mathbb{R}$ and $u \in (0, t)$,

$$\mathfrak{h}_t(x) \geq \mathfrak{h}(z) + \mathcal{L}(0, z; x, t) \geq \mathfrak{h}(z) + \mathcal{L}(0, z; u, w) + \mathcal{L}(w, u; x, t)\,,$$

and hence

$$\mathfrak{h}_t(x) \geq \mathfrak{h}_u(w) + \mathcal{L}(w, u; x, t)\,.$$

By (i) and (18), if $w = \mathcal{P}^{x,t}_{Z_t(x),0}(u)$ then

$$\begin{aligned}\mathfrak{h}_t(x) &= \mathfrak{h}(Z_t(x)) + \mathcal{L}(Z_t(x), 0; x, t)\\ &= \mathfrak{h}(Z_t(x)) + \mathcal{L}(Z_t(x), 0; w, u) + \mathcal{L}(w, u; x, t)\\ &\leq \mathfrak{h}_u(w) + \mathcal{L}(w, u; x, t)\,,\end{aligned}$$

and thus, by (ii),

$$\mathfrak{h}_t(x) = \mathfrak{h}_u(w) + \mathcal{L}(w, u; x, t)$$

and

$$\mathfrak{h}_u(w) = \mathfrak{h}(Z_t(x)) + \mathcal{L}(0, Z_t(x); w, u)\,,$$

which concludes the proof of (iii). To prove (iv), assume that $Z_t(y) < Z_t(x)$ for some $x < y$. Then $\mathcal{P}^{y,t}_{Z_t(y),0}$ and $\mathcal{P}^{x,t}_{Z_t(x),0}$ intersects at some space-time point (w, u). By (iii), we have that

$$\mathfrak{h}_t(y) = \mathfrak{h}_u(w) + \mathcal{L}(w, u; y, t) \;\text{ and }\; \mathfrak{h}_u(w) = \mathfrak{h}(Z_t(x)) + \mathcal{L}(Z_t(x), 0; w, u)\,.$$

This shows that

$$\begin{aligned}\mathfrak{h}_t(y) &= \mathfrak{h}_u(w) + \mathcal{L}(w, u; y, t)\\ &= \mathfrak{h}(Z_t(x)) + \mathcal{L}(Z_t(x), 0; w, u) + \mathcal{L}(w, u; y, t)\\ &\leq \mathfrak{h}(Z_t(x)) + \mathcal{L}(Z_t(x), 0; y, t)\,,\end{aligned}$$

where we use the metric composition (3) for the last inequality. Hence, $Z_t(x)$ is also a location that attains the maximum for $\mathfrak{h}_t(y)$, which leads to a contradiction since we assumed that $Z_t(y) < Z_t(x)$ and $Z_t(y)$ is the rightmost point to attain the maximum. The composition property (v) follows directly item (iii). □

Proposition 3 (Argmax Comparison) *If* $x < y$ *and* $Z_t(y; \mathfrak{h}) \leq Z_t(x; \tilde{\mathfrak{h}})$ *then*

$$\mathfrak{h}_t(y; \mathfrak{h}) - \mathfrak{h}_t(x; \mathfrak{h}) \leq \mathfrak{h}_t(y; \tilde{\mathfrak{h}}) - \mathfrak{h}_t(x; \tilde{\mathfrak{h}})\,.$$

Proof Denote $z \equiv Z_t(y;\mathfrak{h})$ and $\tilde{z} \equiv Z_t(x;\tilde{\mathfrak{h}})$. By assumption, $x < y$ and $z \le \tilde{z}$, and hence there exists $(w,u) \in \mathscr{P}_{z,0}^{y,t} \cap \mathscr{P}_{\tilde{z},0}^{x,t}$. Since $(w,u) \in \mathscr{P}_{\tilde{z},0}^{x,t}$, by (iii)-Proposition 2,

$$\mathfrak{h}_t(x;\tilde{\mathfrak{h}}) = \mathfrak{h}_u(w;\tilde{\mathfrak{h}}) + \mathscr{L}(w,u;x,t)\,,$$

and, by (ii)-Proposition 2,

$$\mathfrak{h}_t(y;\tilde{\mathfrak{h}}) \ge \mathfrak{h}_u(w;\tilde{\mathfrak{h}}) + \mathscr{L}(w,u;y,t)\,,$$

that yields to

$$\mathfrak{h}_t(y;\tilde{\mathfrak{h}}) - \mathfrak{h}_t(x;\tilde{\mathfrak{h}}) \ge \mathscr{L}(w,u;y,t) - \mathscr{L}(w,u;x,t)\,.$$

Now $(w,u) \in \mathscr{P}_{z,0}^{y,t}$ and by using Proposition 2 as before, we have

$$\mathfrak{h}_t(y;\mathfrak{h}) = \mathfrak{h}_u(w;\mathfrak{h}) + \mathscr{L}(w,u;y,t) \text{ and } \mathfrak{h}_t(x;\mathfrak{h}) \ge \mathfrak{h}_u(w;\mathfrak{h}) + \mathscr{L}(w,u;x,t)\,,$$

which implies that

$$\mathfrak{h}_t(y;\mathfrak{h}) - \mathfrak{h}_t(x;\mathfrak{h}) \le \mathscr{L}(w,u;y,t) - \mathscr{L}(w,u;x,t)\,,$$

and therefore $\mathfrak{h}_t(y;\mathfrak{h}) - \mathfrak{h}_t(x;\mathfrak{h}) \le \mathfrak{h}_t(y;\tilde{\mathfrak{h}}) - \mathfrak{h}_t(x;\tilde{\mathfrak{h}})$. □

Proposition 4 (Attractiveness) *If* $\mathfrak{h}(y) - \mathfrak{h}(x) \le \tilde{\mathfrak{h}}(y) - \tilde{\mathfrak{h}}(x)$ *for all* $x < y$ *then*

$$\mathfrak{h}_t(y;\mathfrak{h}) - \mathfrak{h}_t(x;\mathfrak{h}) \le \mathfrak{h}_t(y;\tilde{\mathfrak{h}}) - \mathfrak{h}_t(x;\tilde{\mathfrak{h}}) \;\; \forall\, x < y\,, \forall\, t \ge 0\,.$$

Proof Denote again $z \equiv Z_t(y;\mathfrak{h})$ and $\tilde{z} \equiv Z_t(x;\tilde{\mathfrak{h}})$. If $z \le \tilde{z}$ then

$$\mathfrak{h}_t(y;\mathfrak{h}) - \mathfrak{h}_t(x;\mathfrak{h}) \le \mathfrak{h}_t(y;\tilde{\mathfrak{h}}) - \mathfrak{h}_t(x;\tilde{\mathfrak{h}})\,,$$

by Proposition 3. If $z > \tilde{z}$ then, by (i)-Proposition 2,

$$\begin{aligned}
\mathfrak{h}_t(y;\tilde{\mathfrak{h}}) - \mathfrak{h}_t(x;\tilde{\mathfrak{h}}) - \Big(\mathfrak{h}_t(y;\mathfrak{h}) - \mathfrak{h}_t(x;\mathfrak{h})\Big) \\
&= \mathfrak{h}_t(y;\tilde{\mathfrak{h}}) - \big(\tilde{\mathfrak{h}}(\tilde{z}) + \mathscr{L}(\tilde{z},0;x,t)\big) \\
&- \Big(\big(\mathfrak{h}(z) + \mathscr{L}(z,0;y,t)\big) - \mathfrak{h}_t(x;\mathfrak{h})\Big) \\
&= \mathfrak{h}_t(y;\tilde{\mathfrak{h}}) - \big(\tilde{\mathfrak{h}}(z) + \mathscr{L}(z,0;y,t)\big) + \Big(\mathfrak{h}_t(x;\mathfrak{h}) \\
&- \big(\mathfrak{h}(\tilde{z}) + \mathscr{L}(\tilde{z},0;x,t)\big)\Big) \\
&+ \big(\tilde{\mathfrak{h}}(z) - \tilde{\mathfrak{h}}(\tilde{z})\big) - \big(\mathfrak{h}(z) - \mathfrak{h}(\tilde{z})\big)\,.
\end{aligned}$$

Thus, by (6),

$$\mathfrak{h}_t(y;\tilde{\mathfrak{h}}) - \big(\tilde{\mathfrak{h}}(z) + \mathscr{L}(z,0;y,t)\big) \geq 0\,,$$

and

$$\mathfrak{h}_t(x;\mathfrak{h}) - \big(\mathfrak{h}(\tilde{z}) + \mathscr{L}(\tilde{z},0;x,t)\big) \geq 0\,,$$

while, by assumption,

$$\big(\tilde{\mathfrak{h}}(z) - \tilde{\mathfrak{h}}(\tilde{z})\big) - \big(\mathfrak{h}(z) - \mathfrak{h}(\tilde{z})\big) \geq 0\,,$$

since $z > \tilde{z}$. □

2.1 Uniqueness of the Argmax

We finish this section by pointing out how the ideas in [22] can be combined with the fact that the Airy$_2$ process is locally absolutely continuous with respect to Brownian motion [10], to prove a.s. uniqueness of the location of the maxima in (6). Since $\mathfrak{h}(z) + \mathscr{L}(z,s;x,t)$ satisfies (17), it is enough to prove uniqueness of the location of the maximum restrict to a compact set. On the other hand, $\{\mathscr{L}(z,s;x,t) : z \in \mathbb{R}\}$ is distributed as a rescaled Airy$_2$ process minus a parabola (for fixed $x \in \mathbb{R}$ and $0 < s < t$), which is locally absolutely continuous with respect to Brownian motion [10]. Therefore, uniqueness of the location of the maxima in (6) follows from the next proposition, which is similar to Theorem 2 [22], combined with Lemma 2 [22].[2]

Proposition 5 *Let $K \subseteq \mathbb{R}$ be a compact set and $\mathfrak{f} : K \to \mathbb{R}$ be a random upper semicontinuous function. Denote $\mathfrak{f}^a(z) := \mathfrak{f}(z) + az$, $M(\mathfrak{f}) := \max_{z\in K}\mathfrak{f}(z)$, and let*

$$m(a) = \mathbb{E}\left(M\left(\mathfrak{f}^a\right) - M\left(\mathfrak{f}\right)\right)\,.$$

Then $|m(a)| < \infty$ for all $a \in \mathbb{R}$ and a.s. there exists a unique $Z \in K$ such that $M(\mathfrak{f}) = \mathfrak{f}(Z)$ if and only if $m(a)$ is differentiable at $a = 0$. Furthermore, in this case,

$$m'(0) = \mathbb{E}\,Z\,.$$

Proof The first part of the proof is merely analytic and we follow the proof of Lemma 1 [22], where $\mathfrak{f}$ was assumed to be continuous. There are two fundamental

[2] Lemma 2 [22] shows that $m(a)$ is differentiable at $a = 0$ if $\mathfrak{f}$ is a sum of a deterministic function $\mathfrak{h}$ with a Brownian motion.

steps where we used continuity that needs to be adapted to upper semicontinuous functions. Denote

$$Z_1(\mathfrak{f}) := \inf \arg\max_{z\in K} \mathfrak{f}(z) \quad \text{and} \quad Z_2(\mathfrak{f}) := \sup \arg\max_{z\in K} \mathfrak{f}(z)\,.$$

For simple notation we put $M^a \equiv M(\mathfrak{f}^a)$, $M \equiv M(\mathfrak{f})$, $Z_i^a \equiv Z_i(\mathfrak{f}^a)$ and finally $Z_i \equiv Z_i(\mathfrak{f})$. The first step in [22] was to argue that $M = \mathfrak{f}(Z_i)$ and $M^a = \mathfrak{f}(Z_i^a) + aZ_i^a$. But for a upper semicontinuous function, $\arg\max_{z\in K} \mathfrak{f}(z)$ is a compact set, and then $Z_1(\mathfrak{f}), Z_2(\mathfrak{f}) \in \arg\max_K \mathfrak{f}(z)$ (which also holds for $\mathfrak{f}^a$). Thus, we can conclude that

$$M + aZ_i = \mathfrak{f}(Z_i) + aZ_i \le M^a = \mathfrak{f}(Z_i^a) + aZ_i^a \le M + aZ_i^a\,. \tag{20}$$

Since $Z_i, Z_i^a \in K$ and K is compact, by (20), we have that $|m(a)| < \infty$ for all $a \in \mathbb{R}$. The second step in [22] was to prove that

$$\lim_{a\to 0^-} Z_1^a = Z_1 \ \text{ and } \ \lim_{a\to 0^+} Z_2^a = Z_2\,. \tag{21}$$

Indeed, by (20), we have that $Z_1^a \le Z_1$ for all $a < 0$, and if the convergence of Z_1^a to Z_1 does not hold then, by compactness of K, we can find $\tilde{Z}_1 \in K$, $\delta > 0$ and a sequence $a_n \to 0^-$ such that $\lim_{n\to\infty} Z_1^{a_n} = \tilde{Z}_1$ and $\tilde{Z}_1 \le Z_1 - \delta$. But by (20), we also have that

$$0 \le a\left(Z_i^a - Z_i\right) - \left(\mathfrak{f}(Z_i) - \mathfrak{f}(Z_i^a)\right)\,, \ \text{for } i = 1, 2\,,$$

and thus (first inequality)

$$\mathfrak{f}(Z_1) \le \limsup_n \mathfrak{f}(Z_1^{a_n}) \le \mathfrak{f}(\tilde{Z}_1)\,,$$

where we use upper semicontinuity in the second inequality. But this is a contradiction, since Z_1 is the leftmost location to attain the maximum, and hence $\lim_{a\to 0^-} Z_1^a = Z_1$. Since $Z_2^a \ge Z_2$ for all $a > 0$, the proof of $\lim_{a\to 0^+} Z_2^a = Z_2$ is analogous. By (20) again,

$$0 \le (M^a - M) - aZ_i \le a(Z_i^a - Z_i)\,,$$

which implies that

$$0 \ge \frac{M^a - M}{a} - Z_1 \ge Z_1^a - Z_1 \ge -diam(K)\,, \ \text{for } a < 0\,,$$

and

$$0 \le \frac{M^a - M}{a} - Z_2 \le Z_2^a - Z_2 \le diam(K)\,, \ \text{for } a > 0\,,$$

where $diam(K)$ denotes the diameter of K. Since the location of the maximum is a.s. unique if and only if $\mathbb{E}\,(Z_1) = \mathbb{E}\,(Z_2)$ (now we have a random $\mathfrak{f}$), using the inequalities above, (21) and dominated convergence, we see that the location of the maximum of $\mathfrak{f}$ is a.s. unique if and only if $m(a)$ is differentiable at $a = 0$:

$$\mathbb{E}\,(Z_1) = \mathbb{E}\,(Z_2) \iff \lim_{a\to 0^-} \frac{m(a) - m(0)}{a} = \lim_{a\to 0^+} \frac{m(a) - m(0)}{a}\,,$$

which concludes the proof. □

3 Proof of the Theorems

A key step to use comparison (Proposition 3) relies on the control of $Z_t(x; \mathfrak{h})$ (recall (19)) as a function of $\mathfrak{h}$, x and t. Let X be the closest point to the origin such that $\mathfrak{h}(X) > -\infty$ (if a tiebreak occurs we pick the nonnegative one). By assumption, $X \in \mathbb{R}$ is a well defined random variable. Since the location of a maximum is invariant under vertical shifts of $\mathfrak{h}$, if we want to control the location of the maximum, we can assume without loss of generality that $\mathfrak{h}(X) = 0$. By the symmetries (i)-(ii)-(iii), for fixed values of $x \in \mathbb{R}$ and $t > 0$,

$$Z_t(x; \mathfrak{h}) \overset{dist.}{=} t^{2/3} Z_1(0; S_{\gamma_t} T_x \mathfrak{h}) + x\,, \tag{22}$$

where $\gamma_t := t^{1/3}$. By (22), for $x \in [-a, a]$,

$$\begin{aligned}\mathbb{P}\left(|Z_t(x; \mathfrak{h})| > rt^{2/3}\right) &\le \mathbb{P}\left(|Z_1(0; S_{\gamma_t} T_x \mathfrak{h})| > r - |x|t^{-2/3}\right) \\ &\le \mathbb{P}\left(|Z_1(0; S_{\gamma_t} T_x \mathfrak{h})| > r - at^{-2/3}\right)\,. \end{aligned}\tag{23}$$

The right hand side of (23) is bounded by

$$\mathbb{P}\left(\max_{|z|>r-at^{-2/3}} \left\{S_{\gamma_t} T_x \mathfrak{h}(z) + \mathscr{A}(z) - z^2\right\} = \max_{z\in\mathbb{R}} \left\{S_{\gamma_t} T_x \mathfrak{h}(z) + \mathscr{A}(z) - z^2\right\}\right)\,,$$

where $\mathscr{A}(z) := \mathscr{A}(z, 0)$. If we take

$$X_t := \gamma_t^{-2}(X - x)\,,$$

we get that $S_{\gamma_t} T_x \mathfrak{h}(X_t) = \mathfrak{h}(X) = 0$, and the right hand side of (23) is bounded by

$$\mathbb{P}\left(\max_{|z|>r-at^{-2/3}} \left\{S_{\gamma_t} T_x \mathfrak{h}(z) + \mathscr{A}(z) - z^2\right\} \ge \mathscr{A}\,(X_t) - X_t^2\right)\,. \tag{24}$$

In the next lemmas we will use that the Airy$_2$ process $\{\mathscr{A}(z) : z \in \mathbb{R}\}$ is stationary and independent of X, which implies that $\mathscr{A}(X_t) \overset{dist.}{=} \mathscr{A}(0)$, and we can split the probability in (24) as

$$\mathbb{P}\left(\max_{|z|>r-at^{-2/3}} \left\{S_{\gamma_t} T_x \mathfrak{h}(z) + \mathscr{A}(z) - z^2\right\} \geq -L\right) + \mathbb{P}\Big(\mathscr{A}(0) - X_t^2 \leq -L\Big), \tag{25}$$

for any choice of $L > 0$.

Lemma 1 *Let $a, t >$ be fixed. For every $\mathfrak{h} \in$* UC

$$\lim_{r\to\infty} \mathbb{P}\left(|Z_t(\pm a; \mathfrak{h})| > r\right) = 0\,.$$

Proof For the sake of simplicity, we are going to prove it for $t = 1$ and $a = 1$, and $X_1 = X - 1$. Let us pick $L_r = (r-1)^2/4$. Then

$$\lim_{r\to\infty} \mathbb{P}\left(\mathscr{A}(0) - (X-1)^2 \leq -L_r\right) = 0\,,$$

since the random variable $\mathscr{A}(0) - (X-1)^2$ does not depend on r. By (23), (24) and (25), we still need to prove that

$$\lim_{r\to\infty} \mathbb{P}\left(\max_{|z|>r-1} \left\{T_1\mathfrak{h}(z) + \mathscr{A}(z) - z^2\right\} \geq -L_r\right) = 0\,.$$

If $r > 2$ and $|z| > r-1$ then $|z| > r/2 > 1$ and $\frac{r}{4}|z| - z^2 \leq -z^2/2$. Hence, if $T_1\mathfrak{h}(z) \leq \frac{r}{4}|z|$ then

$$T_1\mathfrak{h}(z) - z^2 \leq \frac{r}{4}|z| - z^2 \leq -z^2/2\,,$$

which shows that

$$\mathbb{P}\left(\max_{|z|>r-1} \left\{T_1\mathfrak{h}(z) + \mathscr{A}(z) - z^2\right\} \geq -L_r\right) \leq \psi\left(r/4; T_1\mathfrak{h}\right) + \mathbb{P}\left(\max_{|z|>r-1} \left\{\mathscr{A}(z) - \frac{z^2}{2}\right\} \geq -L_r\right),$$

where

$$\psi(r; \mathfrak{h}) := 1 - \mathbb{P}\left(\mathfrak{h}(z) \leq r|z|\,,\ \forall\,|z| \geq 1\right)\,.$$

By (b)-Proposition 2.13 [12], there exist constants $c_1, c_2 > 0$ such that for all $r > c_1$,

$$\mathbb{P}\left(\max_{|z|>r}\left\{A(z)-\frac{z^2}{2}\right\} > -\frac{r^2}{4}\right) \leq e^{-c_2 r^3},$$

which shows that

$$\lim_{r\to\infty} \mathbb{P}\left(|Z_1(1;\mathfrak{h})| > r\right) = 0,$$

as soon as we prove that,

$$\lim_{r\to\infty} \psi(r;\mathfrak{h}) = 0 \text{ for all } \mathfrak{h} \in \mathsf{UC}.$$

But for every probability measure on UC, we have that

$$\mathbb{P}\left(\exists\, r > 0 \text{ s. t. } \mathfrak{h}(z) \leq r(1+|z|)\ \forall z \in \mathbb{R}\right) = 1,$$

and if $r_1 < r_2$ then

$$\{\mathfrak{h}(z) \leq r_1(1+|z|)\ \forall z \in \mathbb{R}\} \subseteq \{\mathfrak{h}(z) \leq r_2(1+|z|)\ \forall z \in \mathbb{R}\},$$

which implies that

$$\lim_{r\to\infty} \mathbb{P}\left(\mathfrak{h}(z) \leq r(1+|z|)\ \forall z \in \mathbb{R}\right) = 1.$$

Since $\frac{r}{2}(1+|z|) \leq r|z|$ for all $|z| \geq 1$ we have that

$$\{\mathfrak{h}(z) \leq \frac{r}{2}(1+|z|)\ \forall z \in \mathbb{R}\} \subseteq \{\mathfrak{h}(z) \leq r|z|\ \forall |z| \geq 1\},$$

and therefore, $\lim_{r\to\infty} \psi(r;\mathfrak{h}) = 0$. □

Lemma 2 *Under* (14)*, there exists a real function* ϕ_1*, which does not depend on* $a > 0$ *or* $t > 0$*, such that for all* $t \geq \max\{c^3, a^{3/2}\}$ *we have*

$$\mathbb{P}\left(|Z_t(\pm a;\mathfrak{h})| > rt^{2/3}\right) \leq \phi_1(r) \quad \textit{and} \quad \lim_{r\to\infty} \phi_1(r) = 0.$$

Proof We use again (23), (24) and (25). Pick $L_r = (r-1)^2/4$ and $t \geq \max\{c^3, a^{3/2}\}$. Then (recall that $\gamma_t = t^{1/3}$)

$$\gamma_t^{-4}(X-a)^2 \leq 2\gamma_t^{-4}\left(X^2+a^2\right) \leq 2\left(\frac{X^2}{c^4}+1\right),$$

and thus,

$$\mathbb{P}\left(\mathscr{A}(0) \leq -L_r + \gamma_t^{-4}(X-a)^2\right) \leq \mathbb{P}\left(\mathscr{A}(0) \leq -L_r + 2\left(\frac{X^2}{c^4}+1\right)\right).$$

The right hand side of the above inequality is a function of r that does not depend on $a > 0$ or $t > 0$, and goes to zero as r goes to infinity. To control the first term in the right hand side of (25) we note that, if $S_{\gamma_t}\mathfrak{h}(z) \leq \frac{r}{4}|z|$ for all $|z| \geq 1$, then

$$S_{\gamma_t} T_a \mathfrak{h}(z) = S_{\gamma_t}\mathfrak{h}\left(z + at^{-2/3}\right) \leq \frac{r}{4}|z + at^{-2/3}| \leq \frac{r}{4}\left(|z| + at^{-2/3}\right),$$

as soon as $|z + at^{-2/3}| \geq 1$. This needs to hold for all $|z| > r - at^{-2/3}$ in order to upper bound the maximum over all such z's. However, for $r > 3$ and $|z| > r - at^{-2/3}$ (recall that $t \geq \max\{c^3, a^{3/2}\}$) we certainly have that $|z + at^{-2/3}| \geq 1$. Therefore, if $S_{\gamma_t}\mathfrak{h}(z) \leq \frac{r}{4}|z|$ for all $|z| \geq 1$, then

$$\begin{aligned}\max_{|z|>r-at^{-2/3}} \left\{S_{\gamma_t} T_a \mathfrak{h}(z) + \mathscr{A}(z) - z^2\right\} &\leq \max_{|z|>r-at^{-2/3}} \left\{\frac{r}{4}\left(|z| + at^{-2/3}\right) + \mathscr{A}(z) - z^2\right\} \\ &\leq \max_{|z|>r-1} \left\{\frac{r}{4}\left(|z|+1\right) + \mathscr{A}(z) - z^2\right\}.\end{aligned}$$

To ensure that $\frac{r}{4}(|z|+1) - z^2 \leq -\frac{z^2}{2}$ for $|z| > r-1$ we take $r > 4$. Thus, we can conclude that for $r > 4$ and $t \geq \max\{c^3, a^{3/2}\}$ we have that

$$\mathbb{P}\left(\max_{|z|>r-at^{-2/3}} \left\{S_{\gamma_t} T_a \mathfrak{h}(z) + \mathscr{A}(z) - z^2\right\} \geq -L_r\right)$$

is bounded by

$$\psi(r/4) + \mathbb{P}\left(\max_{|z|>r-1} \left\{\mathscr{A}(z) - \frac{z^2}{2}\right\} \geq -L_r\right).$$

which is only a function of r and it concludes the proof of Lemma 2. □

In order to use Proposition 3 we tilt the initial profile $\mathfrak{b}$ as follows. For $\mu \geq 0$ denote

$$\mathfrak{h}_t^{\pm\mu}(\cdot) \equiv \mathfrak{h}_t(\cdot; \mathfrak{b}^{\pm\mu}), \text{ where } \mathfrak{b}^{\pm\mu}(z) = \pm\mu z + \mathfrak{b}(z), \tag{26}$$

and $\mathfrak{b}$ is given by (9). Hence, for all $x < y$,

$$\mathfrak{b}^{-\mu}(y) - \mathfrak{b}^{-\mu}(x) \leq \mathfrak{b}(y) - \mathfrak{b}(x) \leq \mathfrak{b}^{\mu}(y) - \mathfrak{b}^{\mu}(x). \tag{27}$$

Recall (19) and let

$$Z_t^{\pm\mu}(x) = Z_t\left(x;\, \mathfrak{b}^{\pm\mu}\right) .$$

Then

$$Z_t^{\mu}(x) \stackrel{dist.}{=} Z_t^0(x) + \frac{\mu}{2}t \stackrel{dist.}{=} t^{2/3}Z_1^0(0) + x + \frac{\mu}{2}t \,. \tag{28}$$

The next step is to construct an event $E_t(\mu)$ where we can sandwich the local increments of $\mathfrak{h}_t$ in between the local increments $\mathfrak{h}_t^{\pm\mu}$, and this is the point where we use Proposition 3. Define the event

$$E_t(\mu) = \left\{ Z_t(a;\mathfrak{h}) \le Z_t^{+\mu}(-a) \text{ and } Z_t(-a;\mathfrak{h}) \ge Z_t^{-\mu}(a) \right\} . \tag{29}$$

By (iv)-Proposition 2, on the event $E_t(\mu)$, for $x < y$ and $x, y \in [-a, a]$,

$$Z_t(y;\mathfrak{h}) \le Z_t(a;\mathfrak{h}) \le Z_t^{\mu}(-a) \le Z_t^{\mu}(x) \,,$$

and

$$Z_t^{-\mu}(y) \le Z_t^{-\mu}(a) \le Z_t(-a;\mathfrak{h}) \le Z_t(x;\mathfrak{h}) \,.$$

Therefore, by Proposition 3, on the event $E_t(\mu)$, if $x < y$ and $x, y \in [-a, a]$, then

$$\mathfrak{h}_t^{-\mu}(y) - \mathfrak{h}_t^{-\mu}(x) \le \mathfrak{h}_t(y;\mathfrak{h}) - \mathfrak{h}_t(x;\mathfrak{h}) \le \mathfrak{h}_t^{\mu}(y) - \mathfrak{h}_t^{\mu}(x) \,. \tag{30}$$

3.1 Proof of Theorem 1

We want to control the Hölder semi-norm for $\beta \in [0, 1/2)$ (we omit the dependence on the domain and on the initial profile $\mathfrak{h}$),

$$\|\mathfrak{h}_t\|_\beta \equiv \|\mathfrak{h}_t\|_{\beta,[-a,a]} := \sup_{x,y\in[-a,a]\,,\,x\neq y} \frac{|\mathfrak{h}_t(x) - \mathfrak{h}_t(y)|}{|x-y|^\beta} \,.$$

By (30), on the event $E_t(\mu)$ (29),

$$\|\mathfrak{h}_t\|_\beta \le \max\left\{ \|\mathfrak{h}_t^{\mu}\|_\beta \,,\ \|\mathfrak{h}_t^{-\mu}\|_\beta \right\} ,$$

and hence,

$$\mathbb{P}\left(\|\mathfrak{h}_t\|_\beta > A\right) \leq \mathbb{P}\left(\left\{\|\mathfrak{h}_t\|_\beta > A\right\} \cap E_t(\mu)\right) + \mathbb{P}\left(E_t(\mu)^c\right)$$
$$\leq \mathbb{P}\left(\|\mathfrak{h}_t^{\mu}\|_\beta > A\right) + \mathbb{P}\left(\|\mathfrak{h}_t^{-\mu}\|_\beta > A\right) + \mathbb{P}\left(E_t(\mu)^c\right) .$$

Since $\|\mathfrak{h}_t^{\pm\mu}\|_\beta = \|\Delta\mathfrak{h}_t^{\pm\mu}\|_\beta$ and $\Delta\mathfrak{h}_t^{\pm\mu}$ are drifted Brownian motions (10),

$$\lim_{A\to\infty} \mathbb{P}\left(\|\mathfrak{h}_t^{\pm\mu}\|_\beta > A\right) = \lim_{A\to\infty} \mathbb{P}\left(\|\Delta\mathfrak{h}_t^{\pm\mu}\|_\beta > A\right) = 0\,,$$

which yields to

$$0 \leq \limsup_{A\to\infty} \mathbb{P}\left(\|\mathfrak{h}_t\|_\beta > A\right) \leq \mathbb{P}\left(E_t(\mu)^c\right) .$$

We picked $\mu > 0$ arbitrary and

$$\mathbb{P}\left(E_t(\mu)^c\right) \leq \mathbb{P}\left(|Z_t(a;\mathfrak{h})| > \frac{\mu}{4}t\right) + \mathbb{P}\left(Z_t^{\mu}(-a) \leq \frac{\mu}{4}t\right) + \mathbb{P}\left(Z_t^{-\mu}(a) \geq -\frac{\mu}{4}t\right) .$$

By (28) and Lemma 1, this implies that $\mathbb{P}\left(E_t(\mu)^c\right) \to 0$ as $\mu \to \infty$. Therefore

$$\lim_{A\to\infty} \mathbb{P}\left(\|\mathfrak{h}_t\|_\beta > A\right) = 0\,,$$

which finishes the proof of (12).

To prove convergence of

$$S_{\sqrt{\epsilon}}\Delta\mathfrak{h}_t(x;\mathfrak{h}) = \epsilon^{-1/2}\left(\mathfrak{h}_t(\epsilon x) - \mathfrak{h}_t(0)\right) ,$$

to Brownian motion (13), we consider the event $E_t(\mu)$ (29) again with $a = 1$ (we will choose μ later as a suitable function of ϵ). Given a compact set $K \subseteq \mathbb{R}$ we take $\epsilon > 0$ such that $\epsilon K \subseteq [-1, 1]$. Thus, by (30), on the event $E_t(\mu)$ (29), if $x < y$ and $x, y \in K$, then

$$\mathfrak{h}_t^{-\mu}(\epsilon y) - \mathfrak{h}_t^{-\mu}(\epsilon x) \leq \mathfrak{h}_t(\epsilon y;\mathfrak{h}) - \mathfrak{h}_t(\epsilon x;\mathfrak{h}) \leq \mathfrak{h}_t^{\mu}(\epsilon y) - \mathfrak{h}_t^{\mu}(\epsilon x)\,. \tag{31}$$

Denote the modulus of continuity of a function $\mathfrak{f}$ by

$$\omega(\mathfrak{f},\delta) := \sup_{x,y\in K\,,\,x\neq y\,,\,|x-y|\leq\delta} |\mathfrak{f}(x) - \mathfrak{f}(y)|\,.$$

By (31), on the event $E_t(\mu)$,

$$\omega\left(S_{\sqrt{\epsilon}}\Delta\mathfrak{h}_t,\delta\right) \leq \max\left\{\omega\left(S_{\sqrt{\epsilon}}\Delta\mathfrak{h}_t^{-\mu},\delta\right)\,,\ \omega\left(S_{\sqrt{\epsilon}}\Delta\mathfrak{h}_t^{\mu},\delta\right)\right\} . \tag{32}$$

We note that, for every $\mu \in \mathbb{R}$,

$$S_{\sqrt{\epsilon}}\Delta\mathfrak{h}_t^{\mu}(x) \overset{dist.}{=} \mu\epsilon^{1/2}x + \mathfrak{b}(x) \text{ (as process in } x \in \mathbb{R}) , \tag{33}$$

and we want to tune $\mu = \mu_\epsilon$ in order to have

$$\mathbb{P}\left(E_t(\mu)^c\right) \to 0 \text{ and } \mu\epsilon^{1/2} \to 0 , \text{ as } \epsilon \to 0 .$$

By choosing $\mu_\epsilon = \epsilon^{-1/4}$ we have both (using (28) and Lemma 1), and by (32) and (33), for every $\eta > 0$,

$$\mathbb{P}\left(\omega\left(S_{\sqrt{\epsilon}}\Delta\mathfrak{h}_t, \delta\right) > \eta\right) \leq 2\mathbb{P}\left(\omega\left(\mathfrak{b}, \delta\right) > \eta - \delta\epsilon^{1/4}\right) + \mathbb{P}\left(E_t(\mu_\epsilon)^c\right) .$$

This shows that for every $\eta_1, \eta_2 > 0$ there exist $\delta > 0$ and $\epsilon_0 > 0$ such that

$$\mathbb{P}\left(\omega\left(S_{\sqrt{\epsilon}}\Delta\mathfrak{h}_t, \delta\right) > \eta_1\right) < \eta_2 , \ \forall \epsilon < \epsilon_0 .$$

Since $S_{\sqrt{\epsilon}}\Delta\mathfrak{h}_t(0) = 0$, this implies that the sequence of probability measures in C induced by $S_{\sqrt{\epsilon}}\Delta\mathfrak{h}_t$ is tight. On the other hand, by picking $x = 0$ in (31), $\mu_\epsilon = \epsilon^{-1/4}$ and then using (33), we see that the finite dimensional distributions of $S_{\sqrt{\epsilon}}\Delta\mathfrak{h}_t$ are converging, as $\epsilon \to 0$, to those of $\mathfrak{b}$, which finishes the proof of (13).

3.2 *Proof of Theorem 2*

Recall that

$$\mathscr{A}(x, y) = \mathscr{L}(x, y) + (x - y)^2 ,$$

where $\mathscr{L}(x, y) := \mathscr{L}(x, 0; y, 1)$, and it is sufficient to prove the analog result for $\mathscr{L}$. Since $\Delta\mathscr{L}(0, 0) = 0$, to prove tightness we only need to control the modulus of continuity of the two-dimensional scalar field $\mathscr{L}$. Now we can write

$$\begin{aligned}\epsilon^{-1/2}\left(\mathscr{L}(\epsilon x_2, \epsilon y_2) - \mathscr{L}(\epsilon x_1, \epsilon y_1)\right) &= \epsilon^{-1/2}\left(\mathscr{L}(\epsilon x_2, \epsilon y_2) - \mathscr{L}_1(\epsilon x_2, \epsilon y_1)\right) \\ &+ \epsilon^{-1/2}\left(\mathscr{L}(\epsilon x_2, \epsilon y_1) - \mathscr{L}(\epsilon x_1, \epsilon y_1)\right) .\end{aligned}$$

By the symmetry $\{\mathscr{L}(x, y)\}_{(x,y)\in\mathbb{R}^2} \overset{dist.}{=} \{\mathscr{L}(y, x)\}_{(x,y)\in\mathbb{R}^2}$, it is sufficient to control the supremum of

$$\epsilon^{-1/2}\left(\mathscr{L}(\epsilon x, \epsilon y_2) - \mathscr{L}(\epsilon x, \epsilon y_1)\right) ,$$

over all $(y_1, x), (y_2, x) \in K$ with $|y_1 - y_2| \le \delta$, where K is a fixed compact subset of $\mathbb{R}^2$. Recall that the directed landscape can be expressed as

$$\mathscr{L}(x, y) = \mathfrak{h}_1(y; \mathfrak{d}_x)\,, \text{ where } \mathfrak{d}_x(z) = \begin{cases} 0 & \text{for } z = x \\ -\infty & \text{for } z \neq x\,. \end{cases}$$

Notice also that $Z_1(y; \mathfrak{d}_x) = x$ for all $y \in \mathbb{R}$. Given $K \subseteq \mathbb{R}^2$ compact there exists ϵ_0 such that $\epsilon|x|, \epsilon|y| \le 1$ for all $(x, y) \in K$ and for all $\epsilon < \epsilon_0$. Hence

$$|Z_1(\epsilon y; \mathfrak{d}_{\epsilon x})| = \epsilon|x| \le 1\,, \text{ for all } (x, y) \in K\,,$$

and, on the event that

$$Z_t^{-\mu}(1) \le -1 < 1 \le Z_t^{\mu}(-1)\,, \tag{34}$$

(as in (31)) we have that for all $(x, y_1) \in K$ and $(x, y_2) \in K$, with $y_1 < y_2$,

$$\begin{aligned} \epsilon^{-1/2}\Big(\mathfrak{h}_1^{-\mu}(\epsilon y_2) - \mathfrak{h}_1^{-\mu}(\epsilon y_1)\Big) &\le \epsilon^{-1/2}\Big(\mathfrak{h}_1(\epsilon y_2; \mathfrak{d}_{\epsilon x}) - \mathfrak{h}_1(\epsilon y_1; \mathfrak{d}_{\epsilon x})\Big) \\ &\le \epsilon^{-1/2}\Big(\mathfrak{h}_1^{\mu}(\epsilon y_2) - \mathfrak{h}_1^{\mu}(\epsilon y_1)\Big)\,. \end{aligned}$$

For $\mu = \mu_\epsilon = \epsilon^{-1/4}$, (34) occurs with high probability as $\epsilon \to 0$, and under (34), for all $x \in \mathbb{R}$ such that $(x, y) \in K$ for some $y \in \mathbb{R}$, we have that

$$\omega\left(S_{\sqrt{\epsilon}}\Delta\mathfrak{h}_t(\cdot; \mathfrak{d}_{\epsilon x}), \delta\right) \le \max\left\{\omega\left(S_{\sqrt{\epsilon}}\Delta\mathfrak{h}_t^{-\mu}, \delta\right)\,,\, \omega\left(S_{\sqrt{\epsilon}}\Delta\mathfrak{h}_t^{\mu}, \delta\right)\right\}\,.$$

From here one can follow the proof of Theorem 1 to conclude tightness and marginal local Brownian behaviour. From the same argument, one can get 1/2-Holder regularity of the Airy Sheet.

To prove independence we have to change the comparison set up, and we do it by splitting the space-time directed landscape at time $s = 1/2$. For $x, y \in \mathbb{R}$ consider

$$Z_{1/2}(x, y) = \mathscr{P}_{x,0}^{y,1}(1/2)\,,$$

i.e. the location at time $s = 1/2$ of the rightmost geodesic between $(x, 0)$ and $(y, 1)$. Thus, by metric composition (3),

$$\begin{aligned} \mathscr{L}(x, y) &= \max_{z\in\mathbb{R}}\{\mathscr{L}(x, 0; z, 1/2) + \mathscr{L}(z, 1/2; y, 1)\} \\ &= \mathscr{L}(x, 0; Z_{1/2}, 1/2) + \mathscr{L}(Z_{1/2}, 1/2; y, 1)\,. \end{aligned}$$

As in the proof of (iv)-Proposition 2, we have monotonicity of geodesics as follows: for all $x_1 \le x_2$ and $y_1 \le y_2$ then

$$\mathscr{P}_{x_1,0}^{y_1,1}(s) \le \mathscr{P}_{x_2,0}^{y_2,1}(s)\,,\ \forall\, s \in [0,1]\,,$$

and, in particular,

$$Z_{1/2}(x_1, y_1) \le Z_{1/2}(x_2, y_2)\,. \tag{35}$$

Let

$$\bar{\mathfrak{h}}_{1/2+}(y;\mathfrak{h}) := \max_{z\in\mathbb{R}} \{\mathfrak{h}(z) + \mathscr{L}(z, 1/2; y, 1)\}$$

and

$$\bar{\mathfrak{h}}_{1/2-}(x;\mathfrak{h}) := \max_{z\in\mathbb{R}} \{\mathfrak{h}(z) + \mathscr{L}(x, 0; z, 1/2)\}\,.$$

Then, by metric composition (3),

$$\mathscr{L}(x, y) = \bar{\mathfrak{h}}_{1/2+}(y; \bar{\mathfrak{h}}_{+,x}) \text{ and } \mathscr{L}(x, 0) = \bar{\mathfrak{h}}_{1/2-}(x; \bar{\mathfrak{h}}_{-})\,,$$

where

$$\bar{\mathfrak{h}}_{+,x}(z) = \mathscr{L}(x, 0; z, 1/2) = \mathfrak{h}_{1/2}(z; \mathfrak{d}_x) \text{ and } \bar{\mathfrak{h}}_{-}(z) = \mathscr{L}(z, 1/2; 0, 1)\,.$$

Therefore,

$$\begin{aligned}\mathscr{L}(x, y) - \mathscr{L}(0, 0) &= \mathscr{L}(x, y) - \mathscr{L}(x, 0) + \mathscr{L}(x, 0) - \mathscr{L}(0, 0)\\ &= \Delta\bar{\mathfrak{h}}_{1/2+}(y; \bar{\mathfrak{h}}_{+,x}) + \Delta\bar{\mathfrak{h}}_{1/2-}(x; \bar{\mathfrak{h}}_{-})\,.\end{aligned}$$

The trick now is to pick $\mathfrak{b}_1$ and $\mathfrak{b}_2$, two independent copies of $\mathfrak{b}$, and then apply the coupling method to compare simultaneously $\Delta\bar{\mathfrak{h}}_{1/2+}(y; \bar{\mathfrak{h}}_{+,x})$ with $\Delta\bar{\mathfrak{h}}_{1/2+}(y; \mathfrak{b}_1^{\mu})$, and $\Delta\bar{\mathfrak{h}}_{1/2-}(y; \bar{\mathfrak{h}}_{-})$ with $\Delta\bar{\mathfrak{h}}_{1/2-}(y; \mathfrak{b}_2^{\mu})$. By time independence and stationarity (2) of the directed landscape, we clearly have that $\bar{\mathfrak{h}}_{1/2+}(\cdot; \mathfrak{b}_1^{\mu})$ and $\bar{\mathfrak{h}}_{1/2-}(\cdot; \mathfrak{b}_2^{\mu})$ are independent processes, and

$$\Delta\bar{\mathfrak{h}}_{1/2+}(\cdot; \mathfrak{b}_1^{\mu}) \overset{dist.}{=} \mathfrak{b}^{\mu} \overset{dist.}{=} \Delta\bar{\mathfrak{h}}_{1/2-}(\cdot; \mathfrak{b}_2^{\mu})\,.$$

Let

$$\bar{Z}_{1/2+}(y, \mathfrak{h}) := \max \arg\max_{z\in\mathbb{R}} \{\mathfrak{h}(z) + \mathscr{L}(z, 1/2; y, 1)\}\,,$$

and

$$\bar{Z}_{1/2-}(x,\mathfrak{h}) := \max \arg\max_{z\in\mathbb{R}} \{\mathfrak{h}(z) + \mathscr{L}(x,0;z,1/2)\} \; .$$

Hence,

$$\bar{Z}_{1/2+}(y,\bar{\mathfrak{h}}_{+,x}) = Z_{1/2}(x,y) \text{ and } \bar{Z}_{1/2-}(x,\bar{\mathfrak{h}}_{-}) = Z_{1/2}(x,0) \, .$$

Let

$$\bar{E}_{1/2+}(\mu) := \left\{ \bar{Z}_{1/2+}(-1,\mathfrak{b}_1^{\mu}) \geq Z_{1/2}(1,1) \text{ and } \bar{Z}_{1/2+}(1,\mathfrak{b}_1^{-\mu}) \leq Z_{1/2}(-1,-1) \right\} \; ,$$

and

$$\bar{E}_{1/2-}(\mu) := \left\{ \bar{Z}_{1/2-}(-1,\mathfrak{b}_2^{\mu}) \geq Z_{1/2}(1,0) \text{ and } \bar{Z}_{1/2+}(1,\mathfrak{b}_2^{-\mu}) \leq Z_{1/2}(-1,0) \right\} \; .$$

For a compact set $K \subseteq \mathbb{R}^2$, chose ϵ_0 so that $\epsilon|x|, \epsilon|y| \leq 1$ for all $(x, y) \in K$ and $\epsilon < \epsilon_0$. By (35),

$$Z_{1/2}(-1,-1) \leq Z_{1/2}(\epsilon x,\epsilon y) \leq Z_{1/2}(1,1)$$

and

$$Z_{1/2}(-1,0) \leq Z_{1/2}(\epsilon x,0) \leq Z_{1/2}(1,0) \; .$$

Denote

$$\bar{\mathfrak{h}}^{\pm\mu}_{+1/2}(\cdot) \equiv \bar{\mathfrak{h}}_{+1/2}(\cdot;\mathfrak{b}_1^{\pm\mu}) \text{ and } \bar{\mathfrak{h}}^{\pm\mu}_{-1/2}(\cdot) \equiv \bar{\mathfrak{h}}_{-1/2}(\cdot;\mathfrak{b}_2^{\pm\mu}) \, .$$

On the event $\bar{E}_{1/2+}(\mu)$, for all $(x, y) \in K$, if $0 < y$ then

$$\bar{\mathfrak{h}}^{-\mu}_{1/2+}(\epsilon y) - \bar{\mathfrak{h}}^{-\mu}_{1/2+}(0) \leq \bar{\mathfrak{h}}_{1/2+}(\epsilon y;\mathfrak{d}_{\epsilon x}) - \bar{\mathfrak{h}}_{1/2+}(0;\mathfrak{d}_{\epsilon x}) \leq \bar{\mathfrak{h}}^{\mu}_{1/2+}(\epsilon y) - \bar{\mathfrak{h}}^{\mu}_{1/2+}(0) \, ,$$

while if $y < 0$ then

$$\bar{\mathfrak{h}}^{\mu}_{1/2+}(\epsilon y) - \bar{\mathfrak{h}}^{\mu}_{1/2+}(0) \leq \bar{\mathfrak{h}}_{1/2+}(\epsilon y;\mathfrak{d}_{\epsilon x}) - \bar{\mathfrak{h}}_{1/2+}(0;\mathfrak{d}_{\epsilon x}) \leq \bar{\mathfrak{h}}^{-\mu}_{1/2+}(\epsilon y) - \bar{\mathfrak{h}}^{-\mu}_{1/2+}(0) \, .$$

On the event $\bar{E}_{1/2-}(\mu)$ for all $(x, y) \in K$, if $0 < y$ then

$$\bar{\mathfrak{h}}^{-\mu}_{1/2-}(\epsilon y) - \bar{\mathfrak{h}}^{-\mu}_{1/2-}(0) \leq \bar{\mathfrak{h}}_{1/2-}(\epsilon y;\bar{\mathfrak{h}}_{-}) - \bar{\mathfrak{h}}_{1/2-}(0;\bar{\mathfrak{h}}_{-}) \leq \bar{\mathfrak{h}}^{\mu}_{1/2-}(\epsilon y) - \bar{\mathfrak{h}}^{\mu}_{1/2-}(0) \, ,$$

while if $y < 0$ then

$$\bar{\mathfrak{h}}^{\mu}_{1/2-}(\epsilon y) - \bar{\mathfrak{h}}^{\mu}_{1/2-}(0) \leq \bar{\mathfrak{h}}_{1/2-}(\epsilon y; \bar{\mathfrak{h}}_{-}) - \bar{\mathfrak{h}}_{1/2-}(0; \bar{\mathfrak{h}}_{-}) \leq \bar{\mathfrak{h}}^{-\mu}_{1/2-}(\epsilon y) - \bar{\mathfrak{h}}^{-\mu}_{1/2-}(0) .$$

Thus, for $\mu = \mu_\epsilon = \epsilon^{-1/4}$, on the event $\bar{E}_{1/2+}(\mu) \cap \bar{E}_{1/2-}(\mu)$, one can approximate the finite dimensional distributions of $(S_{\sqrt{\epsilon}} \Delta \bar{\mathfrak{h}}_{1/2+}, S_{\sqrt{\epsilon}} \Delta \bar{\mathfrak{h}}_{1/2-})$ using the finite dimensional distributions of $(\mathfrak{b}_1, \mathfrak{b}_2)$ (as in the proof of Theorem 1). Since

$$\mathbb{P}\left(\bar{E}_{1/2+}(\mu) \cap \bar{E}_{1/2-}(\mu)\right) \to 1 \text{ as } \epsilon \to 0 ,$$

this finishes the proof Theorem 2.

3.3 *Proof of Theorem 3*

Recall (26) and (27). By Proposition 4 (attractiveness),

$$\Delta \mathfrak{h}^{-\mu}_t(x) \leq \Delta \mathfrak{h}^0_t(x) \leq \Delta \mathfrak{h}^{+\mu}_t(x) , \text{ for } x \geq 0 ,$$

and

$$\Delta \mathfrak{h}^{+\mu}_t(x) \leq \Delta \mathfrak{h}^0_t(x) \leq \Delta \mathfrak{h}^{-\mu}_t(x) , \text{ for } x \leq 0 .$$

Furthermore,[3]

$$0 \leq \Delta \mathfrak{h}^{+\mu}_t(x) - \Delta \mathfrak{h}^{-\mu}_t(x) \leq \Delta \mathfrak{h}^{+\mu}_t(a) - \Delta \mathfrak{h}^{-\mu}_t(a) , \forall x \in [0, a] ,$$

and

$$0 \leq \Delta \mathfrak{h}^{-\mu}_t(x) - \Delta \mathfrak{h}^{+\mu}_t(x) \leq \Delta \mathfrak{h}^{-\mu}_t(-a) - \Delta \mathfrak{h}^{+\mu}_t(-a) , \forall x \in [-a, 0] .$$

By time invariance (10), $\Delta \mathfrak{h}^{\pm\mu}_t(x)$ is a two-sided Brownian motion with drift $\pm\mu$. Hence

$$\mathbb{E}\left(\Delta \mathfrak{h}^{+\mu}_t(a) - \Delta \mathfrak{h}^{-\mu}_t(a)\right) = \mathbb{E}\left(\Delta \mathfrak{h}^{-\mu}_t(-a) - \Delta \mathfrak{h}^{+\mu}_t(-a)\right) = 2\mu a .$$

Consider the event $E_t(\mu)$ (29). By (30),

$$\Delta \mathfrak{h}^{-\mu}_t(x) \leq \Delta \mathfrak{h}_t(x; \mathfrak{h}) \leq \Delta \mathfrak{h}^{+\mu}_t(x) , \text{ for } x \in [0, a] ,$$

[3]It also follows from Proposition 4 that $\Delta \mathfrak{h}^{+\mu}_t(x) - \Delta \mathfrak{h}^{-\mu}_t(x)$ is a nondecreasing function of x.

and

$$\Delta\mathfrak{h}_t^{+\mu}(x) \le \Delta\mathfrak{h}_t(x;\mathfrak{h}) \le \Delta\mathfrak{h}_t^{-\mu}(x)\,, \text{ for } x \in [-a, 0]\,.$$

Thus, if $E_t(\mu)$ occurs then both $\Delta\mathfrak{h}_t^0(\cdot) = \Delta\mathfrak{h}_t(\cdot;\mathfrak{b})$ and $\Delta\mathfrak{h}_t(\cdot;\mathfrak{h})$ are sandwiched by $\Delta\mathfrak{h}_t^{\pm\mu}(\cdot)$, which implies the following uniform control on the distance between $\Delta\mathfrak{h}_t(\cdot;\mathfrak{h})$ and $\Delta\mathfrak{h}_t(\cdot;\mathfrak{b})$:

$$\sup_{x\in[-a,a]} |\Delta\mathfrak{h}_t(x;\mathfrak{h}) - \Delta\mathfrak{h}_t(x;\mathfrak{b})| \le I_t(a)\,,$$

where

$$0 \le I_t(a) = \Delta\mathfrak{h}_t^{+\mu}(a) - \Delta\mathfrak{h}_t^{-\mu}(a) + \Delta\mathfrak{h}_t^{-\mu}(-a) - \Delta\mathfrak{h}_t^{+\mu}(-a)\,.$$

Therefore, using Markov inequality and that $\mathbb{E}\,(I_t(a)) = 4\mu a$,

$$\begin{aligned}\mathbb{P}\left(\sup_{x\in[-a,a]} |\Delta\mathfrak{h}_t(x;\mathfrak{h}) - \Delta\mathfrak{h}_t(x;\mathfrak{b})| > \eta\sqrt{a}\right) &\le \mathbb{P}\left(E_t(\mu)^c\right) + \frac{\mathbb{E}\,(I_t(a))}{\eta\sqrt{a}}\\ &= \mathbb{P}\left(E_t(\mu)^c\right) + \frac{4\mu\sqrt{a}}{\eta}\,.\end{aligned}$$

In order to make this inequality useful, we have to chose $\mu = \mu_t$ in such way that

$$\mathbb{P}\left(E_t(\mu)^c\right) \to 0 \text{ and } \mu\sqrt{a} \to 0\,, \text{ as } t \to \infty$$

(we allow $a = a_t$ as well). For $t \ge a^{3/2}$ we have that $\pm a t^{-2/3}$ does not play any rule in the asymptotic analysis of $E_t(\mu)$ (recall (28)). By Lemma 2, we know that

$$\mathbb{P}\left(|Z_t(\pm a;\mathfrak{h})| > r t^{2/3}\right) \to 0\,, \text{ as } r \to \infty\,,$$

(uniformly in t). Thus, by (28), $E_t(\mu)$ should occur with high probability, as soon as $\pm\mu t^{1/3} \to \pm\infty$. By setting $\mu = r(4t^{1/3})^{-1}$, for some $r = r_t \to \infty$, then

$$4\mu\sqrt{a} = r(a t^{-2/3})^{1/2}\,.$$

A natural choice is $r_t = (a t^{-2/3})^{-\delta}$ with $\delta \in (0, 1/2)$, and for the sake of simplicity we take $\delta = 1/4$, which yields to

$$\mathbb{P}\Big(\sup_{x\in[-a,a]} |\Delta\mathfrak{h}_t(x;\mathfrak{h}) - \Delta\mathfrak{h}_t(x;\mathfrak{b})| > \eta\sqrt{a}\Big) \le \mathbb{P}\left(E_t(\mu)^c\right) + \frac{1}{\eta r_t}\,. \tag{36}$$

Therefore, Theorem 3 is a consequence of (36) and Lemma 3 below.

Lemma 3 *Let* $\mu := r(4t^{1/3})^{-1}$. *Then, under assumption* (14)*, there exists a function* ϕ*, that does not depend on* $a, t > 0$*, such that for all* $t \geq \max\{c^3, a^{3/2}\}$

$$\mathbb{P}\left(E_t(\mu)^c\right) \leq \phi(r) \ \ \textit{and} \ \ \lim_{r\to\infty}\phi(r) = 0\,.$$

Proof By the definition of $E_t(\mu)$,

$$E_t(\mu)^c \cap \left\{|Z_t(a;\mathfrak{h})| \leq \frac{r}{16}t^{2/3}\right\} \subseteq \left\{Z_t^{\mu}(-a) \leq \frac{r}{16}t^{2/3}\right\} \cup \left\{Z_t^{-\mu}(a) \geq -\frac{r}{16}t^{2/3}\right\},$$

and hence, $\mathbb{P}\left(E_t(\mu)^c\right)$ is bounded by

$$\mathbb{P}\left(|Z_t(a;\mathfrak{h})| > \frac{r}{16}t^{2/3}\right) + \mathbb{P}\left(Z_t^{\mu}(-a) \leq \frac{r}{16}t^{2/3}\right) + \mathbb{P}\left(Z_t^{-\mu}(a) \geq -\frac{r}{16}t^{2/3}\right). \tag{37}$$

By Lemma 2, we only need to show that there exists a function ϕ_2, that does not depend on $a, t > 0$, such that for all $t \geq \max\{c^3, a^{3/2}\}$

$$\max\left\{\mathbb{P}\left(Z_t^{\mu}(-a) \leq \frac{r}{16}t^{2/3}\right), \mathbb{P}\left(Z_t^{-\mu}(a) \geq -\frac{r}{16}t^{2/3}\right)\right\} \leq \phi_2(r),$$

and $\lim_{r\to\infty}\phi_2(r) = 0$. Since $\mu := r(4t^{1/3})^{-1}$ and $t \geq a^{3/2}$, by (28),

$$\begin{aligned}\mathbb{P}\left(Z_t^{\mu}(-a) \leq \frac{r}{16}t^{2/3}\right) &= \mathbb{P}\left(Z_1^0(0) \leq -\frac{r}{16} + at^{-2/3}\right)\\ &\leq \mathbb{P}\left(Z_1^0(0) \leq -\left(\frac{r}{16} - 1\right)\right),\end{aligned}$$

and

$$\begin{aligned}\mathbb{P}\left(Z_t^{-\mu}(a) \geq -\frac{r}{16}t^{2/3}\right) &= \mathbb{P}\left(Z_1^0(0) \geq \frac{r}{16} - at^{-2/3}\right)\\ &\leq \mathbb{P}\left(Z_1^0(0) \geq \frac{r}{16} - 1\right),\end{aligned}$$

which allows us to take $\phi_2(r) := \mathbb{P}\left(|Z_1^0(0)| > \frac{r}{16} - 1\right)$. Therefore, together with (37), this shows that

$$\mathbb{P}\left(E_t(\mu)^c\right) \leq \phi_1(r) + 2\phi_2(r),$$

and finishes the proof of Lemma 3. ∎

Acknowledgments This research was supported in part by the International Centre for Theoretical Sciences (ICTS) during a visit for participating in the program—Universality in random structures: Interfaces, Matrices, Sandpiles (Code: ICTS/urs2019/01), and also by the CNPQ grants 421383/2016-0, 302830/2016-2 and 305356/2019-4, and by the FAPERJ grant E-26/203.048/2016. The author would like to thank Patrik Ferrari, Daniel Remenik for useful comments and enlightening discussions concerning this subject, and to an anonymous referee for the review of the article.

References

1. Amir, G., Corwin, I., Quastel, J.: Probability distribution of the free energy of the continuum directed random polymer in 1 + 1 dimensions. Commun. Pure Appl. Math. **64**, 466–537 (2011)
2. Baik, J., Deift, P.A., Johansson, K.: On the distribution of the length of the longest increasing subsequence of random permutations. J. Am. Math. Soc. **12**, 1119–1178 (1999)
3. Balázs, M., Seppäläinen, T.: Fluctuation bounds for the asymmetric simple exclusion process. ALEA Lat. Am. J. Probab. Math. Stat. **6**, 1–24 (2009)
4. Balázs, M., Cator,E.A., Seppäläinen, T.: Cube root fluctuations for the corner growth model associated to the exclusion process. Electron. J. Probab. **11**, 1094–1132 (2006)
5. Balázs, M., Quastel, J., Seppäläinen, T.: Scaling exponent for the Hopf-Cole solution of KPZ/Stochastic Burgers. J. Am. Math. Soc. **24**, 683–708 (2011)
6. Borodin, A., Ferrari, P.L., Prähofer, M., Sasamoto, T.: Fluctuation properties of the TASEP with periodic initial configuration. J. Stat. Phys. **129**, 1055–1080 (2007)
7. Cator, E.A., Groeneboom, P.: Second class particles the cube root asymptotics for Hammersley's process. Ann. Probab. **34**, 1273–1295 (2006)
8. Cator, E.A., Pimentel, L.P.R.: On the local fluctuations of last-passage percolation models. Stoch. Proc. Appl. **125**, 879–903 (2012)
9. Corwin, I.: The Kardar-Parisi-Zhang equation and universality. Random Matrices: Theory Appl. (2012). https://doi.org/10.1142/S2010326311300014
10. Corwin, I., Hammond, A.: Brownian Gibbs property for Airy line ensembles. Invent. Math. **195**, 441–508 (2014)
11. Corwin, I., Quastel, J., Remenik, D.: Renormalization fixed point of the KPZ universality class. J. Stat. Phys. **160**, 815–834 (2015)
12. Corwin, I., Liu, Z., Wang, D.: Fluctuations of TASEP and LPP with general initial data. Ann. Appl. Probab. **26**, 2030–2082 (2016)
13. Dauvergne, D., Ortmann, J., Virág, B.: The directed landscape. Available via arXiv. https://arxiv.org/pdf/1812.00309.pdf
14. Ferrari, P.L., Occelli, A.: Universality of the GOE Tracy-Widom distribution for TASEP with arbitrary particle density. Electron. J. Probab. (2018). https://doi.org/10.1214/18-EJP172
15. Harris, T.E.: Additive set-valued Markov processes and graphical methods. Ann. Probab. **6**, 355–378 (1978)
16. Johansson, K.: Shape fluctuations and random matrices. Commun. Math. Phys. **209**, 437–476 (2000)
17. Johansson, K.: Discrete polynuclear growth and determinantal processes. Commun. Math. Phys. **242**, 277–239 (2003)
18. Kardar, M., Parisi, G., Zhang, Y.C.: Dynamic scaling of growing interfaces. Phys. Rev. Lett. **56**, 889–892 (1986)
19. Matetski, K., Quastel, J., Remenik, D.: The KPZ fixed point. Available via arXiv. https://arxiv.org/pdf/1701.00018.pdf
20. Nica, M., Quastel, J., Remenik, D.: One-sided reflected Brownian motions and the KPZ fixed point. Available via arXiv. https://arxiv.org/pdf/2002.02922.pdf

21. O' Connell, N., Yor, M.: Brownian analogues of Burke's theorem. Stoch. Proc. Appl. **96**, 285–304 (2001)
22. Pimentel, L.P.R.: On the location of the maximum of a continuous stochastic process. J. Appl. Probab. **173**, 152–161 (2014)
23. Pimentel, L.P.R.: Local behavior of Airy processes. J. Stat. Phys. **173**, 1614–1638 (2018)
24. Pimentel, L.P.R.: Ergodicity of the KPZ fixed point. Available via arXiv. https://arxiv.org/pdf/1708.06006.pdf
25. Prähofer, M., Spohn, H.: Scale invariance of the PNG droplet and the Airy process. J. Stat. Phys. **108**, 1071–1106 (2002)
26. Seppäläinen, T.: Scaling for a one-dimensional directed polymer with boundary conditions. Ann. Probab. **40**, 19–73 (2012)
27. Takeuchi, K.A., Sano, M., Sasamoto, T., Spohn, H.: Growing interfaces uncover universal fluctuations behind scale invariance. Sci. Rep. (2011). https://doi.org/10.1038/srep00034

How Can the Appropriate Objective and Predictive Probabilities Get into Non-collapse Quantum Mechanics?

Roberto H. Schonmann

Abstract It is proved that in non-collapse quantum mechanics the state of a closed system can always be expressed as a superposition of states all of which describe histories that conform to Born's probability rule. This theorem allows one to see Born probabilities in non-collapse quantum mechanics as an appropriate predictive tool, implied by the theory, provided an appropriate version of the superposition principle is included in its axioms

Keywords Non-collapse quantum mechanics · Everett · Born's rule · Origin of probability in quantum mechanics

1 Introduction

This is a shorter version of the paper [17], where the reader will find a much more detailed and thorough discussion of the relevance of the theorem introduced here, as well as further comparison of the role of probabilities in collapse and non-collapse quantum mechanics. This version is being written in memory of Vladas Sidoravicius, whose premature death shocked and saddened his friends and colleagues, and whose interests focused on probability theory not only in the abstract, but especially as it relates to physics. Vladas' passing happened close in time to that of Harry Kesten, good friend and mentor to both of us and to so many others. This paper is also dedicated to his memory.

For mathematicians who may need an introduction to quantum mechanics, I recommend the text [11]. (Chapters 1 and 3 suffice for the purposes of this paper.)

This paper deals with an important aspect of what is known as the "measurement problem in quantum mechanics". In standard quantum mechanics the state of a system (which is a vector in a Hilbert space) evolves in two distinct and

R. H. Schonmann (✉)
Department of Mathematics, University of California at Los Angelesss, Los Angeles, CA, USA
e-mail: rhs@math.ucla.edu

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_30

incompatible fashions, and it is unclear when each one applies. When it is not being observed it evolves in a deterministic, continuous way, according to certain unitary transformations. But when observed, the system evolves in a probabilistic, discontinuous way (called a collapse, or reduction of the state), jumping to a new state according to a probabilistic prescription called Born's rule (we will refer to this as the collapse axiom, or the probability axiom). But what constitutes an "observation"? After all the "observers" (whether we are able to include in this class humans, other animals, robots, photographic plates, ...) should be considered as part of the system, so that "observations" should not have distinct physics. Non-collapse quantum mechanics (first introduced in [9]) proposes that the collapse axiom be eliminated from quantum mechanics, and claims that we would still have the same experiences that we predict from standard quantum mechanics. Instead of collapses happening, the system always evolves in the deterministic unitary fashion, and this implies that at the end of each experiment all the possible outcomes materialize, including one version of the "observers" (possibly humans) associated to each outcome, perceiving and recording that outcome and no other. This accounts for our observation of collapses as illusions, so to speak. But then, what accounts for them following Born's probability rule, rather than some other probability rule, or no probability rule at all? This is the focus of this paper (see below for references and some comments on the extensive work already available on this fundamental issue).

Before we proceed, a few words about terminology. We will use the expression "Born-rule collapse quantum mechanics" for the standard quantum mechanics theory, as presented in our textbooks, including the assumption that measurements lead to collapses of the state of the system according to Born's rule. "Collapse quantum mechanics" will be used for a broader set of theories, in which the collapses follow some probability distribution that may or not be the one given by Born's rule. And by "non-collapse quantum mechanics" we simply mean that we eliminate the assumption of collapse when measurements are performed. In non-collapse quantum mechanics, we do not include the words "measurement" or "observation" in the axioms of the theory, and use them only informally when applying the theory to explain and predict our experiences.

Readers who want an introduction to non-collapse quantum mechanics will benefit from the classic [8], where papers by those who first proposed and advertised it as a (better) alternative to collapse quantum mechanics are collected. The subject is not standard in textbooks geared to physicists, or mathematicians, but is standard in texts concerned with the philosophy of quantum mechanics; see, e.g., [2–4, 21] and [12]. For a positive appraisal of the theory, written for the general scientific public, see, e.g., [19]. For expositions for the general public, see, e.g., [5, 20] and [6]. For a recent collection of mostly philosophical discussions see [16]. And for some among the many research papers on the subject, see, e.g., [1, 7, 18] and [13], which also provide extensive additional references.

Our concern here is with the origin of our perception of Born-rule probabilities in a theory, non-collapse quantum mechanics, in which everything is deterministic and, in particular, no probabilities are introduced in its axioms. A great deal has been written about this problem, e.g., in the references cited in the last paragraph and references therein, with opinions ranging from "the problem is solved" (sometimes by the authors themselves), to "the problem is hopeless and the proposed solutions all flawed". This project was motivated by my dissatisfaction with the previously proposed solutions, especially with the current trend of treating the probabilities in non-collapse quantum mechanics as subjective ones ([7, 13, 18, 21]; see for instance Chapter 6 of [12] for a criticism). I hope nevertheless to convince the reader that the theorem stated and proved here provides a solution to this puzzle and explains how our perception of probabilities, as given by Born's rule, emerges in non-collapse quantum mechanics, if we include in its axioms an appropriate version of the superposition principle. I propose even that the puzzle be turned around: If collapses do happen, why do they happen precisely with the same rule that comes out of quantum mechanics without collapse?

In Sect. 2 we will state the theorem alluded to in the abstract, in a mathematically self-contained fashion, but without emphasizing the corresponding physics, which will then be briefly discussed in Sect. 3. (For a longer discussion the reader is referred to [17].) To help the reader keep in mind what is planned, we include next a few words of introduction on how the mathematical setting in Sect. 2 is motivated by collapse quantum mechanics.

We will be working in the Heisenberg picture (operators evolve in time, rather than states), as applied to a closed system (possibly the whole universe). Associated to the system there is a Hilbert space $\mathscr{H}$ (not assumed in this paper to be necessarily separable). The state of the system is given at any time by a non-null vector in $\mathscr{H}$ (with non-null scalar multiples of a vector corresponding to the same state). This state does not change with time except when there is a collapse. Collapses are associated with measurements and with their corresponding self-adjoint operators (which in the Heisenberg picture are time dependent). In each collapse, the state immediately after the collapse is a projection of the state immediately before the collapse on a subspace (a subset of $\mathscr{H}$ closed linearly and topologically) chosen at random, according to a specified probability law (in the standard case, Born's rule), from among the eigenspaces of that operator, one eigenspace for each possible outcome of the experiment. (To avoid unnecessary mathematical complications, and on physical grounds, we are assuming that every experiment can only have a finite number of possible outcomes.) To each subspace of $\mathscr{H}$ there is associated a projection operator (self-adjoint idempotent operator on $\mathscr{H}$) that projects on that subspace. If initially the state was a vector $\psi \in \mathscr{H}$, then immediately after a collapse the state can be expressed as Proj ψ, where Proj is the composition of the projections that took place after each collapse, up to and including this last one.

It is natural to represent all the possible ways in which the system can evolve using a rooted (oriented) tree. The root vertex of the tree will correspond to the beginning of times for the system under study, and the other vertices will either correspond to collapse events, or be terminal vertices (vertices of degree 1) that signal that no further experiment is performed along a branch of the tree. (In the interesting cases the tree will be infinite. One can think of terminal vertices as uncommon in the tree, possibly even absent.) The projections associated to the possible outcomes in the collapses, as described at the end of the last paragraph, will then be indexed by the edges of the tree. The tree does not have to be homogeneous, as, e.g., decisions on what experiments to perform in a lab may depend on the outcomes of previous experiments. More interesting and dramatic examples of non-homogeneity of the tree occur if one thinks of some major human decisions being made by use of "quantum coins", i.e., outcomes of experiments performed for this purpose (depending on these decisions the future of humanity may take quite different turns).

After stating the theorem in Sect. 2 and then briefly discussing its relevance in Sect. 3, we will prove it in Sect. 4.

2 The Theorem

Let $(\mathbf{V}, \mathbf{E})$ be a tree with vertex set $\mathbf{V}$, including a singled out vertex called the root vertex, and edge set $\mathbf{E}$. We assume that the root vertex has a single edge incident to it and call it the root edge. Such a tree will be called an edge-rooted tree. We orient the root edge from the root vertex to its other end, and we give an orientation to every edge in the tree, so that each vertex other than the root vertex has exactly one edge oriented towards it. If e is the edge oriented towards vertex v and $e_1, \ldots, e_n$ are the edges incident to v and oriented away from it, we call $e_1, \ldots, e_n$ the children of e, and we refer to $\{e_1, \ldots e_n\}$ as a set of siblings and to e as their parent. (The advantage of using such "family" language, even if a bit funny, is that the terminology becomes easy to remember and easy to extend.) Childless edges will be called terminal edges, and the vertices to which terminal edges point will be called terminal vertices. Each edge belongs to a generation defined inductively by declaring the generation of the root edge as 1, and the generation of the children of the edges of generation i to be $i + 1$. It will be convenient to declare that childless edges that belong to generation i also belong to generations $i + 1, i + 2, \ldots$ A partial history is a finite sequence of edges $(e_1, e_2, \ldots, e_n)$, where each e_i is a child of e_{i-1}, $i = 2, \ldots, n$. A complete history (or just a history) is either a partial history in which e_1 is the root edge and the last edge is a terminal edge, or an infinite sequence of edges $(e_1, e_2, \ldots)$, where e_1 is the root edge and each e_i is a child of e_{i-1}, $i = 2, \ldots$.

Definition 1 A tree-structured set of projections on a Hilbert space $\mathscr{H}$ is a collection of such projections, $\mathscr{P} = \{Proj_e : e \in \mathbf{E}\}$, where the index set $\mathbf{E}$ is the set of edges of an edge-rooted tree, and the following conditions are satisfied:

1. If e is the root edge, then $Proj_e$ is the identity operator.
2. If $e_1, \ldots, e_n$ are the children of e, then $\sum_{i=1}^{n} Proj_{e_i} = Proj_e$.

We write $\mathscr{H}_e = Proj_e \mathscr{H}$ for the subspace associated to $Proj_e$. The first condition means that $\mathscr{H}_e = \mathscr{H}$ when e is the root edge, while the second one means that the subspaces $\mathscr{H}_{e_i}$ associated to a set of siblings $\{e_1, \ldots, e_n\}$ are orthogonal to each other and their linear span is the subspace $\mathscr{H}_e$ associated to the parent e.

Implicit in the definition of a tree-structured set of projections $\mathscr{P}$ is the associated edge-rooted tree $(\mathbf{V}, \mathbf{E})$. The set of histories on this tree, denoted Ω, is the sample space on which one defines Born's probabilities (and alternative ones) associated to $\mathscr{P}$. Recall that, informally speaking, an element $\omega \in \Omega$ is a sequence of edges starting from the root and having each of its elements succeeded by one of its children, either with no end, or ending at a terminal edge. Abusing notation, we will write $e \in \omega$ for the statement that the edge e is an element of the sequence ω. For each $e \in \mathbf{E}$, we define $\Omega_e = \{\omega : e \in \omega\}$, the set of histories that go through e. Unions of finitely many sets Ω_e define an algebra of sets (a class of sets that is closed with respect to complements, finite unions and finite intersections) that we denote by $\mathscr{A}$. (This statement requires a proof, which is easily obtained by noting that every set $A \in \mathscr{A}$ can be written as a union over sets Ω_e with all e in the same generation, and that A^c is then the union of the sets Ω_e over the other e belonging to this same generation. This shows closure under complements. Closure under unions is immediate and De Morgan's law then provides closure under intersections.) The smallest σ-algebra that contains $\mathscr{A}$ will be denoted by $\mathscr{B}$.

Born's probabilities are defined on the measure space $(\Omega, \mathscr{B})$ and, in addition to $\mathscr{P}$, depend on a vector $\psi \in \mathscr{H}\setminus\{0\}$. (In the theorem below, ψ is arbitrary, but in all our applications it will be the initial state of our system. In collapse quantum mechanics, ψ will be chosen as the state, in the Heisenberg picture, before collapses. In non-collapse quantum mechanics, ψ will be chosen as the unchanging state, in the Heisenberg picture.) Born's probability corresponding to ψ will be denoted by $\mathbb{P}_\psi$. It is described informally by imagining a walker that moves on the edges of the tree. The walker starts at the root vertex of the tree and then moves in the direction of the orientation, deciding at each vertex where to go in a probabilistic fashion, with edges chosen with probability proportional to norm-squared, i.e., when at a vertex that separates a parent e from its children, the walker chooses child e' with probability $||\mathrm{Proj}_{e'}\psi||^2/||\mathrm{Proj}_e\psi||^2$, independently of past choices. If ever at a terminal vertex, the walker stops. A simple inductive computation shows that this is equivalent to the statement

$$\mathbb{P}_\psi(\Omega_e) = \frac{||\mathrm{Proj}_e\psi||^2}{||\psi||^2}, \quad \text{for each } e \in \mathbf{E}. \tag{1}$$

It is standard to show that (1) extends in a unique fashion to $\mathscr{A}$ and then to $\mathscr{B}$, defining in this way a unique probability measure on $(\Omega, \mathscr{B})$. Actually, for our purposes it will be important to observe that this standard procedure yields even more. For each $\psi \in \mathscr{H}$, the extension of the probability measure is to a larger measure space, $(\Omega, \mathscr{M}_\psi)$, where $\mathscr{M}_\psi \supset \mathscr{B}$, completes $\mathscr{B}$ with respect to the measure $\mathbb{P}_\psi$, meaning in particular that if $A \in \mathscr{B}$, $\mathbb{P}_\psi(A) = 0$ and $B \subset A$, then also $B \in \mathscr{M}_\psi$ and $\mathbb{P}_\psi(B) = 0$. We should note that all that is needed to implement this extension is contained in two facts about the non-negative numbers $p_e = \mathbb{P}_\psi(\Omega_e)$, which are similar to conditions 1 and 2 in Definition 1: $p_e = 1$, when e is the root edge, and $\sum_{i=1,\ldots,n} p_{e_i} = p_e$, when $e_1, \ldots, e_n$ are the children of e. (In obtaining the extension of $\mathbb{P}_\psi$ to the algebra $\mathscr{A}$ as a premeasure, the only non-trivial claim that has to be checked is that if $A \in \mathscr{A}$ is described in two distinct ways as finite disjoint unions of sets Ω_e, then the sum of the p_e over these sets is the same for both descriptions. And this is not difficult, if one realizes that it is possible to compare both representations to a third one, in which all the sets Ω_e have all e in the same sufficiently large generation. The extension from a premeasure on $\mathscr{A}$ to a measure on $\mathscr{M}_\psi$ is an application of Carathéodory's Extension Theorem; see Sections 1 and 2 of Chapter 12 in [15], or Section 4 of Chapter 1 in [10].)

Before stating our theorem, we need to introduce a few more definitions, which will play a fundamental role in this paper. Given $\phi \in \mathscr{H}$ and $\omega \in \Omega$, we say that ϕ persists on ω if for each $e \in \omega$, $\text{Proj}_e \phi \neq 0$. Otherwise we say that ϕ terminates on ω. We set now

$$\Omega(\phi) = \{\omega \in \Omega : \phi \text{ persists on } \omega\}, \tag{2}$$

and

$$\Omega^c(\phi) = \Omega \backslash \Omega(\phi) = \{\omega \in \Omega : \phi \text{ terminates on } \omega\}. \tag{3}$$

Keep in mind that the choice of $\mathscr{P}$ is implicit in the definitions in the last two paragraphs. We omitted it from the notation, but should not forget that Ω, $\mathbb{P}_\psi$, $\Omega(\phi)$, etc., depend on this choice.

Theorem 1 *Let $\mathscr{H}$ be a Hilbert space and $\mathscr{P}$ be a tree-structured set of projections on $\mathscr{H}$. For any $\psi \in \mathscr{H}\backslash\{0\}$ and $A \subset \Omega$, the following are equivalent.*

(1) $\mathbb{P}_\psi(A) = 0$.
(2.i) There exist $\phi_1, \phi_2, \ldots$ orthogonal to each other, such that $\psi = \sum \phi_i$ and $\Omega(\phi_i) \subset A^c$, for each i.
(2.ii) There exist $\zeta_1, \zeta_2, \ldots$ such that $\zeta_n \to \psi$ and $\Omega(\zeta_n) \subset A^c$, for each n.

Note that we are not, a priori, making any assumption of measurability on A. But if we assume that one of (2.i), (2.ii) is true, then we learn from the theorem that $A \in \mathscr{M}_\psi$ (and $\mathbb{P}_\psi(A) = 0$). On the other hand, assuming that (1) holds means assuming that $A \in \mathscr{M}_\psi$ (and $\mathbb{P}_\psi(A) = 0$).

The first two propositions stated and proved in Sect. 4 will add mathematical structure to the content of Theorem 1, and allow it to be restated in a very compact form in display (9).

3 Relevance of the Theorem

The theorem stated in the previous section holds for any Hilbert space $\mathscr{H}$ and any choice of tree-structured set of projections $\mathscr{P}$ on it. The arbitrariness of $\mathscr{P}$ should be kept in mind as our discussion returns to Physics. When we consider collapse quantum mechanics, there is a special choice of $\mathscr{P}$, namely the one described in the introduction: vertices (other than the root vertex and terminal ones) correspond to experiments and edges (other than the root edge) correspond to the possible outcomes in each experiment. In the case of collapse quantum mechanics, and with this choice of $\mathscr{P}$, Ω is the set of possible histories that could materialize from the collapses. And in the special case of Born-rule collapse quantum mechanics, statement (1) in Theorem 1 means that event A is (probabilistically) precluded from happening. In the case of non-collapse quantum mechanics there is in principle no special choice of $\mathscr{P}$. But, as the reader may have anticipated, for the purpose of comparing non-collapse to collapse quantum mechanics, via Theorem 1, it is natural to choose precisely the same $\mathscr{P}$. We will observe below, as the reader may have also anticipated, that with this choice, the equivalence between statements (1) and (2.i) in the theorem implies (modulo plausible postulates on how the theories provide predictions) that quantum mechanics without collapse gives raise to the same predictions as Born-rule collapse quantum mechanics.

In collapse quantum mechanics only one history $\omega = (e_1, e_2, e_3, \ldots)$ materializes. In the Heisenberg picture that we are considering, the state of our system is initially some $\psi \in \mathscr{H}$, but it changes at each collapse, following the path $(\mathrm{Proj}_{e_1}\psi, \mathrm{Proj}_{e_2}\psi, \mathrm{Proj}_{e_3}\psi, \ldots) = (\psi, \mathrm{Proj}_{e_2}\psi, \mathrm{Proj}_{e_3}\psi, \ldots)$. In non-collapse quantum mechanics in the Heisenberg picture, ψ never changes. Everett [9] was the first to make the observation that this would still be compatible with our perception that collapses happen. As observers who are part of the system (otherwise we would not be able to interact with the experimental arrangement and observe it), the particles that form our bodies and in particular our brains must follow the same quantum mechanics that describes the rest of the system that we are observing. So that at the end of an experiment we can be described as being in a superposition of states, each one with a brain that encodes a different outcome for this experiment. All the possible outcomes materialize, and versions of the human observers, entangled to each possible experimental outcome, are included in this superposition.

The non-collapse view of quantum mechanics has the significant advantage of eliminating the mystery of collapse: How can systems behave differently when they are being "measured"? It yields a much simpler and consistent theory. One of the main hurdles that prevents its acceptance is probably psychological, as it affects substantially our sense of identity and of our reality. But other than this, probably the greatest obstacle to its acceptance is the issue addressed (once more) in this paper: even accepting Everett's observation that we will see collapses even if they do not happen, the question remains of why it is that we perceive them happening as if they were produced according to Born's probability rule. I will not discuss here the various previous approaches to this problem, and rather refer the reader to the recent papers [1, 13, 18], references therein and papers in the collection [16] for background and recent ideas. In [17] I argue at some length why Theorem 1 presents an answer. Here I will summarize the idea.

Stating that collapses do happen according to Born's probabilities can only have meaning if we add some postulate telling us how this leads to predictions. I assume that in collapse quantum mechanics, the predictive power derived from the collapse axiom is fully contained in the following postulate.

Prediction Postulate of Collapse Quantum Mechanics (PPCQM) *In making predictions in collapse quantum mechanics, events of probability 0 can be deemed as sure not to happen.*

If it is accepted that this postulate covers the full predictive power of the probability axiom in collapse quantum mechanics, and in particular of Born's rule in Born-rule collapse quantum mechanics, then Theorem 1 tells us that non-collapse quantum mechanics will yield the same predictions as Born-rule collapse quantum mechanics, provided we accept the following postulate for non-collapse quantum mechanics.

Prediction Postulate of Non-Collapse Quantum mechanics (PPNCQM) *In making predictions in non-collapse quantum mechanics, if the state of our system is a superposition of states all of which exclude a certain event (i.e., if (2.i) of Theorem 1 holds for this event A), then this event can be deemed as sure not to happen.*

This postulate can be seen as a version of the superposition principle of quantum mechanics, and does not include probabilities in its statement. Theorem 1 therefore provides an explanation of how probabilities emerge in non-collapse quantum mechanics, and why they are given by Born's rule.

The observation above provides an answer to the question in the title of this paper, but it also raises a fundamental question. Can one formulate non-collapse quantum mechanics in a precise and consistent fashion that provides a clear notion for what is reality in the theory (provides a precise ontology for the theory) and is compatible with the PPNCQM? In current work in progress I hope to provide an affirmative answer.

One important consequence of the observations above is that not only objective predictive probabilities emerge in non-collapse quantum mechanics (from the non-probabilistic PPNCQM), but that they are precisely the ones supported by experimental observation, namely, Born's rule probabilities. This point is discussed at some length in Section 4 of [17]. It implies that non-collapse quantum mechanics with the PPNCQM included would be falsified by data that indicated collapses with a (significantly) different probability law.

Theorem 1 and the discussion in this section help dismiss an old and important misconception associated to non-collapse quantum mechanics. That the "natural" probability distribution that it entails is some sort of "branch counting" or "uniform" one. (The quotation marks are used because in the case of non-homogeneous trees, there is ambiguity in such wording. In the case of a homogeneous tree in which each edge has b children, this probability distribution is well defined by setting $\mathbb{P}(\Omega_e) = b^{-n+1}$, when e is an edge in the n-th generation.) Typically this probability distribution will produce predictions at odds with those produced by using Born's rule. And this has been used as an argument against non-collapse quantum mechanics. But while the use of Born's rule to make predictions in non-collapse quantum mechanics is shown here to be equivalent to the PPNCQM, one cannot find any good reason why a "branch counting rule" would be the appropriate tool for this purpose. There is a tradition of saying something like: "since all branches are equally real, a branch counting probability distribution is implied". But this is a meaningless sentence. What would "real, but not equally real" mean? "Equally real" (whatever it may mean) does not imply equally likely in any predictive sense. For instance all the teams competing for a soccer World Cup are "equally real", but if we want to predict who will win the cup, there is no reason for using a uniform distribution.

4 Proof of the Theorem

The orientation that was introduced on the tree (**V**, **E**) induces a partial order on the set of edges: for any two edges we write $e' \leq e''$ if there is a partial history that starts with e' and ends with e''. We write $e' < e''$ if $e' \leq e''$ and $e' \neq e''$. If neither $e' \leq e''$, nor $e'' \leq e'$, then we say that e' and e'' are not comparable.

Definition 1 has some simple consequences. If e'' is a child of e', then $\mathscr{H}_{e''} \subset \mathscr{H}_{e'}$. By induction along a partial history line, this extends to:

$$\text{If } e' \leq e'', \text{ then } \mathscr{H}_{e''} \subset \mathscr{H}_{e'}. \tag{4}$$

In contrast, if e' and e'' are siblings, then $\mathscr{H}_{e'} \perp \mathscr{H}_{e''}$. By induction along partial history lines, this extends to:

$$\text{If } e' \text{ and } e'' \text{ are not comparable, then } \mathscr{H}_{e'} \perp \mathscr{H}_{e''}. \tag{5}$$

Proposition 1 *For every* $\phi_1, \phi_2 \in \mathscr{H}$*,* $\Omega(\phi_1 + \phi_2) \subset \Omega(\phi_1) \cup \Omega(\phi_2)$*, or equivalently* $\Omega^c(\phi_1) \cap \Omega^c(\phi_2) \subset \Omega^c(\phi_1 + \phi_2)$.

Proof Suppose $\omega \in \Omega^c(\phi_1) \cap \Omega^c(\phi_2)$. Then there are $e_1, e_2 \in \omega$ such that $\mathrm{Proj}_{e_1}\phi_1 = \mathrm{Proj}_{e_2}\phi_2 = 0$. As ω is a history, e_1 and e_2 are comparable. Let e be the larger between e_1 and e_2. Also because ω is a history, for $i = 1, 2$ we have now, from (4), $\mathscr{H}_e \subset \mathscr{H}_{e_i}$ and hence $\mathrm{Proj}_e\phi_i = 0$. Therefore $\mathrm{Proj}_e(\phi_1 + \phi_2) = \mathrm{Proj}_e\phi_1 + \mathrm{Proj}_e\phi_2 = 0$, which means that $\omega \in \Omega^c(\phi_1 + \phi_2)$. □

For each $A \subset \Omega$ we define the following two sets (T stands for "truth" and F for "falsehood"):

$$T(A) = \{\phi \in \mathscr{H} : \Omega(\phi) \subset A\}, \tag{6}$$

and

$$F(A) = T(A^c) = \{\phi \in \mathscr{H} : \Omega(\phi) \subset A^c\} = \{\phi \in \mathscr{H} : A \subset \Omega^c(\phi)\}. \tag{7}$$

Proposition 2 *For every $A \subset \Omega$, $T(A)$ and $F(A)$ are vector spaces.*

Proof Since $F(A) = T(A^c)$, it suffices to prove the statement for $T(A)$. Suppose $\phi_1, \phi_2 \in T(A)$, a_1, a_2 scalars. Then, for $i = 1, 2$, $\Omega(a_i\phi_i) = \Omega(\phi_i)$, if $a_i \neq 0$, and $\Omega(a_i\phi_i) = \emptyset$, if $a_i = 0$. In any case $\Omega(a_i\phi_i) \subset \Omega(\phi_i) \subset A$. From Proposition 1 we obtain $\Omega(a_1\phi_1 + a_2\phi_2) \subset \Omega(a_1\phi_1) \cup \Omega(a_2\phi_2) \subset A$, which means $a_1\phi_1 + a_2\phi_2 \in T(A)$. □

We can rephrase Statement (2.ii) in Theorem 1 as

$$\psi \in \overline{F(A)}, \tag{8}$$

where the bar denotes topological closure in the Hilbert space $\mathscr{H}$.

The equivalence of (2.ii) and the apparently stronger statement (2.i) in Theorem 1, can be obtained, in a standard fashion, by applying the Gram-Schmidt orthonormalization procedure (see p. 46 of [14], or p. 167 of [10]) to the vectors $\zeta_1, \zeta_2 - \zeta_1, \zeta_3 - \zeta_2, \ldots$ to produce an orthonormal system with the same span. Proposition 2 assures us that this orthonormal system will be contained in $F(A)$, since the ζ_i are. The vectors $\phi_1, \phi_2, \phi_3, \ldots,$ are then obtained by expanding ψ in this orthonormal system.

The proof of Theorem 1 is now reduced to showing that for any $\psi \in \mathscr{H}\backslash\{0\}$ and $A \subset \Omega$,

$$\mathbb{P}_\psi(A) = 0 \iff \psi \in \overline{F(A)}. \tag{9}$$

The class $\mathscr{A}_\sigma$ of subsets of Ω obtained by countable unions of elements of $\mathscr{A}$ will play a major role in the proof of (9). Every $A \in \mathscr{A}_\sigma$ is a union of sets in the countable class $\{\Omega_e\colon e \in \mathbf{E}\}$. But since $\Omega_{e''} \subset \Omega_{e'}$, whenever $e' \leq e''$, we will avoid redundancies in this union by writing it as

$$A = \bigcup_{e \in \mathbf{E}(A)} \Omega_e, \tag{10}$$

where

$$\mathbf{E}(A) = \{e \in \mathbf{E} : \Omega_e \subset A \text{ and there is no } e' \in \mathbf{E} \text{ such that } e' < e \text{ and } \Omega_{e'} \subset A\}. \tag{11}$$

Any two distinct elements of $\mathbf{E}(A)$ are not comparable. And since $\Omega_{e'} \cap \Omega_{e''} = \emptyset$, whenever e' and e'' are not comparable, (10) is a disjoint union. Moreover, using (5) we see that $\{\mathscr{H}_e : e \in \mathbf{E}(A)\}$ is a countable collection of orthogonal subspaces of $\mathscr{H}$. We will associate to A their direct sum (the topological closure of the linear span of vectors in these $\mathscr{H}_e$), which we denote by

$$\mathscr{H}(A) = \bigoplus_{e \in \mathbf{E}(A)} \mathscr{H}_e. \tag{12}$$

If $\mathscr{S}$ is a subspace of $\mathscr{H}$ and $\phi \in \mathscr{H}$, we will use the notation $\mathrm{Proj}(\phi|\mathscr{S})$ to denote the projection of ϕ on $\mathscr{S}$. For instance $\mathrm{Proj}(\phi|\mathscr{H}_e) = \mathrm{Proj}_e\phi$.

Lemma 1 *For any $\phi \in \mathscr{H}$ and $A \in \mathscr{A}_\sigma$,*

(i) For any $e \in \mathbf{E}$, $Proj_e\phi = 0 \iff \Omega_e \subset \Omega^c(\phi)$.
(ii) $\mathscr{H}^\perp(A) = F(A)$.
(iii) $\Omega^c(\phi) \in \mathscr{A}_\sigma$.
(iv) $\phi \in \mathscr{H}^\perp(\Omega^c(\phi))$.
(v) $||Proj(\phi|\mathscr{H}(A))||^2 = ||\phi||^2\, \mathbb{P}_\phi(A)$, *if* $\phi \neq 0$.

Proof

(i) The implication ($\Longrightarrow$) is clear. To prove ($\Longleftarrow$) suppose that $\mathrm{Proj}_e\phi \neq 0$. Then either e is a terminal edge, or it has a child e' with $\mathrm{Proj}_{e'}\phi \neq 0$. Repeating inductively this reasoning, we produce a history ω such that $e \in \omega$ and ϕ persists on ω. Hence $\Omega_e \not\subset \Omega^c(\phi)$.

(ii)

$$\mathscr{H}^\perp(A) = \bigcap_{e\in\mathbf{E}(A)} \mathscr{H}_e^\perp = \bigcap_{e\in\mathbf{E}(A)} \{\phi \in \mathscr{H} : \Omega_e \subset \Omega^c(\phi)\}$$
$$= \{\phi \in \mathscr{H} : A \subset \Omega^c(\phi)\} = F(A),$$

where in the first equality we used the definition (12) of $\mathscr{H}(A)$, in the second equality we used part (i) of the lemma, in the third equality we used (10), and in the fourth equality we used (7)

(iii) $\Omega^c(\phi) = \cup\{\Omega_e : e \in \mathbf{E},\ \mathrm{Proj}_e\phi = 0\}$. And this set belongs to $\mathscr{A}_\sigma$, since this union is countable.

(iv) Thanks to part (iii) of the lemma, we can take $A = \Omega^c(\phi)$ in part (ii) of the lemma. Using then (7), we obtain

$$\mathscr{H}^\perp(\Omega^c(\phi)) = F(\Omega^c(\phi)) = \{\phi' \in \mathscr{H} : \Omega^c(\phi) \subset \Omega^c(\phi')\} \ni \phi.$$

(v)

$$||\mathrm{Proj}(\phi|\mathscr{H}(A))||^2 = \sum_{e\in\mathbf{E}(A)} ||\mathrm{Proj}_e(\phi)||^2 = \sum_{e\in\mathbf{E}(A)} ||\phi||^2\mathbb{P}_\phi(\Omega_e)$$
$$= ||\phi||^2\mathbb{P}_\phi(\cup_{e\in\mathbf{E}(A)}\Omega_e) = ||\phi||^2\mathbb{P}_\phi(A),$$

where in the first equality we used the definition (12) of $\mathscr{H}(A)$, in the second equality we used (1), in the third equality we used the disjointness of the sets involved, and in the fourth equality we used (10).

□

We will use some consequences of Carathéodory's theorem that extends the measure $\mathbb{P}_\psi$ from $\mathscr{A}$ to $\mathscr{M}_\psi$ (see Sections 1 and 2 of Chapter 12 in [15], or Section 4 of Chapter 1 in [10]). Given $\psi \in \mathscr{H}$, define the outer measure of any set $A \subset \Omega$ by

$$\mathbb{P}^*_\psi(A) = \inf\{\mathbb{P}_\psi(A') : A' \in \mathscr{A}_\sigma,\ A \subset A'\}, \tag{13}$$

and define also

$$\mathscr{M}_\psi \ = \ \{A \subset \Omega : \text{for all } S \subset \Omega,\ \mathbb{P}^*_\psi(A \cap S) + \mathbb{P}^*_\psi(A^c \cap S) = \mathbb{P}^*_\psi(S)\}. \tag{14}$$

Then it follows from Carathéodory's Extension Theorem that $\mathscr{M}_\psi$ is a σ-algebra that extends $\mathscr{B}$ and $\mathbb{P}^*_\psi(A) = \mathbb{P}_\psi(A)$ for every $A \in \mathscr{M}_\psi$, in particular for every $A \in \mathscr{B}$ and therefore for every $A \in \mathscr{A}_\sigma$. It also follows that $\mathbb{P}^*_\psi(A) = 0$ implies $A \in \mathscr{M}_\psi$ and is necessary and sufficient for $\mathbb{P}_\psi(A) = 0$.

The next two lemmas prove each one of the directions of the equivalence (9), completing the proof of Theorem 1.

Lemma 2 *For any $\psi \in \mathscr{H}\backslash\{0\}$ and $A \subset \Omega$,*

$$\psi \in \overline{F(A)} \implies \mathbb{P}_\psi(A) = 0.$$

Proof If $\psi \in \overline{F(A)}$, there are $\zeta_n \in F(A)$ such that $\zeta_n \to \psi$. Set $B_n = \Omega^c(\zeta_n)$. From (7) and Lemma 1(iii) we have $A \subset B_n \in \mathscr{A}_\sigma$. Using (13) and Lemma 1(v), we obtain

$$0 \ \le \ \mathbb{P}^*_\psi(A) \ \le \ \mathbb{P}_\psi(B_n) \ = \ \frac{||\mathrm{Proj}(\psi|\mathscr{H}(B_n))||^2}{||\psi||^2}.$$

But since Lemma 1(iv) tells us that $\zeta_n \in \mathscr{H}^\perp(B_n)$, we can write

$$\begin{aligned} ||\mathrm{Proj}(\psi|\mathscr{H}(B_n))||^2 \ &= \ ||\mathrm{Proj}(\psi - \zeta_n|\mathscr{H}(B_n)) \ + \ \mathrm{Proj}(\zeta_n|\mathscr{H}(B_n))||^2 \\ &= \quad ||\mathrm{Proj}(\psi - \zeta_n|\mathscr{H}(B_n))||^2 \ \le \ ||\psi - \zeta_n||^2. \end{aligned}$$

Since n is arbitrary, the two displays combined give

$$0 \ \le \ \mathbb{P}^*_\psi(A) \ \le \ \lim_{n\to\infty} \frac{||\psi - \zeta_n||^2}{||\psi||^2} \ = \ 0,$$

proving that $\mathbb{P}^*_\psi(A) = 0$ and hence $A \in \mathscr{M}_\psi$ and $\mathbb{P}_\psi(A) = 0$. $\square$

Lemma 3 *For any* $\psi \in \mathscr{H}\setminus\{0\}$ *and* $A \subset \Omega$,

$$\mathbb{P}_\psi(A) = 0 \implies \psi \in \overline{F(A)}.$$

Proof If $\mathbb{P}_\psi(A) = 0$, (13) tells us that there are $A_n \in \mathscr{A}_\sigma$ such that $A \subset A_n$ and $\mathbb{P}_\psi(A_n) \to 0$. Set $\xi_n = \text{Proj}(\psi|\mathscr{H}^\perp(A_n))$. Then $\xi_n \in \mathscr{H}^\perp(A_n) = F(A_n) \subset F(A)$, where the equality is Lemma 1(ii), and in the last step we are using (7). Therefore, using Lemma 1(v), we obtain

$$||\xi_n - \psi||^2 = ||\text{Proj}(\psi|\mathscr{H}(A_n))||^2 = ||\psi||^2\,\mathbb{P}_\psi(A_n) \to 0,$$

as $n \to \infty$. This shows that (ξ_n) is a sequence in $F(A)$ that converges to ψ and therefore $\psi \in \overline{F(A)}$. □

Acknowledgments I am grateful to Marek Biskup, Michael Gutperle, Ander Holroyd, Jim Ralston, Pierrre-François Rodriguez and Sheldon Smith for enlightening discussions. Special thanks go to Jim Ralston and Maria Eulalia Vares for their careful reading of the proof.

References

1. Aguirre, A., Tegmark, M.: Born in an infinite universe: a cosmological interpretation of quantum mechanics. Phys. Rev. D **84**, 105002 (2011)
2. Albert, D.: Quantum Mechanics and Experience. Harvard University Press, Cambridge (1992)
3. Barrett, J.: The Quantum Mechanics of Minds and Worlds. Oxford University Press, Oxford (1999)
4. Bricmont, J.: Making Sense of Quantum Mechanics Springer, Berlin (2016)
5. Bruce, C.: Schrödinger Rabbits. The Many Worlds of the Quantum. Joseph Henry Press, Washington (2004)
6. Carroll, S.: Something Deeply Hidden. Dutton (2019)
7. Deutsch, D.: Quantum theory of probability and decisions. Proc. R. Soc. A **455**, 3129–3137 (1999)
8. De Witt, B., Grahan, N. (eds.): The Many Worlds Interpretation of Quantum Mechanics. Princeton University Press, Princeton (1973)
9. Everett, H.: "Relative state" formulation of quantum mechanics. Rev. Mod. Phys. **29**, 454–462 (1957)
10. Folland, G.B.: Real Analysis. Modern Techniques and Their Application. Wiley, London (1984)
11. Hall, B.C.: Quantum Theory for Mathematicians. Springer, Berlin (2013)
12. Maudlin, T.: Philosophy of Physics. Quantum Theory. Princeton University Press, Princeton (2019)
13. McQueen, K.J., Vaidman, L.: In defense of the self-location uncertainty account of probability in the many-worlds interpretation. Stud. Hist. Philos. Sci. B: Stud. Hist. Philos. Mod. Phys. **66**, 14–23 (2019)
14. Reed, M., Simon, B.: Functional Analysis. Revised and Enlarged Edition. Academic Press, New York (1980)
15. Royden, H.L.: Real Analysis, 3rd edn. Macmillan Publishing Company, New York (1986)

16. Saunders, S., Barrett, J., Kent, A., Wallace, D. (eds.): Many Worlds? Oxford University Press, Oxford (2010)
17. Schonmann, R.H.: A theorem and a remark with the purpose of comparing the role and origin of probabilities in non-collapse and in collapse quantum mechanics (2019). Preprint
18. Sebens, T., Carroll, S.M.: Self-locating uncertainty and the origin of probability in Everettian quantum mechanics. Br. J. Philos. Sci. **69(1)**, 25–74 (2018)
19. Tegmark, M.: Many lives in many worlds. Nature **448(5)**, 23–24 (2007)
20. Tegmark, M.: Our Mathematical Universe. My Quest for the Ultimate Nature of Reality. Doubleday, New York (2014)
21. Wallace, D.: The Emergent Multiverse: Quantum Theory According to the Everett Interpretation. Oxford University Press, Oxford (2012)

On One-Dimensional Multi-Particle Diffusion Limited Aggregation

Allan Sly

Dedicated to the memory of Vladas Sidoravicius, colleague, mentor and friend.

Abstract We prove that the one dimensional Multi-Particle Diffusion Limited Aggregation model has linear growth whenever the particle density exceeds 1 answering a question of Kesten and Sidoravicius. As a corollary we prove linear growth in all dimensions d when the particle density is at least 1.

Keywords Interacting particle systems · Diffusion Limited Aggregation

1 Introduction

In the Diffusion Limited Aggregation (DLA) model introduced by Witten and Sanders [7] particles arrive from infinity and adhere to a growing aggregate. It produces beautiful fractal-like pictures of dendritic growth but mathematically it remains poorly understood. We consider a variant, multiparticle DLA, where the aggregate sits in an infinite Poisson cloud of particles which adhere when they hit the aggregate, a model which has been studied in both physics [6] and mathematics [3, 5]. Again one is interested in the growth of the aggregate and its structure.

In the model, initially, there is a collection of particles whose locations are given by a mean K Poisson initial density on $\mathbb{Z}^d$. The particles each move independently according to rate 1 continuous time random walks on $\mathbb{Z}^d$. We follow the random evolution of an aggregate $\mathscr{D}_t \subset \mathbb{Z}^d$ where at time 0 an aggregate is placed at the origin $\mathscr{D}_0 = \{0\}$ to which other particles adhere according the following rule. When

A. Sly (✉)
Princeton University, Fine Hall, Princeton, New Jersey
e-mail: allansly@princeton.edu

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_31

a particle at $v \notin \mathscr{D}_{t-}$ attempts to move onto the aggregate at time t, it stays in place and instead is added to the aggregate so $\mathscr{D}_t = \mathscr{D}_{t-} \cup \{v\}$ and the particle no longer moves. Any other particles at v at the time are also frozen in place.

We will mainly focus on the one dimension setting and in Sect. 5 will discuss how to boost the results to higher dimensions. In this case the aggregate is simply a line segment and the processes on the positive and negative axes are independent so we simply restrict our attention to the rightmost position of the aggregate at time t which we denote X_t. When a particle at time t at position $X_{t-} + 1$ attempts to take a step to the left it is incorporated into the aggregate along with any other particles.

It was proved by Kesten and Sidoravicius [3] that X_t grows like $\sqrt{t}$ when $K < 1$. Indeed there simply are not enough particles around for it to grow faster. They conjectured, however, that when $K > 1$ then it should grow linearly. Our main result confirms this conjecture.

Theorem 1 *For all $K > 1$ the limit $\lim_t \frac{1}{t} X_t$ exists almost surely and is a positive constant.*

We also give a simple extension of these results to higher dimensions and prove the following corollary.

Corollary 1 *In all dimensions $d \geq 2$ when $K > 1$ the diameter of the aggregate grows linearly in t, that is for some positive constant $\delta > 0$*

$$\liminf_t \frac{1}{t} Diam(\mathscr{D}_t) > \delta \text{ a.s.}$$

Previously Sidoravicius and Stauffer [5] studied the case of $d \geq 2$ in a slightly different variant where particles instead perform a simple exclusion process. They showed that for densities close to 1, that there is a positive probability that the aggregate grows with linear speed. Also in Sect. 5 we describe how for $d \geq 2$ the upper bound on the threshold can be reduced further below 1, for example to $\frac{5}{6}$ when $d = 2$. However, strikingly Eldan [2] conjectured that the critical value is always 0, that is the aggregate grows with linear speed for all $K > 0$. We are inclined to agree with this conjecture but our methods do not suggest a way of reaching the threshold. A better understanding of the growth of the standard DLA seems to be an important starting point.

2 Basic Results

We will analyse the function valued process Y_t given by,

$$Y_t(s) := \begin{cases} X_t - X_{t-s} & 0 \leq s \leq t \\ \infty & s > t. \end{cases} \tag{1}$$

Let $\mathscr{F}_t$ denote the filtration generated by X_t. We let $S(t)$ denote the infinitesimal rate at which X_t increases given $\mathscr{F}_t$. Given $\mathscr{F}_t$ the number of particles at $X_t + 1$ is conditionally Poisson with intensity given by the probability that a random walker at $X_t + 1$ at time t was never located in the aggregate. Each of the particles jumps to the left at rate $\frac{1}{2}$ so with W_t denoting an independent continuous time random walk,

$$S(t) = \frac{1}{2} K \mathbb{P}[\max_{0 \le s \le t} W_s - Y_t(s) \le 0 \mid Y_t].$$

Note that $S(t)$ is increasing as a function of Y_t. Indeed we could realise X_t as follows, let Π be a Poisson process on $[0, \infty)^2$ and then

$$X_t = \Pi(\{(x, y) : 0 \le x \le t, 0 \le y \le S(x)\}.$$

Since both X_t and Y_t are increasing functions of Π we can make use of the FKG property.

Also note that Y_t is a function valued Markov process in t. Its infinitesimal rate change can be described as follows, if there is no new particle in $[t, t + dt]$ then

$$Y_{t+dt}(s) = \begin{cases} Y_t(s - dt), & s > dt \\ 0, & s \le dt. \end{cases}$$

If a new particle arrives at time then $Y_t(s) = Y_{t-}(s) + 1$. Since new particles arrive at rate $S(t)$, which is itself a function of Y_t the process is Markovian.

We now observe an important monotonicity property of this process. Suppose that we have two copies of the process X_t and X'_t such that at time t_* we have that $Y_{t_*}(s) \ge Y'_{t_*}(s)$ for all $s \ge 0$. Then the infinitesimal rates will satisfy $S(t_*) \ge S'(t_*)$. Suppose that we couple the processes to use the same point process Π after time t_*. Let t_1 be the first time after t_* that either X_t or X'_t encounters a new particle. For $t_* \le t < t_1$ we must have that $Y_t \ge Y'_t$ and so we also have that $S(t) \ge S(t')$. Hence, since S is larger in the first chain, at time t_1 we either have that both processes encounter a new particle or X does and X' does not. In either case $Y_{t_1} \ge Y'_{t_1}$. Applying this inductively we will have that $Y_t \ge Y'_t$ for all time $t \ge t_*$. In other words the Markov process Y_t is a stochastically monotone Markov process.

Since $Y_0(s) = +\infty$ for all $s > 0$ we have that the initial value of Y is the maximal value and $Y_0 \ge Y_t$ for all t. Applying the Markov property this means that Y_s stochastically dominated Y_{t+s} for all $t, s \ge 0$, in other words Y_t is stochastically decreasing.

Most of our analysis will involve estimating $S(t)$ and using that to control the evolution of Y_t and show that it does not become too small for too long. Let $M_t = \max_{0 \le s \le t} W_s$ be the maximum process of W_t.

Lemma 1 *For any $i \geq 0$ we have that*

$$S(t) \geq \frac{K}{2}\mathbb{P}[M_{2^i} = 0] \prod_{i'=i}^{\infty} \mathbb{P}[M_{2^{i'+1}} \leq Y_t(2^{i'}) \mid Y_t]$$

Proof We have

$$\begin{aligned} S(t) &\geq \frac{K}{2}\mathbb{P}[\max_{0\leq s\leq t} W_s - Y_t(s) \leq 0 \mid Y_t] \\ &\geq \frac{K}{2}\mathbb{P}[M_{2^i} = 0, \forall i' \geq i \ M_{2^{i'+1}} \leq Y_t(2^{i'}) \mid Y_t] \\ &\geq \frac{K}{2}\mathbb{P}[M_{2^i} = 0] \prod_{i'\geq i} \mathbb{P}[M_{2^{i'+1}} \leq Y_t(2^{i'}) \mid Y_t] \end{aligned}$$

where the final inequality follows from the FKG inequality.

By the reflection principle we have that for any integer $j \geq 0$,

$$\mathbb{P}[M_t \geq j] = \mathbb{P}[W_t \geq j] + \mathbb{P}[W_t \geq j+1].$$

Thus asymptotically we have that

$$\mathbb{P}[M_t = 0] \approx \frac{1}{\sqrt{2\pi}} t^{-1/2} \tag{2}$$

Now let T_j be the first hitting time of j. Since $\cosh(s) - 1 \leq s^2$ for $0 \leq s \leq 1$ we have that for $t \geq 1$,

$$\mathbb{E}[e^{\frac{1}{\sqrt{t}} W_{t\wedge T_j}}] \leq \mathbb{E}[e^{\frac{1}{\sqrt{t}} W_t}] = e^{(\cosh(\frac{1}{\sqrt{t}})-1)t} \leq e^1,$$

and hence by Markov's inequality

$$\mathbb{P}[M_t \geq jt^{1/2}] \leq \mathbb{P}[e^{\frac{1}{\sqrt{t}} W_{t\wedge T_{j\sqrt{t}}}} = e^j] \leq e^{1-j}. \tag{3}$$

Plugging the above equations into Lemma 1 we get the following immediate corollary.

Corollary 2 *There exists i^* such that the following holds. Suppose that $i \geq i^*$ and for all $i' \geq i$ we have $j_{i'} = Y_t(2^{i'})2^{-i'/2}$. Then*

$$S(t) \geq \frac{K}{10} 2^{-i/2} \prod_{i'=i}^{\infty} (1 - e^{1-\max\{1, j_{i'}/\sqrt{2}\}})$$

Next we check that provided $S(t)$ remains bounded below during an interval then we get a comparable lower bound on the speed of X_t.

Lemma 2 *We have that for all* $\rho \in (0, 1)$ *there exists* $\psi(\rho) > 0$ *such that for all* $\Delta > 0$,

$$\mathbb{P}[\min_{s\in[t,t+\Delta]} S(s) \geq \gamma, X_{t+\Delta} - X_t \leq \rho\Delta\gamma \mid Y_t] \leq \exp(-\psi(\rho)\Delta\gamma)$$

The function ψ *satisfies* $\psi(\frac{1}{2}) \geq \frac{1}{10}$.

Proof Using the construction of the process in terms of Π we have that

$$\begin{aligned}\mathbb{P}[\min_{s\in[t,t+\Delta]} S(s) \geq \gamma, X_{t+\Delta} - X_t \leq \rho\Delta\gamma \mid Y_t] &\leq \mathbb{P}[\Pi([t, t + \Delta] \times [0, \gamma]) \leq \rho\Delta\gamma] \\ &= \mathbb{P}[\text{Poisson}(\Delta\gamma) \leq \rho\Delta\gamma]\end{aligned}$$

Now if $N \sim \text{Poisson}(\Delta\gamma)$ then $\mathbb{E}e^{-\theta N} = \exp((e^{-\theta} - 1)\Delta\gamma)$ and so by Markov's inequality

$$\mathbb{P}[N \leq \rho\Delta\gamma] = \mathbb{P}[e^{-\theta N} \geq e^{-\theta\rho\Delta\gamma}] \leq \frac{\exp((e^{-\theta} - 1)\Delta\gamma)}{\exp(-\theta\rho\Delta\gamma)} = \exp((\theta\rho + e^{-\theta} - 1)\Delta\gamma).$$

Setting $f_\rho(\theta) = -(\theta\rho + e^{-\theta} - 1)$ and

$$\psi(\rho) = \sup_{\theta\geq 0} f_\rho(\theta)$$

it remains to check that $\psi(\rho) > 0$. This follows from the fact that $f_\rho(0) = 0$ and $f_\rho'(0) = 1 - \rho > 0$. Since $f_{\frac{1}{2}}(\frac{1}{2}) \geq \frac{1}{10}$ we have that $\psi(\frac{1}{2}) \geq \frac{1}{10}$.

3 Proof of Positive Speed

In this section we aim to show that Y_t does not become too small in order to show that X_t continues to progress. We say that Y_t is *permissive* at time t and at scale i if $Y_t(2^i) \geq 10i2^{i/2}$. Our approach, will be to consider functions

$$y_\alpha(s) = \begin{cases} 0 & s \leq \alpha^{-3/2} \\ \min\{\alpha(s - \alpha^{-3/2}), s^{1/2}\log_2 s\} & s \geq \alpha^{-3/2}. \end{cases}$$

and show that if $Y_t(s) \geq y_\alpha(s)$ holds for α, then it is likely that $Y_{t'}(s) \geq y_{\alpha'}(s)$ for some specified $t' > t, \alpha' > \alpha$ and thus show that if Y_t starts to get too small, it will have a positive drift and will usually not stay too small for too long. We define events $\mathscr{R}$ as

$$\mathscr{R}(t, s, \gamma) = \{X_{t+s} - X_t \geq \gamma s\},$$

to measure the speed of the aggregate in an interval of time.

Lemma 3 *For all* $\epsilon > 0$ *there exists* $0 < \alpha_\star(\epsilon) \leq 1$ *such that for all* $0 < \alpha < \alpha_\star$,

$$\mathbb{P}[\max_{s \geq 0} W_s - y_\alpha((s - \alpha^{-4/3}) \vee 0) \leq 0] \geq 2(1-\epsilon)\alpha.$$

Proof When $\alpha_\star(\epsilon)$ is small enough we have that for $\alpha^{-3/2} \leq s \leq \alpha^{-2}$,

$$\alpha(s - \alpha^{-4/3} - \alpha^{-3/2}) \leq (s - \alpha^{-4/3})^{1/2} \log_2(s - \alpha^{-4/3})$$

Hence with $\xi = \xi_\alpha = \alpha^{-4/3} + \alpha^{-3/2}$ if we set

$$\mathscr{A} = \{\max_{s \geq 0} W_s - \alpha((s - \xi) \vee 0) \leq 0\}$$

and

$$\mathscr{B} = \{\max_{s \geq \alpha^{-2}} W_s - (s - \alpha^{-4/3})^{1/2} \log_2(s - \alpha^{-4/3}) \leq 0\}$$

then

$$\mathbb{P}[\max_{s \geq 0} W_s - y_\alpha((s - \alpha^{-4/3}) \vee 0) \leq 0] \geq \mathbb{P}[\mathscr{A}, \mathscr{B}] \geq \mathbb{P}[\mathscr{A}]\mathbb{P}[\mathscr{B}].$$

where the second inequality follows by the FKG inequality since $\mathscr{A}$ and $\mathscr{B}$ are both decreasing events for W_s. For large s, we have $s^{1/2} \log_2 s \leq 2(s/2)^{1/2} \log_2(s/2)$ and so

$$\begin{aligned}
\mathbb{P}[\mathscr{B}] &\geq \mathbb{P}[\max_{s \geq \alpha^{-2}} W_s - \tfrac{1}{2} s^{1/2} \log_2(\tfrac{1}{2} s) \leq 0] \\
&\geq \mathbb{P}[\forall i \geq \lfloor \log_2(\alpha^{-2}) \rfloor, M_{2^{i+1}} \leq \tfrac{1}{2}(i-1)2^{i/2}] \\
&\geq \prod_{i \geq \lfloor \log_2(\alpha^{-2}) \rfloor} \mathbb{P}[M_{2^{i+1}} \leq \tfrac{1}{2}(i-1)2^{i/2}] \\
&\geq \prod_{i \geq \lfloor \log_2(\alpha^{-2}) \rfloor} (1 - \exp(1 - (i-1)2^{-3/2}))
\end{aligned}$$

where the third inequality follows from the FKG inequality and the final inequality is by Eq. (3). Thus as $\alpha \to 0$ we have that $\mathbb{P}[\mathscr{B}] \to 1$ so it is sufficient to show that for small enough α that $\mathbb{P}[\mathscr{A}] \geq 2\alpha(1-\epsilon/2)$. By the reflection principle for $a \leq 1$,

$$\mathbb{P}[M_t \geq 1, W_t = a] = \mathbb{P}[M_t \geq 1, W_t = 2-a] = \mathbb{P}[W_t = 2-a] = \mathbb{P}[W_t = a-2],$$

and so

$$\begin{aligned}\mathbb{P}[M_t \geq 1] &= \sum_{a>1} \mathbb{P}[M_t \geq 1, W_t = a] + \sum_{a\leq 1} \mathbb{P}[M_t \geq 1, W_t = a] \\ &= \sum_{a>1} \mathbb{P}[W_t = a] + \sum_{a\leq 1} \mathbb{P}[W_t = a-2] \\ &= 1 - \mathbb{P}[W_t = 0] - \mathbb{P}[W_t = 1].\end{aligned}$$

Hence by the Local Central Limit Theorem,

$$\lim_t \sqrt{t}\mathbb{P}[M_t = 0] = \lim_t \sqrt{t}\,(\mathbb{P}[W_t = 0] + \mathbb{P}[W_t = 1]) = \frac{2}{\sqrt{2\pi}}.$$

Also we have for $a \leq 0$,

$$\mathbb{P}[W_t{=}a, M_t = 0]{=}\mathbb{P}[W_t = a]{-}\mathbb{P}[W_t = a, M_t \geq 1] = \mathbb{P}[W_t = a]{-}\mathbb{P}[W_t = a{-}2]$$

and so the law of W_t conditioned on $M_t = 0$ satisfies,

$$\begin{aligned}\lim_t \mathbb{P}[\frac{1}{\sqrt{t}} W_t \leq x \mid M_t = 0] &= \lim_t \frac{\sum_{a=-\infty}^{x\sqrt{t}} \mathbb{P}[W_t = a] - \mathbb{P}[W_t = a-2]}{\mathbb{P}[M_t = 0]} \\ &= \lim_t \frac{\mathbb{P}[W_t = x\sqrt{t}] + \mathbb{P}[W_t = x\sqrt{t} - 1]}{\mathbb{P}[M_t = 0]} \\ &= \lim_t \frac{2\frac{1}{\sqrt{2\pi}} e^{-x^2/2}}{\frac{2}{\sqrt{2\pi}}} = e^{-x^2/2}\end{aligned}$$

where $x \leq 0$ and hence is the negative of the Rayleigh distribution. Now let $Z_t = W_t - \alpha t$ and $U_t = e^{\theta Z_t}$. Then

$$\mathbb{E}U_t = \exp((\cosh(\theta) - 1 - \alpha\theta)t).$$

As $f_\alpha(\theta) = \cosh(\theta) - 1 - \alpha\theta$ is strictly convex, it has two roots, one of which is at $\theta = 0$. Let θ_α be the non-zero root of f_α. Since

$$f_\alpha(\theta) = -\alpha\theta + \frac{1}{2}\theta^2 + O(\theta^4)$$

for small α we have that $\theta_\alpha = 2\alpha + O(\alpha^2)$. Then with $\theta = \theta_\alpha$ we have that $U_t = e^{\theta_\alpha Z_t}$ is a martingale. Let $T = \min_t Z_t > 0$ and so by the Optional Stopping Theorem,

$$\mathbb{E}[U_T \mid Z_0 = x] = e^{\theta_\alpha x}.$$

Also since $U_T \in [0, 1]$ if $T < \infty$ so

$$\mathbb{E}[U_T \mid Z_0 = x] \geq \mathbb{P}[T < \infty \mid Z_0 = x]$$

so

$$\mathbb{P}[T < \infty \mid Z_0 = z] \leq e^{-\theta_\alpha}.$$

Thus we have that as $\alpha \to 0$,

$$\begin{aligned}
\mathbb{P}[\mathscr{A}] &= \mathbb{P}[\max_{s\geq 0} W_s - \alpha((s-\xi) \vee 0) \leq 0] \\
&= \sum_{x=-\infty}^{0} \mathbb{P}[M_\xi = 0, W_\xi = x]\mathbb{P}[T = \infty \mid Z_0 = x] \\
&\geq \mathbb{P}[M_\xi = 0] \sum_{x=-\infty}^{0} \mathbb{P}[W_\xi = x \mid M_\xi = 0](1 - e^{\theta_\alpha x}) \\
&\geq \frac{2 + o(1)}{\sqrt{2\pi\xi}} \sum_{x=-\infty}^{0} \mathbb{P}[W_\xi = x \mid M_\xi = 0](-2\alpha x) \\
&\to \frac{2 + o(1)}{\sqrt{2\pi\xi}} \cdot (2\alpha)\sqrt{\xi}\sqrt{\frac{\pi}{2}} = 2\alpha,
\end{aligned}$$

since the mean of the Rayleigh distribution is $\sqrt{\frac{\pi}{2}}$. This completes the lemma.

Lemma 4 *For all $K > 1$ there exists $i_\star(K)$ such that if $i \geq i_\star$ and Y_T is permissive at all levels i and above then with*

$$\alpha = \frac{1}{80} 2^{-i/2}$$

we have that

$$\mathbb{P}[\inf_s Y_{T+2^i}(s) - y_\alpha(s) \geq 0 \mid Y_T] \geq 1 - \exp(-2^{i/10}).$$

Proof Since $Y_T(2^{i'}) \geq 10i'2^{i'/2}$ for all $i' \geq i$ if we set

$$\tilde{y}(s) = \begin{cases} 0 & s < 2^{i+1}, \\ 10j2^{j/2} & s \in [2^{j+1}, 2^{j+2}), \ \ j \geq i. \end{cases}$$

then since $Y_{T+u}(s) \geq Y_T(s-u)$ we have that

$$\inf_{0 \leq u \leq 2^i} \inf_{s \geq 0} Y_{T+u}(s) - \tilde{y}(s) \geq 0. \tag{4}$$

By Corollary 2 for all $t \in [0, 2^i]$

$$S(t) \geq \frac{1}{10} 2^{-(i+1)/2} \prod_{i'=i}^{\infty} (1 - e^{1-\max\{1,5(i'-1)\}}) \geq \frac{1}{20} 2^{-i/2},$$

where the second inequality holds provided that $i_\star(K)$ is sufficiently large. Defining $\mathscr{D}$ as the event that X_t moves at rate at least $\frac{1}{40}2^{-i/2}$ for each interval $(\ell 2^{2i/3}, (\ell+1)2^{2i/3}]$,

$$\mathscr{D} := \bigcap_{\ell=0}^{2^{i/3}-1} \mathscr{R}(T + \ell 2^{2i/3}, 2^{2i/3}, \frac{1}{40} 2^{-i/2})$$

by Lemma 2 we have that

$$\mathbb{P}[\mathscr{D}] \geq 1 - 2^{i/3} \exp(-\frac{1}{10} \cdot 2^{2i/3} \cdot \frac{1}{40} 2^{-i/2}) \geq 1 - \exp(-2^{i/10})$$

where the last inequality holds provided that $i_\star(K)$ is sufficiently large. We claim that on the event $\mathscr{D}$, we have that $Y_{T+2^i}(s) \geq y_\alpha(s)$ for all s. For $s \geq 2^{i+1}$ this holds since by Eq. (4) we have that

$$Y_{T+2^i}(s) \geq \tilde{y}(s) \geq s^{1/2} \log_2 s \geq y_\alpha(s).$$

For $0 \leq s \leq 2^i$, on the event $\mathscr{D}$,

$$Y_{T+2^i}(s) \geq \lfloor s 2^{-2i/3} \rfloor 2^{2i/3} \frac{1}{40} 2^{-i/2} \geq \max\{0, s - \alpha^{-3/2}\} \frac{1}{40} 2^{-i/2} \geq y_\alpha(s),$$

and for $2^i \leq s \leq 2^{i+1}$

$$Y_{T+2^i}(s) \geq Y_{T+2^i}(2^i) \geq 2^i \cdot \frac{1}{40} 2^{-i/2} \geq y_\alpha(2^{i+1}).$$

Thus for all $s \geq 0$, $Y_{T+2^i}(s) \geq y_\alpha(s)$ which completes the proof.

Lemma 5 *For all $K > 1$, there exists $\Delta(K)$ and $\chi(K) > 0$ such that if $0 \le \alpha \le \Delta$ and $\inf_s Y_T(s) - y_\alpha(s) = 0$ then*

$$\mathbb{P}\Big[\mathscr{R}\big(T, \alpha^{-4/3}, \frac{\alpha(K+1)}{2}\big)^c \mid Y_T\Big] \le \exp\Big(-\chi(K)\alpha^{-1/3}\Big).$$

Proof With $\alpha_\star(\epsilon)$ defined as in Lemma 3 set $\Delta(K) = \alpha_\star(\frac{K-1}{3K})$. Then for $0 \le t \le \alpha^{-4/3}$

$$\begin{aligned} S(T+t) &= \frac{K}{2}\mathbb{P}[\max_{s\ge 0} W_s - Y_{T+t}(s) \le 0 \mid Y_{T+t}] \\ &\ge \frac{K}{2}\mathbb{P}[\max_{s\ge 0} W_s - y_\alpha((s-\alpha^{-4/3}) \vee 0) \le 0] \\ &\ge \frac{K}{2} 2\big(1 - \frac{K-1}{3K}\big)\alpha = \frac{\alpha(2K+1)}{3} \end{aligned}$$

where the first inequality follows from the fact that

$$Y_{T+t}(s) \ge Y_T((s-\alpha^{-4/3}) \vee 0) \ge y_\alpha((s-\alpha^{-4/3}) \vee 0)$$

and the second inequality follows from Lemma 3. Now take $\rho = \frac{3K+3}{4K+2} < 1$ and with ψ defined in Lemma 2 set $\chi(K) = \psi(\rho)$. Then since

$$\inf_{0\le t\le \alpha^{-4/3}} S(T+t) \ge \frac{\alpha(2K+1)}{3} = \rho^{-1}\frac{\alpha(K+1)}{2}$$

by Lemma 2 we have that

$$\mathbb{P}\Big[\mathscr{R}\big(T, \alpha^{-4/3}, \frac{\alpha(K+1)}{2}\big)^c \mid Y_T\Big] \le \exp\Big(-\chi(K)\alpha^{-1/3}\Big).$$

This result is useful because of the following claim.

Claim For some $0 \le \alpha \le \frac{1}{2}$ suppose that $\inf_s Y_T(s) - y_\alpha(s) = 0$. Then for an $0 \le t \le \alpha^{-3/2}$ and $\gamma \ge 1$ on the event $\mathscr{R}\big(T, t, \alpha\gamma\big)$ we have that $\inf_s Y_{T+t}(s) - y_\alpha(s) = 0$.

Proof Since $y_\alpha(s) =$ for $0 \le s \le \alpha^{-3/2}$ it is sufficient to check $s \ge \alpha^{-3/2}$. Then

$$\begin{aligned} Y_{T+t}(s) &= Y_T(s-t) + X_{T+t} - X_t \\ &\ge Y_T(s-t) + \alpha\gamma t \\ &\ge y_\alpha(s-t) + \alpha\gamma t \\ &\ge y_\alpha(s) - \alpha t + \alpha\gamma t \ge y_\alpha(t), \end{aligned}$$

where the first inequality is by the event $\mathscr{R}(T, t, \alpha\gamma)$, the second is by assumption and the third is since $\frac{d}{ds} y_\alpha(s)$ is uniformly bounded above by α.

Lemma 6 *For all* $K > 1$, *there exists* $\Delta(K)$ *and* $\chi(K) > 0$ *such that if* $0 \le \alpha \le \Delta$ *and* $\inf_s Y_T(s) - y_\alpha(s) = 0$ *then*

$$\mathbb{P}\Big[\inf_s Y_{T+\alpha^{-3}}(s) - y_{\frac{\alpha(K+1)}{2}}(s) \ge 0 \mid Y_T\Big] \ge 1 - \alpha^{-5/3} \exp\Big(-\chi(K)\alpha^{-1/3}\Big).$$

Proof Let $\mathscr{D}_\ell$ denote the event,

$$\mathscr{D}_\ell = \mathscr{R}(T + \ell\alpha^{-4/3}, \alpha^{-4/3}, \frac{\alpha(K+1)}{2}).$$

By Claim 3 and induction if $\bigcap_{\ell'=0}^{\ell-1} \mathscr{D}_{\ell'}$ holds then $\inf_s Y_{T+\ell\alpha^{-4/3}}(s) - y_\alpha(s) = 0$. Thus by Lemma 5 we have that

$$\mathbb{P}\Big[\mathscr{D}_\ell \mid \bigcap_{\ell'=0}^{\ell-1} \mathscr{D}_{\ell'}, Y_T\Big] \ge 1 - \exp\Big(-\chi(K)\alpha^{-1/3}\Big)$$

and so with $\mathscr{D}^* = \bigcap_{\ell=0}^{\alpha^{-5/3}-1} \mathscr{D}_\ell$,

$$\mathbb{P}\big[\mathscr{D}^* \mid Y_T\big] \ge 1 - \alpha^{-5/3} \exp\Big(-\chi(K)\alpha^{-1/3}\Big).$$

Now suppose that the event $\mathscr{D}^*$ holds and assume that $\Delta(K)$ is small enough so that for all $0 \le \alpha \le \Delta(K)$ the following hold:

- $\alpha^{-4/3} \le (\frac{\alpha(K+1)}{2})^{-3/2}$,
- $\frac{K+1}{2}\alpha^{-2} \ge (2\alpha^{-3})^{1/2} \log_2(2\alpha^{-3})$,
- $\forall s \ge \alpha^{-3}, \ \min\{\alpha(s - \alpha^{-3/2}), s^{1/2} \log_2 s\} = s^{1/2} \log_2 s$,
- $\forall s \ge \alpha^{-3}, \ \min\{\frac{\alpha(K+1)}{2}(s - (\frac{\alpha(K+1)}{2})^{-3/2}), s^{1/2} \log_2 s\} = s^{1/2} \log_2 s$,
- $\inf_{s \ge 2\alpha^{-3}} -s^{1/2} \log_2 s + (s - \alpha^{-3})^{1/2} \log_2(s - \alpha^{-3}) + \frac{K+1}{2}\alpha^{-2} \ge 0$.

It is straightforward to check that all of these hold for sufficiently small α. For all $(\frac{\alpha(K+1)}{2})^{-3/2} \le s \le \alpha^{-3}$ that

$$\begin{aligned} Y_{T+\alpha^{-3}}(s) &\ge \lfloor s\alpha^{4/3} \rfloor \alpha^{-4/3} \cdot \frac{\alpha(K+1)}{2} \\ &\ge \left(s - \left(\frac{\alpha(K+1)}{2}\right)^{3/2}\right) \frac{\alpha(K+1)}{2} \\ &\ge y_{\frac{\alpha(K+1)}{2}}(s). \end{aligned}$$

For $\alpha^{-3} \le s \le 2\alpha^{-3}$,

$$\begin{aligned} Y_{T+\alpha^{-3}}(s) &\ge Y_T(s-\alpha^{-3}) + \frac{K+1}{2}\alpha^{-2} \\ &\ge y_\alpha(s-\alpha^{-3}) + \frac{K+1}{2}\alpha^{-2} \\ &\ge (2\alpha^{-3})^{1/2}\log_2(2\alpha^{-3}) = y_{\frac{\alpha(K+1)}{2}}(2\alpha^{-3}). \end{aligned}$$

Finally, for $s \ge 2\alpha^{-3}$,

$$\begin{aligned} Y_{T+\alpha^{-3}}(s) &\ge y_\alpha(s-\alpha^{-3}) + \frac{K+1}{2}\alpha^{-2} \\ &= y_{\frac{\alpha(K+1)}{2}}(s) - s^{1/2}\log_2 s + (s-\alpha^{-3})^{1/2}\log_2(s-\alpha^{-3}) + \frac{K+1}{2}\alpha^{-2} \\ &\ge y_{\frac{\alpha(K+1)}{2}}(s). \end{aligned}$$

Combining the previous 3 equations implies that $Y_{T+\alpha^{-3}}(s) \ge y_{\frac{\alpha(K+1)}{2}}(s)$ for all s and hence

$$\mathbb{P}\Big[\inf_s Y_{T+\alpha^{-3}}(s) - y_{\frac{\alpha(K+1)}{2}}(s) \ge 0 \mid Y_T\Big] \ge \mathbb{P}[\mathscr{D}^*] \ge 1-\alpha^{-5/3}\exp\Big(-\chi(K)\alpha^{-1/3}\Big).$$

Lemma 7 *For all $K > 1$, there exists $i^*(K)$ such that the following holds. If $i \ge i^*$ and Y_T is permissive for all $i' > i$ then*

$$\mathbb{P}\Big[\min_{s\in[4^i, 2e^{2^{i/10}}]} Y_{T+s}(2^i) \le 10i2^{i/2} \mid \mathscr{F}_T\Big] \le 3e^{-2^{i/10}},$$

that is Y_{T+s} is permissive at scale i for all $s \in [2^i, 2e^{2^{i/10}}]$.

Proof We choose $i^*(K)$ large enough so that,

$$20i^*2^{-i^*/2}\left(\frac{1+K}{2}\right) \le \Delta(K)$$

where $\Delta(K)$ was defined in Lemma 6. Set $t_0 = 2^{i+1}$ and $\alpha_0 = \frac{1}{80}2^{-(i+1)/2}$. We define $\alpha_\ell = \left(\frac{K+1}{2}\right)^\ell \alpha_0$ and $t_\ell = t_{\ell-1} + \alpha_{\ell-1}^{-3}$. Define the event $\mathscr{W}_\ell$ as

$$\mathscr{W}_\ell = \Big\{\inf_s Y_{T+t_\ell}(s) - y_{\alpha_\ell}(s) \ge 0\Big\}.$$

By Lemma 4 we have that

$$\mathbb{P}[\mathscr{W}_0 \mid \mathscr{F}_T] \geq 1 - \exp(-2^{(i+1)/10}),$$

and by Lemma 6 we have that

$$\mathbb{P}\Big[\mathscr{W}_\ell \mid \bigcap_{\ell'=0}^{\ell-1} \mathscr{W}_{\ell'} \mid \mathscr{F}_T\Big] \geq 1 - \alpha_{\ell-1}^{-5/3} \exp\Big(-\chi(K)\alpha_{\ell-1}^{-1/3}\Big).$$

Now choose L to be the smallest integer such that $\alpha_L \geq 20i2^{-i/2}$. So $L = \lceil \frac{\log(1600 i 2^{1/2})}{\log((K+1)/2)} \rceil$ which is bounded above by i provided that $i^*(K)$ is sufficiently large and $\alpha_L \leq \Delta(K)$. Thus

$$\begin{aligned}
\mathbb{P}[\mathscr{W}_L \mid \mathscr{F}_T] &\geq 1 - \exp(-2^{i/10}) - \sum_{\ell=0}^{L-1} \alpha_{\ell-1}^{-5/3} \exp\Big(-\chi(K)\alpha_{\ell-1}^{-1/3}\Big) \\
&\geq 1 - \exp(-2^{i/10}) - i(20i2^{-i/2})^{-5/3} \exp\Big(-\chi(K)(20i2^{-i/2})^{-1/3}\Big) \\
&\geq 1 - 2\exp(-2^{i/10})
\end{aligned}$$

where the final inequality holds for i is sufficiently large. Now let $\mathscr{D}_k$ denote the event,

$$\mathscr{D}_k = \mathscr{R}(T + t_L + k\alpha_L^{-4/3}, \alpha_L^{-4/3}, \alpha_L).$$

By Claim 3 on the event $\mathscr{W}_L$ and $\bigcap_{k'=0}^{k-1} \mathscr{D}_{k'}$ we have

$$\inf_s Y_{T+t_L+k\alpha_L^{-4/3}}(s) - y_{\alpha_L}(s) \geq 0.$$

Thus by Lemma 5 we have that

$$\mathbb{P}[\mathscr{D}_k \mid \mathscr{F}_T, \mathscr{W}_L, \bigcap_{k'=0}^{k-1} \mathscr{D}_{k'}] \geq 1 - \exp\Big(-\chi(K)\alpha_L^{-1/3}\Big).$$

Let $\mathscr{D}^*$ be the event

$$\mathscr{D}^* = \left\{ \mathscr{W}_L, \bigcap_{k'=0}^{e^{2^{i/10}}-1} \mathscr{D}_{k'} \right\}.$$

Then for i sufficiently large since $\alpha_L \leq 20iK2^{-i/2}$,

$$\mathbb{P}[\mathscr{D}^* \mid \mathscr{F}_T] \geq 1-2\exp(-2^{i/10})-\exp\left(2^{i/10}-\chi(K)\alpha_L^{-1/3}\right) \geq 1-3\exp(-2^{i/10}).$$

One the event $\mathscr{D}^*$ we have that for all $t_L + 2^i \leq s \leq \alpha_L^{-4/3}e^{2^{i/10}}$ that

$$Y_{T+s} \geq \alpha_L(2^i - 2\alpha_L^{-4/3}) \geq 10i2^{i/2}.$$

By construction $t_L = 2^{i+1} + \sum_{\ell=0}^{L-1} \alpha_\ell^{-3} \leq 4^i$ and hence

$$\mathbb{P}\left[\min_{s\in[4^i,2e^{2^{i/10}}]} Y_{T+s}(2^i) \leq 10i2^{i/2} \mid \mathscr{F}_T\right] \leq \mathbb{P}[(\mathscr{D}^*)^c \mid \mathscr{F}_T] \leq 3e^{-2^{i/10}}.$$

Corollary 3 *For all $K > 1$, there exists $i^*(K)$ such if $i \geq i^*$ then*

$$\mathbb{P}\left[\min_{s\in[0,e^{2^{i/10}}]} Y_s(2^i) \leq 10i2^{i/2}\right] \leq 3e^{-2^{i/10}},$$

Proof We can apply Lemma 7 to time $T = 0$ since it is permissive at all levels and hence have that

$$\mathbb{P}\left[\min_{s\in[4^i,2e^{2^{i/10}}]} Y_s(2^i) \leq 10i2^{i/2}\right] \leq 3e^{-2^{i/10}},$$

Since Y_t is stochastically decreasing in t we have that

$$\mathbb{P}\left[\min_{0\leq t\leq e^{2^{i/10}}} Y_t(2^i) \leq 10i2^{i/2} - i2^i\right] \leq \mathbb{P}\left[\min_{s\in[4^i,4^i+e^{2^{i/10}}]} Y_s(2^i) \leq 10i2^{i/2}\right] \leq 3e^{-2^{i/10}},$$

which completes the corollary.

Lemma 8 *For all $K > 1$, there exists $i^*(K)$ such that*

$$\inf_t \mathbb{P}\left[\forall i \geq i^*, Y_t(2^i) \geq 10i2^{i/2}\right] \geq \frac{1}{2}.$$

Proof Take $i^*(K)$ as in Lemma 7 and suppose that $I \geq i^*$. Let $\mathscr{D}_I$ denote the event that Y_t is permissive for all levels $i \geq I$ and all $t \in [0, e^{2^{I/10}}]$. By Corollary 3 we have that

$$\mathbb{P}[\mathscr{D}_I^c] \leq \sum_{i\geq I} 3e^{-2^{i/10}} \leq 4e^{-2^{I/10}}.$$

Next set $t_0 = \frac{1}{2}e^{2^{I/10}}$ and let $t_k = t_{k-1} + 4^{I-k}$. Let $\mathscr{H}_k$ denote the event that Y_t is permissive at level $I - k$ for all $t \in [t_k, t_k + e^{2^{(I-k)/10}}]$. By Lemma 7 then for $0 \leq k \leq I - i^*$,

$$\mathbb{P}[\mathscr{H}_k^c, \cap_{k'=1}^{k-1} \mathscr{H}_{k'}, \mathscr{D}_{i^*}] \leq 3e^{-2^{(I-k)/10}}.$$

Thus, provided i^* is large enough,

$$\mathbb{P}[\cap_{k'=1}^{I-i^*} \mathscr{H}_{k'}, \mathscr{D}_{i^*}] \geq 1 - 4e^{-2^{I/10}} - \sum_{k'=1}^{I-i^*} 3e^{-2^{(I-k)/10}} \geq \frac{1}{2}.$$

Let $\tau = \tau_I = t_{I-i^*}$. Then for all $I \geq i^*$,

$$\mathbb{P}\left[\forall i \geq i^*, Y_{\tau_I}(2^i) \geq 10i2^{i/2}\right] \geq \frac{1}{2}.$$

since Y_t is stochastically decreasing in t and $\tau_I \to \infty$ as $I \to \infty$,

$$\inf_t \mathbb{P}\left[\forall i \geq i^*, Y_t(2^i) \geq 10i2^{i/2}\right] \geq \frac{1}{2}.$$

Theorem 2 *For $K > 1$ there exists a random function $Y^*(s)$ such that Y_t converges weakly to Y^* in finite dimensional distributions. Furthermore, with*

$$\alpha^* = \frac{K}{2}\mathbb{E}\left[\mathbb{P}[\max_{s\geq 0} W_s - Y^*(s) \leq 0 \mid Y^*]\right],$$

we have that $\frac{1}{t}X_t$ converges in probability to $\alpha^* > 0$.

Proof Since Y_t is stochastically decreasing it must converge in distribution to some limit Y^*. By Claim 2

$$\mathbb{P}\left[\frac{K}{2}\mathbb{P}[\max_{s\geq 0} W_s - Y^*(s) \leq 0 \mid Y^*] \geq \frac{K}{10}2^{-i/2}\prod_{i=i^*}^{\infty}(1 - e^{1-\max\{1,10i/\sqrt{2}\}})\right] \geq \frac{1}{2},$$

and so $\alpha^* = \lim_t \mathbb{E}S(t) > 0$. To show convergence in probability fix $\epsilon > 0$. For some large enough L,

$$\mathbb{E}[\frac{1}{L}X_L] = \frac{1}{L}\int_0^L \mathbb{E}S(t)dt \leq \alpha^* + \epsilon/2.$$

Let $N_k = \mathbb{E}[X_{kL} - X_{(k-1)L} \mid \mathscr{F}_{(k-1)L}]$ and $R_k = X_{kL} - X_{(k-1)L} - N_k$. By monotonicity

$$\frac{1}{L} N_k \leq \mathbb{E}[\frac{1}{L} X_L] \leq \alpha^* + \epsilon/2.$$

The sequence R_k are martingale differences with uniformly bounded exponential moments (since it is bounded from below by $-(\alpha^* + \epsilon/2)$ and stochastically dominated by a Poisson with mean LK). Thus

$$\lim_n \frac{1}{n} \sum_{k=1}^{n} R_k = 0 \text{ a.s. .}$$

It follows that almost surely $\limsup_t \frac{1}{t} X_t \leq \alpha^*$. Since X_t is stochastically dominated by Poisson(Kt) we have that $\mathbb{E}[(\frac{1}{t} X_t)^2] \leq K^2 + K/t$ and so is uniformly bounded. Hence since $\lim \mathbb{E} \frac{1}{t} X_t \to \alpha^*$ it follows that we must have that $\frac{1}{t} X_t$ converges in distribution to α^*.

4 Regeneration Times

In order to establish almost sure convergence to the limit we define a series of regeneration times. We select some small $\alpha(K) > 0$, and say an integer time t is a regeneration time if

1. The function Y_t satisfies $\inf_s Y_t(s) - y_\alpha(s) \geq 0$.
2. For J_t the set of particles to the right of the aggregate at time t, their trajectories $\{\zeta_j(s)\}_{j \in J_t}$ on $(-\infty, t]$ satisfy

$$\inf_s \zeta_j(t-s) - (X_t - y_\alpha(s)) > 0. \tag{5}$$

Let $0 \leq T_1 < T_2 < \ldots$ denote the regeneration times and let $\mathfrak{R}$ denote the set of regeneration times.

Lemma 9 *For all $K > 1$, there exists $\delta(K) > 0$ such that,*

$$\inf_{t \in \mathbb{N}} \mathbb{P}[t \in \mathfrak{R}] \geq \delta.$$

Proof Let $\mathscr{D}_t$ be the event that $\inf_s Y_t(s) - y_\alpha(s) \geq 0$. Provided that $\alpha(K)$ is small enough by Lemmas 4 and 8 we have that

$$\mathbb{P}[\mathscr{D}_t] \geq \frac{1}{3}.$$

As the density of particles to the right of X_t is increasing in Y_t it is, therefore greatest when $t = 0$ and so $\mathbb{P}[t \in \mathfrak{R} \mid \mathscr{D}_t]$ is minimized at $t = 0$. Let w_ℓ be defined as the probability

$$w_\ell = \mathbb{P}[\max_{s \geq 0} W_s - y_\alpha(s) > \ell].$$

Then $K w_\ell$ is the Poisson density of particles at ℓ which fail to satisfy (5) for time $t = 0$. For $0 \leq \ell < \alpha^{-4}$ we simply bound $w_\ell \leq 1$ so let us consider $\ell \geq \alpha^{-4}$. Then

$$\begin{aligned} w_\ell &\leq 1 - \mathbb{P}[M_\ell \leq \ell, \forall i \geq \lfloor \log_2(\ell) \rfloor : \ M_{2^{i+1}} \leq \ell + i2^{i+1}] \\ &\leq 1 - \mathbb{P}[M_\ell \leq \ell] \prod_{i \geq \lfloor \log_2(\ell) \rfloor} \mathbb{P}[M_{2^{i+1}} \leq \ell + i2^{i+1}] \\ &\leq 1 - (1 - e^{1-\ell^{1/2}}) \prod_{i \geq \lfloor \log_2(\ell) \rfloor} (1 - e^{1-i/\sqrt{2}}) \\ &\leq e^{1-\ell^{1/2}} + \sum_{i \geq \lfloor \log_2(\ell) \rfloor} e^{1-i/\sqrt{2}} \end{aligned}$$

where the third inequality is be the FKG inequality and the final inequality is by Eq. (3). Then we have that

$$\begin{aligned} \sum_{\ell \geq \alpha^{-4}} w_\ell &\leq \sum_{\ell \geq \alpha^{-4}} e^{1-\ell^{1/2}} + \sum_{\ell \geq \alpha^{-4}} \sum_{i \geq \lfloor \log_2(\ell) \rfloor} e^{1-i/\sqrt{2}} \\ &\leq \sum_{\ell \geq \alpha^{-4}} e^{1-\ell^{1/2}} + \sum_i 2^{i+1} e^{1-i/\sqrt{2}} < \infty, \end{aligned}$$

since $2e^{-1/\sqrt{2}} < 1$. Hence $\sum_{\ell=0}^{\infty} w_\ell < \infty$ and so since the number of particles which fail to satisfy (5) at time 0 is distributed as a Poisson with mean $K \sum_{\ell=0}^{\infty} w_\ell$,

$$\mathbb{P}[0 \in \mathfrak{R} \mid \mathscr{D}_0] = \mathbb{P}[\text{Poisson}(K \sum_{\ell=0}^{\infty} w_\ell) = 0] > 0.$$

Thus there exists $\delta > 0$ such that $\inf_{t \in \mathbb{N}} \mathbb{P}[t \in \mathfrak{R}] \geq \delta$.

We can now establish our main result.

Proof (Theorem 1) By Lemma 9 there is a constant density of regeneration times so the expected inter-arrival time is finite. By Theorem 2 the process X_t travels at speed α^*, at least in probability. By the Strong Law of Large Numbers for renewal-reward processes this convergence must also be almost sure.

5 Higher Dimensions

Our approach gives a simple way of proving positive speed in higher dimensions although not down to the critical threshold. Simulations for small K in two dimensions produce pictures which look very similar to the classical DLA model. Surprisingly, however, Eldan [2] conjectured that the critical value for $d \geq 2$ is 0! That is to say that despite the simulations there is linear growth in of the aggregate for all densities of particles and that these simulations are just a transitory effect reflecting that we are not looking at large enough times. We are inclined to agree but our techniques will only apply for larger values of K. A better understanding of the notoriously difficult classical DLA model may be necessary, for instance that the aggregate has dimension smaller than 2.

Let us now assume that $K > 1$. In the setting of $\mathbb{Z}^d$ it will be convenient for the sake of notation to assume that the particles perform simple random walks with rate d which simply speeds the process be a factor of d. The projection of the particles in each co-ordinate is then a rate 1 walk. We let U_t be the location of the rightmost particle in the aggregate (if there are multiple rightmost particles take the first one) at time t and let X_t denote its first coordinate. We then define $Y_t(s)$ according to (1) as before. We call a particle with path $\big(Z_1(t), \ldots, Z_d(t)\big)$ *conforming* at time t if $Z_1(s) > X_s$ for all $s \leq t$. By construction conforming particles cannot be part of the aggregate and conditional on X_t form a Poisson process with intensity depending only on the first coordinate.

Let e_i denote the unit vector in coordinate i. The intensity of conforming particles at time t at $U_t + e_1$ is then simply

$$K\mathbb{P}[\max_{0\leq s\leq t} W_s - Y_t(s) \leq 0 \mid Y_t].$$

where W_s is an independent simple random walk. Similarly the rate at which conforming particles move from $U_t + e_1$ to Y_t thus forming a new rightmost particle is

$$S(t) = \frac{1}{2}K\mathbb{P}[\max_{0\leq s\leq t} W_s - Y_t(s) \leq 0 \mid Y_t],$$

the same as the formula we found in the one dimensional case. Of course by restricting to conforming particles we are restricting ourselves and so the rate at which X_t increments is strictly larger than $S(t)$. Since $S(t)$ is increasing as a function of X_t (through Y_t) we can stochastically dominate the one dimensional case by the higher dimensional process which establishes Corollary 1.

Let us now briefly describe how to improve upon $K = 1$. In the argument above we are being wasteful in two regards, first by only considering conforming particles and secondly by considering only a single rightmost particle. If there are two rightmost particles then the rate at which X_t increases doubles. The simplest way to get such a new particle is for a conforming particle at $U_t + e_1 \pm e_i$ to jump first to $U_t \pm e_i$ and then to U_t. There are $(2d-2)$ such location and the first move occurs at rate $S(t)$ and the second at has probability $1/(2d)$ to move in the correct direction and takes time exponential with rate d. After this sequence of events the rate at which X_t increments becomes $2S(t)$.

In Lemma 2, on which the whole proof effectively rests, we show that for $\rho < 1$ if $S(s) \geq \gamma$ for $s \in [t, t+\Delta]$ then with exponentially high probability $X_{t+\Delta} - X_t \geq \rho\Delta\gamma$ for any $\rho < 1$ which is intuitively obvious since X_t grows at rate $S(s) \geq \gamma$. We can improve our lower bound on K by increasing the range of ρ for which this holds for small values of γ.

Define the following independent random variables

$$V_1 \sim \operatorname{Exp}(\gamma),\, V_2 \sim \operatorname{Exp}(\frac{2d-2}{2d}\gamma),\, V_3 \sim \operatorname{Exp}(d),\, V_4 \sim \operatorname{Exp}(\gamma)$$

where we interpret V_1 as the time until the first conforming particle hits U_t. We will view V_2 as the waiting time for a conforming particle to move from $U_t + \pm e_i + e_1$ to $U_t \pm e_i$ for some $2 \leq i \leq d$ and we further specify that their next step will move directly to U_t which thins the process by a factor $\frac{1}{2d}$. Let V_3 be the time until its next move. On the event $V_2 + V_3 < V_1$ there is an additional rightmost particle before one has been added to the right of U_t. Now let V_4 be the first time a conforming particle reaches this new rightmost site. So the time for X_t to increase is stochastically dominated by

$$T = \min\{V_1, V_2 + V_3 + V_4\}.$$

Now using the memoryless property of exponential random variables,

$$\begin{aligned}
\mathbb{E}T &= \mathbb{E}V_1 - \mathbb{E}\left[(V_1 - (V_2 + V_3 + V_4))I(V_1 \geq V_2 + V_3 + V_4)\right] \\
&= \frac{1}{\gamma}(1 - \mathbb{P}[V_1 \geq V_2 + V_3 + V_4])
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{P}[V_1 \geq V_2 + V_3 + V_4] &= \mathbb{P}[V_1 \geq V_2]\mathbb{P}[V_1 \geq V_2 + V_3 \mid V_1 \geq V_2] \\
&\quad \times \mathbb{P}[V_1 \geq V_2 + V_3 + V_3 \mid V_1 \geq V_2 + V_3] \\
&= \frac{\frac{2d-2}{2d}\gamma}{\gamma + \frac{2d-2}{2d}\gamma} \frac{d}{\gamma + d} \frac{\gamma}{2\gamma} \\
&= \frac{d-1}{2(2d-1)} \frac{d}{\gamma + d}
\end{aligned}$$

In the proof we need only to consider the case where γ is close to 0 and

$$\lim_{\gamma \to 0} \gamma \mathbb{E} T = \frac{3d - 1}{4d - 2}.$$

Having X_t growing at rate $\gamma \frac{4d-2}{3d-1}$ corresponds in the proof to linear growth provided that $K > \frac{3d-1}{4d-2}$. In the case for $d = 2$ this means $K > \frac{5}{6}$. We are still being wasteful in several ways and expect that a more careful analysis would yield better bounds that tend to 0 as $d \to \infty$. However, we don't believe that this approach alone is sufficient to show that the critical value of K is 0 when $d \geq 2$. For that more insight into the local structure is likely needed along with connections to standard DLA.

6 Open Problems

In the one dimensional case the most natural open questions concern the behaviour of X_t for densities close to 1. Rath and Sidoravicius made a series of predictions in [4] including that the speed should be approximately $\frac{1}{2}(K - 1)$ when K is slightly above 1. Perhaps of most interest is what is the exponent of growth for X_t when $K = 1$. Here the heuristics of [4] suggest that it may grow as $t^{2/3}$. In a related model, Dembo and Tsai [1] established $t^{2/3}$ at the critical density.

In higher dimensions the main open problem is to establish Eldan's conjecture of linear growth for all K. Another natural question is to prove a shape theorem for the aggregate.

Acknowledgments The author would like to that Vladas Sidoravicius for suggesting the problem useful discussions and for his comments on a draft of the paper and his hospitality at NYU Shanghai. In this and many other papers, Vladas had a profound impact on my research direction. This work was supported by an NSF career award (DMS-1352013), a Simons Investigator grant and a MacArthur Fellowship.

References

1. Dembo, A., Tsai, L.-C.: Criticality of a randomly-driven front. Arch. Ration. Mech. Anal. **233**, 643–699 (2019)
2. Eldan, R.: Personal Communication (2016)
3. Kesten, H., Sidoravicius, V.: A problem in one-dimensional diffusion-limited aggregation (DLA) and positive recurrence of Markov chains. Ann. Probab. **36**(5), 1838–1879 (2008)
4. Sidoravicius, V., Rath, B.: One-dimensional multi-particle DLA–a PDE approach. arXiv preprint:1709.00484 (2017)
5. Sidoravicius, V., Stauffer, A.: Multi-particle diffusion limited aggregation. Invent. Math. **218**, 491–571 (2019)
6. Voss, R.F.: Multiparticle fractal aggregation. J. Stat. Phys. **36**(5–6), 861–872 (1984)
7. Witten Jr, T.A., Sander, L.M.: Diffusion-limited aggregation, a kinetic critical phenomenon. Phys. Rev. Lett. **47**(19), 1400 (1981)

On the C^1-Property of the Percolation Function of Random Interlacements and a Related Variational Problem

Alain-Sol Sznitman

In memory of Vladas Sidoravicius

Abstract We consider random interlacements on $\mathbb{Z}^d$, $d \geq 3$. We show that the percolation function that to each $u \geq 0$ attaches the probability that the origin does not belong to an infinite cluster of the vacant set at level u, is C^1 on an interval $[0, \widehat{u})$, where $\widehat{u}$ is positive and plausibly coincides with the critical level u_* for the percolation of the vacant set. We apply this finding to a constrained minimization problem that conjecturally expresses the exponential rate of decay of the probability that a large box contains an excessive proportion ν of sites that do not belong to an infinite cluster of the vacant set. When u is smaller than $\widehat{u}$, we describe a regime of "small excess" for ν where all minimizers of the constrained minimization problem remain strictly below the natural threshold value $\sqrt{u_*} - \sqrt{u}$ for the variational problem.

Keywords Random interlacements · Percolation function · Variational problem

MSC (2010) 60K35, 35A15, 82B43

1 Introduction

In this work we consider random interlacements on $\mathbb{Z}^d$, $d \geq 3$, and the percolation of the vacant set of random interlacements. We show that the percolation function θ_0 that to each level $u \geq 0$ attaches the probability that the origin does not belong to an *infinite* cluster of $\mathcal{V}^u$, the vacant set at level u of the random interlacements, is C^1 on an interval $[0, \widehat{u})$, where $\widehat{u}$ is positive and plausibly coincides with the critical level

A.-S. Sznitman (✉)
Department of Mathematics, ETH Zürich, Zürich, Switzerland
e-mail: sznitman@math.ethz.ch

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_32

u_* for the percolation of $\mathscr{V}^u$, although this equality is presently open. We apply this finding to a constrained minimization problem that for $0 < u < u_*$ conjecturally expresses the exponential rate of decay of the probability that a large box contains an excessive proportion ν bigger than $\theta_0(u)$ of sites that do not belong to the infinite cluster of $\mathscr{V}^u$. When $u > 0$ is smaller than $\widehat{u}$ and ν close enough to $\theta_0(u)$, we show that all minimizers φ of the constrained minimization problem are $C^{1,\alpha}$-functions on $\mathbb{R}^d$, for all $0 < \alpha < 1$, and their supremum norm lies strictly below $\sqrt{u}_* - \sqrt{u}$. In particular, the corresponding "local level" functions $(\sqrt{u} + \varphi)^2$ do not reach the critical value u_*.

We now discuss our results in more details. We consider random interlacements on $\mathbb{Z}^d$, $d \geq 3$, and refer to [4] or [6] for background material. For $u \geq 0$, we let $\mathscr{I}^u$ stand for the random interlacements at level u and $\mathscr{V}^u = \mathbb{Z}^d \backslash \mathscr{I}^u$ for the vacant set at level u. A key object of interest is the percolation function

$$\theta_0(u) = \mathbb{P}[0 \overset{u}{\nleftrightarrow} \infty], \text{ for } u \geq 0, \tag{1}$$

where $\{0 \overset{u}{\nleftrightarrow} \infty\}$ is a shorthand for the event $\{0 \overset{\mathscr{V}^u}{\nleftrightarrow} \infty\}$ stating that 0 does not belong to an infinite cluster of $\mathscr{V}^u$. One knows from [14] and [13] that there is a critical value $u_* \in (0, \infty)$ such that θ_0 equals 1 on $(u_{*,\infty})$ and is smaller than 1 on $(0, u_*)$. And from Corollary 1.2 of [16], one knows that the non-decreasing left-continuous function θ_0 is continuous except maybe at the critical value u_*.

With an eye towards applications to a variational problem that we discuss below, see (9), we are interested in proving that θ_0 is C^1 on some (hopefully large) neighborhood of 0. With this goal in mind, we introduce the following definition. Given $0 \leq \alpha < \beta < u_*$, we say that NLF$(\alpha, \beta)$, the no large finite cluster property on $[\alpha, \beta]$, holds when

$$\begin{aligned} &\text{there exists } L_0(\alpha, \beta) \geq 1,\ c_0(\alpha, \beta) > 0,\ \gamma(\alpha, \beta) \in (0, 1] \text{ such that} \\ &\text{for all } L \geq L_0 \text{ and } u \in [\alpha, \beta],\ \mathbb{P}[0 \overset{u}{\longleftrightarrow} \partial B_L,\ 0 \overset{u}{\nleftrightarrow} \infty] \leq e^{-c_0 L^\gamma}, \end{aligned} \tag{2}$$

where $B_L = B(0, L)$ is the closed ball for the sup-norm with center 0 and radius L, ∂B_L its internal boundary (i.e. the subset of sites in B_L that are neighbors of $\mathbb{Z}^d \backslash B_L$), and the notation is otherwise similar to (1). We then set

$$\widehat{u} = \sup\{u \in [0, u_*)\, ;\ \text{NLF}(0, u) \text{ holds}\}. \tag{3}$$

One knows from Corollary 1.2 of [7] that $\widehat{u}$ is positive:

$$\widehat{u} \in (0, u_*]\,. \tag{4}$$

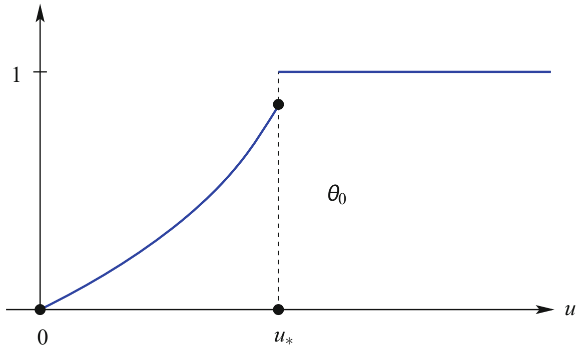


Fig. 1 A heuristic sketch of the graph of θ_0 (with a possible but not expected jump at u_*)

It is open, but plausible, that $\widehat{u} = u_*$ (see also [8] for related progress in the context of level-set percolation of the Gaussian free field). Our first main result is:

Theorem 1

$$\textit{The function } \theta_0 \textit{ is } C^1 \textit{ on } [0, \widehat{u}) \textit{ and} \tag{5}$$

$$\theta_0' \textit{ is positive on } [0, \widehat{u}). \tag{6}$$

Incidentally, let us mention that in the case of Bernoulli percolation the function corresponding to θ_0 is known to be C^∞ in the supercritical regime, see Theorem 8.92 of [10]. However, questions pertaining to the sign of the second derivative (in particular the possible convexity of the corresponding function in the supercritical regime) are presently open. Needless to say that in our case the shape of the function θ_0 is not known (and the sketch in Fig. 1 conceivably misleading).

Our interest in Theorem 1 comes in conjunction with an application to a variational problem that we now describe. We consider

$$D \text{ the closure of a smooth bounded domain, or of an open sup-norm ball, of } \mathbb{R}^d \text{ that contains } 0. \tag{7}$$

Given u and ν such that

$$0 < u < u_* \text{ and } \theta_0(u) \le \nu < 1, \tag{8}$$

we introduce the constrained minimization problem

$$I_{u,\nu}^D = \inf\Big\{\frac{1}{2d}\int_{\mathbb{R}^d} |\nabla\varphi|^2 dz;\, \varphi \ge 0, \varphi \in C_0^\infty(\mathbb{R}^d),\ \fint_D \theta_0\big((\sqrt{u}+\varphi)^2\big)\, dz > \nu\Big\}, \tag{9}$$

where $C_0^\infty(\mathbb{R}^d)$ stands for the set of smooth compactly supported functions on $\mathbb{R}^d$ and $⨍_D \dots dz$ for the normalized integral $\frac{1}{|D|}\int \dots dz$ with $|D|$ the Lebesgue measure of D (see also below (10) for the interpretation of φ).

The motivation for the variational problem (9) lies in the fact that it conjecturally describes the large deviation cost of having a fraction at least ν of sites in the large discrete blow-up $D_N = (ND) \cap \mathbb{Z}^d$ of D that are not in the infinite cluster $\mathscr{C}_\infty^u$ of $\mathscr{V}^u$. One knows by the arguments of Remark 6.6 2) of [15] that

$$\liminf_N \frac{1}{N^{d-2}} \log \mathbb{P}[| D_N \backslash \mathscr{C}_\infty^u | \geq \nu \, |D_N|] \geq -I_{u,\nu}^D \quad \text{for } u, \nu \text{ as in (8)} \tag{10}$$

(with $|A|$ standing for the number of sites in A for A subset of $\mathbb{Z}^d$).

It is presently open whether the lim inf can be replaced by a limit and the inequality by an equality in (10), i.e. if there is a matching asymptotic upper bound. If such is the case, there is a direct interest in the introduction of a notion of minimizers for (9). Indeed, $(\sqrt{u} + \varphi)^2(\frac{\cdot}{N})$ can be interpreted as the slowly varying local levels of the tilted interlacements that enter the derivation of the lower bound (10) (see Section 4 and Remark 6.6 2) of [15]). In this perspective, it is a relevant question whether minimizers φ reach the value $\sqrt{u_*} - \sqrt{u}$. The regions where they reach the value $\sqrt{u_*} - \sqrt{u}$ could potentially reflect the presence of droplets secluded from the infinite cluster $\mathscr{C}_\infty^u$ and taking a share of the burden of creating an excess fraction ν of sites of D_N that are not in $\mathscr{C}_\infty^u$ (see also the discussion at the end of Sect. 3).

The desired notion of minimizers for (9) comes in Theorem 2 below. For this purpose we introduce the right-continuous modification $\overline{\theta}_0$ of θ_0:

$$\overline{\theta}_0(u) = \begin{cases} \theta_0(u), & \text{when } 0 \leq u < u_*, \\ 1, & \text{when } u \geq u_*. \end{cases} \tag{11}$$

Clearly, $\overline{\theta}_0 \geq \theta_0$ and it is plausible, but presently open, that $\overline{\theta}_0 = \theta_0$. We recall that $D^1(\mathbb{R}^d)$ stands for the space of locally integrable functions f on $\mathbb{R}^d$ with finite Dirichlet energy that decay at infinity, i.e. such that $\{|f| > a\}$ has finite Lebesgue measure for all $a > 0$, see Chapter 8 of [11], and define for D, u, ν as in (7), (8)

$$\overline{J}_{u,\nu}^D = \inf\Big\{\frac{1}{2d}\int_{\mathbb{R}^d} |\nabla\varphi|^2\, dz; \varphi \geq 0, \varphi \in D^1(\mathbb{R}^d),\ ⨍_D \overline{\theta}_0\big((\sqrt{u}+\varphi)^2\big)\, dz \geq \nu\Big\}. \tag{12}$$

Since $\overline{\theta}_0 \geq \theta_0$ and $D^1(\mathbb{R}^d) \supseteq C_0^\infty(\mathbb{R}^d)$, we clearly have $\overline{J}_{u,\nu}^D \leq I_{u,\nu}^D$. But in fact:

Theorem 2 *For D, u, ν as in (7), (8), one has*

$$\overline{J}_{u,\nu}^D = I_{u,\nu}^D. \tag{13}$$

In addition, the infimum in (12) is attained:

$$\overline{J}^D_{u,\nu} = \min\Big\{\frac{1}{2d}\int_{\mathbb{R}^d} |\nabla\varphi|^2\,dz;\ \varphi \ge 0,\ \varphi \in D^1(\mathbb{R}^d),\ \fint_D \overline{\theta}_0\big((\sqrt{u}+\varphi)^2\big)\,dz \ge \nu\Big\}. \tag{14}$$

and any minimizer φ in (14) satisfies

$$\begin{aligned} &0 \le \varphi \le \sqrt{u}_* - \sqrt{u} \text{ a.e.,}\\ &\varphi \text{ is harmonic outside } D, \text{ and } \operatorname*{ess\,sup}_{z\in\mathbb{R}^d} |z|^{d-2}\varphi(z) < \infty. \end{aligned} \tag{15}$$

Thus, Theorem 2 provides a notion of minimizers for (9), the variational problem of interest. Its proof is given in Sect. 3. Additional properties of (14) and the corresponding minimizers can be found in Remark 1. We refer to Chapter 11 §3 of [1] for other instances of non-smooth variational problems.

In Sect. 4 we bring into play the C^1-property of θ_0 and show

Theorem 3 *If $u_0 \in (0, u_*)$ is such that*

$$\theta_0 \text{ is } C^1 \text{ on a neighborhood of } [0, u_0], \tag{16}$$

then for any $u \in (0, u_0)$ there are $c_1(u, u_0, D) < \theta_0(u_*) - \theta_0(u)$ and $c_2(u, u_0) > 0$ such that

$$\begin{aligned} &\text{for } \nu \in [\theta_0(u), \theta_0(u) + c_1], \text{ any minimizer } \varphi \text{ in (14) is } C^{1,\alpha} \text{ for all}\\ &0 < \alpha < 1, \text{ and } 0 \le \varphi \le \big\{c_2\big(\nu - \theta_0(u)\big)\big\} \wedge (\sqrt{u}_0 - \sqrt{u})\ (< \sqrt{u}_* - \sqrt{u}).\\ &\text{Here } C^{1,\alpha} \text{ stands for the } C^1\text{-functions with } \alpha\text{-Hölder continuous partial}\\ &\text{derivatives.} \end{aligned} \tag{17}$$

In view of Theorem 1 the above Theorem 3 applies to any $u_0 < \widehat{u}$ (with $\widehat{u}$ as in (3)). It describes a regime of "small excess" for ν where minimizers do not reach the threshold value $\sqrt{u}_* - \sqrt{u}$. In the proof of Theorem 3 we use the C^1-property to write an Euler-Lagrange equation for the minimizers, see (90), and derive a bound in terms of $\nu - \theta_0(u)$ of the corresponding Lagrange multipliers, see (91). It is an interesting open problem whether a regime of "large excess" for ν can be singled out where some (or all) minimizers of (14) reach the threshold value $\sqrt{u}_* - \sqrt{u}$ on a set of positive Lebesgue measure. We refer to Remark 2 for some simple minded observations related to this issue.

Finally, let us state our convention about constants. Throughout we denote by $c, c', \widetilde{c}$ positive constants changing from place to place that simply depend on the dimension d. Numbered constants $c_0, c_1, c_2, \dots$ refer to the value corresponding to their first appearance in the text. Dependence on additional parameters appears in the notation.

2 The C^1-Property of θ_0

The main object of this section is to prove Theorem 1 stated in the Introduction. Theorem 1 is the direct consequence of the following Lemma 1 and Proposition 1. We let $g(\cdot,\cdot)$ stand for the Green function of the simple random walk on $\mathbb{Z}^d$.

Lemma 1 *For $0 \le u < u_*$, one has*

$$\liminf_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \big(\theta_0(u+\varepsilon) - \theta_0(u)\big) \ge \big(1 - \theta_0(u)\big) \frac{1}{g(0,0)}. \tag{18}$$

Proposition 1 *For any $0 \le \alpha < \beta < u_*$ such that NLF(α,β) holds (see (2)),*

$$\theta_0 \textit{ is } C^1 \textit{ on } \Big[\alpha, \frac{\alpha+\beta}{2}\Big]. \tag{19}$$

As we now explain, Theorem 1 follows immediately. By Proposition 1 and a covering argument, one see that θ_0 is C^1 on $[0,\widehat{u})$. Then, by Lemma 1, one finds that $\theta_0' > 0$ on $[0,\widehat{u})$, and Theorem 1 follows.

There remains to prove Lemma 1 and Proposition 1.

Proof of Lemma 1 Consider $u \ge 0$ and $\varepsilon > 0$ such that $u+\varepsilon < u_*$. Then, denoting by $\mathscr{I}^{u,u+\varepsilon}$ the collection of sites of $\mathbb{Z}^d$ that are visited by trajectories of the interlacement with level lying in $(u, u+\varepsilon]$, we have

$$\begin{aligned}
\theta_0(u+\varepsilon) - \theta_0(u) &= \mathbb{P}[0 \overset{u}{\longleftrightarrow} \infty,\ 0 \overset{u+\varepsilon}{\not\longleftrightarrow} \infty] \\
&\ge \mathbb{P}[0 \overset{u}{\longleftrightarrow} \infty,\ 0 \in \mathscr{I}^{u,u+\varepsilon}] \\
&\overset{\text{independence}}{=} \big(1-\theta_0(u)\big)\, \mathbb{P}[0 \in \mathscr{I}^{u,u+\varepsilon}] \\
&= \big(1-\theta_0(u)\big)\big(1 - e^{-\varepsilon/g(0,0)}\big).
\end{aligned} \tag{20}$$

Dividing by ε both members of (20) and letting ε tend to 0 yields (18). This proves Lemma 1. □

We now turn to the proof of Proposition 1. An important tool is Lemma 2 below. We will use Lemma 2 to gain control over the difference quotients of θ_0, as expressed in (27) or (37) below. The claimed C^1-property of θ_0 on $[\alpha, \frac{\alpha+\beta}{2}]$ will then quickly follow, see below (37). To prove (27) with Lemma 2, we define an increasing sequence of levels u_i, $1 \le i \le i_\eta$ so that $u_1 = u'$ (in Proposition 1) and $u_i - u$ doubles as i increases by one unit, until it reaches η (of (27)), and in essence apply Lemma 2 repeatedly to compare the successive difference quotients of θ_0 between u and u_i, see (32) till (36).

Proof of Proposition 1 We consider $0 \le \alpha, \beta < u_*$ such that NLF(α,β) holds (see (0.2)), and set

$$c_3\,(\alpha,\beta) = 2/c_0\,. \tag{21}$$

As mentioned above, an important tool in the proof of Proposition 1 is provided by

Lemma 2 *Consider* $u < u' \le u''$ *in* $[\alpha, \frac{\alpha+\beta}{2}]$ *such that*

$$u'' - u \le e^{-\frac{1}{c_3} L_0^\gamma} \; (\le 1), \tag{22}$$

and set

$$\Delta' = \frac{1}{u' - u} \big(\theta_0(u') - \theta_0(u)\big) \;\; \textit{and} \;\; \Delta'' = \frac{1}{u'' - u} \big(\theta_0(u'') - \theta_0(u)\big), \tag{23}$$

as well as $L' \ge L'' \ge L_0$ *(with* L_0 *as in (2)) via*

$$L' = \Big(c_3 \;\log\; \frac{1}{u' - u}\Big)^{1/\gamma} \;\; \textit{and} \;\; L'' = \Big(c_3 \;\log\; \frac{1}{u'' - u}\Big)^{1/\gamma}. \tag{24}$$

Then, with cap$(\cdot)$ *denoting the simple random walk capacity, one has*

$$|\Delta' - e^{(u''-u')\,\mathrm{cap}(B_{L'})}\, \Delta''| \le 3(u'' - u)\big(1 + \mathrm{cap}(B_{L'})^2\big)\, e^{(u''-u')\,\mathrm{cap}(B_{L'})}. \tag{25}$$

Let us first admit Lemma 2 and conclude the proof of Proposition 1 (i.e. that θ_0 is C^1 on $[\alpha, \frac{\alpha+\beta}{2}]$). We introduce

$$\eta_0 = \frac{1}{4}\Big(\frac{\beta - \alpha}{2} \wedge e^{-\frac{1}{c_3} L_0^\gamma}\Big) \;\Big(\le \frac{1}{4}\Big). \tag{26}$$

We will use Lemma 2 to show that

$$\text{when } 0 < \eta \le \eta_0, \text{ then for all } u < u' \text{ in } \Big[\alpha, \frac{\alpha+\beta}{2}\Big] \text{ with } u' \le u + \eta, \text{ one has}$$
$$\Big|\frac{1}{u' - u}\big(\theta_0(u') - \theta_0(u)\big) - \frac{1}{\eta}\big(\theta_0(u + \eta) - \theta_0(u)\big)\Big| \le c(\alpha, \beta)\,\sqrt{\eta}. \tag{27}$$

Once (27) is established, Proposition 1 will quickly follow (see below (37)). For the time being we will prove (27). To this end we set

$$u_i = 2^{i-1}(u' - u) + u, \text{ for } 1 \le i \le i_\eta, \text{ where } i_\eta = \max\{i \ge 1, u_i \le u + \eta\} \tag{28}$$

(note that $u_1 = u'$), as well as

$$\Delta_i = \frac{1}{u_i - u}\big(\theta_0(u_i) - \theta_0(u)\big) \;\text{ and }\; L_i = \Big(c_3 \;\log\; \frac{1}{u_i - u}\Big)^{\frac{1}{\gamma}} \; \big(\overset{(26)}{\ge} L_0\big),$$
$$\text{for } 1 \le i \le i_\eta. \tag{29}$$

We also define

$$\delta_i = (u_i - u)\,\mathrm{cap}(B_{L_i}) \ \text{ and } \ \widetilde{\delta}_i = 6(u_i - u) + 6\delta_i\,\mathrm{cap}(B_{L_i}),\ \text{ for } \ 1 \le i \le i_\eta. \tag{30}$$

We will apply (25) of Lemma 2 to $u' = u_i$, $u'' = u_{i+1}$, when $1 \le i < i_\eta$, and to $u' = u_{i_\eta}$, $u'' = u + \eta$. We note that for $1 \le i \le i_\eta$, we have $\delta_i \le c(\alpha,\beta)\sqrt{u_i - u}$ and $\widetilde{\delta}_i \le c(\alpha,\beta)\sqrt{u_i - u}$ so that

$$\text{for } 1 \le j \le i_\eta,\ \sum_{1 \le i \le j} \delta_i \le c(\alpha,\beta)\sqrt{u_j - u} \text{ and } \sum_{1 \le i \le j} \widetilde{\delta}_i \le c(\alpha,\beta)\sqrt{u_j - u}. \tag{31}$$

The application of (25) to $u'' = u_{i+1}$, $u' = u_i$, for $1 \le i < i_\eta$ and the observation that $u_{i+1} - u_i = u_i - u$ yield the inequality

$$\begin{aligned}&|\Delta_i - e^{\delta_i}\Delta_{i+1}| \le c\,\widetilde{\delta}_i\, e^{\delta_i}, \text{ for } 1 \le i < i_\eta, \text{ so that}\\ &\Big|e^{\sum_{\ell<i}\delta_\ell}\Delta_i - e^{\sum_{\ell<i+1}\delta_\ell}\Delta_{i+1}\Big| \le c\, e^{\sum_{\ell<i+1}\delta_\ell}\,\widetilde{\delta}_i,\ \text{ for } 1 \le i < i_\eta.\end{aligned} \tag{32}$$

Hence, adding these inequalities, we find that

$$\Big|\Delta_1 - e^{\sum_{\ell<i_\eta}\delta_\ell}\Delta_{i_\eta}\Big| \le c \sum_{1 \le i < i_\eta} e^{\sum_{\ell<i+1}\delta_\ell}\,\widetilde{\delta}_i \overset{(31),\eta\le\frac{1}{4}}{\le} c(\alpha,\beta)\sqrt{\eta}\,. \tag{33}$$

Then, the application of (25) to $u'' = u + \eta$ and $u' = u_{i_\eta}$, noting that $u + \eta - u_{i_\eta} \le u_{i_\eta} - u$, yields

$$\Big|\Delta_{i_\eta} - e^{(u+\eta-u_{i_\eta})\,\mathrm{cap}(B_{L_{i_\eta}})}\,\frac{1}{\eta}\,\big(\theta_0(u+\eta) - \theta_0(u)\big)\Big| \le \widetilde{\delta}_{i_\eta}\, e^{\delta_{i_\eta}} \le c(\alpha,\beta)\sqrt{\eta}. \tag{34}$$

Multiplying both members of (34) by $e^{\sum_{\ell<i_\eta}\delta_\ell}$ and using (33) and (31) as well, we thus find

$$\begin{aligned}&\Big|\frac{1}{u'-u}\,\big(\theta_0(u') - \theta_0(u)\big) - e^{\sum_{\ell<i_\eta}\delta_\ell + (u+\eta-u_{i_\eta})\,\mathrm{cap}(B_{L_{i_\eta}})}\,\frac{1}{\eta}\,\big(\theta_0(u+\eta) - \theta_0(u)\big)\Big|\\ &\le c(\alpha,\beta)\sqrt{\eta}\end{aligned} \tag{35}$$

and the term inside the exponential is at most $c(\alpha,\beta)\sqrt{\eta}$.

Applying (35) with the choice $\eta = \eta_0$, see (26), one obtains that

$$\sup_{\alpha \le u < u' \le \frac{\alpha+\beta}{2},\, u' \le u + \eta_0} \frac{1}{u'-u}\,\big(\theta_0(u') - \theta_0(u)\big) \le c(\alpha,\beta). \tag{36}$$

Coming back to (35), with the help of the observation below (35) and the inequality $e^a - 1 \le c'(\alpha, \beta)a$ for $0 \le a \le c(\alpha, \beta)$, one obtains the claim (27).

We will now see how the C^1-property of θ_0 on $[\alpha, \frac{\alpha+\beta}{2}]$ (i.e. Proposition 1) follows. We note that for $v, w \in [\alpha, \frac{\alpha+\beta}{2}]$ with $0 < |v - w| \le \eta (\le \eta_0)$, the claim (27) applied to $u = v \wedge w$ and $u' = v \vee w$ yields that

$$\Big|\frac{1}{w-v}\,\big(\theta_0(w) - \theta_0(v)\big) - \frac{1}{\eta}\,\big(\theta_0(v \wedge w + \eta) - \theta_0(v \wedge w)\big)\Big| \le c(\alpha, \beta)\,\sqrt{\eta}. \tag{37}$$

Letting $\Gamma(\cdot)$ stand for the modulus of continuity of θ_0 on the interval $[\alpha, \frac{\alpha+\beta}{2}] \subseteq [0, u_*)$, we find that for $v, w \in [\alpha, \frac{\alpha+\beta}{2}]$ with $0 < |v - w| \le \eta\ (\le \eta_0)$, one has

$$\begin{aligned}&\Big|\frac{1}{w-v}\,\big(\theta_0(w) - \theta_0(v)\big) - \frac{1}{\eta}\,\big(\theta_0(v + \eta) - \theta_0(v)\big)\Big| \le \\ &c(\alpha, \beta)\,\sqrt{\eta} + \frac{2}{\eta}\,\Gamma(|w - v|).\end{aligned} \tag{38}$$

The above inequality implies that for any $v \in [\alpha, \frac{\alpha+\beta}{2}]$, when $w \in [\alpha, \frac{\alpha+\beta}{2}]$ tends to v, the difference quotients $\frac{1}{w-v}\,(\theta_0(w) - \delta_0(v))$ are Cauchy. Thus, letting w tend to v, we find that

$$\begin{aligned}&\theta_0 \text{ is differentiable on } [\alpha, \tfrac{\alpha+\beta}{2}], \text{ and for } 0 < \eta \le \eta_0 \text{ and } v \in [\alpha, \tfrac{\alpha+\beta}{2}], \\ &\Big|\theta_0'(v) - \frac{1}{\eta}\,\big(\theta_0(v + \eta) - \theta_0(v)\big)\Big| \le c(\alpha, \beta)\,\sqrt{\eta}.\end{aligned} \tag{39}$$

As a result we see that θ_0' is the uniform limit on $[\alpha, \frac{\alpha+\beta}{2}]$ of continuous functions, and as such θ_0' is continuous. This is the claimed C^1-property of Proposition 1. The last missing ingredient is the

Proof of Lemma 2 We introduce the notation for $v \ge 0$ and $L \ge 1$

$$\theta_{0,L}(v) = \mathbb{P}[0 \overset{v}{\longleftrightarrow\!\!\!\!\!\!/\;\;} \partial B_L], \tag{40}$$

and the approximations of Δ' and Δ'' in (23)

$$\widetilde{\Delta}' = \frac{1}{u'-u}\,\big(\theta_{0,L'}(u') - \theta_{0,L'}(u)\big),\ \ \widetilde{\Delta}'' = \frac{1}{u''-u}\,\big(\theta_{0,L''}(u'') - \theta_{0,L''}(u)\big), \tag{41}$$

where we recall that $L' \ge L''\ (\ge L_0)$ are defined in (24). Note that

$$\begin{aligned}\Delta' &= \frac{1}{u'-u}\,\big(\mathbb{P}[0 \overset{u'}{\longleftrightarrow\!\!\!\!\!\!/\;\;} \infty] - \mathbb{P}[0 \overset{u}{\longleftrightarrow\!\!\!\!\!\!/\;\;} \infty]\big) = \frac{1}{u'-u}\,\mathbb{P}[0 \overset{u}{\longleftrightarrow} \infty, 0 \overset{u'}{\longleftrightarrow\!\!\!\!\!\!/\;\;} \infty], \\ \widetilde{\Delta}' &= \frac{1}{u'-u}\,\mathbb{P}[0 \overset{u}{\longleftrightarrow} \partial B_{L'}, 0 \overset{u'}{\longleftrightarrow\!\!\!\!\!\!/\;\;} \partial B_{L'}],\end{aligned} \tag{42}$$

and as we now explain

$$\Delta' - \widetilde{\Delta}' = \frac{1}{u'-u}\big(\mathbb{P}[0 \stackrel{u'}{\longleftrightarrow} \partial B_{L'},\ 0 \stackrel{u'}{\not\longleftrightarrow} \infty] - \mathbb{P}[0 \stackrel{u}{\longleftrightarrow} \partial B_{L'}, 0 \stackrel{u}{\not\longleftrightarrow} \infty]\big). \tag{43}$$

Indeed, by (42), (42), one has

$$\begin{aligned}
\Delta' - \widetilde{\Delta}' = {} & \frac{1}{u'-u}\big(\mathbb{P}[0 \stackrel{u}{\longleftrightarrow} \infty, 0 \stackrel{u'}{\not\longleftrightarrow} \infty] - \mathbb{P}[0 \stackrel{u}{\longleftrightarrow} \partial B_{L'},\ 0 \stackrel{u'}{\not\longleftrightarrow} \partial B_{L'}]\big) = \\
& \frac{1}{u'-u}\ \big(\mathbb{P}[0 \stackrel{u}{\longleftrightarrow} \partial B_{L'}, 0 \stackrel{u'}{\not\longleftrightarrow} \infty] - \mathbb{P}[0 \stackrel{u}{\longleftrightarrow} \partial B_{L'},\ 0 \stackrel{u}{\not\longleftrightarrow} \infty,\ 0 \stackrel{u'}{\not\longleftrightarrow} \infty] \\
& - \mathbb{P}[0 \stackrel{u}{\longleftrightarrow} \partial B_{L'}, 0 \stackrel{u'}{\not\longleftrightarrow} \partial B_{L'}]\big) = \frac{1}{u'-u}\ \big(\mathbb{P}[0 \stackrel{u}{\longleftrightarrow} \partial B_{L'}, 0 \stackrel{u'}{\not\longleftrightarrow} \partial B_{L'}] \\
& + \mathbb{P}[0 \stackrel{u}{\longleftrightarrow} \partial B_{L'}, 0 \stackrel{u'}{\longleftrightarrow} \partial B_{L'}, 0 \stackrel{u'}{\not\longleftrightarrow} \infty] \\
& - \mathbb{P}[0 \stackrel{u}{\longleftrightarrow} \partial B_{L'}, 0 \stackrel{u}{\not\longleftrightarrow} \infty] - \mathbb{P}[0 \stackrel{u}{\longleftrightarrow} \partial B_{L'}, 0 \stackrel{u'}{\not\longleftrightarrow} \partial B_{L'}]) = \\
& \frac{1}{u'-u}\ \big(\mathbb{P}[0 \stackrel{u'}{\longleftrightarrow} \partial B_{L'}, 0 \stackrel{u'}{\not\longleftrightarrow} \infty] - \mathbb{P}[0 \stackrel{u}{\longleftrightarrow} \partial B_{L'}, 0 \stackrel{u}{\not\longleftrightarrow} \infty]\big),
\end{aligned} \tag{44}$$

whence (43). Clearly, one also has similar identities as in (42)–(43) for Δ'' and $\widetilde{\Delta}''$.

We now proceed with the proof of (25). By (43), we have

$$\begin{aligned}
|\Delta' - \widetilde{\Delta}'| \le {} & \frac{1}{u'-u}\ \max\big\{\mathbb{P}[0 \stackrel{u'}{\longleftrightarrow} \partial B_{L'}, 0 \stackrel{u'}{\not\longleftrightarrow} \infty],\ \mathbb{P}[0 \stackrel{u}{\longleftrightarrow} \partial B_{L'}, 0 \stackrel{u}{\not\longleftrightarrow} \infty]\big\} \\
& \stackrel{(2)}{\le} \frac{1}{u'-u}\ e^{-c_0\, L'^{\gamma}} \stackrel{(24),(21)}{=} u' - u,
\end{aligned} \tag{45}$$

and likewise we have

$$|\Delta'' - \widetilde{\Delta}''| \le u'' - u. \tag{46}$$

We will now compare $\widetilde{\Delta}'$ and $\widetilde{\Delta}''$. We first recall that when Z is a Poisson-distributed random variable with parameter $\lambda > 0$, then one has

$$P[Z \ge 2] = 1 - e^{-\lambda} - \lambda\, e^{-\lambda} = \int_0^{\lambda} s\, e^{-s}\, ds \le \frac{\lambda^2}{2}\ . \tag{47}$$

If $N_{u,u'}(B_{L'})$ stands for the number of trajectories in the interlacements with labels in $(u, u']$ that reach $B_{L'}$ (this is a Poisson$((u'-u)\,\mathrm{cap}(B_{L'}))$-distributed random variable), we find by (42) that

$$\begin{aligned}
\widetilde{\Delta}' = {} & \frac{1}{u'-u}\ \big(\mathbb{P}[0 \stackrel{u}{\longleftrightarrow} \partial B_{L'}, 0 \stackrel{u'}{\not\longleftrightarrow} \partial B_{L'},\ N_{u,u'}(B_{L'}) = 1] + \\
& \mathbb{P}[0 \stackrel{u}{\longleftrightarrow} \partial B_{L'}, 0 \stackrel{u'}{\not\longleftrightarrow} \partial B_{L'},\ N_{u,u'}(B_{L'}) \ge 2]\big).
\end{aligned} \tag{48}$$

If we consider an independent random walk $X_{\cdot}$ with initial distribution $\overline{e}_{B_{L'}}$, where $\overline{e}_{B_{L'}}$ stands for the normalized equilibrium measure of $B_{L'}$, and write $\widehat{\mathscr{V}}^u = \mathscr{V}^u \backslash (\text{range } X)$, we find from (48), (47) that

$$\Big|\widetilde{\Delta}' - \text{cap}(B_{L'})\, e^{-(u'-u)\,\text{cap}(B_{L'})}\, P_{\overline{e}_{B_{L'}}} \otimes \mathbb{P}[0 \overset{u}{\longleftrightarrow} \partial B_{L'},\ 0 \overset{\widehat{\mathscr{V}}^u}{\not\longleftrightarrow} \partial B_{L'}]\Big| \leq \frac{1}{2}\,(u'-u)\,\text{cap}(B_{L'})^2 \tag{49}$$

(this formula is close in spirit to Theorem 1 of [5]). Then, we note that

$$\begin{aligned} &\text{cap}(B_{L'})\, e^{-(u'-u)\,\text{cap}(B_{L'})}\, P_{\overline{e}_{B_{L'}}} \otimes \mathbb{P}[0 \overset{u}{\longleftrightarrow} \partial B_{L'},\ 0 \overset{\widehat{\mathscr{V}}^u}{\not\longleftrightarrow} \partial B_{L'}] = \\ &\frac{1}{u''-u}\, e^{(u''-u')\,\text{cap}(B_{L'})} \mathbb{P}[N_{u,u''}(B_{L'}) = 1]\, P_{\overline{e}_{B_{L'}}} \otimes \mathbb{P}[0 \overset{u}{\longleftrightarrow} \partial B_{L'}, 0 \overset{\widehat{\mathscr{V}}^u}{\not\longleftrightarrow} \partial B_{L'}] = \\ &\frac{1}{u''-u}\, e^{(u''-u')\,\text{cap}(B_{L'})} \big(\mathbb{P}[0 \overset{u}{\longleftrightarrow} \partial B_{L'}, 0 \overset{u''}{\not\longleftrightarrow} \partial B_{L'}] \\ &- \mathbb{P}[0 \overset{u}{\longleftrightarrow} \partial B_{L'}, 0 \overset{u''}{\not\longleftrightarrow} \partial B_{L'},\ N_{u,u''}(B_{L'}) \geq 2]\big). \end{aligned} \tag{50}$$

Inserting this identity into (49) and using (47) once again, we find that

$$\begin{aligned} &\Big|\widetilde{\Delta}' - \frac{1}{u''-u}\, e^{(u''-u')\,\text{cap}(B_{L'})} \mathbb{P}[0 \overset{u}{\longleftrightarrow} \partial B_{L'}, 0 \overset{u''}{\not\longleftrightarrow} \partial B_{L'}]\Big| \leq \\ &\frac{1}{2}\,(u'-u)\,\text{cap}(B_{L'})^2 + \frac{1}{2}\,(u''-u)\,\text{cap}(B_{L'})^2\, e^{(u''-u')\,\text{cap}(B_{L'})} \leq \\ &(u''-u)\,\text{cap}(B_{L'})^2\, e^{(u''-u')\,\text{cap}(B_{L'})}. \end{aligned} \tag{51}$$

Note that $L'' \leq L'$ and a similar calculation as (44) yields the identity

$$\begin{aligned} &\frac{1}{u''-u}\, \mathbb{P}[0 \overset{u}{\longleftrightarrow} \partial B_{L'}, 0 \overset{u''}{\not\longleftrightarrow} \partial B_{L'}] - \widetilde{\Delta}'' = \\ &\frac{1}{u''-u}\, \big(\mathbb{P}[0 \overset{u''}{\longleftrightarrow} \partial B_{L''}, 0 \overset{u''}{\not\longleftrightarrow} \partial B_{L'}] - \mathbb{P}[0 \overset{u}{\longleftrightarrow} \partial B_{L''},\ 0 \overset{u}{\not\longleftrightarrow} \partial B_{L'}]\big) \end{aligned} \tag{52}$$

(u'' plays the role of u', L'' the role of L', and L' the role of ∞ in (43)). The application of (2) with L'' as in (24) now yields

$$\Big|\frac{1}{u''-u}\, \mathbb{P}[0 \overset{u}{\longleftrightarrow} \partial B_{L'},\ 0 \overset{u''}{\not\longleftrightarrow} \partial B_{L'}] - \widetilde{\Delta}''\Big| \leq u''-u. \tag{53}$$

Coming back to (51), we find that

$$|\widetilde{\Delta}' - e^{(u''-u')\,\text{cap}(B_{L'})}\, \widetilde{\Delta}''| \leq (u''-u)\big(1 + \text{cap}(B_{L'})^2\big)\, e^{(u''-u')\,\text{cap}(B_{L'})}. \tag{54}$$

Using (45), (46), it then follows that

$$|\Delta' - e^{(u''-u')\,\mathrm{cap}(B_{L'})}\,\Delta''| \le 3(u''-u)\big(1+\mathrm{cap}(B_{L'})^2\big)\,e^{(u''-u')\,\mathrm{cap}(B_{L'})}. \tag{55}$$

This completes the proof of (25) and hence of Lemma 2. □

With this last ingredient the proof of Proposition 1 is now complete. □

3 The Variational Problem

The main object of this section is to prove Theorem 2 that provides a notion of minimizers for the variational problem (9), see (13)–(15). At the end of the section, the Remark 1 contains additional information on the variational problem, in particular when D, see (7), is star-shaped or a ball.

Proof of Theorem 2 We will first prove (14) and (15). We consider D, u, ν as in (7), (8) and $\overline{J}^D_{u,\nu}$ defined in (12). We let $\varphi_n \ge 0$ in $D^1(\mathbb{R}^d)$, $n \ge 0$, stand for a minimizing sequence of (12). Then, by Theorem 8.6, p. 208 and Corollary 9.7, p. 212 of [11], we can extract a subsequence still denoted by φ_n and find $\varphi \ge 0$ in $D^1(\mathbb{R}^d)$ such that $\frac{1}{2d}\int_{\mathbb{R}^d} |\nabla\varphi|^2 dz \le \liminf_n \frac{1}{2d}\int_{\mathbb{R}^d} |\nabla\varphi_n|^2 dz = \overline{J}^D_{u,\nu}$ and $\varphi_n \to \varphi$ a.e. and in $L^2_{\mathrm{loc}}(\mathbb{R}^d)$. Then, one has

$$\begin{aligned}\fint_D \overline{\theta}_0\big((\sqrt{u}+\varphi)^2\big)\,dz &\ge \fint_D \limsup_n \overline{\theta}_0\big((\sqrt{u}+\varphi_n)^2\big)\,dz \\ &\overset{\text{reverse Fatou}}{\ge} \limsup_n \fint_D \overline{\theta}_0\big((\sqrt{u}+\varphi_n)^2\big)\,dz \ge \nu.\end{aligned} \tag{56}$$

This shows that φ is a minimizer for the variational problem in (12) and (14) is proved. If φ is a minimizer for (12), note that $\widetilde{\varphi} = \varphi \wedge (\sqrt{u}_* - \sqrt{u}) \in D^1(\mathbb{R}^d)$, and using Theorem 6.17, p. 152 of [11], $\varphi - \widetilde{\varphi} = (\varphi - (\sqrt{u}_* - \sqrt{u}))_+$ and $\widetilde{\varphi}$ are orthogonal in $D^1(\mathbb{R}^d)$. In addition, one has $\overline{\theta}_0((\sqrt{u}+\widetilde{\varphi})^2) = \overline{\theta}_0((\sqrt{u}+\varphi)^2)$ so that $\widetilde{\varphi}$ is a minimizer for (12) as well. It follows that $\varphi = \widetilde{\varphi}$ (otherwise φ would not be a minimizer). With analogous arguments, one sees that the infimum defining $\overline{J}^D_{u,\nu}$ in (12) remains the same if one omits the condition $\varphi \ge 0$ in the right member of (12). Then, using smooth perturbations in $\mathbb{R}^d \backslash D$ of a minimizer φ for (12), one finds that φ is harmonic outside D and tends to 0 at infinity (see Remark 5.10 1) of [15] for more details). In addition, see the same reference, $|z|^{d-2}\varphi(z)$ is bounded at infinity and hence everywhere since φ is bounded. This completes the proof of (15).

We now turn to the proof of (13). As already stated above Theorem 2, we know by direct inspection that $I^D_{u,\nu} \ge \overline{J}^D_{u,\nu}$. Thus, we only need to show that

$$\overline{J}^D_{u,\nu} \ge I^D_{u,\nu}. \tag{57}$$

To this end, we consider a minimizer φ for $\overline{J}^D_{u,\nu}$ and know that (15) holds. As we now explain, if $\psi \geq 0$ belongs to $C_0^\infty(\mathbb{R}^d)$ and $\psi > 0$ on D, then one has

$$\fint_D \theta_0\big((\sqrt{u} + \varphi + \psi)^2\big)\, dz > \nu. \tag{58}$$

We consider two cases to argue (58). Letting m_D stand for the normalized Lebesgue measure on D, either

$$m_D(\varphi < \sqrt{u}_* - \sqrt{u}) = 0 \ \text{ or} \tag{59}$$

$$m_D(\varphi < \sqrt{u}_* - \sqrt{u}) > 0. \tag{60}$$

In the first case (59), then $\varphi \geq \sqrt{u}_* - \sqrt{u}$ a.e. on D so that the left member of (58) equals 1 and (58) holds since $\nu < 1$ by (8). In the second case (60), since θ_0 is strictly increasing on $[0, u_*)$ (cf. Lemma 1), one has

$$\begin{aligned}
&\int_D \theta_0\big((\sqrt{u} + \varphi + \psi)^2\big)\, dz = \\
&\int_{D\cap\{\varphi<\sqrt{u}_*-\sqrt{u}\}} \theta_0\big((\sqrt{u} + \varphi + \psi)^2\big)\, dz + \\
&\int_{D\cap\{\varphi\geq\sqrt{u}_*-\sqrt{u}\}} \theta_0\big((\sqrt{u} + \varphi + \psi)^2\big)\, dz > \\
&\int_{D\cap\{\varphi<\sqrt{u}_*-\sqrt{u}\}} \theta_0\big((\sqrt{u} + \varphi)^2\big)\, dz + |D \cap \{\varphi \geq \sqrt{u}_* - \sqrt{u}\}| = \\
&\int_D \overline{\theta}_0\big((\sqrt{u} + \varphi)^2\big)\, dz \geq \nu\, |D|,
\end{aligned} \tag{61}$$

and (58) follows. We have thus proved (58). Using multiplication by a smooth compactly supported $[0, 1]$-valued function and convolution, we can construct a sequence $\varphi_n \geq 0$ in $C_0^\infty(\mathbb{R}^d)$, which approximates $\varphi + \psi$ in $D^1(\mathbb{R}^d)$ and such that φ_n converges to $\varphi + \psi$ a.e. on D. Then, we have

$$\begin{aligned}
\nu \overset{(58)}{<} \fint_D \theta_0\big((\sqrt{u} + \varphi + \psi)^2\big)\, dz &\leq \fint_D \liminf_n \theta_0\big((\sqrt{u} + \varphi_n)^2\big)\, dz \\
&\overset{\text{Fatou}}{\leq} \liminf_n \fint_D \theta_0\big((\sqrt{u} + \varphi_n)^2\big)\, dz.
\end{aligned} \tag{62}$$

Hence, for infinitely many n, one has $I^D_{u,\nu} \leq \frac{1}{2d} \int |\nabla \varphi_n|^2\, dz$, so that

$$I^D_{u,\nu} \leq \frac{1}{2d} \int_{\mathbb{R}^d} |\nabla(\varphi + \psi)|^2\, dz. \tag{63}$$

If we now let ψ tend to 0 in $D^1(\mathbb{R}^d)$ and recall that $\frac{1}{2d}\int_{\mathbb{R}^d} |\nabla\varphi|^2\,dz = \overline{J}^{\,D}_{u,\nu}$, we find (57). This completes the proof of Theorem 2. $\square$

Remark 1

(1) Note that for D as in (7) and $0 < u < u_*$, the non-decreasing map

$$\nu \in [\theta_0(u), 1) \longrightarrow I^D_{u,\nu} \overset{\text{Theorem 2}}{=} \overline{J}^{\,D}_{u,\nu} \text{ is continuous.} \tag{64}$$

Indeed, by definition of $I^D_{u,\nu}$ in (9), the map is right continuous. To see that the map is also left continuous, consider $\nu \in (\theta_0(u), 1)$ and a sequence ν_n smaller than ν increasing to ν. If φ_n is a corresponding sequence of minimizers for (14), by the same arguments as above (56), we can extract a subsequence still denoted by φ_n and find $\varphi \geq 0$ in $D^1(\mathbb{R}^d)$ so that $\frac{1}{2d}\int_{\mathbb{R}^d} |\nabla\varphi|^2\,dz \leq \liminf_n \int_{\mathbb{R}^d} |\nabla\varphi_n|^2\,dz = \lim_n \overline{J}^{\,D}_{u,\nu_n}$ and $\varphi_n \to \varphi$ a.e. Using the reverse Fatou inequality as in (56), we then have

$$\begin{aligned} \fint_D \overline{\theta}_0\big((\sqrt{u}+\varphi)^2\big)\,dz &\geq \fint_D \limsup_n \overline{\theta}_0\big((\sqrt{u}+\varphi_n)^2\big)\,dz \\ &\geq \limsup_n \fint_D \overline{\theta}_0\big((\sqrt{u}+\varphi_n)^2\big)\,dz \geq \limsup_n \nu_n = \nu. \end{aligned} \tag{65}$$

This shows that $\overline{J}^{\,D}_{u,\nu} \leq \lim_n \overline{J}^{\,D}_{u,\nu_n}$ and completes the proof of (64).

(2) If D in (7) is star-shaped around $z_* \in D$ (that is, when $\lambda(z - z_*) + z_* \in D$ for all $z \in D$ and $0 \leq \lambda \leq 1$), then for u, ν as in (8), one has the additional fact

$$\text{any minimizer } \varphi \text{ in (14) satisfies } \fint_D \overline{\theta}_0\big((\sqrt{u}+\varphi)^2\big)\,dz = \nu, \text{ and} \tag{66}$$

$$\overline{J}^{\,D}_{u,\nu} = \min\Big\{\frac{1}{2d}\int_{\mathbb{R}^d} |\nabla\varphi|^2\,dz;\ \varphi \geq 0, \varphi \in D^1(\mathbb{R}^d),\ \fint_D \overline{\theta}_0\big((\sqrt{u}+\varphi)^2\big)\,dz = \nu\Big\}. \tag{67}$$

Indeed, if φ is a minimizer of (14), one sets for $0 < \lambda < 1$, $\varphi_\lambda(z) = \varphi(z_* + \frac{1}{\lambda}(z - z_*))$. Then, one has $\int_{\mathbb{R}^d} |\nabla\varphi_\lambda|^2\,dz = \lambda^{d-2}\int_{\mathbb{R}^d} |\nabla\varphi|^2\,dz$, and, with $D_\lambda \supseteq D$, the image of D under the dilation with center z_* and ratio λ^{-1}, one finds $\fint_D \overline{\theta}_0((\sqrt{u}+\varphi_\lambda)^2)\,dz = \fint_{D_\lambda} \overline{\theta}_0\,((\sqrt{u}+\varphi)^2)\,dz \geq \lambda^d \fint_D \overline{\theta}_0((\sqrt{u}+\varphi)^2)\,dz$. Thus $\fint_D \overline{\theta}_0((\sqrt{u}+\varphi)^2)\,dz \geq \nu$ must actually equal ν, otherwise the consideration of φ_λ for $\lambda < 1$ close to 1 would contradict the fact that φ is a minimizer for (14). This proves (66) and (67) readily follows.

Incidentally, note that due to (66), (67),

$$\text{the map in (64) is strictly increasing.} \tag{68}$$

Indeed, otherwise there would be $\nu < \nu'$ with $\overline{J}^D_{u,\nu} = \overline{J}^D_{u,\nu'}$, and corresponding minimizers φ, φ' as in (67). But then φ' would contradict (66). The claim (68) thus follows.

(3) If D satisfying (7) is a closed Euclidean ball of positive radius in $\mathbb{R}^d$, given a minimizer φ of (14), we can consider its symmetric decreasing rearrangement φ^* relative to the center of D, see Chapter 3 §3 of [11]. One knows that $\varphi^* \in D^1(\mathbb{R}^d)$ and $\int_{\mathbb{R}^d} |\nabla \varphi^*|^2\, dz \le \int_{\mathbb{R}^d} |\nabla \varphi|^2\, dz$, see p. 188–189 of the same reference. As we now explain:

$$\varphi^* \text{ is a minimizer of (14) as well.} \tag{69}$$

The argument is a (small) variation on Remark 5.10 2) of [15]. With m_D the normalized Lebesgue measure on D, one has $m_D(\varphi \ge s) \le m_D(\varphi^* \ge s)$ for all s in $\mathbb{R}$. Setting $\overline{\theta}_0^{-1}(a) = \inf\{t \ge 0; \overline{\theta}_0(t) \ge a\}$, for $0 \le a \le 1$, we see that for $0 \le t \le 1$, $\{\overline{\theta}_0((\sqrt{u} + \varphi)^2) \ge t\} = \{\varphi \ge \sqrt{\overline{\theta}_0^{-1}(t)} - \sqrt{u}\}$, and a similar identity holds with φ^* in place of φ. Hence, we have

$$\begin{aligned}
\nu &\le ⨏_D \overline{\theta}_0\big((\sqrt{u} + \varphi)^2\big)\, dt = \int_0^1 m_D\big(\overline{\theta}_0\big((\sqrt{u} + \varphi)^2\big) \ge t\big)\, dt \\
&= \int_0^1 m_D\big(\varphi \ge \sqrt{\overline{\theta}_0^{-1}(t)} - \sqrt{u}\big)\, dt \le \int_0^1 m_D\big(\varphi^* \ge \sqrt{\overline{\theta}_0^{-1}(t)} - \sqrt{u}\big)\, dt \\
&= \int_0^1 m_D\big(\overline{\theta}_0\big((\sqrt{u} + \varphi)^2\big) \ge t\big) dt = ⨏_D \overline{\theta}_0\big((\sqrt{u} + \varphi^*)^2\big)\, dz\,.
\end{aligned} \tag{70}$$

Thus, φ^* is a minimizer of (14) as well, and the claim (69) follows. Incidentally, note that D is clearly star-shaped so that (64) and (68) hold. □

With Theorem 2 we have a notion of minimizers for the variational problem corresponding to (9). As mentioned in the Introduction, it is a natural question whether there is a strengthening of the asymptotics (10): is it the case that

$$\lim_N \frac{1}{N^{d-2}} \log \mathbb{P}[|D_N \backslash \mathscr{C}^u_\infty| \ge \nu\, |D_N|] = J^D_{u,\nu} \overset{\text{Theorem 2}}{=} \overline{J}^D_{u,\nu}\ ? \tag{71}$$

Given a minimizer φ in (14), the function $(\sqrt{u} + \varphi)^2(\frac{\cdot}{N})$ can heuristically be interpreted as describing the slowly varying local levels of the tilted interlacements that enter the derivation of the lower bound (10) for (71), see Section 4 of [15]. Hence, the special interest in analyzing whether the minimizers φ for (14) reach the value $\sqrt{u_*} - \sqrt{u}$. Indeed, if φ remains smaller than $\sqrt{u_*} - \sqrt{u}$ the local level function $(\sqrt{u} + \varphi)^2$ remains smaller than u_*, and so with values in the percolative regime of the vacant set of random interlacements. On the other hand, the presence of a region where $\varphi \ge \sqrt{u_*} - \sqrt{u}$ raises the question of the possible occurrence of droplets secluded from the infinite cluster of the vacant set that would take part in the creation of an excessive fraction ν of sites of D_N outside the infinite cluster of

$\mathscr{V}^u$ (somewhat in the spirit of the Wulff droplet in the case Bernoulli percolation or for the Ising model, see [2, 3]).

4 An Application of the C^1-Property of θ_0 to the Variational Problem

The main object of this section is to prove Theorem 3 of the Introduction that describes a regime *of small excess* ν for which all minimizers of the variational problem (14) remain strictly below the threshold value $\sqrt{u}_* - \sqrt{u}$. At the end of the section, the Remark 2 contains some simple observations concerning the existence of minimizers reaching the threshold value $\sqrt{u}_* - \sqrt{u}$.

We consider D as in (7), and as in (16)

$$u_0 \in (0, u_*) \text{ such that } \theta_0 \text{ is } C^1 \text{ on a neighborhood of } [0, u_0]. \tag{72}$$

To prove Theorem 3, we will replace θ_0 by a suitable C^1-function $\widetilde{\theta}$, which agrees with θ_0 on $[0, u_0]$, see Lemma 3, and show that for $0 < u < u_0$ and $\nu \geq \theta_0(u)$ the variational problem $\widetilde{J}^D_{u,\nu}$ attached to $\widetilde{\theta}$, see (86) and Lemma 5, has minimizers that satisfy an Euler-Lagrange equation, see (90), involving a Lagrange multiplier that can be bounded from above and below in terms of $\nu - \theta_0(u)$, see (91). Using such tools, we will derive properties such as stated in (17) for the minimizers of $\widetilde{J}^D_{u,\nu}$ and show that they coincide with the minimizers of the original problem $\overline{J}^D_{u,\nu}$ in (14) when $0 < u < u_0$ and ν is close to $\theta_0(u)$, see below (99).

Proof of Theorem 3 Recall u_0 as in (72). Our fist step is

Lemma 3 *There exist non-negative functions $\widetilde{\theta}$ and $\widetilde{\gamma}$ on $\mathbb{R}_+$ such that*

$$\theta_0 = \widetilde{\theta} - \widetilde{\gamma}, \tag{73}$$

$$\text{the function } \widetilde{\eta}(b) = \widetilde{\theta}(b^2) \text{ is } C^1 \text{ on } \mathbb{R}, \tag{74}$$

$$\widetilde{\eta}' \text{ is bounded and uniformly continuous on } \mathbb{R}, \tag{75}$$

$$\widetilde{\eta}' \text{ is uniformly positive on each interval } [a, +\infty), a > 0, \tag{76}$$

$$\widetilde{\gamma} = 0 \text{ on } [0, u_0] \text{ and } \widetilde{\gamma} > 0 \text{ on } (u_0, \infty). \tag{77}$$

Proof By assumption there is $u_1 \in (u_0, u_*)$ such that θ_0 is C^1 on a neighborhood of $[0, u_1]$ with a uniformly positive derivative on $[0, u_1]$ by Lemma 1. We set $u_2 = \max\{u_*, 4\}$, so that $u_0 < u_1 < u_2$. We then define $\widetilde{\theta}(v) = \theta_0(v)$ on $[0, u_0]$, $\widetilde{\theta}(v) = \theta_0(v) + a(v - u_0)^2$ on $[u_0, u_1]$, where $a > 0$ is chosen so that $\widetilde{\theta}(u_1) = 1$ $(\geq \theta_0(u_*) > \theta_0(u_1))$, and $\widetilde{\theta}(v) = \sqrt{v}$ (≥ 2) on $[u_2, \infty)$. In particular, $\widetilde{\eta}(b) = b$ for $b \geq \sqrt{u}_2$. Then, any choice of $\widetilde{\theta}$ on $[u_1, u_2]$ that is C^1 on $[u_1, u_2]$ with right derivative $\theta_0'(u_1)$

at u_1, left derivative $\frac{1}{2\sqrt{u_2}}$ at u_2, and uniformly positive derivative on $[u_1, u_2]$, leads to functions $\widetilde{\theta}$, $\widetilde{\gamma}$ that satisfy (73)–(77).

We select functions fulfilling (73)–(77) and from now on we view

$$\widetilde{\theta} \text{ (and hence } \widetilde{\gamma}) \text{ as fixed and solely depending on } u_0. \tag{78}$$

For the results below up until the end of the proof of Theorem 3, the only property of u_0 that matters is that u_0 is positive and a decomposition of θ_0 satisfying (73)–(77) has been selected. In particular, if such a decomposition can be achieved in the case of $u_0 = u_*$, the results that follow until the end of the proof of Theorem 3, with the exception of the last inequality (17) (part of the claim at the end of the proof), remain valid. This observation will be useful in Remark 2 at the end of this section.

With $u \in (0, u_0)$, D as in (7), and $\widetilde{\eta}$ as in (74), we now introduce the map:

$$\widetilde{A} : \varphi \in D^1(\mathbb{R}^d) \to \widetilde{A}(\varphi) = \fint_D \widetilde{\eta}(\sqrt{u} + \varphi)\, dz \in \mathbb{R}. \tag{79}$$

We collect some properties of $\widetilde{A}$ in the next

Lemma 4

$$|\widetilde{A}(\varphi + \psi) - \widetilde{A}(\varphi)| \le c(u_0)\, \|\psi\|_{L^1(m_D)},\ \text{for } \varphi, \psi \in D^1(\mathbb{R}^d) \tag{80}$$

(recall m_D stands for the normalized Lebesgue measure on D).

$$\widetilde{A} \text{ is a } C^1\text{-map and } A'(\varphi), \text{ its differential at } \varphi \in D^1(\mathbb{R}^d), \text{ is the linear form } \psi \in D^1(\mathbb{R}^d) \to \fint_D \widetilde{\eta}'(\sqrt{u} + \varphi)\, \psi\, dz = A'(\varphi)\, \psi. \tag{81}$$

$$\text{For any } \varphi \ge 0,\ A'(\varphi) \text{ is non-degenerate.} \tag{82}$$

Proof The claim (80) is an immediate consequence of the Lipschitz property of $\widetilde{\eta}$ resulting from (75). We then turn to the proof of (81). For φ, ψ in $D^1(\mathbb{R}^d)$, we set

$$\begin{aligned} \Gamma = {} & \widetilde{A}(\varphi + \psi) - \widetilde{A}(\varphi) - \fint_D \widetilde{\eta}'(\sqrt{u} + \varphi)\, \psi\, dz = \\ & \int_0^1 ds \fint_D \big(\widetilde{\eta}'(\sqrt{u} + \varphi + s\,\psi) - \widetilde{\eta}'(\sqrt{u} + \varphi)\big)\, \psi\, dz. \end{aligned} \tag{83}$$

With the help of the uniform continuity and boundedness of $\widetilde{\eta}'$, see (75), for any $\delta > 0$ there is a $\rho > 0$ such that for any φ, ψ in $D^1(\mathbb{R}^d)$

$$\begin{aligned} |\Gamma| &\leq \int_D (\delta + 2\|\widetilde{\eta}'\|_\infty 1\{|\psi| \geq \rho\})\, |\psi|\, dz \\ &\leq \delta \|\psi\|_{L^1(m_D)} + \frac{2}{\rho}\, \|\widetilde{\eta}'\|_\infty\, \|\psi\|^2_{L^2(m_D)}. \end{aligned} \tag{84}$$

Since the $D^1(\mathbb{R}^d)$-norm controls the $L^2(m_D)$-norm, see Theorem 8.3, p. 202 of [11], we see that for any $\varphi \in D^1(\mathbb{R}^d)$, $\Gamma = o(\|\psi\|_{D^1(\mathbb{R}^d)})$, as $\psi \to 0$ in $D^1(\mathbb{R}^d)$. Hence, $\widetilde{A}$ is differentiable with differential given in the second line of (81). In addition, with $\delta > 0$ and $\rho > 0$ as above, for any φ, γ, ψ in $D^1(\mathbb{R}^d)$

$$\begin{aligned} &\Big| \int_D \big(\widetilde{\eta}'\big(\sqrt{u} + \varphi + \gamma\big) - \widetilde{\eta}'(\sqrt{u} + \varphi)\big)\, \psi\, dz \Big| \leq \\ &\int_D (\delta + 2\|\widetilde{\eta}'\|_\infty\, 1\{|\gamma| \geq \rho\})\, |\psi|\, dz \\ &\leq \delta \|\psi\|_{L^1(m_D)} + \frac{2}{\rho}\, \|\widetilde{\eta}'\|_\infty\, \|\gamma\|_{L^2(m_D)}\, \|\psi\|_{L^2(m_D)}. \end{aligned} \tag{85}$$

This readily implies that $\widetilde{A}$ is C^1 and completes the proof of (81). Finally, (82) follows from (76) and the fact that $u > 0$. This completes the proof of Lemma 4.

Recall that $u \in (0, u_0)$. We now define the auxiliary variational problem

$$\begin{aligned} &\widetilde{J}^D_{u,\nu} = \min\Big\{\frac{1}{2d} \int_{\mathbb{R}^d} |\nabla\varphi|^2 dz;\ \varphi \geq 0, \varphi \in D^1(\mathbb{R}^d),\, \widetilde{A}(\varphi) \geq \nu\Big\}, \\ &\text{for } \nu \geq \widetilde{\theta}(u)\ (\overset{(77)}{=}\ \theta_0(u)\,). \end{aligned} \tag{86}$$

In the next lemma we collect some useful facts about this auxiliary variational problem and its minimizers. We denote by G the convolution with the Green function of $\frac{1}{2d}\,\Delta$ (i.e. $\frac{d}{2\pi^{d/2}}\,\Gamma(\frac{d}{2} - 1)\,|\cdot|^{-(d-2)}$ with $|\cdot|$ the Euclidean norm on $\mathbb{R}^d$).

Lemma 5 *For D as in (7), $u \in (0, u_0)$, $\nu \geq \widetilde{\theta}(u)$ $(= \theta_0(u))$, one has*

$$\widetilde{J}^D_{u,\nu} = \min\Big\{\frac{1}{2d} \int_{\mathbb{R}^d} |\nabla\varphi|^2 dz;\ \varphi \geq 0, \varphi \in D^1(\mathbb{R}^d),\, \widetilde{A}(\varphi) = \nu\Big\}. \tag{87}$$

Moreover, one can omit the condition $\varphi \geq 0$ without changing the above value, and

$$\textit{any minimizer of (86) satisfies } \widetilde{A}(\varphi) = \nu. \tag{88}$$

In addition, when $\nu = \widetilde{\theta}(u)$, $\widetilde{\varphi} = 0$ *is the only minimizer of (86) and when* $\nu > \theta_0(u)$, *for any minimizer* $\widetilde{\varphi}$ *of (86)*

$$\begin{aligned}&\widetilde{\varphi}\,(\geq 0) \textit{ is } C^{1,\alpha} \textit{ for all } \alpha \in (0,1), \textit{ harmonic outside } D, \textit{ with} \\ &\sup_z |z|^{d-2}\varphi(z) < \infty,\end{aligned} \tag{89}$$

and there exists a Lagrange multiplier $\widetilde{\lambda} > 0$ *such that*

$$\widetilde{\varphi} = \widetilde{\lambda}\, G(\widetilde{\eta}'(\sqrt{u} + \widetilde{\varphi})\, 1_D), \;\; \textit{with} \tag{90}$$

$$c'(u_0, D)\,\big(\nu - \theta_0(u)\big) \leq \widetilde{\lambda} \leq c(u, u_0, D)\,\big(\nu - \theta_0(u)\big) \tag{91}$$

(recall that $\theta_0(u) = \widetilde{\theta}(u)$*).*

Proof We begin by the proof of (87), (88). For $\varphi \in D^1(\mathbb{R}^d)$, we write $\mathscr{D}(\varphi)$ as a shorthand for $\frac{1}{2d}\int_{\mathbb{R}^d} |\nabla\varphi|^2 dz$. Note that $\lim_{b\to\infty} \widetilde{\eta}(b) = \infty$ by (76), so that the set in the right member of (86) is not empty. Taking a minimizing sequence φ_n in (86), we can extract a subsequence still denoted by φ_n and find $\varphi \in D^1(\mathbb{R}^d)$ such that $\mathscr{D}(\varphi) \leq \liminf_n \mathscr{D}(\varphi_n)$ and $\varphi_n \to \varphi$ in $L^1(m_D)$ (see Theorem 8.6, p. 208 of [11]). By (80) of Lemma 4, we find that $\widetilde{A}(\varphi) \geq \nu$. Hence, φ is a minimizer of (86).

Now, for any minimizer φ of (86), if $\widetilde{A}(\varphi) > \nu$, then for some $\lambda \in (0,1)$ close to 1, $\widetilde{A}(\lambda\varphi) \geq \nu$. Moreover, φ is not the zero function (since $\widetilde{A}(\varphi) > \nu$), and $\mathscr{D}(\lambda\varphi) = \lambda^2\mathscr{D}(\varphi) < \mathscr{D}(\varphi)$. This yields a contradiction and (88), (87) follow.

Also, if one removes the condition $\varphi \geq 0$ in (88), one notes that for any φ in $D^1(\mathbb{R}^d)$, $\mathscr{D}(|\varphi|) \leq \mathscr{D}(\varphi)$ and $\widetilde{A}(|\varphi|) \geq \widetilde{A}(\varphi)$. So, the infimum obtained by removing the condition $\varphi \geq 0$ is at least $\widetilde{J}^D_{u,\nu}$ and hence equal to $\widetilde{J}^D_{u,\nu}$. The claim of Lemma 5 below (87) follows.

When $\nu = \widetilde{\theta}(u)$, $\widetilde{J}^D_{u,\nu} = 0$ and $\varphi = 0$ is the only minimizer. We now assume $\nu > \widetilde{\theta}(u)$ and will prove (89), (90). For $\widetilde{\varphi} \geq 0$ in $D^1(\mathbb{R}^d)$ a minimizer of (87), one finds using smooth perturbations in $\mathbb{R}^d \backslash D$ (see Remark 5.10 1) of [15] for similar arguments) that $\widetilde{\varphi}$ is a non-negative harmonic function in $\mathbb{R}^d \backslash D$ that vanishes at infinity and that $|z|^{d-2}\,\widetilde{\varphi}(z)$ is bounded at infinity. By (81), (82) of Lemma 4, $\widetilde{\varphi}$ satisfies an Euler-Lagrange equation (see Remark 5.10 4) of [15] for a similar argument) and for a suitable Lagrange multiplier $\widetilde{\lambda}$, one has (90) (and necessarily $\widetilde{\lambda} > 0$). Since $\widetilde{\eta}'$ is bounded by (21), it follows from (90) that $\widetilde{\varphi}$ is $C^{1,\alpha}$ for all $\alpha \in (0,1)$, see for instance (4.8), p. 71 of [9]. This proves (89), (90).

There remains to prove (91). We have (recall that $\theta_0(u) = \widetilde{\theta}(u)$)

$$\nu - \theta_0(u) = \fint_D \widetilde{\eta}(\sqrt{u} + \widetilde{\varphi}) - \widetilde{\eta}\,(\sqrt{u})\, dz. \tag{92}$$

By (75), we see that

$$\begin{aligned} \nu - \theta_0(u) \le\ & \|\widetilde{\eta}'\|_\infty \fint_D \widetilde{\varphi}\, dz \overset{(90)}{=} \widetilde{\lambda}\, \|\widetilde{\eta}'\|_\infty \fint_D G\big(\widetilde{\eta}'(\sqrt{u} + \widetilde{\varphi})\, 1_D\big)\, dz \\ \le\ & \widetilde{\lambda}\, \|\widetilde{\eta}'\|^2_\infty \fint_D G(1_D)\, dz = c(u_0, D)\, \widetilde{\lambda}. \end{aligned} \tag{93}$$

On the other hand, by (76), we see that

$$\begin{aligned} \nu - \theta_0(u) \ge\ & \inf_{[\sqrt{u},\infty)} \widetilde{\eta}' \fint_D \widetilde{\varphi}\, dz \overset{(90)}{=} \widetilde{\lambda} \inf_{[\sqrt{u},\infty)} \widetilde{\eta}' \fint_D G\big(\widetilde{\eta}'(\sqrt{u} + \varphi^2)\, 1_D\big)\, dz \\ \ge\ & \widetilde{\lambda}\big(\inf_{[\sqrt{u},\infty)} \widetilde{\eta}'\big)^2 \fint_D G(1_D)\, dz = c(u, u_0, D)\, \widetilde{\lambda}. \end{aligned} \tag{94}$$

The claim (91) now follows from (93) and (94). This concludes the proof of Lemma 5. □

We now continue the proof of Theorem 3. Given $u \in (0, u_0)$ and $\nu \ge \widetilde{\theta}(u)$ $(= \theta_0(u))$, we see by Lemma 5 that any minimizer $\widetilde{\varphi}$ for (87) satisfies (90) for a suitable $\widetilde{\lambda}$ satisfying (91), so that

$$\|\widetilde{\varphi}\|_\infty \le \widetilde{\lambda}\, \|\widetilde{\eta}'\|_\infty\, \|G 1_D\|_\infty \overset{(91),(75)}{\le} c_2(u, u_0, D)\big(\nu - \theta_0(u)\big). \tag{95}$$

In particular, we find that

$$\begin{aligned} &\text{for } \theta_0(u) \le \nu \le \theta_0(u) + c_1(u, u_0, D)(< 1), \text{ any minimizer } \widetilde{\varphi} \\ &\text{for (87) satisfies } 0 \le \widetilde{\varphi} \le (\sqrt{u}_0 - \sqrt{u}) \wedge \big\{c_2\big(\nu - \theta_0(u)\big)\big\}. \end{aligned} \tag{96}$$

We will now derive the consequences for the basic variational problem of interest $\overline{J}^D_{u,\nu}$, see (12), (14). By (73), (77) and the definition of $\overline{\theta}_0$ (see (11)), we find that $\widetilde{\theta} \ge \overline{\theta}_0$, so that

$$\text{for all } u \in (0, u_0) \text{ and } \nu \in [\theta_0(u), 1),\ \overline{J}^D_{u,\nu} \ge \widetilde{J}^D_{u,\nu}. \tag{97}$$

Moreover, when $\nu \in [\theta_0(u), \theta_0(u) + c_1]$ (with c_1 as in (96)), any minimizer $\widetilde{\varphi}$ for (87) is bounded by $\sqrt{u}_0 - \sqrt{u}$, and hence satisfies as well $\fint_D \overline{\theta}_0((\sqrt{u} + \widetilde{\varphi})^2)\, dz \ge \nu$ (in fact an equality by (88)). We thus find that

$$\begin{aligned} &\overline{J}^D_{u,\nu} = \widetilde{J}^D_{u,\nu} \text{ for all } \nu \in [\theta_0(u) + c_1], \text{ and any minimizer } \widetilde{\varphi} \text{ of } \widetilde{J}^D_{u,\nu} \text{ in (87)} \\ &\text{is a minimizer of } \overline{J}^D_{u,\nu} \text{ in (14).} \end{aligned} \tag{98}$$

Now for ν as above, consider φ a minimizer of (14). Then, we have $\mathscr{D}(\varphi) = \overline{J}^D_{u,\nu} = \widetilde{J}^D_{u,\nu}$, and since $\widetilde{\theta} \geq \overline{\theta}_0$, we find that

$$\widetilde{A}(\varphi) = \fint_D \widetilde{\theta}\big((\sqrt{u} + \varphi)^2\big)\, dz \geq \fint_D \overline{\theta}_0\big((\sqrt{u} + \varphi)^2\big)\, dz \geq \nu. \tag{99}$$

This show that φ is a minimizer for (86), hence for (87) by (88). We thus find that when $\nu \in [\theta_0(u), \theta_0(u) + c_1]$, the set of minimizers of (14) and (87) coincide and the claim (17) now follows from Lemma 5. This concludes the proof of Theorem 3.

With Theorem 3 we have singled out a regime of "small excess" for ν such that all minimizers φ for $\overline{J}^D_{u,\nu}$ in (14) stay below the maximal value $\sqrt{u}_* - \sqrt{u}$. In the remark below we make some simple observations about the possible existence of a regime where some minimizers in (14) reach the threshold $\sqrt{u}_* - \sqrt{u}$.

Remark 2

(1) If θ_0 is discontinuous at u_* (a not very plausible possibility), then $\theta_0(u_*) < 1$, and for any $\nu \in (\theta_0(u_*), 1)$ any minimizer for (14) must reach the threshold value $\sqrt{u}_* - \sqrt{u}$ on a set of positive Lebesgue measure due to the constraint in (14).

(2) If θ_0 is continuous and its restriction to $[0, u_*]$ is C^1 with uniformly positive derivative (corresponding to a "mean field" behavior of the percolation function θ_0), then a decomposition as in Lemma 3 can be achieved with now $u_0 = u_*$. As mentioned below (78), the facts established till the end of Theorem 3 (with the exception of the last inequality of (17)) remain valid in this context. In particular, if for some $u \in (0, u_*)$ and $\nu \in (\theta_0(u), 1)$ there is a minimizer $\widetilde{\varphi}$ for $\widetilde{J}^D_{u,\nu}$ in (87) such that $\|\widetilde{\varphi}\|_\infty = \sqrt{u}_* - \sqrt{u}$, then $\widetilde{\varphi}$ is a minimizer for $\overline{J}^D_{u,\nu}$ in (14) and it reaches the threshold value $\sqrt{u}_* - \sqrt{u}$. In the toy example where $\widetilde{\eta}$ is affine on $[\sqrt{u} + \infty)$ and $0 < \widetilde{\eta}(\sqrt{u}) < \widetilde{\eta}(\sqrt{u}_*) = 1$, such a $\nu < 1$ and $\widetilde{\varphi}$ (which satisfies (90)) are for instance easily produced. □

The above remark naturally raises the question of finding some plausible assumptions on the behavior of the percolation function θ_0 close to u_* (if the behavior mentioned in Remark 2 2) is not pertinent, see for instance Figure 4 of [12] for the level-set percolation of the Gaussian free field, when $d = 3$) and whether such assumptions give rise to a regime for u, ν, ensuring that minimizers of $\overline{J}^D_{u,\nu}$ in (14) achieve the maximal value $\sqrt{u}_* - \sqrt{u}$ on a set of positive measure. But there are many other open questions. For instance, what can be said about the number of minimizers for (14)? Is the map $\nu \to \overline{J}^D_{u,\nu}$ in (64) convex? An important question is of course whether the asymptotic lower bound (10) can be complemented by a matching asymptotic upper bound.

References

1. Ambrosetti, A., Malchiodi, A.: Nonlinear Analysis and Semilinear Elliptic Problems, vol. 104, pp. 1–316. Cambridge University Press, Cambridge (2007)
2. Bodineau, T.: The Wulff construction in three and more dimensions. Commun. Math. Phys. **207**, 197–229 (1999)
3. Cerf, R.: Large deviations for three dimensional supercritical percolation. Astérisque 267, Société Mathématique de France (2000)
4. Černý, J.,Teixeira, A.: From random walk trajectories to random interlacements. Sociedade Brasileira de Matemática **23**, 1–78 (2012)
5. de Bernardini, D., Popov, S.: Russo's formula for random interlacements. J. Stat. Phys. **160**(2), 321–335 (2015)
6. Drewitz, A., Ráth, B., Sapozhnikov, A.: An Introduction to Random Interlacements. SpringerBriefs in Mathematics, Berlin (2014)
7. Drewitz, A., Ráth, B., Sapozhnikov, A.: Local percolative properties of the vacant set of random interlacements with small intensity. Ann. Inst. Henri Poincaré Probab. Stat. **50**(4), 1165–1197 (2014)
8. Duminil-Copin, H., Goswami, S., Rodriguez, P.-F. , Severo, F.: Equality of critical parameters for percolation of Gaussian free field level-sets. Preprint. Also available at arXiv:2002.07735
9. Gilbarg, D.,Trudinger, N.S.: Elliptic Partial Differential Equations of Second Order. Grundlehren der Mathematischen Wissenschaften, vol. 224. Springer, Berlin/New York (1983)
10. Grimmett, G.: Percolation, 2nd edn. Springer, Berlin (1999)
11. Lieb, E., Loss, M.: Analysis. Graduate Studies in Mathematics, vol. 14, 2nd edn. AMS, New York (2001)
12. Marinov, V.I., Lebowitz, J.L.: Percolation in the harmonic crystal and voter model in three dimensions. Phys. Rev. E **74**, 031120 (2006)
13. Sidoravicius, V., Sznitman, A.S.: Percolation for the vacant set of random interlacements. Commun. Pure Appl. Math. **62**(6), 831–858 (2009)
14. Sznitman, A.S.: Vacant set of random interlacements and percolation. Ann. Math. **171**, 2039–2087 (2010)
15. Sznitman, A.S.: On bulk deviations for the local behavior of random interlacements. Preprint. Also available at arXiv:1906.05809
16. Teixeira, A.:On the uniqueness of the infinite cluster of the vacant set of random interlacements. Ann. Appl. Probab. **19**(1), 454–466 (2009)

On Clusters of Brownian Loops in d Dimensions

Wendelin Werner

This paper is dedicated to the memory of Vladas Sidoravicius.[1]

Abstract We discuss random geometric structures obtained by percolation of Brownian loops, in relation to the Gaussian Free Field, and how their existence and properties depend on the dimension of the ambient space. We formulate a number of conjectures for the cases $d = 3, 4, 5$ and prove some results when $d > 6$.

Keywords Brownian loop-soups · Gaussian free field · Percolation

MSC Class 60K35, 82B43, 60G60

1 Introduction

Field theory has been remarkably successful in describing features of many models of statistical physics at their critical points. In that approach, the focus is put on correlation functions between the values taken by the field at a certain number of given points in space. In many instances, these functions correspond to experimentally measurable macroscopic quantities (such as for instance the global magnetization in the Ising model).

[1]The content of this paper corresponds to the last of my talks that Vladas attended in 2017 and 2018. Like so many of us in the mathematical community, I remember and miss his enthusiasm as well as his contagious, warm and charming smile.

W. Werner (✉)
Department of Mathematics, ETH Zürich, Zürich, Switzerland
e-mail: wendelin.werner@math.ethz.ch

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8_33

Some of these correlation functions can also be directly related to features of conjectural (and sometimes physically relevant) random fractal geometric objects; for instance, a 2-point function $F(x_1, x_2)$ can describe the asymptotic behaviour as $\varepsilon \to 0$ of the probability that x_1 and x_2 are both in the ε-neighbourhood of some "random cluster" in a statistical physics model—and the critical exponent that describes the behaviour of F as $y \to x$ is then related to the fractal dimension of the scaling limits of those clusters. This type of more concrete geometric interpretation is however not instrumental in the field-theoretical set-up (and for some fields, there is actually no underlying geometric object). It remained for a long time rather hopeless to go beyond this aforementioned partial description of these geometric structures via correlation functions, due to the lack of other available mathematical tools to define such random geometric objects in the continuum.

In the very special case of two-dimensions (which is related to Conformal Field Theory (CFT) on the field theory side), this has changed with Oded Schramm's construction of Schramm–Loewner Evolutions (SLE processes) in [32]. These are concrete random curves in the plane defined via some mathematical conformally invariant growth mechanism, and that are conjectured to be relevant for most critical systems in two dimensions. The Conformal Loop Ensembles (CLE) that were subsequently introduced in [34, 35] are random collection of loops, or equivalently random connected fractal sets that are built using variants of SLE, and that describe the (conjectural) scaling limit of the joint law of all clusters in critical lattice models. It should be stressed that all these SLE-based developments rely on conformal invariance in a crucial manner, so that they are specific to the two-dimensional case.

In this study of two-dimensional and conformal invariant random structures, the following two random objects have turned out to be very closely related to the SLE and CLE:

- The Gaussian Free Field (GFF): As shown in a series of work by Schramm-Sheffield, Dubédat and Miller-Sheffield starting with [9, 27, 33], this random generalized function essentially turns out to host (in a deterministic way) most SLE-based structures. There exists for instance a procedure that allows to deterministically draw a CLE, starting from a sample of a GFF. In particular, it was pointed out by Miller and Sheffield (see [5] and the references therein) that the CLE_4 appears naturally as a collection of generalized level lines of the GFF. Of course, it should be recalled that the GFF is also an elementary and fundamental building block in field theory.
- The Brownian loop-soups: This object, introduced in [21], is a Poissonian cloud of Brownian loops in a domain D. If, as proposed in [38] and shown in [35], one considers *clusters of Brownian loops*, and their outer boundaries, one constructs also a CLE_κ where $\kappa = \kappa(c)$ varies between $8/3$ and 4 as the intensity c of the loop-soup varies between 0 and 1. This intensity plays the role of the *central charge* in the CFT language.

There is actually a close relation between these two constructions of CLE_4 (via the GFF or via the Brownian loop-soup with intensity $c = 1$), see [30] and the references therein. We will come back to this later, but roughly speaking, starting from a sample of a Brownian loop-soup, one can construct a GFF in such a way that the Brownian loop-soup clusters can be interpreted as "excursion sets" of the GFF, a little bit like the excursion intervals away from 0 of one-dimensional Brownian motion, see [4] and the references therein.

The starting point of the present paper is the observation that both the Brownian loop-soup and the Gaussian Free Field can be defined in any dimension. This leads naturally to wonder what natural random fractal subsets of d-dimensional space for $d \geq 3$ can be built using these special and natural objects. In particular, one can guess that just as in two dimensions, clusters of Brownian loops (for a loop-soup of intensity $c = 1$) will have an interesting geometry, and argue that they should be fundamental structures within a GFF sample. A first immediate reaction is however to be somewhat cautious or even sceptical. Indeed, Brownian loops in dimensions 4 and higher are simple loops, and no two loops in a Brownian loop-soup will intersect, so that a Brownian loop-soup cluster will a priori consist only of one single isolated simple loop. But, as we shall explain in the present paper, things are more subtle, and, if properly defined, it should still be possible to agglomerate these disjoint Brownian loops into interesting clusters when the dimension of the space is 4 and 5.

The structure of the present paper is the following: We will first review some basic facts about Lupu's coupling of the GFF and loop-soups on cable graphs. After discussing heuristically some general aspects of their scaling limits and reviewing the known results in $d = 2$, we will make conjectures about the cases $d = 3, 4, 5$. Then, we will state and derive some results for $d > 6$.

We conclude this introduction with the following remark: It is interesting that this loop-soup approach to the GFF bears many similarities with the random walk representations of fields as initiated by Symanzik [37] and further developed by many papers, including by Simon [36], the celebrated work by Brydges, Fröhlich and Spencer [7] or Dynkin [11]. Their motivation was actually to understand/describe "interacting fields" (i.e., beyond the free field!) via their correlation functions; given that the correlation functions of the GFF are all explicit, there was then not much motivation to study it further, while the question of existence and constructions of non-Gaussian fields was (and actually still is) considered to be an important theoretical challenge.

2 Background: Lupu's Coupling on Cable-Graphs

A crucial role will be played here by the cable-graph GFF and the cable-graph loop-soup, that have been introduced by Titus Lupu in [24, 25]. Let us briefly review their main features in this section, and we refer to those papers for details.

In this section, we consider $\mathscr{D}$ to be a fixed connected (via nearest-neighbour connections) subset of $\mathbb{Z}^d$ (the case of subsets of $\delta\mathbb{Z}^d$ is then obtained simply by scaling space by a factor δ) on which the discrete Green's function is finite. We can for instance take $\mathscr{D}$ to be all (or any connected subset) of $\mathbb{Z}^d$ when $d \geq 3$, or a bounded subset of $\mathbb{Z}^2$. The set $\partial\mathscr{D}$ is the set of points that is at distance exactly 1 of $\mathscr{D}$. The Green's function $G(x, y) = G_{\mathscr{D}}(x, y)$ is the expected number of visits of y made by a simple random walk in $\mathbb{Z}^d$ starting from x before exiting $\mathscr{D}$ (if this exit time is finite, otherwise count all visits of y).

The cable graph $\mathscr{D}_c$ associated to $\mathscr{D}$ is the set consisting of the union of $\mathscr{D}$ with all edges (viewed as open intervals of length 1) that have at least one endpoint in $\mathscr{D}$. One can define also Brownian motion on the cable graph (that behaves like one-dimensional Brownian motion on the edges and in an isotropic way when it is at a site of $\mathscr{D}$). One can then also define the Green's function $G_{\mathscr{D}_c}$ for this Brownian motion (this time, the boundary conditions correspond to a killing when it hits $\partial\mathscr{D}$) and note that its values on $\mathscr{D} \times \mathscr{D}$ coincide with that of the discrete Green's function $G_{\mathscr{D}}$ for the discrete random walk.

One can then on the one hand define the Gaussian Free Field (GFF) on the cable graph $(\phi(x))_{x\in\mathscr{D}}$ as a centred Gaussian process with covariance given by the Green's function $G_{\mathscr{D}_c}$ on the cable graph. This is a random continuous function on $\mathscr{D}_c$ that generalizes Brownian motion (or rather Brownian bridges) to the case where the time-line is replaced by the graph $\mathscr{D}_c$. The process $(\phi^2(x))_{x\in\mathscr{D}_c}$ is then called a *squared GFF* on $\mathscr{D}_c$. The connected components of $\{x \in \mathscr{D}_c,\ \phi(x) \neq 0\}$ are called the *excursion sets* of ϕ (or equivalently of ϕ^2).

On the other hand, one can also define a natural Brownian loop measure on Brownian loops on $\mathscr{D}_c$, and then the Brownian loop-soups which are Poisson point processes with intensity given by a multiple c of this loop measure. In all the sequel, we will always work with Brownian loop-soups with intensity equal to $c = 1$ (in the normalization that is for instance described in [40]—in the Le Jan-Lupu normalization that differs by a factor 2, this would be the loop-soup with intensity $\alpha = 1/2$), which is the one for which one can make the direct relation to the GFF. Let us make two comments about this loop-soup $\mathscr{L}$ on the cable-graph:

(i) When one considers a given point on the cable-graph, it will be almost surely visited by an infinite number of small Brownian loops in the loop-soup. However, it turns out that there almost surely exist exceptional points in the cable-graph that are visited by no loop in the loop-soup (what follows will actually show that the set $\mathscr{Z}$ of such points has Hausdorff dimension 1/2). Another equivalent way to define these sets is to first consider *clusters of Brownian loops*: We say that two loops γ and γ' in a loop-soup belong to the same loop-soup cluster, if one can find a finite chain of loops $\gamma_0 = \gamma, \gamma_1, \ldots, \gamma_n = \gamma'$ in $\mathscr{L}$ such that $\gamma_j \cap \gamma_{j-1} \neq \emptyset$ for $j = 1, \ldots, n$. Then, loop-soup clusters are exactly the connected components of $\mathscr{D}_c \setminus \mathscr{Z}$.

(ii) Just in the same way in which the occupation time measure of one-dimensional Brownian motion has a continuous density with respect to Lebesgue measure (the *local time* of Brownian motion, see e.g. [31]), each Brownian loop γ will have an occupation time measure with a finite intensity ℓ_γ on the cable graph, so that for all sets A, the total time spent by γ in A is equal to $\int_A \ell_\gamma(x)dx$ where dx denote the one-dimensional Lebesgue measure on $\mathscr{D}_c$. One can then define the "cumulative" occupation time density Γ of the loop-soup as $\Gamma :=$

$\sum_{\gamma \in \mathscr{L}} \ell_\gamma$. This is a continuous function on the cable-graph, that is equal to 0 on all points of $\partial \mathscr{D}$. Simple properties of Brownian local time show that $\mathscr{Z} = \{x \in \mathscr{D}_c,\ \Gamma(x) = 0\}$.

Lupu's coupling between the cable-graph loop-soup and the GFF can now be stated as follows.

Proposition 1 (Le Jan [22] and Lupu [24]) *Suppose that one starts with a Brownian loop-soup $\mathscr{L}$ on the cable-graph $\mathscr{D}_c$. Then the law of its total occupation time density Γ is that of (a constant multiple) of a squared GFF. Furthermore, if one then defines the function $U = \sqrt{\Gamma}$ and tosses i.i.d. $\pm$ fair coins ε_j (one for each excursion set K_j of Γ), then if we write $\varepsilon(x) = \varepsilon_j$ for $x \in K_j$, the function $(\varepsilon(x)U(x))_{x \in \mathscr{D}_c}$ is distributed exactly like (a constant multiple of) a GFF on the cable-graph.*

In the sequel, we will always implicitly assume that a GFF ϕ on a cable-system is coupled to a loop-soup $\mathscr{L}$ in this way. We can note that the excursion sets of ϕ are then exactly the loop-soup clusters of $\mathscr{L}$.

We see that in this setting, the only contribution to the correlation between $\phi(x)$ and $\phi(y)$ comes from the event that x and y are in the same loop-soup cluster (we denote this event by $x \leftrightarrow y$), i.e., one has

$$E[\phi(x)\phi(y)] = E[\varepsilon(x)\varepsilon(y) \times |\phi(x)| \times |\phi(y)|] = E[|\phi(x)| \times |\phi(y)| \times 1_{x \leftrightarrow y}]$$

for all x, y in $\mathscr{D}_c$. In the last expression, all quantities are functions of the loop-soup only (and do not involve the ε_j coin tosses). Similarly, all higher order correlation functions and moments can be expressed only in terms of the cable-graph loop-soup.

Conversely, since the law of the GFF is explicit and the correlations between $\varepsilon(x) = \mathrm{sgn}(\phi(x))$ is given in term of cable-graph loop-soup connection events, one gets explicit formulas for those connection probabilities. For instance, Proposition 1 immediately shows that for all x, y in $\mathscr{D}_c$,

$$E[\mathrm{sgn}(\phi(x))\mathrm{sgn}(\phi(y))] = E[\varepsilon(x)\varepsilon(y)] = P[x \leftrightarrow y],$$

from which one readily deduces that:

Corollary 2 (Part of Proposition 5.2 in [24]) *For all $x \neq y$ in $\mathscr{D}_x$,*

$$P[x \leftrightarrow y] = \arcsin \frac{G(x, y)}{\sqrt{G(x, x)G(y, y)}}.$$

In particular, if one considers the cable-graph loop-soup in $\mathbb{Z}^d$ for $d \geq 3$, we see that

$$P[0 \leftrightarrow x] \sim \frac{C}{\|x\|^{d-2}} \tag{1}$$

for some constant C as $x \to \infty$.

We can note that in this case of $\mathbb{Z}^d$ for $d \geq 3$, $P[0 \leftrightarrow x]$ and $E[|\phi(0)||\phi(x)|1_{0\leftrightarrow x}]$ are comparable when $x \to \infty$. Loosely speaking, this means that when one conditions on $0 \leftrightarrow x$ (and lets $x \to \infty$), the number of small Brownian loops (say of diameter between 1 and A for a fixed A) that pass through the origin does not blow up (this type of considerations can easily be made rigorous—the conditional law of $|\phi(0)|$ in fact remains tight as $x \to \infty$).

Remark Throughout this paper, we will always work with loop-soups defined under the very special intensity $c = 1$ that makes its occupation time related to the GFF as described above. Understanding features of the "percolation phase transition" when the loop-soup intensity varies is a question that will not be discussed here (see [8, 10, 23] and the references therein for results in this direction).

3 The Fine-Mesh and Continuum Limit

When D is a connected subset of $\mathbb{R}^d$, in which the continuum Green's function $G_D(x, y)$ is finite when $x \neq y$ (one can for instance think of D to be the unit disk in $\mathbb{R}^2$, or the whole of $\mathbb{R}^d$ when $d \geq 3$), instead of sampling a Brownian loop-soup or a continuum GFF directly in D, we will consider a Brownian loop-soup and a GFF defined on the cable-graph of a connected fine-grid approximation of D in $\delta\mathbb{Z}^d$. For instance (this slightly convoluted definition is just to avoid issues with "thin" boundary pieces), if z_0 is a given point in D, when δ is small enough, we can choose D_δ to be the connected component of the set of points in $\delta\mathbb{Z}^d$ that are at distance at least δ from the complement of D, and that contains the points that are at distance less than δ from z_0. One can then consider its cable graph $D_{\delta,c}$ and the corresponding GFF and loop-soups as in the previous section (just scaling space by a factor δ).

We now discuss what happens in the fine-mesh limit (when $\delta \to 0$). To avoid confusion, we will use the following terminology:

- The *cable-graph loop-soup* and the *cable-graph clusters* will respectively be the soup of Brownian loops defined on the cable graph $D_{\delta,c}$ and the corresponding collection of clusters.
- The *Brownian loop-soup* will be the usual continuum Brownian loop-soup in D. The clusters that are created via intersecting Brownian loops will be called *Brownian loop-soup clusters*.

Now, when the mesh of the lattice δ goes to 0, one can consider the joint limit in distribution of the cable-graph loop-soup, of the corresponding cable-graph clusters and of the cable-graph GFF, and make the following observations:

(i) *About the limit of the loop-soup*. If one sets any positive macroscopic cut-off a, then the law of the loops in the cable-graph loop-soup which have a diameter greater than a does converge to that of the loops with diameter greater than a in a (continuum) Brownian loop-soup in D. This follows from rather standard approximations of Brownian motion by random walks (see [20] for this particular instance). So, in that sense, the scaling limit of the cable graph loop-soup is just the Brownian loop-soup in D. By Skorokhod's representation theorem, we can also view a Brownian loop-soup in D as an almost sure limit of cable-graph loop-soups.

(ii) *About the limit of the cable-graph GFF*. The cable-graph GFF does converge in law to the continuum GFF, because the correlation functions of the cable-graph GFF converge to those of the continuum GFF (all this is due to elementary consideration on Gaussian processes). It should however be stressed that the continuum GFF is not a random function anymore (see for instance [40]) so that this weak convergence has to be understood in the appropriate function space.

(iii) *A warning when $d \geq 4$*. While the GFF and the Brownian loop-soup are well-defined in any dimension, it is possible to make sense neither of the (renormalized) square of the GFF nor of the (renormalized) total occupation time measure of the Brownian loop when $d \geq 4$. This is due to the fact that the total occupation time of the Brownian loops of diameter in $[2^{-n}, 2^{-n+1}]$ inside a box of size 1 will have a second moment of the order of a constant times $2^{n(d-4)}$, which is not summable as soon as $d \geq 4$ (so that the fluctuations of the occupation times of the very small loops will outweigh those of the macroscopic ones). Since the relation between the cable-graph GFF and the cable-graph loop-soup did implicitly involve the square of the cable-graph GFF, this indicates that some caution is needed when one tries to tie a direct relation between the continuum GFF and the Brownian loop-soup in $\mathbb{R}^d$ when $d \geq 4$.

Despite (iii), one can nevertheless always study the joint limit of the coupled cable-graph GFF and cable-graph loop-soup (and its clusters). The correlation functions of the cable-graph GFF do provide information on the structure of the cable graph clusters, and therefore on their behaviour as $\delta \to 0$, as illustrated by Corollary 2. One key point is that the scaling limit of the cable graph clusters (if they exist) might be strictly larger than the Brownian loop-soup clusters. Indeed, cable graph clusters may contain loops of macroscopic size (say, some of the finitely many loops of diameter greater than some cut-off value a), but they will also contain many small loops, for instance of diameter comparable to the mesh-size δ, or to δ^b for some positive power b. All these small loops do disappear from the loop-soup in the scaling limit if one uses the procedure described in (i), but (just as critical percolation does create macroscopic clusters made of union of edges of size equal to the mesh-size, while each individual edge does "disappear" in the scaling limit) their cumulative effect in terms of contributing to create macroscopic cable graph clusters does not necessarily vanish.

In the fine-mesh limit, there a priori appear to be four possible likely scenarios (for presentation purposes, we will consider in the remaining of this section that D is the hypercube $(0, 1)^d$):

- Case 0. There is no limiting joint law for the cable graph clusters when $\delta \to 0$. This should for instance be the case when the number of macroscopic cable graph clusters in D_δ tends to infinity as $\delta \to 0$. We will come back to this interesting case later. In the remaining cases 1, 2a and 2b, we will assume that the number of cable graph clusters of diameter greater than any fixed a remains tight, and that their joint law has a scaling limit as $\delta \to 0$.
- Case 1: The limit of the family of macroscopic cable graph clusters is exactly the family of macroscopic clusters Brownian loop-soup clusters. This means that in this case, the effect of the microscopic loops disappears as δ vanishes.
- Case 2: The limit of the cable graph clusters consists of macroscopic Brownian loops that are somehow agglomerated together also by the effect of the microscopic loops (i.e., the limit of the cable graph clusters are strictly larger than the clusters of macroscopic Brownian loops). Here, the limit of the cable graph clusters would consist of a combination of macroscopic effects and microscopic effects. There are actually two essentially different subcases:
  - Case 2a: The glueing procedure does involve additional randomness (i.e., randomness that is not present in the Brownian loop soup).
  - Case 2b: The glueing procedure of how to agglomerate the macroscopic loops is a deterministic function of these macroscopic loops (i.e., the limit of the cable-graph clusters is a deterministic function of the corresponding Brownian loop-soup).

Let us summarize here already the conjectures that we will state more precisely in the next sections. We will conjecture that each of the four cases 0, 1, 2a and 2b do occur for some value of the dimension. More specifically, in dimension $d = 2$, it is known that Case 1 holds, and we believe that this should also the case when $d = 3$, although a proof of this fact appears to remain surprisingly elusive at this point. So, in those lower dimensions, only the macroscopic (in the scaling limit, Brownian) loops prevail to construct the excursion sets of the GFF. For intermediate dimensions, microscopic loops will start to play an important role: As we will try to explain, it is natural to expect that Case 2b holds for $d = 4$ and that Case 2a holds for $d = 5$. These are two quite fascinating instances, with an actual interplay between microscopic and macroscopic features.

In higher dimensions, one can adapt some ideas that have been developed in the context of (ordinary) high-dimensional percolation to show that Case 0 holds. There is no excursion decomposition of the continuum GFF anymore, but a number of instructive features can be highlighted. A "typical" large cable graph cluster will actually contain no macroscopic Brownian loop (even though some exceptional clusters will contain big Brownian loops). Hence, this loop-soup percolation provides a simple percolation-type model that somehow explains "why" general high-dimensional critical percolation models should exhibit "Gaussian behaviour.

Indeed, the collection of all these cable-graph clusters is actually very similar to that of ordinary percolation (as they are constructed using only small loops of vanishingly small size at macroscopic level). This in turn sheds some light onto some of the lace-expansion ideas.

We will now discuss separately the different dimensions. We will first briefly review what is known and proved when $d = 2$ and mention the conjectures for $d = 3$. We will then heuristically discuss the cases $d = 4$ and $d = 5$ and make some further conjectures, based on some analogies with features of critical percolation within Conformal Loop Ensembles. Finally, we will state and prove some results in the case where the dimension is greater than 6. We note that we will (as often in these percolation questions) not say anything about the "critical" case $d = 6$ here.

4 Low and Intermediate Dimensions

4.1 Low Dimensions

4.1.1 Review of the Two-Dimensional Case

This is the case where the behaviour of the scaling limit of cable-graph loop-soup clusters is by now essentially fully understood. Indeed, in this case, one has an additional direct good grip on features of the continuum GFF that are built on its coupling with the SLE_4 curves (as initiated in [33]) and the CLE_4 loop ensembles. The paper [35] provides an explicit description of the Brownian loop-soup clusters as CLE_4 loops, so that one can deduce some explicit formulas (such as in [41]) for the laws of these clusters. These formulas turn out to match exactly the ones that appear in the scaling limit of cable-graph clusters (in the spirit of the formulas by Le Jan [22]), so that one can conclude (this is one of the main results of [26]) that the scaling limit of the cable-graph loop-soup clusters are exactly the Brownian loop-soup clusters (see also some earlier discussion of this problem without the cable-graph insight in [6]).

It is then actually possible to push this further: One important result in [3, 4] is that if one associates to each Brownian loop-soup cluster C_j a particular "natural" measure μ_j supported on C_j (which is a deterministic function of this cluster C_j), then, if (ε_j) are i.i.d. ± 1 fair coin flips, the sum $\sum_j \varepsilon_j \mu_j$ (viewed as an L^2 limit) is actually a continuum GFF. In other words, the Brownian loop-soup clusters provide indeed a loop-soup based "excursion decomposition" of the continuum GFF despite the fact that the GFF is not a continuous function (it is only a generalized function).

4.1.2 Conjectural Behaviour in Dimension 3

When $d = 3$, one can recall that Brownian paths (and loops) have many double points (the Hausdorff dimension of the set of double points in actually equal to

1). Hence, a Brownian loop in a Brownian loop-soup will almost surely intersect infinitely many other Brownian loops in this loop-soup. From this, one can actually deduce that the Hausdorff dimension Δ of the Brownian loop-soup clusters is almost surely greater than 2 ($0-1$ law arguments show that this dimension actually always takes the same constant value). On the other hand, Corollary 2 can be used to prove that Δ can not be larger than $5/2$. It is natural to conjecture that:

Conjecture A *Just as in two dimensions, the scaling limit of the cable-graph loop-soup clusters in three dimensions should exactly be the collection of Brownian loop-soup clusters. The dimension Δ of these clusters should be equal to* $5/2$.

One difficulty in proving this conjecture is to be able to exclude the somewhat absurd-looking scenario that in the limit $\delta \to 0$, there might exist infinitely many disjoint dense (and "very skinny") cable-graph loop-soup clusters.

4.1.3 A Further Open Question

When $d = 2$, it is known that the obtained loop-soup clusters are in fact a deterministic function of the continuum GFF (based on the fact that their boundaries are level lines of this GFF in the sense of [27]), so that this "excursion decomposition" of the GFF is indeed unique (see [3, 4]).

Let us also recall that when $d = 2$ and $d = 3$, it is possible to define the (renormalized) square of the continuum GFF (or equivalently, the renormalized total occupation time measure of the loop-soup), see for instance [30] and the references therein. Let us now mention a related open question (also to illustrate that some questions remain also in the two-dimensional case).

Open Question B *In dimension $d = 2$ and $d = 3$: Are the (scaling limits of the) loop-soup clusters a deterministic function of this (renormalized) square of the continuum GFF? If not, what is the missing randomness?*

In dimension $d = 3$: Are the (scaling limits of the) loop-soup clusters a deterministic function of the continuum GFF?

4.2 Intermediate Dimensions

4.2.1 Some a Priori Estimates

Again, the cable-graph loop-soup clusters do not proliferate in the $\delta \to 0$ limit, then it is to be expected, based on estimates such as Corollary 2 that the dimension of the scaling limits would be $\Delta = 1 + (d/2)$. In particular, when $d = 4$ and $d = 5$, if one adds another independent macroscopic Brownian loop to an existing loop-soup, this additional loop will almost surely intersect infinitely many of these limits of cable-graph clusters. From this, it is easy to deduce that a limit of cable-graph clusters would actually contain infinitely many Brownian loops.

Recall however that a Brownian loop is almost surely a simple loop and that almost surely, any two loops in the loop-soup will be disjoint, so that Brownian loop-soup clusters will all consist of just one loop each (and therefore have Hausdorff dimension equal to 2).

Finally, self-similarity of the construction suggests that Brownian loops will be part of the scaling limit of the cable graph loop-soups at every scale, and that if one removes all Brownian loops of size greater than a say, then as $a \to 0$, the size of the largest limiting cluster will also vanish. In other words, the "macroscopic" loops are instrumental in the construction of the Brownian loop-soup clusters.

Let us summarize part of this in terms of a concrete conjecture.

Conjecture C *When $d = 4$ and $d = 5$, the limit in distribution of the cable-graph clusters in $(0, 1)^d \cap \delta\mathbb{Z}^d$ does exist, and it is supported on families of clusters of fractal dimension $1 + (d/2)$ with the property that for all small a, the number of clusters of diameter greater than a is finite.*

The main additional heuristic question that we will now discuss is whether the disjoint Brownian loops in the loop-soup get agglomerated into these scaling limit of cable-graph clusters in a deterministic manner or not (i.e., are the scaling limit of the cable-graph clusters a deterministic function the collection of Brownian loops or not?).

4.2.2 Background and Analogy with CLE Percolation

It is worthwhile to draw an analogy with one aspect of the papers [28, 29] about the existence of a non-trivial "critical percolation" model in a random fractal domain. Here, one should forget that CLE_κ for $\kappa \in (8/3, 4]$ is related to loop-soups or to the GFF, and one should view it as an example of random fractal "carpet" in the square $[0, 1]^2$. The CLE_κ carpet K_κ in $[0, 1]^2$ is obtained by removing from this square a countable collection of simply connected sets, that are all at positive distance from each other. It can be therefore be thought of as a conformal randomized version of the Sierpinski carpet. The following features are relevant here:

- The larger κ is, the smaller the CLE_κ tends to be. It is actually possible (this follows immediately from the CLE construction via loop-soups in [35]) to couple them in a decreasing way i.e., $K_\kappa \subset K_{\kappa'}$ when $8/3 < \kappa' \le \kappa \le 4$.
- There is one essential difference between CLE_κ for $\kappa < 4$ and CLE_4: When $\kappa < 4$, there exists a positive $u(\kappa)$ such that for all $a > 0$, the probability that there exists two holes in K_κ that have diameter greater than a and are at distance less than ε from each other does decay (at least) as $\varepsilon^{u+o(1)}$ as $\varepsilon \to 0$. This property fails to hold for CLE_4 (this probability will decay logarithmically) which intuitively means that exceptional bottlenecks are more likely in that CLE_4 case.

One of the results of [28] is the construction of a process that can be interpreted as a critical percolation process within the random set K_κ. One can view this either as defining a collection of clusters that live within K_κ, or if one looks at the dual picture, as a collection of clusters that "glue" the different CLE loops together (in the original percolation picture, the loops and their interior are "closed" and in the dual one, they are now "open"). In this dual picture, this does therefore construct a natural way to randomly regroup these holes (or their outer boundaries, that are SLE-type loops) into clusters.

One of the results of [29] is that this percolation/clustering procedure is indeed random (i.e., the obtained clusters are not a deterministic function of the CLE_κ) as long a $\kappa < 4$. On the other hand, it is shown in [28] that no non-trivial clustering mechanism can work for CLE_4.

4.2.3 Conjectures

The complement of a Brownian loop-soup in $(0, 1)^d$ for $d \geq 4$ has some similarities with the previous CLE_κ case. It is the complement of a random collection of disjoint simple loops, with a fractal structure. When $d \geq 5$, the "space" in-between the loops is much larger than in the 4-dimensional case, in the sense that the probability that two macroscopic loops are ε-close decays like a power of ε, while it only decays in a logarithmic fashion in 4 dimensions. Further analogies can also be made, that lead to:

Conjecture D *When $d = 5$, we conjecture that "critical percolation" in the space defined by "contracting all the loops in a loop-soup" (or equivalently, percolation that tries to glue together the loops in a loop-soup) should exist and be non-trivial. In other words, by observing the Brownian loop-soup only, one does not know which Brownian loops do belong to the same clusters.*

When $d = 4$, we conjecture that the glueing mechanism is deterministic. In other words, by observing the Brownian loop-soup only, one knows which Brownian loops do belong to the same clusters.

Let us finally conclude with the same question as for $d = 3$:

Open Question E *When $d = 4, 5$: In the scaling limit (taking the joint limit of the cable-graph clusters and of the GFF), are the limits of the cable-graph clusters determined by the limiting GFF?*

5 High Dimensions ($d > 6$)

5.1 General Features

As opposed to the cases $d = 3, 4, 5$ where most features are conjectural, it is possible to derive a number of facts when the dimension of the ambient space becomes large enough (note that we will not discuss the somewhat complex case

$d = 6$ here). As opposed to the lower-dimensional cases, these results do not say anything about geometric structures within the continuum GFF, but they provide insight into the asymptotic behaviour of the cable-graph loop-soup clusters in $\mathbb{Z}^d$ (or in large boxes in $\mathbb{Z}^d$). Actually, when the dimension of the space is large enough, we expect that the Brownian loop-soup in $\mathbb{R}^d$ (appearing as the scaling limit of the cable-graph loop-soup) and the GFF (appearing as the limit of the cable-graph GFF constructed using the cable-graph loop-soup) become asymptotically independent.

It is worth first recalling some of the results about usual (finite-range) critical percolation in high dimensions (see [1, 12–18] and the references therein). A landmark result in the study of those models is that when d is large enough, the "two-point function" (i.e., the probability that two points x and y belong to the same cluster) behaves (up to a multiplicative constant) like $1/\|x - y\|^{d-2}$ as $\|x - y\| \to \infty$. This is known to hold for (sufficiently) spread-out percolation in $\mathbb{Z}^d$ for $d > 6$, and in the case of usual nearest-neighbour percolation for $d \geq 11$. The existing proofs are based on the lace-expansion techniques (that have also been successfully applied to other models than percolation) as developed in this context by Hara and Slade [13, 15–17]). This estimate is then the key to the following subsequent statements that we describe in rather loose terms here (see Aizenman [1]): If one considers a finite-range percolation model restricted to $[-N, N]^d$, for which the two-point function estimates is shown to hold, then as $N \to \infty$:

- Clusters with large diameter (say, greater than $N/2$) will proliferate as $N \to \infty$—their number will be greater than $N^{d-6+o(1)}$ with high probability.
- With high probability, no cluster will have more than $N^{4+o(1)}$ points in it.

Note also that the geometry of large clusters can be related to superbrownian excursions.

As we shall explain now, similar results hold true for the loop-soup clusters in the cable-graph of $\mathbb{Z}^d$ when $d > 6$. The general feature is that the behaviour of the two-point function in this case is given for free by Corollary 2, so that the difficult lace-expansion ideas are not needed here. One just has to adjust ideas such as developed by Aizenman in [1] on how to extract further information from the estimate on the two-point function.

5.2 Some Results

Let us now explain how to adapt some arguments of [1] to the case of loop-soup percolation. It is convenient to work in the following setting: We define Λ_N to be the set of integer lattice points in $[-N, N]^d$, and $\Lambda_{N,c}$ the cable graph associated to it. We will consider the cable-graph loop-soup on $\Lambda_{N,c}$ and study its clusters and connectivity properties. We denote by n_0 the number of cable-graph clusters that contain at least one point of Λ_N, and we order them using some deterministic rule as $C_1, \ldots, C_{n_0}$. We denote by $|C|$ the number of points of Λ_N that lie in a set C, and when $x \in \Lambda_N$, we call $C(x)$ the cluster that contains x. In the sequel, $x \leftrightarrow y$ will always denote the event that x and y are connected via the cable-graph loop-soup in $\Lambda_{N,c}$ (the dependency on N will always be implicit). Note that for all $k \geq 1$,

$$E[|C(x)|^k] = \sum_{y_1,\ldots,y_k \in \Lambda_N} P[x \leftrightarrow y_1, \ldots, x \leftrightarrow y_k].$$

and also that

$$E[\sum_{n \leq n_0} |C_n|^{k+1}] = \sum_{x \in \Lambda_N} E[|C(x)|^k|].$$

Corollary 2 then implies (using simple bounds on the Green's function in a box) immediately that there exist constants v_1, v_2 such that for all sufficiently large N,

$$v_1 N^2 \leq \min_{x \in \Lambda_{N/2}} E[|C(x)|] \leq \max_{x \in \Lambda_{N/2}} E[|C(x)|] \leq \max_{x \in \Lambda_N} E[|C(x)|] \leq v_2 N^2$$

and then summing over x in Λ_N and in $\Lambda_{N/2}$, one gets the existence of v_3, v_4 such that for all large N,

$$v_3 N^{d+2} \leq E[\sum_{n \leq n_0} |C_n|^2] \leq v_4 N^{d+2}.$$

Let us first show the following analogue of (4.10) in [1]:

Proposition 3 *For some fixed large c_0, with probability going to* 1 *as $N \to \infty$, no loop-soup cluster (in Λ_N) contains more than $c_0 N^4 \log N$ points.*

Proof This is based on the fact that the Aizenman-Newman diagrammatic procedure [2] used in [1] to bound the moments of $|C(x)|$ can be adapted to this loop-soup percolation setting. Let us first explain this in some detail the case of the second moment. As mentioned above, one has

$$E[|C(x)|^2] = \sum_{y_1, y_2 \in \Lambda_N} P[x \leftrightarrow y_1, x \leftrightarrow y_2].$$

When $x \leftrightarrow y_1, x \leftrightarrow y_2$ both occur, then it means that for some loop γ in the cable-system loop-soup the events $\gamma \leftrightarrow x$, $\gamma \leftrightarrow y_1$ and $\gamma \leftrightarrow y_2$ occur disjointly (i.e., using disjoint sets of loops—the loops may overlap, but each event is realized using different loops); we call $\mathscr{T}$ this event. [To see this, one can first choose a "minimal" chain of loops that join x to y_1 (this means that one can not remove any these loops from the chain without disconnecting x to y_1) and then use a second "minimal" chain of loops that join y_2 to this first chain. The loop γ will be the loop of the first chain that this second chain joins y_2 to.]

In particular, it means that for at least one loop γ in the cable-system loop-soup, one can find integer points x_0, x_1 and x_2 in Λ_N that are at distance at most 1 from γ

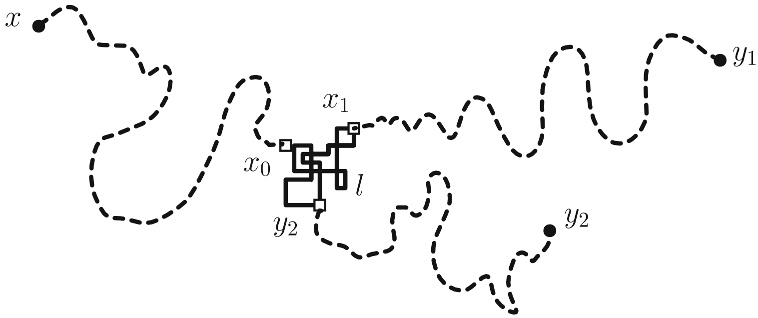


Fig. 1 Compared to the "usual" Aizenman–Newman tree expansion, one has also to sum over the loops that play the role of nodes of the tree, but this additional sum converges

such that $x \leftrightarrow x_0$, $y_1 \leftrightarrow x_1$ and $y_2 \leftrightarrow x_2$ occur disjointly (see Fig. 1). We are going to treat differently the case where γ visits at least two points of $\mathbb{Z}^d$ from the case where it visits less than two points.

Let us introduce some notation and make some further preliminary comments: For each cable-system loop γ that visits at least two integer points, one can look at its trace on $\mathbb{Z}^d$ that we denote by $l(\gamma)$, which is a discrete loop in Λ_N. Note that the collection L of all $l(\gamma)$'s for γ in the loop-soup $\mathscr{L}$ is a discrete random walk loop-soup in Λ_N, and that when an integer point is at distance at most 1 from γ, it is also at distance at most 1 from $l(\gamma)$. If $|l| \geq 2$ denotes the number of steps of the discrete loop $l(\gamma)$, there are therefore at most $|l| \times (2d+1)$ possibilities for each of x_0, x_1 and x_2.

For each given x, y_1 and y_2, we can now use the BK inequality to bound $P[x \leftrightarrow y_1, x \leftrightarrow y_2]$ by the sum of the contributions described in (a) and (b) below:

(a) The sum over all x_0, x_1 and x_2 that are all at distance at least 2 from each other of the product

$$P[x_0 \leftrightarrow x]P[x_1 \leftrightarrow y_1]P[x_2 \leftrightarrow y_2].$$

This sum corresponds to the contribution to the event $\mathscr{T}$ of the cases where γ visits at most one point of $\mathbb{Z}^d$. Note that for a given x_0, there are at most $(2d+1)^2$ choices (corresponding to the two steps or less needed to go from x_0 to x_1) for x_1 and $(2d+1)^2$ choices for x_2.

(b) The sum over all discrete loops l with $|l| \geq 2$ steps, of the sum over all x_0, x_1, x_2 that lie at distance at most 1 of l, of the product

$$P[l \in L]P[x_0 \leftrightarrow x]P[x_1 \leftrightarrow y_1]P[x_2 \leftrightarrow y_2].$$

This sum corresponds to the case where the loop γ in the event $\mathscr{T}$ visits at least two integer points (and we sum over all possible choices for $l(\gamma)$).

Equation (1) shows the existence of a constant w_0 independent of N, such that for all $y, y' \in \Lambda_N$ (as it is easier to create a connection in $\mathbb{Z}^d$ than in Λ_N), $P[y \leftrightarrow y'] \leq w_0/(1 + \|y - y'\|^{d-2})$; it follows immediately (summing over all y' that are in $y + \Lambda_{2N}$) that for some constant w_1, for all $N \geq 1$ and all $y \in \Lambda_N$,

$$\sum_{y' \in \Lambda_N} P[y \leftrightarrow y'] \leq w_1 N^2, \tag{1}$$

which is an inequality that we will now repeatedly use. For each choice of x_0, x_1 and x_2 (and possibly l if we are in the case (a)), if we now sum over all choices of y_1 and y_2 in Λ_N, we can use (1) to see that

$$E[|C(x)|^2] \leq \sum_{x_0 \in \Lambda_N} P[x_0 \leftrightarrow x](2d+1)^4 (w_1 N^2)^2$$
$$+ \sum_{(x_0, l) \in \mathscr{U}} \left[P[x \leftrightarrow x_0] \times P[l \in L] \times (|l|(2d+1))^2 \times (w_1 N^2)^2 \right]$$

where $\mathscr{U}$ is the set of pairs (x_0, l) satisfying (i)–(iii) where (i) $x_0 \in \Lambda_N$, (ii) the discrete loop l has at least 2 steps, and (iii) x_0 is at distance at most 1 from l; the term $(2d+1)^4$ comes from the bound on the number of possible choices for x_1 and x_2 for a given x_0 in (a), and the term $(|l|(2d+1))^2$ comes from the possible choices for x_1 and x_2 in (b) for a given discrete loop l with $|l| \geq 2$ steps).

The first sum over x_0 is bounded $(2d+1)^4 w_1^3 N^6$ (using (1) again). For the second one, we can first note that for each given x_0, the expected number of discrete loops of length m in a loop-soup in the whole of $\mathbb{Z}^d$ that pass through x_0 is given by the total mass of such loops under the discrete loop-measure, which is in turn expressed in terms of the probability that a random walk started from x_0 is back at x_0 after m steps (see for instance [19, 40] for such elementary considerations on loop-measures), which is bounded by some constant w_2 times $m^{-d/2}$. Hence, if we regroup the sum over all loops with the same length m, we see that the second sum over (x_0, l) in $\mathscr{U}$ is bounded by

$$\sum_{x_0 \in \Lambda_N} \left[(2d+1) P[x \leftrightarrow x_0] w_1^2 N^4 \sum_{m \geq 2} [w_2 m^{-d/2} (m(2d+1))^2] \right].$$

The key point is now that when $d/2 - 2 > 1$, i.e., $d > 6$, then $\sum_m m^{2-d/2} < \infty$, so that finally, we see that this sum over (x_0, l) in $\mathscr{U}$ is bounded by some constant times

$$N^4 \sum_{x_0 \in \Lambda_N} P[x \leftrightarrow x_0]$$

which in turn is also bounded by some constant times N^6 (using (1) again). Together with the bound for the sum in (a), we can therefore conclude that for some constant w_3, for all $N \geq 2$ and all $x \in \Lambda_N$,

$$E[|C(x)|^2] \leq w_3 N^6.$$

In summary, we see that $d > 6$ is also the threshold at which the extended nature of the Brownian loops does not essentially influence the estimates compared to finite-range percolation.

Similarly, for any $k \geq 3$, by enumerating trees, and expanding in a similar way (this time, one has to sum over $k-1$ loops in the loop-soup that will be the nodes of the tree) using the Aizenman-Newman enumeration ideas, one obtains the existence of constants w_4 and w_5 such that for all N, x and k,

$$E[|C(x)|^k] \leq w_4 k! w_5^k N^{4k-2}. \tag{2}$$

If we then finally sum over all x in Λ_N, we get that

$$E[\sum_{n \leq n_0} |C_n|^{k+1}] = \sum_{x \in \Lambda_n} E[|C(x)|^k] \leq w_4 k! w_5^k N^{d+4k-2}.$$

In particular, if M denotes $\max |C_n|$, we get an upper bound for $E[M^{k+1}]$ from which one readily deduces the proposition by using Markov's inequality and choosing the appropriate k (of the order of a constant times $\log N$).

Let us now turn to the proliferation of large clusters:

Proposition 4 *With probability that tends to* 1 *as N tends to infinity, there exist more than $N^{d-6}/\log^2 N$ disjoint loop-soup clusters with diameter greater than $N/2$.*

The proof proceeds along the same lines as the analogous result (4.8) in [1]:

Proof One can for instance define B_1 and B_2 to be the boxes obtained by shifting $\Lambda_{N/4}$ along the first-coordinate axis by $-N/2$ and $N/2$ respectively. Each of the two boxes has circa $(N/2)^d$ points in it, they at distance at least $N/4$ from $\partial \Lambda_N$, and they are at distance circa $N/2$ from each other. Now, Corollary 2 readily shows that if we define

$$X := \sum_n |C_n \cap B_1| \times |C_n \cap B_2|,$$

then for some positive finite constant b_1,

$$E[X] = E[\sum_{x_1 \in B_1, x_2 \in B_2} 1_{x_1 \leftrightarrow x_2}] \sim b_1 N^{d+2}$$

as $N \to \infty$. On the other hand, one can bound the second moment

$$E[X^2] = \sum_{x_1, y_1 \in B_1, x_2, y_2 \in B_2} P[\mathscr{E}(x_1, x_2, y_1, y_2)]$$

where $\mathscr{E}(x_1, x_2, y_1, y_2) := \{x_1 \leftrightarrow x_2, y_1 \leftrightarrow y_2\}$ using the following remark (call the *truncation lemma* in [1]): To check if $\mathscr{E}$ holds, one can first discover $C(x_1)$. If it does contain x_2, y_1 and y_2 (we call this event $\mathscr{E}_1$), then we know already that $\mathscr{E}$ holds. The only other scenario (we call this event $\mathscr{E}_2 = \mathscr{E} \setminus \mathscr{E}_1$) for which $\mathscr{E}$ holds is that $y_1 \in C(x_1)$, that neither y_1 nor y_2, are in $C(x_1)$, and then that for the remaining loop-soup percolation in the complement of $C(x_1)$ in the cable-graph, y_1 is connected to y_2. Clearly,

$$P[\mathscr{E}_2] = P[\mathscr{E}] - P[\mathscr{E}_1] \leq P[x_1 \leftrightarrow y_1] P[x_2 \leftrightarrow y_2]$$

(the first probability in the product is an upper bound for the probability that $y_1 \in C(x_1)$ and that neither y_1 nor y_2 are in $C(x_1)$, and the second probability is an upper bound for the conditional probability that $x_2 \leftrightarrow y_2$ in the remaining domain). Summing this inequality over all x_1, x_2, y_1, y_2, and using (2) one immediately gets that

$$\begin{aligned} E[X^2] - E[X]^2 &= \sum_{x_1, y_1 \in B_1, x_2, y_2 \in B_2} [P[\mathscr{E}(x_1, x_2, y_1, y_2)] - P[x_1 \leftrightarrow y_1] P[x_2 \leftrightarrow y_2]] \\ &\leq \sum_{x_1, x_2, y_1, y_2 \in \Lambda_N} P[\mathscr{E}_1(x_1, x_2, y_1, y_2)] \\ &= E[\sum_{n \leq n_0} |C_n|^4] \\ &\leq b_2 N^{d+10} \end{aligned}$$

for some constant b_2 independent of N. Combining this bound of the variance of X with the estimate of its mean (and noting that $d + 10 < 2(d + 2)$ because $d > 6$), we see that for all ε,

$$P[X \in [(b_1 - \varepsilon) N^{d+2}, (b_1 + \varepsilon) N^{d+2}]] \to 1$$

as $N \to \infty$. If X denotes the number of clusters that intersect both B_1 and B_2, noting that with high probability, all quantities $|C_n \cap B_1|$ and $|C_n \cap B_2|$ are smaller than $c_0 N^4 \log N$ (because of Proposition 3), we deduce that with a probability that goes to 1 as $N \to \infty$,

$$X \geq \frac{(b_1/2) \times N^{d+2}}{(c_0 N^4 \log N)^2} = (b_1/2c_0^2) \times \frac{N^{d-6}}{\log^2 N}.$$

5.3 Some Final Comments

We conclude with the following comments: On the one hand, we have seen that when $N \to \infty$, there will typically be a large number of large clusters (say of diameter greater than $N/2$), but on the other hand, only a tight number of Brownian loops of diameter comparable to N. In fact, when $a \in (0, d)$, the N^a-th largest Brownian loop will have a diameter of the order of $N \times N^{-a/d+o(1)}$. This means for instance that an overwhelming fraction of the numerous large clusters will contain no loop of diameter greater than N^b for $b > 6/d$. In other words, if we remove all loops of diameter greater than N^b, one will still have at least $N^{d-6+o(1)}$ large clusters, and the estimates for the two-point function will actually remain valid. If we fix $b \in (6/d, 1)$, since N^b is also much smaller than the size N of the box, we can interpret this cable-graph loop-soup percolation with cut-off as a critical (or near-critical) percolation model: If we scale everything down by a factor N: We are looking at a Poissonian family of small sets, and for the chosen parameters one observes macroscopic clusters (as $N \to \infty$).

We plan to discuss further aspects of loop-soup cluster percolation and the structure of the GFF in high dimensions in forthcoming work. In particular, when $d \geq 9$, the relation with the integrated superbrownian excursions can be made more precise.

Acknowledgement This work was supported by the Swiss National Science Foundation (SNF) grant #175505.

References

1. Aizenman, M.: On the number of incipient spanning clusters. Nucl. Phys. B **485**, 551–582 (1997)
2. Aizenman, M., Newman, C.: Tree graph inequalities and critical behaviour in percolation models. J. Stat. Phys. **36**, 107–143 (1984)
3. Aru, J., Lupu, T., Sepúlveda, A.: First passage sets of the 2D continuum Gaussian free field. Probab. Theory Rel. Fields **176**, 1303–1355 (2020).
4. Aru, J., Lupu, T., Sepúlveda, A.: The first passage sets of the 2D Gaussian free field: convergence and isomorphisms. Commun. Math. Phys. **375**, 1885–1929 (2020)
5. Aru, J., Sepúlveda, A., Werner, W.: On bounded-type thin local sets of the two-dimensional Gaussian free fields. J. Inst. Math. Jussieu **18**, 591–618 (2019)
6. van de Brug, T., Camia, F., Lis, M.: Random walk loop soups and conformal loop ensembles. Probab. Theory Related Fields **166**, 553–584 (2016)
7. Brydges, D., Fröhlich, J., Spencer, T.: The random walk representation of classical spin systems and correlation inequalities. Commun. Math. Phys. **83**, 123–150 (1982)
8. Alvez, C., Sapozhnikov, A.: Decoupling inequalities and supercritical percolation for the vacant set of random walk loop soup. Electron. J. Probab. **24**, 34 (2019)
9. Dubédat, J.: SLE and the free field: partition functions and couplings. J. Am. Math. Soc. **22**, 995–1054 (2009)
10. Duminil-Copin, H., Goswami, S., Raoufi, A., Severo, F., Yadin, A.: Existence of phase transition for percolation using the Gaussian Free Field. Duke Math. J. **169**, 3539–3563 (2018)

11. Dynkin, E.B.: Gaussian and non-Gaussian random fields associated with Markov processes. J. Funct. Anal. **55**, 344–376 (1984)
12. Fitzner, R., van der Hofstad, R.: Mean-field behavior for nearest-neighbor percolation in $d > 10$. Electron. J. Probab. **22**, 65 (2017)
13. Hara, T.: Decay of correlations in nearest-neighbor self-avoiding walk, percolation, lattice trees and animals. Ann. Probab. **36**, 530–593 (2008)
14. Hara, T., van der Hofstad, R., Slade, G.: Critical two-point functions and the lace expansion for spread-out high-dimensional percolation and related models. Ann. Probab. **31**, 349–408 (2003)
15. Hara, T., Slade, G.: Mean-field critical behaviour for percolation in high dimensions. Commun. Math. Phys. **128**, 333–391 (1990)
16. Hara, T., Slade, G.: The scaling limit of the incipient infinite cluster in high-dimensional percolation. I. Critical Exponents. J. Stat. Phys. **99**, 1075–1168 (2000)
17. Hara, T., Slade, G.: The scaling limit of the incipient infinite cluster in high-dimensional percolation. II. Integrated super-Brownian excursion. Lett. Math. Phys. **41**, 1244–1293 (2000)
18. Heydenreich, M., van der Hofstad, R.: Progress in high-dimensional percolation and random graphs. CRM Short Courses. Springer, Berlin (2017)
19. Lawler, G.F., Limic, V.: Random Walk: A Modern Introduction. Cambridge Studies in Advanced Mathematics, vol. 123. Cambridge University Press, Cambridge (2010)
20. Lawler, G.F., Trujillo Ferraras, J.: Random walk loop soup. Trans. Am. Math. Soc. **359**, 767–787 (2007)
21. Lawler, G.F., Werner, W.: The Brownian loop-soup. Probab. Theory Related Fields **128**, 565–588 (2004)
22. Le Jan, Y.: Markov paths, loops and fields. In: 2008 St-Flour Summer School. L.N. Math., vol. 2026. Springer, Berlin (2011)
23. Lupu, T.: Loop percolation on discrete half-plane. Electron. Commun. Probab. **21**, 9 (2016)
24. Lupu, T.: From loop clusters and random interlacements to the free field. Ann. Probab. **44**, 2117–2146 (2016)
25. Lupu, T.: Poisson ensembles of loops of one-dimensional diffusions. Mém. SMF **158**, 1–158 (2018)
26. Lupu, T.: Convergence of the two-dimensional random walk loop-soup clusters to CLE. J. Eur. Math. Soc. **21**, 1201–1227 (2019)
27. Miller, J., Sheffield, S.: Imaginary geometry I: interacting SLEs. Probab. Theory Related Fields **164**, 553–705 (2016)
28. Miller, J., Sheffield, S., Werner, W.: CLE Percolations. Forum Math. Pi **5**, 102 (2017)
29. Miller, J., Sheffield, S., Werner, W.: Non-simple SLE curves are not determined by their range. J. Eur. Math. Soc. **22**, 669–716 (2020)
30. Qian, W., Werner, W.: Decomposition of Brownian loop-soup clusters. J. Eur. Math. Soc. **21**, 3225–3253 (2019)
31. Revuz, D., Yor, M.: Continuous martingales and Brownian motion. Grundlehren Math. Wiss., vol. 293 Springer, Berlin (1999)
32. Schramm, O.: Scaling limits of loop-erased random walks and uniform spanning trees. Israel J. Math. **118**, 221–288 (2000)
33. Schramm, O., Sheffield, S.: A contour line of the continuum Gaussian free field. Probab. Theory Rel. Fields **157**, 47–80 (2013)
34. Sheffield, S.: Exploration trees and conformal loop ensembles. Duke Math. J. **147**, 79–129 (2009)
35. Sheffield, S., Werner, W.: Conformal loop ensembles: the Markovian characterization and the loop-soup construction. Ann. Math. **176**, 1827–1917 (2012)
36. Simon, B.: $P(\Phi)_2$ Euclidean (Quantum) Field Theory. Princeton University Press, Princeton (1974)
37. Symanzik, K.: Euclidean quantum field theory. In: Jost, R. (ed.) Local Quantum Theory. Academic Press, New York (1969)
38. Werner, W.: SLEs as boundaries of clusters of Brownian loops. C. R. Math. Acad. Sci. Paris **337**, 481–486 (2003).

39. Werner, W.: in preparation
40. Werner, W., Powell, E.: Lecture notes on the Gaussian Free Field (2020). arXiv
41. Werner, W., Wu, H.: From CLE_κ to $\text{SLE}_\kappa(\rho)$. Electr. J. Probab. **18**, 20 (2013)

Vladas, 3rd from left in bottom row, organizer, Workshop in Mathematical Physics (Mambucaba, Brasil, 2001)

M. E. Vares et al. (eds.), *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, Progress in Probability 77,
https://doi.org/10.1007/978-3-030-60754-8

Vladas, in bottom row, conference organizer (Beijing, 2017)