

# Aging of the Metropolis dynamics on the random energy model

Jiří Černý<sup>1</sup> · Tobias Wassmer<sup>1</sup>

Received: 20 February 2015 / Revised: 23 September 2015 / Published online: 8 December 2015  
© Springer-Verlag Berlin Heidelberg 2015

**Abstract** We study the Metropolis dynamics of the simplest mean-field spin glass model, the random energy model. We show that this dynamics exhibits aging by showing that the properly rescaled time change process between the Metropolis dynamics and a suitably chosen ‘fast’ Markov chain converges in distribution to a stable subordinator. The rescaling might depend on the realization of the environment, but we show that its exponential growth rate is deterministic.

**Mathematics Subject Classification** 82C44 · 82D30 · 60K37

## 1 Introduction

This paper studies the out-of-equilibrium behavior of the Metropolis dynamics on the random energy model (REM). Our main goal is to answer one of the remaining important open questions in the field, namely whether this dynamics exhibits aging, and, if yes, whether its aging behavior admits the usual description in terms of stable Lévy processes.

Aging is one of the main features appearing in the long-time behavior of complex disordered systems (see e.g. [12] for a review). It was for the first time observed experimentally in the anomalous relaxation patterns of the residual magnetization of spin glasses (e.g. [18, 29]). One of the most influential steps in the theoretical modeling of the aging phenomenon is the introduction of the so-called trap models by

---

✉ Jiří Černý  
jiri.cerny@univie.ac.at

Tobias Wassmer  
tobias.wassmer@univie.ac.at

<sup>1</sup> Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria

Bouchaud [11] and Bouchaud and Dean [13]. These models, while being sufficiently simple to allow analytical treatment, reproduce the characteristic power law decay seen experimentally.

Since then a considerable effort has been made in putting the predictions obtained from the trap models to a solid basis, that is to derive these predictions from underlying spin-glass dynamics. The first attempt in this direction was made in [6–8] where it was shown that, for a very particular Glauber-type dynamics, at time scales very close to the equilibration, a well chosen two-point correlation function converges to that given by Bouchaud's trap model.

With the paper [9], where the same type of dynamics was studied in a more general framework and on a broader range of time scales, it emerged that aging establishes itself by the fact that scaling limits of certain additive functionals of Markov chains are stable Lévy processes, and that the convergence of the two-point correlation functions is just a manifestation of the classical arcsine law for stable subordinators.

The Glauber-type dynamics used in those papers, sometimes called random hopping time (RHT) dynamics, is however rather simple and is often considered as 'non-realistic', mainly because its transition rates do not take into account the energy of the target state. Its advantage is that it can be expressed as a time change of a simple random walk on the configuration space of the spin glass, which allows for a certain decoupling of the randomness of the dynamics from the randomness of the Hamiltonian of the spin glass, making its rigorous studies more tractable.

For more realistic Glauber-type dynamics of spin glasses, like the so-called Bouchaud's asymmetric dynamics or the Metropolis dynamics, such decoupling is not possible. As a consequence, these dynamics are far less understood.

Recently, some progress has been achieved in the context of the simplest mean-field spin glass model, the REM. First, in [31], the Bouchaud's asymmetric dynamics have been considered in the regime where the asymmetry parameter tends to zero with the size of the system. Building on the techniques started in [32], this paper confirms the predictions of Bouchaud's trap model in this regime. Second, the Metropolis dynamics have been studied in [26], for a truncated version of the REM, using the techniques developed for the RHT dynamics in [24, 25], again confirming Bouchaud's predictions.

The weak asymmetry assumption of [31] and the truncation of [26] have both the same purpose. They aim at overcoming some specific features of the asymmetry and recovering certain features of the RHT dynamics. Our aim in this work is to get rid of these simplifications and treat the non-modified REM with the usual Metropolis dynamics.

Let us also mention that Bouchaud's asymmetric dynamics (and implicitly the Metropolis one) is rather well understood in the context of trap models on  $\mathbb{Z}^d$ , see [3, 14, 27], where it is possible to exploit the connections to the random conductance model with unbounded conductances [4]. Finally, the Metropolis dynamics on the complete graph was considered in [25].

Before stating our main result, let us briefly recall the general scheme for proving aging in terms of convergence to stable Lévy processes. The actual spin glass dynamics,  $X = (X_t)_{t \geq 0}$ , which is reversible with respect to the Gibbs measure of the Hamiltonian, is compared to another Markov chain  $Y = (Y_t)_{t \geq 0}$  on the same space, which is an

‘accelerated’ version of  $X$  and whose stationary measure is uniform. The process  $Y$  is typically easier to understand, e.g. it is a simple random walk for the RHT dynamics, and the original process  $X$  can be written as its time change,

$$X(t) = Y(S^{-1}(t)), \tag{1.1}$$

for the right continuous inverse  $S^{-1}$  of a certain additive functional  $S$  of the Markov chain  $Y$ , called the ‘clock process’. The aim is then to show convergence of the properly rescaled clock process  $S$  to an increasing stable Lévy process, that is to a stable subordinator.

We now state our main result. We consider the unmodified REM, as introduced in [19,20]. The state space of this model is the  $N$ -dimensional hypercube  $\mathbb{H}_N = \{-1, 1\}^N$ , and its Hamiltonian is a collection  $(E_x)_{x \in \mathbb{H}_N}$  of i.i.d. standard Gaussian random variables defined on some probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . The non-normalized Gibbs measure  $\tau_x = e^{\beta\sqrt{N}E_x}$  at inverse temperature  $\beta > 0$  gives the equilibrium distribution of the system.

The Metropolis dynamics on the REM is the continuous-time Markov chain  $X = (X_t)_{t \geq 0}$  on  $\mathbb{H}_N$  with transition rates

$$r_{xy} = e^{-\beta\sqrt{N}(E_x - E_y)} \mathbf{1}_{\{x \sim y\}} = \left(1 \wedge \frac{\tau_y}{\tau_x}\right) \mathbf{1}_{\{x \sim y\}}, \quad x, y \in \mathbb{H}_N. \tag{1.2}$$

Here,  $x \sim y$  means that  $x$  and  $y$  are neighbors on  $\mathbb{H}_N$ , that is they differ in exactly one coordinate.

As explained above, we will compare the Metropolis chain  $X$  with another ‘fast’ Markov chain  $Y = (Y_t)_{t \geq 0}$  with transition rates

$$q_{xy} = \frac{\tau_x \wedge \tau_y}{1 \wedge \tau_x} \mathbf{1}_{\{x \sim y\}}, \quad x, y \in \mathbb{H}_N. \tag{1.3}$$

It can be easily checked using the detailed balance conditions that  $Y$  is reversible and that its equilibrium distribution is

$$\nu_x = \frac{1 \wedge \tau_x}{Z_N}, \quad x \in \mathbb{H}_N,$$

where  $Z_N = \sum_{x \in \mathbb{H}_N} (1 \wedge \tau_x)$ . Finally, since  $r_{xy} = (1 \vee \tau_x)^{-1} q_{xy}$ ,  $X$  can be written as a time change of  $Y$  as in (1.1) with the clock process  $S$  being given by

$$S(t) = \int_0^t (1 \vee \tau_{Y_s}) \, ds. \tag{1.4}$$

For the rest of the paper we mostly only deal with the process  $Y$  and the clock process  $S$ . With exception of this introduction and Sect. 8, the actual Metropolis dynamics  $X$  does not appear. For a fixed environment  $\tau = (\tau_x)_{x \in \mathbb{H}_N}$ , let  $P_\nu^\tau$  denote the law of the process  $Y$  started from its stationary distribution  $\nu$ , and let  $D([0, T], \mathbb{R})$  be the

space of  $\mathbb{R}$ -valued cadlag functions on  $[0, T]$ . We denote by  $\beta_c = \sqrt{2 \log 2}$  the (static) critical temperature of the REM. Our main result is the following theorem.

**Theorem 1.1** *Let  $\alpha \in (0, 1)$  and  $\beta > 0$  be such that*

$$\frac{1}{2} < \frac{\alpha^2 \beta^2}{\beta_c^2} < 1, \tag{1.5}$$

and define

$$g_N = e^{\alpha \beta^2 N} (\alpha \beta \sqrt{2\pi N})^{-\frac{1}{\alpha}}. \tag{1.6}$$

Then there are random variables  $R_N$  which depend on the environment  $(E_x)_{x \in \mathbb{H}_N}$  only, such that for every  $T > 0$  the rescaled clock processes

$$S_N(t) = g_N^{-1} S(t R_N), \quad t \in [0, T],$$

converge in  $\mathbb{P}$ -probability as  $N \rightarrow \infty$ , in  $P_v^\tau$ -distribution on the space  $D([0, T], \mathbb{R})$  equipped with the Skorokhod  $M_1$ -topology, to an  $\alpha$ -stable subordinator  $V_\alpha$ . The random variables  $R_N$  satisfy

$$\lim_{N \rightarrow \infty} \frac{\log R_N}{N} = \frac{\alpha^2 \beta^2}{2}, \quad \mathbb{P}\text{-a.s.} \tag{1.7}$$

Let us make a few remarks on this result.

1. The result of Theorem 1.1 confirms that the predictions of Bouchaud’s trap model hold for the Metropolis dynamics on the REM, at least at the level of scaling limits of clock processes. It also compares directly to the results obtained for the symmetric (RHT) dynamics in Theorem 3.1 of [9]. The scales  $g_N$  and  $R_N$  are (up to sub-exponential prefactors) the same as previously. Similarly as in [9], the right inequality of the assumption (1.5) is completely natural, beyond it  $Y$  ‘feels’ the finiteness of  $\mathbb{H}_N$  and aging is not expected to occur. The left inequality in (1.5) is technical, it ensures that the relevant deep traps are well separated (cf. Lemma 2.1), introducing certain simplifications in the proof. Basing also on Theorem 2.1 of [15] (which is an improved version of Theorem 3.1 of [9]), we believe that this bound might be improved to  $\alpha^2 \beta^2 / \beta_c^2 > 0$ , by further exploiting our method. Finally, as previously, note that (1.5) is satisfied also for  $\beta < \beta_c$  for appropriate  $\alpha$ , hence aging can occur above the critical temperature at sufficiently short observation scales.
2. Our choice of the fast chain  $Y$  is rather unusual. In view of the previous papers [3,31], it would be natural to take instead the chain  $\tilde{Y}$  with uniform invariant measure and with transition rates  $\tau_x \wedge \tau_y$ , that is without the correction  $1 \wedge \tau_x$  which appears in the denominator of (1.3). This choice has, however, some deficiencies. On the heuristic level,  $\tilde{Y}$  is not an acceleration of  $X$ , since it is much slower than  $X$  on sites with very small Gibbs measure  $\tau_x \ll 1$ . These sites, which are irrelevant for the statics, then ‘act as traps’ on  $\tilde{Y}$ , making them relevant for the dynamics,

which is undesirable. On the technical level, the trapping on sites with small Gibbs measure has the consequence that the mixing time of  $\tilde{Y}$  is very large.

Our choice of the fast chain  $Y$  runs as fast as  $X$  on the sites with small Gibbs measure and thus does not have this deficiency. Moreover, since  $\nu_x = Z_N^{-1}$  whenever  $E_x \geq 0$ , the equilibrium distribution of the fast chain  $Y$  is still uniform on the relevant deep traps, so the clock process  $S$  retains its usual importance for aging. Remark also that in order to overcome the difficulties with the slow mixing of the chain  $\tilde{Y}$ , [31] truncate the Hamiltonian of the REM at 0 which effectively sets  $\tau_x \geq 1$  for all  $x \in \mathbb{H}_N$ . We prefer to retain the full REM and use the modified fast chain  $Y$  instead. Finally, [26] uses the discrete skeleton of  $X$  as the base chain, which has some interesting features but introduces similar undesirable effects.

3. We view Theorem 1.1 as our primary aging statement. The reason for this is that it seems hard, without extending the paper considerably, to derive aging results in terms of the usual two-point functions as e.g. in [9,26], or in terms of the age process as in [22,31]. Such derivation usually requires some knowledge of the fast chain  $Y$  that goes over the  $M_1$ -convergence of the clock processes. This knowledge is typically automatically obtained in the previous approaches. The strength (or the weakness) of our method is that we do not need to obtain such finer knowledge to show the clock process convergence.

Without the finer knowledge, it is possible to consider only non-standard (and to some extent artificial) two-point functions or age processes. We give two examples of such results in the concluding Sect. 8.

4. A rather unusual feature of Theorem 1.1 is the fact that the scaling  $R_N$  is random, it depends on the random environment. This is again a consequence of our technique. Claim (1.7) in Theorem 1.1 however shows that at least the exponential growth of  $R_N$  is deterministic. The random scale  $R_N$  is explicitly defined in (2.10). We will see that its definition depends on a somewhat free choice of an auxiliary parameter, but nevertheless the final result does not depend on this parameter. This property (and also other heuristic considerations) makes us conjecture that  $R_N$  should actually satisfy a law of large numbers,

$$\lim_{N \rightarrow \infty} h(N) e^{-\alpha^2 \beta^2 N/2} R_N = 1, \quad \mathbb{P}\text{-a.s.},$$

for some deterministic function  $h(N)$  growing (or decaying) at most sub-exponentially.

Observe also that the randomness of  $R_N$  should not have, at least heuristically, any influence on the aging behavior of the process  $X$ .  $R_N$  is the time scale of the auxiliary process  $Y$ . The only physically relevant scale is  $g_N$ , giving the observation time-scale for the Metropolis chain  $X$ . At least partially, this is confirmed by the aging results of Sect. 8.

5. The mode of convergence in Theorem 1.1 is not optimal, one would rather like to obtain the convergence in  $P_v^\tau$ -distribution for  $\mathbb{P}$ -almost every environment, which is usually called ‘quenched’ convergence. Actually, Theorem 1.1 can be strengthened slightly to a statement which is somewhere between  $\mathbb{P}$ -a.s. convergence and convergence in  $\mathbb{P}$ -probability. Namely, the statement holds for a.e. realiza-

tion of sites with ‘small’  $\tau_x$ , but only in probability over sites with ‘large’  $\tau_x$ , cf. Remark 6.4.

6. Our proof of Theorem 1.1 strongly exploits the i.i.d. structure of the Hamiltonian of the REM. At present we do not know if it is possible to combine our techniques with those used for the RHT dynamics of the  $p$ -spin model in [5, 10].

We proceed by commenting on the proof of Theorem 1.1, concentrating mainly on its novelties. The general strategy so far to prove such a result has been to first reduce the problem to the clock process restricted to a set of deep traps which govern the behavior of the original clock process. The different methods then all more or less aim at dividing the contribution of consequently found deep traps into essentially i.i.d. blocks. For example in [9] or [3], this is achieved by controlling the hitting probabilities of deep traps, proving that they are hit essentially uniformly in exponentially distributed times, and controlling the time the chain spends at the deep traps by a sharp control of the Green function. Similar rather precise estimates on hitting probabilities and/or Green function are necessary in other approaches. Using this i.i.d. structure, one can then show convergence of the clock process by standard methods, e.g. computing the Laplace transform.

The method used in this paper is slightly inspired by the general approach taken in [17, 22]. There, models of trapped random walks on  $\mathbb{Z}^d$  are considered where little information about the discrete skeleton as well as the waiting times of a continuous-time Markov chain is available, and minimal necessary conditions for convergence of the clock process are found. Taking up this idea, instead of analyzing in detail the behavior of the fast chain  $Y$ , we extract the minimal amount of information needed to show convergence of the clock process. In particular, we do not need any exact control of hitting probabilities and Green functions of deep traps, as most previous work did.

The first step in our proof is standard, namely that the main contribution to the clock process comes from a small set of vertices with large Gibbs measure  $\tau_x$ , the so-called deep traps. More precisely, denoting the set of deep traps by  $\mathcal{D}_N$  (see Sect. 2 for details), we will show that the clock process  $S$  can be well approximated by the ‘clock process of the deep traps’

$$S_{\mathcal{D}}(t) = \int_0^t (1 \vee \tau_{Y_s}) \mathbf{1}_{\{Y_s \in \mathcal{D}_N\}} ds. \quad (1.8)$$

Then it remains to show that in fact  $g_N^{-1} S_{\mathcal{D}}(t R_N)$  converges to a stable subordinator.

To this end, we will in some sense invert the standard procedure described above. Instead of approximating the clock process by an i.i.d. block structure and then use the Laplace transform to show convergence, we will first compute a certain conditional Laplace transform using some special properties of the Metropolis dynamics. Then we analyze what is actually needed in order to show convergence of the unconditional Laplace transform.

A bit more detailed, this will be done as follows. Under condition (1.5), the deep traps are almost surely well separated. This fact and the fact that the definition (1.3) contains the factor  $\tau_x \wedge \tau_y$  imply that the transition rates  $q_{xy}$  of the fast chain  $Y$  do

not depend on the energies  $E_x$  of the deep traps. Therefore, one can condition on the location of all deep traps and the energies  $E_x$  of the non-deep traps, which determines the law  $P_\nu^\tau$  of  $Y$ , and take the expectation over the energies of the deep traps. We call this a ‘quasi-annealed’ expectation, and denote it by  $\mathbb{E}_{\mathcal{D}}$  for the moment. Let  $\ell_t(x)$  denote the local time of the fast chain  $Y$  (see Sect. 2 for details). As  $\mathbb{E}_{\mathcal{D}}$  is simply an expectation over i.i.d. random variables, the quasi-annealed Laplace transform of the rescaled clock process of the deep traps given  $Y$  can be computed. It essentially behaves like

$$\mathbb{E}_{\mathcal{D}} \left[ e^{-\lambda \frac{1}{\varepsilon_N} S_{\mathcal{D}}(tR_N)} \mid Y \right] \approx \exp \left\{ -\mathcal{K} \lambda^\alpha \varepsilon_N \sum_{x \in \mathcal{D}_N} \ell_{tR_N}(x)^\alpha \right\}. \quad (1.9)$$

Here,  $\varepsilon_N$  is a deterministic sequence tending to 0 as  $N \rightarrow \infty$ . The above approximation shows that the only object related to  $Y$  we have to control is the local-time functional  $\varepsilon_N \sum_{x \in \mathcal{D}_N} \ell_{tR_N}(x)^\alpha$ .

We will show that this a priori non-additive functional of  $Y$  actually behaves in an additive way, namely that it converges to  $t$  as  $N \rightarrow \infty$ , under  $P_\nu^\tau$  for  $\mathbb{P}$ -a.e. environment  $\tau$ . For this convergence to hold it is sufficient to have some weak bounds on the mean hitting time of deep traps as well as some control on the mixing of the chain  $Y$  together with an appropriate choice of the scale  $R_N$  that depends on the environment.

Using standard methods we then strengthen the quasi-annealed convergence to quenched convergence (in the sense of Theorem 1.1).

To conclude the introduction, let us comment on how our method might be extended. The key argument in the computation of the quasi-annealed Laplace transform, namely the fact that the chain  $Y$  is independent of the depth of the deep traps, seems very specific for the Metropolis dynamics. However, by adapting the method appropriately and using network reduction techniques, we believe that one could also treat Bouchaud’s asymmetric dynamics and Metropolis dynamics in the regime where the left-hand side inequality of (1.5) fails, i.e. there are neighboring deep traps.

The rest of the paper is structured as follows. Detailed definitions and notations used through the paper are introduced in Sect. 2. In Sect. 3 we analyze the mixing properties of the fast chain  $Y$ , which will be crucial at several points later. In Sect. 4 we give bounds on the mean hitting time of deep traps and on the normalizing scale  $R_N$ . Using these bounds and the results on the mixing of  $Y$ , we show concentration of the local time functional  $\varepsilon_N \sum_{x \in \mathcal{D}_N} \ell_{tR_N}(x)^\alpha$  in Sect. 5. We prove convergence of the rescaled clock process of the deep traps in Sect. 6 with the above mentioned computation of the quasi-annealed Laplace transform, using the concentration of the local time functional. We treat the shallow traps in Sect. 7 by showing that their contribution to the clock process can be neglected. Finally, in Sect. 8, we prove Theorem 1.1 and present some additional aging results, as already mentioned. In Appendix we give the proof of a technical result which is used to bound the expected hitting times in Sect. 4.

## 2 Definitions and notation

In this section we introduce some notation used through the paper and recall a few useful facts. We use  $\mathbb{H}_N$  to denote the  $N$ -dimensional hypercube  $\{-1, 1\}^N$  equipped

with the usual distance

$$d(x, y) = \frac{1}{2} \sum_{i=1}^N |x_i - y_i|,$$

and write  $\mathcal{E}_N$  for the set of nearest-neighbor edges  $\mathcal{E}_N = \{\{x, y\} : d(x, y) = 1\}$ .

For given parameters  $\alpha$  and  $\beta$ , let

$$\gamma = \frac{\alpha^2 \beta^2}{\beta_c^2} \in (1/2, 1), \tag{2.1}$$

by condition (1.5) in Theorem 1.1.

Recall from the introduction that  $(E_x : x \in \mathbb{H}_N, N \geq 1)$ , is a family of i.i.d. standard Gaussian random variables defined on some probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . Note that we do not denote the dependence on  $N$  explicitly, but we assume that the space  $(\Omega, \mathcal{F}, \mathbb{P})$  is the same for all  $N$ . For  $\beta > 0$  the non-normalized Gibbs factor  $\tau_x$  is given by  $\tau_x = e^{\beta\sqrt{N}E_x}$ .

Using the standard Gaussian tail approximation,

$$\mathbb{P}[E_x \geq t] = \frac{1}{t\sqrt{2\pi}} e^{-t^2/2}(1 + o(1)) \quad \text{as } t \rightarrow \infty, \tag{2.2}$$

we obtain that  $g_N$ , as defined in Theorem 1.1, satisfies

$$\mathbb{P}[\tau_x > ug_N] = u^{-\alpha} 2^{-\gamma N} (1 + o(1)).$$

This heuristically important computation explains the appearance of stable laws in the distribution of sums of  $\tau_x$ : If we observe  $2^{\gamma N}$  vertices, then finitely many of them have their rescaled Gibbs measures  $\tau_x/g_N$  of order unity, and, moreover, those rescaled Gibbs measures behave like random variables in the domain of attraction of an  $\alpha$ -stable law.

Recall also that  $Y = (Y_t)_{t \geq 0}$  stands for the fast Markov chain whose transition rates  $q_{xy}$  are given in (1.3), and that  $\nu = (\nu_x)_{x \in \mathbb{H}_N}$  denotes the invariant distribution of this chain,  $\nu_x = \frac{1 \wedge \tau_x}{Z_N}$ . For a given environment  $\tau = (\tau_x)_{x \in \mathbb{H}_N}$ , let  $P_x^\tau$  and  $P_\nu^\tau$  denote the laws of  $Y$  started from a vertex  $x$  or from  $\nu$  respectively, and  $E_x^\tau, E_\nu^\tau$  the corresponding expectations.

Note that the normalization factor  $Z_N = \sum_{x \in \mathbb{H}_N} (1 \wedge \tau_x)$  satisfies, for every constant  $\kappa \in (0, 1/2)$ ,

$$\kappa 2^N \leq Z_N \leq 2^N \quad \mathbb{P}\text{-a.s for } N \text{ large enough.} \tag{2.3}$$

Indeed, obviously  $Z_N \leq 2^N$ , and  $Z_N \geq \sum_{x \in \mathbb{H}_N} \mathbf{1}_{\{E_x \geq 0\}}$ . But  $\mathbf{1}_{\{E_x \geq 0\}}$  are i.i.d. Bernoulli random variables, therefore the statement follows immediately by the law of large numbers.



An important role in the study of properties of  $Y$  is played by the conductances defined by

$$c_{xy} = \nu_x q_{xy} = \frac{\tau_x \wedge \tau_y}{Z_N} \quad \text{for } x \sim y. \tag{2.4}$$

Let  $\theta_s$  be the left shift on the space of trajectories of  $Y$ , that is

$$(\theta_s Y)_t = Y_{s+t}. \tag{2.5}$$

Let  $H_x = \inf\{t > 0 : Y_t = x\}$  be the hitting time of  $x$  by  $Y$ ,  $J_1$  the time of the first jump of  $Y$ , and let  $H_x^+ = H_x \circ \theta_{J_1} + J_1 = \inf\{t > J_1 : Y_t = x\}$  be the return time to  $x$  by  $Y$ . Similarly define  $H_A$  and  $H_A^+$  for a set  $A \subset \mathbb{H}_N$ . The local time  $\ell_t(x)$  of  $Y$  is given by

$$\ell_t(x) = \int_0^t \mathbf{1}_{\{Y_s=x\}} ds.$$

Using this notation the clock process  $S$  introduced in (1.4) can be written as

$$S(t) = \int_0^t (1 \vee \tau_{Y_s}) ds = \sum_{x \in \mathbb{H}_N} \ell_t(x) (1 \vee \tau_x).$$

To define the set of deep traps  $\mathcal{D}_N$  and the random scale  $R_N$  mentioned in the introduction we introduce a few additional parameters. For  $\alpha \in (0, 1)$ ,  $\beta > 0$  as in Theorem 1.1 and  $\gamma$  as in (2.1), we fix  $\gamma'$  and  $\alpha'$  such that

$$\frac{1}{2} < \gamma' < \gamma, \quad \text{and} \quad \alpha' = \frac{\beta_c}{\beta} \sqrt{\gamma'}. \tag{2.6}$$

An explicit choice of  $\gamma'$  will be made later in Sect. 5. We define the auxiliary scale

$$g'_N = e^{\alpha' \beta^2 N} (\alpha' \beta \sqrt{2\pi N})^{-\frac{1}{\alpha'}},$$

and set

$$\mathcal{D}_N = \{x \in \mathbb{H}_N : \tau_x \geq g'_N\}.$$

to be the set of deep traps. By the Gaussian tail approximation (2.2) it follows that the density of  $\mathcal{D}_N$  satisfies

$$\mathbb{P}[x \in \mathcal{D}_N] = 2^{-\gamma' N} (1 + o(1)). \tag{2.7}$$

We quote the following observation on the size and sparseness of  $\mathcal{D}_N$ . The sparseness will play a key role in our computation of the quasi-annealed Laplace transform in Sect. 6.

**Lemma 2.1** [9, Lemma 3.7] *For every  $\varepsilon > 0$ ,  $\mathbb{P}$ -a.s. for  $N$  large enough,*

$$|\mathcal{D}_N|2^{(\gamma'-1)N} \in (1 - \varepsilon, 1 + \varepsilon). \tag{2.8}$$

*Moreover, since  $\gamma' > 1/2$ , there exists  $\delta > 0$  such that  $\mathbb{P}$ -a.s. for  $N$  large enough, the separation event*

$$\mathcal{S} = \{ \min\{d(x, y) : x, y \in \mathcal{D}_N\} \geq \delta N \} \tag{2.9}$$

*holds.*

Finally, for the sake of concreteness, let us give the explicit form of the random scale  $R_N$ ,

$$R_N = 2^{(\gamma-\gamma')N} \left( \sum_{x \in \mathcal{D}_N} \frac{E_x^\tau[\ell_{T_{\text{mix}}}(x)^\alpha]}{E_v^\tau[H_x]} \right)^{-1}, \tag{2.10}$$

where  $T_{\text{mix}}$  denotes the mixing time of  $Y$ , a randomized stopping time which we will construct in Sect. 3. The reason for this definition will become apparent when we prove the concentration of the local time functional mentioned in the introduction. Although the definition of  $R_N$  seems arbitrary by the somewhat free choice of the parameter  $\gamma'$ , Theorem 1.1 actually shows that asymptotically  $R_N$  will be independent of  $\gamma'$ .

For the rest of the paper,  $c, c', c''$  will always denote positive constants whose values may change from line to line. We will use the notation  $g = o(1)$  for a function  $g(N)$  that tends to 0 as  $N \rightarrow \infty$ , and  $g = O(f)$  for a function  $g(N)$  that is asymptotically at most of order  $f(N)$ , i.e.  $\lim_{N \rightarrow \infty} |g(N)|/f(N) \leq c$ , for some  $c > 0$ .

### 3 Mixing properties of the fast chain

The fact that the chain  $Y$  mixes fast, namely on a scale polynomial in  $N$ , plays a crucial role in many of our arguments. In this section we analyze the mixing behavior of  $Y$ . We first give a lower bound on the spectral gap  $\lambda_Y$  of  $Y$ , which we then use to construct a strong stationary time  $T_{\text{mix}}$ .

**Proposition 3.1** *There are constants  $\kappa > 0, K > 0, C_0 > 0$ , such that  $\mathbb{P}$ -a.s. for  $N$  large enough,*

$$\lambda_Y \geq \frac{\kappa}{4} N^{-K-1-\beta C_0}.$$

We prove this proposition with help of the Poincaré inequality derived in [21]. To state this inequality, let  $\Gamma$  be a complete set of self-avoiding nearest-neighbor paths on  $\mathbb{H}_N$ , that is for each  $x \neq y \in \mathbb{H}_N$  there is exactly one path  $\gamma_{xy} \in \Gamma$  connecting  $x$  and  $y$ . Let  $|\gamma|$  be the length of the path  $\gamma$ . By [21, Proposition 1', p. 38], using also the reversibility of  $Y$  and recalling the definition (2.4) of the conductances, it follows that

$$\frac{1}{\lambda_Y} \leq \max_{e=\{u,v\} \in \mathcal{E}_N} \left\{ \frac{1}{c_{uv}} \sum_{\gamma_{xy} \in \Gamma: \gamma_{xy} \ni e} |\gamma_{xy}| \nu_x \nu_y \right\}. \tag{3.1}$$

To minimize the right-hand side of (3.1), special care should be taken of the edges whose conductance  $c_{uv} = (\tau_u \wedge \tau_v)/Z_N$  is very small, that is which are incident to vertices with very small  $\tau_u$ . Those ‘bad’ edges should be avoided if possible by paths  $\gamma \in \Gamma$ . They cannot be avoided completely, since  $\Gamma$  should be a complete set of paths. On the other hand, if such edge is the first or the last edge of some path  $\gamma_{xy}$ , its small conductance is canceled by equally small  $\nu_x$  or  $\nu_y$ . Therefore, to apply (3.1) efficiently, one should find a set of paths  $\Gamma$  such that all paths  $\gamma \in \Gamma$  avoid ‘bad’ vertices, except for vertices at both ends of the paths.

In the context of spin glass dynamics this method was used before in [23] to find the spectral gap of the Metropolis dynamics (1.2). Using the same approach, that is using the same set of paths  $\Gamma$  as in [23], we could find a lower bound on the spectral gap of the fast chain  $Y$  of leading order  $\exp\{-c\sqrt{N \log N}\}$ . This turns out to be too small for our purposes, cf. Remark 6.4.

In the next lemma we construct a set of paths  $\Gamma$  that avoids more ‘bad’ vertices, which allows to improve the lower bound on the spectral gap to be polynomial in  $N$ . This is possible by using an embedding of  $\mathbb{H}_N$  into its sub-graph of ‘good’ vertices, i.e. vertices with not too small  $\tau_x$ , which is inspired by similar embeddings in [28].

For a nearest-neighbor path  $\gamma = \{x_0, \dots, x_n\}$ , we call the vertices  $x_1, \dots, x_{n-1}$  the interior vertices of  $\gamma$ , and the edges  $\{x_i, x_{i+1}\}, i = 1, \dots, n - 2$ , the interior edges of  $\gamma$ .

**Lemma 3.2** *There is an integer  $K > 0$  and a constant  $C_0 > 0$ , such that  $\mathbb{P}$ -a.s. for  $N$  large enough there exists a complete set of paths  $\Gamma$ , such that the following three properties hold.*

- (i) *For every path  $\gamma \in \Gamma$ , every interior edge  $e = \{u, v\}$  satisfies*

$$Z_N c_{uv} = \tau_u \wedge \tau_v \geq N^{-\beta C_0}.$$

- (ii)  *$|\gamma| \leq 8N$  for all  $\gamma \in \Gamma$ .*
- (iii) *Every edge  $e \in \mathcal{E}_N$  is contained in at most  $N^K 2^{N-1}$  paths  $\gamma \in \Gamma$ .*

*Proof* For  $C_0 > 0$ , whose value will be fixed later, we say that  $x \in \mathbb{H}_N$  is good if  $\tau_x \geq N^{-\beta C_0}$ , and it is bad otherwise. To construct the complete set of paths  $\Gamma$  satisfying the required properties, we will use the fact that the set of good vertices is very dense in  $\mathbb{H}_N$ . In particular, we will show that

$$\mathbb{P}\text{-a.s. for } N \text{ large enough, every } x \in \mathbb{H}_N \text{ has at least } \frac{1}{2}C_0\sqrt{N} \text{ good neighbors,} \tag{3.2}$$

and

$$\begin{aligned} &\mathbb{P}\text{-a.s. for } N \text{ large enough, for any pair of vertices } x, y \text{ at distance 2 or 3,} \\ &\text{there is a nearest-neighbor path of length at most 7 connecting } x \text{ and } y, \\ &\text{such that all interior vertices of this path are good,} \end{aligned} \tag{3.3}$$

To prove these two claims, note first that for any  $x \in \mathbb{H}_N$ , the probability of being bad is

$$\mathbb{P} \left[ \tau_x < N^{-\beta C_0} \right] = \mathbb{P} \left[ E_x < -C_0 N^{-\frac{1}{2}} \log N \right] = \frac{1}{2} - \int_0^{C_0 N^{-\frac{1}{2}} \log N} \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds.$$

For  $N$  large enough the integrand is larger than  $\frac{3}{10}$ , and it follows that

$$\mathbb{P}[x \text{ is bad}] \leq \frac{1}{2} \left( 1 - \frac{3}{5} C_0 N^{-\frac{1}{2}} \log N \right) =: \frac{1}{2} (1 - q_N).$$

Hence, the number of bad neighbors of a vertex  $x \in \mathbb{H}_N$  is stochastically dominated by a Binomial  $(N, \frac{1}{2}(1 - q_N))$  random variable  $B$ . For  $\lambda > 0$ , the exponential Chebyshev inequality yields

$$\begin{aligned} & \mathbb{P} \left[ x \text{ has more than } N - \frac{1}{2} C_0 \sqrt{N} \text{ bad neighbors} \right] \\ & \leq \mathbb{P} \left[ B \geq N - \frac{1}{2} C_0 \sqrt{N} \right] = \mathbb{P} \left[ e^{\lambda B} \geq e^{\lambda(N - \frac{1}{2} C_0 \sqrt{N})} \right] \\ & \leq e^{-\lambda(N - \frac{1}{2} C_0 \sqrt{N})} \left( 1 + \frac{1}{2} (1 - q_N) (e^\lambda - 1) \right)^N \\ & = e^{-\lambda(N - \frac{1}{2} C_0 \sqrt{N})} \left( \frac{e^\lambda}{2} (1 - q_N + e^{-\lambda} (1 + q_N)) \right)^N \\ & \leq 2^{-N} e^{\frac{\lambda}{2} C_0 \sqrt{N}} (\exp\{-q_N + e^{-\lambda} (1 + q_N)\})^N. \end{aligned}$$

Since  $q_N \rightarrow 0$  as  $N \rightarrow \infty$ , the second term in the braces is bounded by  $2e^{-\lambda}$  for  $N$  large enough. Inserting  $q_N$  and choosing  $\lambda = \log N$ , the above is bounded by

$$\begin{aligned} & 2^{-N} \exp \left\{ \frac{1}{2} C_0 \sqrt{N} \log N - \frac{3}{5} C_0 \sqrt{N} \log N + 2 \right\} \\ & \leq 2^{-N} \exp \left\{ -\frac{1}{10} C_0 \sqrt{N} \log N \right\}, \end{aligned}$$

for  $N$  large enough. With a union bound over all  $x \in \mathbb{H}_N$  and using the Borel–Cantelli lemma, (3.2) follows.

To prove (3.3), we first introduce some notation. For a given vertex  $x$  and  $\{i_1, \dots, i_k\} \subset \{1, \dots, N\}$ , denote by  $x^{i_1 \dots i_k}$  the vertex that differs from  $x$  exactly in coordinates  $i_1, \dots, i_k$ . If two vertices  $x$  and  $y$  are at distance 2, then  $y = x^{kl}$  for some  $k, l \in \{1, \dots, N\}$ . Then for  $\{i, j\} \cap \{k, l\} = \emptyset$  we define the path  $\gamma_{xy}^{ij}$  of length 6 as  $\{x, x^i, x^{ij}, x^{ijk}, x^{ijkl} = y^{ij}, y^j, y\}$ . Similarly, for  $x, y$  with  $d(x, y) = 3$ , we have  $y = x^{klm}$ , and for  $\{i, j\} \cap \{k, l, m\} = \emptyset$  we define the path  $\gamma_{xy}^{ij}$  of length 7 by  $\{x, x^i, x^{ij}, x^{ijk}, x^{ijkl}, x^{ijklm} = y^{ij}, y^j, y\}$ . Observe that for fixed  $x, y$  with

$d(x, y) = 2$  or  $3$  and for different pairs  $i, j$  the innermost 3 or 4 vertices of the paths  $\gamma_{xy}^{ij}$  are disjoint.

We now show that with high probability, for every  $x, y$  at distance 2 or 3, we may find  $i, j$  such that  $\gamma_{xy}^{ij}$  has only good interior vertices. Fix a pair  $x, y \in \mathbb{H}_N$  at distance 2 or 3, and let as above  $k, l$  or  $k, l, m$  be the coordinates in which  $x$  and  $y$  differ. Assume for the moment that both  $x$  and  $y$  have at least  $\frac{1}{2}C_0\sqrt{N}$  good neighbors. Then there are at least  $\frac{1}{4}C_0^2N$  pairs  $i, j$  such that the vertices  $x^i$  and  $y^j$  are good. Moreover, since it is a matter of dealing with a constant number of exceptions, we may tacitly assume that  $i \neq j$ , and  $\{i, j\} \cap \{k, l\} = \emptyset$  or  $\{i, j\} \cap \{k, l, m\} = \emptyset$ , respectively.

The remaining interior vertices  $\{x^{ij}, x^{ijk}, x^{ijkl} = y^{ij}\}$  or  $\{x^{ij}, x^{ijk}, x^{ijkl}, x^{ijklm} = y^{ij}\}$  are all good with probability strictly larger than  $1/2$ , so the probability that one or more of these vertices are bad is bounded by  $15/16$ . Since these 3 or 4 innermost vertices are disjoint for different pairs  $i, j$ , by independence, the probability that among all  $\frac{1}{4}C_0^2N$  pairs  $\{i, j\}$  there is none for which all innermost 3 or 4 vertices of  $\gamma_{xy}^{ij}$  are good is bounded by  $(15/16)^{\frac{1}{4}C_0^2N}$ . Hence, for one fixed pair  $x, y \in \mathbb{H}_N$  at distance 2 or 3, where both  $x$  and  $y$  have at least  $\frac{1}{2}C_0\sqrt{N}$  good neighbors, the probability that there is no path from  $x$  to  $y$  of length 6 or 7 with all interior vertices good is bounded by  $(15/16)^{\frac{1}{4}C_0^2N}$ .

There are less than  $2^N(N^2 + N^3)$  pairs of vertices at distance 2 or 3 respectively, and we know from the proof of (3.2) that with probability larger than  $1 - e^{-c\sqrt{N} \log N}$  every  $x \in \mathbb{H}_N$  has at least  $\frac{1}{2}C_0\sqrt{N}$  good neighbors. It follows that the probability that the event in (3.3) does not happen is bounded by

$$e^{-c\sqrt{N} \log N} + 2^N(N^2 + N^3)(15/16)^{\frac{1}{4}C_0^2N}. \tag{3.4}$$

Choosing  $C_0 > \sqrt{\frac{4 \log 2}{\log 15/16}}$  and applying the Borel–Cantelli lemma implies (3.3).

We now use the density properties (3.2) and (3.3) of good vertices to define a (random) mapping from the hypercube to its sub-graph of good vertices. Let  $\varphi_N(x) : \mathbb{H}_N \rightarrow \mathbb{H}_N$  be given by

$$\varphi_N(x) = \begin{cases} x, & \text{if } x \text{ is good;} \\ x^i, & \text{if } x \text{ and } x^j, j < i, \text{ are bad but } x^i \text{ is good;} \\ x, & \text{if } x \text{ is bad and has no good neighbor.} \end{cases}$$

By (3.2),  $\mathbb{P}$ -a.s. for  $N$  large enough the last option will not be used, and therefore  $\varphi_N$  maps all vertices to good vertices. In this case, for two neighboring vertices  $x, y$ , their good images  $\varphi_N(x)$  and  $\varphi_N(y)$  can either coincide, or be at distance 1, 2, or 3.

Let further

$$\mathcal{P}_N = \{\{x_0, \dots, x_k\} : k \geq 0, d(x_i, x_{i-1}) = 1 \forall i = 1, \dots, k\}$$

be the set of finite nearest-neighbor paths on  $\mathbb{H}_N$ , including paths of length zero, which are just single vertices. We extend the mapping  $\varphi_N$  to the set of edges by associating to every edge  $e \in \mathcal{E}_N$  an element of  $\mathcal{P}_N$ : For an edge  $e = \{x, y\} \in \mathcal{E}_N$ , let  $\varphi_N(e)$  be

- the ‘path’  $\{\varphi_N(x)\}$ , if  $\varphi_N(x)$  is good and  $\varphi_N(x) = \varphi_N(y)$ ;
- the path  $\{\varphi_N(x), \varphi_N(y)\}$ , if both  $\varphi_N(x)$  and  $\varphi_N(y)$  are good and at distance 1;
- the path  $\gamma_{\varphi_N(x), \varphi_N(y)}^{ij}$  with ‘minimal’  $i, j$  such that all vertices of this path are good, if both  $\varphi_N(x)$  and  $\varphi_N(y)$  are good with distance 2 or 3 and such path exists;
- the path  $\{x, y\}$  in any other case.

From (3.2) and (3.3) it follows that  $\mathbb{P}$ -a.s. for  $N$  large enough the last option does not occur and  $\varphi_N$  maps all edges to paths that contain only good vertices.

Finally, we extend  $\varphi_N$  to be a map that sends paths to paths. For  $\gamma = \{x_0, \dots, x_n\} \in \mathcal{P}_N$  we define  $\varphi_N(\gamma)$  to be a concatenation of paths  $\varphi_N(\{x_{i-1}, x_i\}), i = 1, \dots, n$ , with possible loops erased by an arbitrary fixed loop-erasure algorithm. Note that  $\varphi_N$  can make paths shorter or longer, but by construction, for any path  $\gamma \in \mathcal{P}_N$ ,

$$|\varphi_N(\gamma)| \leq 7|\gamma|. \tag{3.5}$$

We can now construct the random set of paths  $\Gamma$  that satisfies the properties of the lemma. We first define a certain canonical set of paths  $\tilde{\Gamma}$ , and then use the mapping  $\varphi_N$  to construct  $\Gamma$  from  $\tilde{\Gamma}$ .

For any pair of vertices  $x \neq y \in \mathbb{H}_N$ , let  $\tilde{\gamma}_{xy}$  be the path from  $x$  to  $y$  obtained by consequently flipping the disagreeing coordinates, starting at coordinate 1. These paths are all of length smaller or equal to  $N$ , and the set  $\tilde{\Gamma} = \{\tilde{\gamma}_{xy} : x \neq y \in \mathbb{H}_N\}$  has the property that any edge  $e$  is used by at most  $2^{N-1}$  paths in  $\tilde{\Gamma}$ . Indeed, if  $e = \{u, v\}$ , then there is a unique  $i$  such that  $u_i \neq v_i$ . By construction,  $e \in \tilde{\gamma}_{xy}$  if

$$\begin{aligned} x &= (x_1, \dots, x_{i-1}, u_i, u_{i+1}, \dots, u_N), \\ y &= (v_1, \dots, v_{i-1}, v_i, y_{i+1}, \dots, y_N). \end{aligned}$$

It follows that a total of  $N - 1$  coordinates of  $x$  and  $y$  are unknown, and so the number of possible pairs  $x, y$  for paths  $\tilde{\gamma}_{xy}$  through  $e$  is bounded by  $2^{N-1}$  (cf. [21, Example 2.2]).

For any pair  $x \neq y \in \mathbb{H}_N$ , let the path  $\gamma_{xy}$  in the set  $\Gamma$  be defined by

$$\gamma_{xy} = \begin{cases} \varphi_N(\tilde{\gamma}_{xy}), & \text{if } x, y \text{ are good,} \\ \{x\} \circ \varphi_N(\tilde{\gamma}_{xy}), & \text{if } x \text{ is bad and } y \text{ is good,} \\ \varphi_N(\tilde{\gamma}_{xy}) \circ \{y\}, & \text{if } x \text{ is good and } y \text{ is bad,} \\ \{x\} \circ \varphi_N(\tilde{\gamma}_{xy}) \circ \{y\}, & \text{if } x, y \text{ are bad,} \end{cases}$$

where ‘ $\circ$ ’ denotes the path concatenation.

It remains to check that this set of paths  $\Gamma$  indeed satisfies the required properties. First, by construction,  $\Gamma$  is complete, that is every path  $\gamma_{xy} \in \Gamma$  connects  $x$  with  $y$  and is nearest-neighbor and self-avoiding. Further, by construction of  $\varphi_N$  and the properties (3.2) and (3.3),  $\mathbb{P}$ -a.s. for  $N$  large enough, all interior vertices of all  $\gamma \in \Gamma$  are good, i.e. (i) is satisfied. Moreover, by (3.5) and the construction of the paths  $\tilde{\gamma} \in \tilde{\Gamma}$ , the paths  $\gamma \in \Gamma$  have length at most  $7N + 2$ , hence (ii) is satisfied for  $N \geq 2$ . Finally,  $\varphi_N$  deforms the paths  $\tilde{\gamma} \in \tilde{\Gamma}$  only locally, so that the number of paths in  $\Gamma$  passing through an edge  $e$  is bounded by the number of paths in  $\tilde{\Gamma}$  passing through

the ball of radius 4 around  $e$ . But this number is bounded by  $2^{N-1}$  times the number of edges in that ball, which is bounded by  $N^K$  for some integer  $K > 0$ . This proves (iii) and thus finishes the proof of the lemma.  $\square$

We can now prove the spectral gap estimate.

*Proof of Proposition 3.1*  $\mathbb{P}$ -a.s. for every  $N$  large enough we can find a complete set of paths  $\Gamma$  such that (i), (ii) and (iii) of Lemma 3.2 and (2.3) hold. By (ii), the expression in (3.1) over which the maximum is taken is bounded from above by

$$\frac{8N}{Z_N} \frac{1}{\tau_u \wedge \tau_v} \sum_{\gamma_{xy} \ni \{u,v\}} (\tau_x \wedge 1)(\tau_y \wedge 1). \tag{3.6}$$

We distinguish three cases for the position of the edge  $\{u, v\}$  in a path  $\gamma_{xy}$ .

- (1) If  $\{u, v\}$  is an interior edge of  $\gamma_{xy}$ , then  $\tau_u \wedge \tau_v$  is larger than  $N^{-\beta C_0}$  by (i) of Lemma 3.2.
- (2) If  $\{u, v\}$  is at the end of the path  $\gamma_{xy}$ , say at  $u = x$ , and  $v$  is an interior vertex of  $\gamma_{xy}$ , then  $\tau_x \wedge \tau_v$  is either larger than  $N^{-\beta C_0}$ , or it is equal to  $\tau_x$  in which case it cancels with  $\tau_x \wedge 1$ . Indeed, if  $\tau_x \wedge \tau_v$  was smaller than  $N^{-\beta C_0}$  and equal to  $\tau_v$ , then  $v$  would be a bad interior vertex of  $\gamma$ , which contradicts (i) of Lemma 3.2.
- (3) If  $\gamma_{xy}$  only consists of the single edge  $\{x, y\}$ , then  $\tau_x \wedge \tau_y$  is either larger than 1, or the term  $\tau_x \wedge \tau_y$  cancels with the smaller one of  $\tau_x \wedge 1$  and  $\tau_y \wedge 1$ .

It follows that for every edge  $\{u, v\}$  the expression (3.6) is bounded from above by

$$\frac{8N}{Z_N} N^{\beta C_0} \#\{\text{paths through } e\}.$$

Since, by (iii) of Lemma 3.2, the number of paths is bounded by  $N^K 2^{N-1}$ , and, by (2.3),  $Z_N \geq \kappa 2^N$ , this completes the proof.  $\square$

In a next step we construct the mixing time  $T_{\text{mix}}$  of the fast chain  $Y$ . To this end, define the mixing scale

$$m_N = \frac{8}{\kappa} N^{K+3+\beta C_0}. \tag{3.7}$$

Then Proposition 3.1 reads  $\lambda_N \geq 2N^2 m_N^{-1}$ .

We assume that our probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  is rich enough so that there exist infinitely many independent random variables distributed uniformly on  $[0, 1]$ , independent of everything else. A randomized stopping time  $T$  is a positive random variable such that the event  $\{T \leq t\}$  depends only on  $\{Y_s : s \leq t\}$ , the environment, and on the values of these additional random variables.

**Proposition 3.3**  $\mathbb{P}$ -a.s. for  $N$  large enough, there exists a randomized stopping time  $T_{\text{mix}}$  with values in  $\{m_N, 2m_N, 3m_N, \dots\}$  such that  $T_{\text{mix}}$  is a strong stationary time for  $Y$ , that is for any (possibly random)  $Y_0 \in \mathbb{H}_N$ ,

- (i)  $P_{Y_0}^\tau [Y_{T_{\text{mix}}} = y] = \nu_y$ ,

- (ii) for any  $k \geq 1$ ,  $P_{Y_0}^\tau [T_{\text{mix}} \geq km_N] = e^{-(k-1)}$ ,
- (iii)  $T_{\text{mix}}$  and  $Y_{T_{\text{mix}}}$  are independent.

*Proof* This construction follows closely [31, Proposition 3.1], with only minor adaptations. Define the following distances from stationarity,

$$s(t) = \min \{s \geq 0 : \forall x, y \in \mathbb{H}_N, P_x^\tau [Y_t = y] \geq (1 - s)v(y)\},$$

$$\bar{d}(t) = \max_{x, y \in \mathbb{H}_N} \|P_x^\tau [Y_t \in \cdot] - P_y^\tau [Y_t \in \cdot]\|_{TV},$$

where  $\|\cdot\|_{TV}$  denotes the total variation distance. Define the time

$$\mathcal{T} = \inf\{t \geq 0 : \bar{d}(t) \leq e^{-1}\}.$$

From [2, Lemmas 4.5, 4.7 and 4.23] we know that

$$\begin{aligned} \bar{d}(t) &\leq e^{-\lfloor t/\mathcal{T} \rfloor}, \\ s(2t) &\leq 1 - (1 - \bar{d}(t))^2, \\ \mathcal{T} &\leq \frac{1}{\lambda_Y} \left(1 + \frac{1}{2} \log \frac{1}{\nu^*}\right), \end{aligned} \tag{3.8}$$

where  $\nu^* = \min_x \nu_x$ . Since  $\mathbb{P}[\tau_x \leq e^{-N^2/4}] \leq ce^{-c'N^3}$ , by the Borel–Cantelli lemma,  $\mathbb{P}$ -a.s. for  $N$  large enough,  $\log \frac{1}{\nu^*} \leq \frac{N^2}{4} + \log Z_N \leq \frac{1}{2}N^2$ . Therefore, by Proposition 3.1 and (3.8),  $\mathbb{P}$ -a.s. for  $N$  large enough,  $\mathcal{T} \leq \frac{1}{4}m_N$ ,  $\bar{d}(\frac{1}{2}m_N) \leq e^{-2}$ , and  $s(m_N) \leq 2e^{-2} \leq e^{-1}$ , which means that for all  $Y_0, y \in \mathbb{H}_N$ ,

$$P_{Y_0}^\tau [Y_{m_N} = y] \geq (1 - e^{-1})\nu_y.$$

We can now define the strong stationary time  $T_{\text{mix}}$  with values in  $\{m_N, 2m_N, \dots\}$ . Let  $U_1, U_2, \dots$  be i.i.d. uniformly on  $[0, 1]$  distributed random variables, independent of anything else. Conditionally on  $Y_0 = x, Y_{m_N} = y$ , let  $T_{\text{mix}} = m_N$  if

$$U_1 \leq \frac{(1 - e^{-1})\nu_y}{P_x^\tau [Y_{m_N} = y]} \quad (\leq 1).$$

Otherwise, we define  $T_{\text{mix}}$  inductively: for every  $k \in \mathbb{N}$ , conditionally on  $T_{\text{mix}} > km_N, Y_{km_N} = z$  and  $Y_{(k+1)m_N} = y$ , let  $T_{\text{mix}} = (k + 1)m_N$  if

$$U_{k+1} \leq \frac{(1 - e^{-1})\nu_y}{P_z^\tau [Y_{m_N} = y]} \quad (\leq 1).$$

By construction, we have for every  $x \in \mathbb{H}_N$ ,

$$P_x^\tau [T_{\text{mix}} = m_N \mid Y_{m_N} = y] = \frac{(1 - e^{-1})\nu_y}{P_x^\tau [Y_{m_N} = y]},$$



and thus

$$P_{Y_0}^\tau [T_{\text{mix}} = m_N, Y_{m_N} = y \mid Y_0 = x] = (1 - e^{-1})v_y.$$

Similarly, we have

$$P_{Y_0}^\tau [T_{\text{mix}} = (k + 1)m_N, Y_{(k+1)m_N} = y \mid T_{\text{mix}} > km_N, Y_{km_N} = x] = (1 - e^{-1})v_y.$$

By induction over  $k$ , we obtain that for any  $k \in \mathbb{N}$  and  $y \in \mathbb{H}_N$ ,

$$P_{Y_0}^\tau [T_{\text{mix}} = km_N, Y_{km_N} = y] = e^{-(k-1)}(1 - e^{-1})v_y,$$

which finishes the proof. □

For future reference we collect here two useful statements that follow directly from the construction of  $T_{\text{mix}}$ .

**Lemma 3.4** *For every  $t > 0$  and  $x \in \mathbb{H}_N$  and every starting distribution  $\rho$ ,*

$$\begin{aligned} P_\rho^\tau [Y_t = x \mid T_{\text{mix}} < t] &= v_x, \\ |P_\rho^\tau [Y_t = x] - v_x| &\leq P_\rho^\tau [T_{\text{mix}} > t] = e^{-t/m_N-1}. \end{aligned}$$

### 4 Bounds on mean hitting time and random scale

In this section we prove bounds on the mean hitting time  $E_v^\tau [H_x]$  of deep traps  $x \in \mathcal{D}_N$ . As a corollary of the proof we will obtain a useful bound on the Green function in deep traps. The bounds on the mean hitting times will further imply bounds on the random scale  $R_N$ , which will imply the claim (1.7) of Theorem 1.1.

**Proposition 4.1** *There exists  $\delta \in (0, 1/6)$ , such that  $\mathbb{P}$ -a.s. for  $N$  large enough,*

$$2^{N-N^{1-\delta}} \leq E_v^\tau [H_x] \leq 2^{N+N^{1-\delta}} \quad \text{for every } x \in \mathcal{D}_N.$$

The proof of Proposition 4.1 is split in two parts.

*Proof of the upper bound* For the upper bound we use [2, Lemma 3.17] which states that

$$E_v^\tau [H_x] \leq \frac{1 - v_x}{\lambda_Y v_x}.$$

Since  $\tau_x \geq 1$  for deep traps  $x \in \mathcal{D}_N$ , this is smaller than  $\frac{Z_N}{\lambda_Y}$ , which by Proposition 3.1 and (2.3) is bounded by  $2^{N+N^{1-\delta}}$ ,  $\mathbb{P}$ -a.s. for  $N$  large enough. □

For the lower bound we will use a version of Proposition 3.2 of [16] which allows to bound the inverse of the mean hitting time  $E_v^\tau [H_x]$  in terms of the effective conductance from  $x$  to a suitable set  $B$ . Recall the definition of the conductances  $c_{xy}$  from (2.4),

and let  $c_x = \sum_{y \sim x} c_{xy}$ . Following the terminology of [30, Chapter 2], we define the effective conductance between a vertex  $x$  and a set  $B$  as

$$\mathcal{C}(x \rightarrow B) = P_x^\tau[H_x^+ > H_B]c_x.$$

By Proposition 8.5, which is a generalization of [16, Proposition 3.2] to arbitrary continuous-time finite-state-space Markov chains,

$$\frac{1}{E_v^\tau[H_x]} \leq \mathcal{C}(x \rightarrow B)\nu(B)^{-2}. \tag{4.1}$$

To apply this bound effectively, we should find a set  $B$  such that  $\mathcal{C}(x \rightarrow B)$  is small and  $\nu(B)$  close to 1. In the next lemma we construct such sets  $B$  for every  $x \in \mathbb{H}_N$ . For these sets we have some control on the conductances connecting  $B$  and  $B^c$ . Using standard network reduction techniques we can then give a bound on the effective conductance  $\mathcal{C}(x \rightarrow B)$ , which when plugged into (4.1) will imply the lower bound on  $E_v^\tau[H_x]$ .

Denote by  $B(x, r) = \{y \in \mathbb{H}_N, d(x, y) \leq r\}$  the ball of radius  $r$  around  $x$ , and by  $\partial B(x, r) = \{y \in \mathbb{H}_N, d(x, y) = r\}$  the sphere of radius  $r$ .

**Lemma 4.2** *For every  $\delta \in (0, 1/6)$ ,  $\mathbb{P}$ -a.s. for  $N$  large enough, there exist radii  $(\rho_x)_{x \in \mathbb{H}_N}$  satisfying  $1 \leq \rho_x \leq N^{3\delta}$ , such that for all  $x \in \mathbb{H}_N$  and for all  $y \in \partial B(x, \rho_x)$ ,  $\tau_y \leq 2^{\frac{1}{2}}N^{1-\delta}$ .*

*Proof* Fix  $\delta \in (0, 1/6)$ . We say that a sphere  $\partial B(x, r)$  is good if  $\tau_y \leq 2^{\frac{1}{2}}N^{1-\delta}$  for all  $y \in \partial B(x, r)$ , otherwise we say that it is bad. Using the Gaussian tail approximation (2.2), we get that

$$\mathbb{P} \left[ \tau_y > 2^{\frac{1}{2}}N^{1-\delta} \right] \leq ce^{-\frac{\log^2 2}{8\beta^2}N^{1-2\delta}}.$$

The size of the sphere  $\partial B(x, r)$  is bounded by  $N^r$ , hence the probability that the sphere  $\partial B(x, r)$  is bad is bounded by

$$N^r \mathbb{P} \left[ \tau_y > 2^{\frac{1}{2}}N^{1-\delta} \right] \leq c \exp \left\{ r \log N - \frac{\log^2 2}{8\beta^2}N^{1-2\delta} \right\}.$$

By independence of the  $\tau_x$ , the probability that for one fixed  $x$  all the spheres  $\partial B(x, r)$ ,  $r = 1, \dots, N^{3\delta}$ , are bad is bounded by

$$\begin{aligned} \prod_{r=1}^{N^{3\delta}} N^r \mathbb{P} \left[ \tau_y > 2^{\frac{1}{2}}N^{1-\delta} \right] &\leq \left( N^{N^{3\delta}} \mathbb{P} \left[ \tau_y > 2^{\frac{1}{2}}N^{1-\delta} \right] \right)^{N^{3\delta}} \\ &\leq \exp \left\{ N^{3\delta} \log c + N^{6\delta} \log N - \frac{\log^2 2}{8\beta}N^{1+\delta} \right\}. \end{aligned}$$

Finally, by a union bound, the probability that among all  $2^N$  vertices in  $\mathbb{H}_N$  there is one for which all spheres  $\partial B(x, r)$ ,  $r = 1, \dots, N^{3\delta}$ , are bad is bounded by

$$2^N \left( N^{N^{3\delta}} \mathbb{P} \left[ \tau_y > 2^{\frac{1}{2}} N^{1-\delta} \right] \right)^{N^{3\delta}} \leq \exp \left\{ N^{3\delta} \log c + N^{6\delta} \log N + N \log 2 - \frac{\log^2 2}{8\beta} N^{1+\delta} \right\}.$$

Since  $\delta < 1/6$  this decays faster than exponentially, and so by the Borel–Cantelli lemma the event occurs  $\mathbb{P}$ -a.s. only for finitely many  $N$ , i.e.  $\mathbb{P}$ -a.s. for  $N$  large enough we can find for every  $x \in \mathbb{H}_N$  a radius  $\rho_x \leq N^{3\delta}$  such that the sphere  $\partial B(x, \rho_x)$  is good.  $\square$

For every  $x \in \mathcal{D}_N$  we define the set  $A_x$  by

$$A_x := \begin{cases} B(x, \rho_x), & \text{if the radius } \rho_x \text{ from Lemma 4.2 exists,} \\ \{x\}, & \text{otherwise.} \end{cases} \tag{4.2}$$

We may now proceed with the proof of the lower bound in Proposition 4.1.

*Proof of the lower bound of Proposition 4.1* We will apply (4.1) with  $B = A_x^c$ . Using Lemma 4.2 and (2.3),  $\mathbb{P}$ -a.s. for  $N$  large enough, for all  $x \in \mathcal{D}_N$  all conductances  $c_{yz} = (\tau_y \wedge \tau_z) / Z_N$  connecting  $A_x$  and  $A_x^c$  are smaller than  $2^{\frac{1}{2}N^{1-\delta}} / (\kappa 2^N)$ . By the parallel law (cf. [30, Chapter 2.3]), the effective conductance between the boundaries of  $A_x$  and  $A_x^c$  is equal to the sum of all the conductances of edges connecting  $A_x$  and  $A_x^c$ , and so  $\mathbb{P}$ -a.s. for  $N$  large enough,

$$\mathcal{C}(\partial A_x \rightarrow \partial A_x^c) = \sum_{\substack{y \in \partial A_x \\ z \in \partial A_x^c}} c_{yz} \leq \kappa^{-1} N^{\rho_x+1} 2^{\frac{1}{2}N^{1-\delta}} 2^{-N}.$$

By Rayleigh’s monotonicity principle (cf. [30, Chapter 2.4]), comparing the effective conductances from  $x$  to  $A_x^c$  before and after setting all the conductances inside  $A_x$  to infinity, it follows that

$$\mathcal{C}(x \rightarrow A_x^c) \leq \mathcal{C}(\partial A_x \rightarrow \partial A_x^c) \leq \kappa^{-1} N^{\rho_x+1} 2^{\frac{1}{2}N^{1-\delta}} 2^{-N}.$$

Since  $\delta < 1/6$  and  $\rho_x \leq N^{3\delta}$ , we have  $N^{\rho_x+1} \leq 2^{\frac{1}{2}N^{1-\delta}}$  for  $N$  large enough, and thus,  $\mathbb{P}$ -a.s. for  $N$  large enough,

$$\mathcal{C}(x \rightarrow A_x^c) \leq c 2^{-N+N^{1-\delta}}. \tag{4.3}$$

Moreover,  $\mathbb{P}$ -a.s. for  $N$  large enough, as  $v_y = (1 \wedge \tau_y) / Z_N \leq 1 / Z_N$ , using (2.3) again,

$$v(A_x^c) = 1 - v(A_x) \geq 1 - Z_N^{-1} |A_x| \geq 1 - c 2^{-N} N^{N^{3\delta}} \xrightarrow{N \rightarrow \infty} 1. \tag{4.4}$$

Plugging (4.3) and (4.4) into (4.1) and readjusting  $\delta$  to accommodate for constants easily yields the required lower bound  $E_v^\tau[H_x] \geq 2^{N-N^{1-\delta}}$ . This completes the proof.  $\square$

As a corollary we get a lower bound on  $E_x^\tau[\ell_{H_{A_x^c}}(x)]$  for the deep traps  $x \in \mathcal{D}_N$ .

**Corollary 4.3** *There are constants  $\delta \in (0, 1/6)$  and  $c > 0$ , such that  $\mathbb{P}$ -a.s. for  $N$  large enough, for all  $x \in \mathcal{D}_N$ , under  $P_x^\tau$  the local time of  $Y$  in  $x$  before leaving  $A_x$ ,  $\ell_{H_{A_x^c}}(x)$ , stochastically dominates an exponential random variable with mean  $c2^{-N^{1-\delta}}$ . In particular,  $\mathbb{P}$ -a.s. for  $N$  large enough,*

$$E_x^\tau \left[ \ell_{H_{A_x^c}}(x) \right] \geq c2^{-N^{1-\delta}}.$$

*Proof* Recall that  $J_1$  denotes the time of the first jump of  $Y$ . The local time at  $x$  before hitting  $A_x^c$  is an exponential random variable with mean equal to

$$E_x^\tau [\#\{\text{visits to } x \text{ before } H_{A_x^c}\}] \cdot E_x^\tau [J_1].$$

The expected number of visits before leaving  $A_x$  is  $P_x^\tau[H_x^+ > H_{A_x^c}]^{-1} = c_x \mathcal{C}(x \rightarrow A_x^c)^{-1}$ . The mean duration of one visit to  $x$  is  $E_x^\tau[J_1] = (\sum_{y \sim x} q_{xy})^{-1}$ . For the deep traps we have  $\tau_x > 1$ , therefore  $\sum_{y \sim x} q_{xy} = \sum_{y \sim x} c_{xy}/\nu_x = Z_N c_x$ . It follows that the local time at  $x$  before hitting  $A_x^c$  is in fact an exponential random variable with mean  $Z_N^{-1} \mathcal{C}(x \rightarrow A_x^c)^{-1}$ . Using the bounds (4.3) and (2.3), the claim follows easily.  $\square$

As a next consequence we give bounds on the random scale  $R_N$  defined in (2.10). Note that this lemma also proves the statement (1.7) about the asymptotic behavior of  $R_N$  in Theorem 1.1.

**Lemma 4.4** *For every  $\varepsilon > 0$ ,  $\mathbb{P}$ -a.s. for  $N$  large enough,*

$$2^{(\gamma-\varepsilon)N} \leq R_N \leq 2^{(\gamma+\varepsilon)N}.$$

*Proof* By Proposition 3.3,  $T_{\text{mix}}/m_N$  is a geometric random variable with parameter  $e^{-1}$ , and thus  $E_x^\tau[\ell_{T_{\text{mix}}}(x)^\alpha] \leq E_x^\tau[T_{\text{mix}}^\alpha] \leq cm_N^\alpha \leq e^{\varepsilon N}$  by (3.7), for every  $\varepsilon > 0$  and  $N$  large enough. Moreover,  $|\mathcal{D}_N| \leq c'2^{(1-\gamma')N}$  by (2.8). Using the lower bound on  $E_v^\tau[H_x]$  from Proposition 4.1, we obtain that for every  $\varepsilon > 0$ ,  $\mathbb{P}$ -a.s. for  $N$  large enough,

$$R_N = 2^{(\gamma-\gamma')N} \left( \sum_{x \in \mathcal{D}_N} \frac{E_x^\tau[\ell_{T_{\text{mix}}}(x)^\alpha]}{E_v^\tau[H_x]} \right)^{-1} \geq 2^{(\gamma-\varepsilon)N}.$$

For the upper bound we need a lower bound on  $E_x^\tau[\ell_{T_{\text{mix}}}(x)^\alpha]$ . Recall the sets  $A_x$  defined in (4.2), and note that

$$E_x^\tau[\ell_{T_{\text{mix}}}(x)^\alpha] \geq E_x^\tau \left[ \mathbf{1}_{\{T_{\text{mix}} \geq H_{A_x^c}\}} \ell_{H_{A_x^c}}(x)^\alpha \right]. \tag{4.5}$$

By Corollary 4.3,  $\mathbb{P}$ -a.s. for  $N$  large enough, the local time at  $x$  before hitting  $A_x^c$  stochastically dominates an exponential random variable with mean  $c2^{-N^{1-\delta}}$ , hence

$$P_x^\tau \left[ \ell_{H_{A_x^c}}(x) \leq 2^{-2N^{1-\delta}} \right] \leq 1 - e^{-c2^{-N^{1-\delta}}} \leq c2^{-N^{1-\delta}}.$$

Moreover, for every  $\varepsilon > 0$ ,  $\mathbb{P}$ -a.s. for  $N$  large enough,

$$P_x^\tau [T_{\text{mix}} < H_{A_x^c}] \leq P_x^\tau [Y_{T_{\text{mix}}} \in A_x] = \nu(A_x) \leq \kappa^{-1} 2^{-N} N^{N^{3\delta}} \leq 2^{-\varepsilon N}.$$

Using the last two observations in (4.5),  $\mathbb{P}$ -a.s. for  $N$  large enough,

$$\begin{aligned} E_x^\tau [\ell_{T_{\text{mix}}}(x)^\alpha] &\geq P_x^\tau \left[ \{T_{\text{mix}} \geq H_{A_x^c}\} \cap \{\ell_{H_{A_x^c}}(x) \geq 2^{-2N^{1-\delta}}\} \right] \left(2^{-2N^{1-\delta}}\right)^\alpha \\ &\geq 2^{-2\alpha N^{1-\delta}} \left( P_x^\tau \left[ \ell_{H_{A_x^c}}(x) \geq 2^{-2N^{1-\delta}} \right] \right. \\ &\quad \left. - P_x^\tau \left[ \{\ell_{H_{A_x^c}}(x) \geq 2^{-2N^{1-\delta}}\} \cap \{T_{\text{mix}} < H_{A_x^c}\} \right] \right) \\ &\geq 2^{-2\alpha N^{1-\delta}} \left( P_x^\tau \left[ \ell_{H_{A_x^c}}(x) \geq 2^{-2N^{1-\delta}} \right] - P_x^\tau [T_{\text{mix}} < H_{A_x^c}] \right) \\ &\geq 2^{-2\alpha N^{1-\delta}} \left( (1 - c'2^{-N^{1-\delta}}) - 2^{-\varepsilon N} \right) \\ &\geq 2^{-\varepsilon N}. \end{aligned}$$

Combining this with  $|\mathcal{D}_N| \geq c2^{(1-\gamma')N}$  by (2.8) and the upper bound on  $E_v^\tau [H_x]$  from Proposition 4.1, we obtain the required upper bound on  $R_N$ .  $\square$

### 5 Concentration of the local time functional

In this section we prove the concentration of the local time functional that appears in the computation of the quasi-annealed Laplace transform of the clock process on the deep traps, as explained in the introduction (cf. (1.9)). We denote this functional by

$$L_N(t) = 2^{(\gamma'-\gamma)N} \sum_{x \in \mathcal{D}_N} \ell_{tR_N}(x)^\alpha.$$

So far we had no restriction on the choice of  $\gamma'$  other than  $1/2 < \gamma' < \gamma$ , see (2.6). We now make an explicit choice as follows. Let  $\varepsilon_0 = \frac{1}{2} \left( (1-\gamma) \wedge (\gamma - \frac{1}{2}) \right)$ , and define  $\gamma' = \gamma - \varepsilon_0$ , such that in particular

$$1 - \gamma \geq 2\varepsilon_0, \tag{5.1}$$

$$\gamma - \gamma' = \varepsilon_0. \tag{5.2}$$

The main result of this section is the following proposition.

**Proposition 5.1** *For every fixed  $t \geq 0$ ,  $\mathbb{P}$ -a.s. for  $N$  large enough,*

$$P_v^\tau \left[ |L_N(t) - t| \geq 2^{-\frac{1}{5}\varepsilon_0 N} \right] \leq c 2^{-\frac{1}{10}\varepsilon_0 N}.$$

*Proof* We approximate  $L_N(t)$  by the sum of essentially independent random variables as follows. Let  $K = \lfloor 2^{\varepsilon_0 N} \rfloor$ . For a fixed  $t > 0$ , define

$$t_k = \frac{t R_N}{K} k, \quad k = 0, \dots, K.$$

Recall the notation (2.5). For every  $x \in \mathcal{D}_N$  and  $k = 1, \dots, K$ , define  $H_x^k = t_{k-1} + H_x \circ \theta_{t_{k-1}}$  to be the time of the first visit to  $x$  after  $t_{k-1}$ , and set

$$\ell_{t,x}^k = \left( \int_{H_x^k \wedge (t_k - 2N^2 m_N)}^{(H_x^k + N^2 m_N) \wedge (t_k - N^2 m_N)} \mathbf{1}_{\{Y_s = x\}} ds \right)^\alpha.$$

The random variable  $\ell_{t,x}^k$  gives ‘roughly’ the  $\alpha$ -th power of the time that  $Y$  spends in  $x$  between  $t_{k-1}$  and  $t_k - N^2 m_N$ , with some suitable truncations. Let further

$$U_N^k(t) = 2^{(\gamma' - \gamma)N} \sum_{x \in \mathcal{D}_N} \ell_{t,x}^k.$$

The next lemma, which we prove later, shows that the sum of the  $U_N^k(t)$ ’s is a good approximation for  $L_N(t)$ .

**Lemma 5.2** *For every  $t > 0$ ,  $\mathbb{P}$ -a.s. for  $N$  large enough,*

$$P_v^\tau \left[ L_N(t) \neq \sum_{k=1}^K U_N^k(t) \right] \leq c 2^{-\frac{1}{2}\varepsilon_0 N}.$$

With Lemma 5.2, the proof of the proposition reduces to understanding of the approximating sum  $\sum_{k=1}^K U_N^k(t)$ . We will compute its expectation and variance under  $P_v^\tau$ . In particular, we will show that there is  $c < \infty$  such that for every  $t > 0$ ,

$$\left| E_v^\tau \left[ \sum_{k=1}^K U_N^k(t) \right] - t \right| \leq c 2^{-2\varepsilon_0 N}, \quad \mathbb{P}\text{-a.s. as } N \rightarrow \infty, \tag{5.3}$$

and

$$\text{Var}_v^\tau \left( \sum_{k=1}^K U_N^k(t) \right) \leq c 2^{-\frac{1}{2}\varepsilon_0 N}, \quad \mathbb{P}\text{-a.s. as } N \rightarrow \infty. \tag{5.4}$$

The statement of the proposition then follows from Lemma 5.2, (5.3) and (5.4) by routine application of the Chebyshev inequality. Indeed,  $\mathbb{P}$ -a.s. for  $N$  large enough,

$$\begin{aligned} &P_v^\tau \left[ |L_N(t) - t| \geq 2^{-\frac{1}{5}\varepsilon_0 N} \right] \\ &\leq P_v^\tau \left[ L_N(t) \neq \sum_{k=1}^K U_N^k(t) \right] + P_v^\tau \left[ \left| \sum_{k=1}^K U_N^k(t) - E_v^\tau \left[ \sum_{k=1}^K U_N^k(t) \right] \right| \geq 2 \cdot 2^{-\frac{1}{5}\varepsilon_0 N} \right] \\ &\leq c2^{-\frac{1}{2}\varepsilon_0 N} + c'2^{-\frac{1}{10}\varepsilon_0 N} \leq c''2^{-\frac{1}{10}\varepsilon_0 N}, \end{aligned}$$

which is the claim of the proposition.

We proceed by computing the expectation (5.3). We will need two lemmas which we show later. The first lemma estimates the probability that a deep trap is visited by the process  $Y$ .

**Lemma 5.3** *For every  $t_N$  such that  $1 \leq t_N \leq 2^N$ , for every  $\varepsilon > 0$ ,  $\mathbb{P}$ -a.s. for  $N$  large enough, for all  $x \in \mathcal{D}_N$ ,*

$$P_v^\tau [H_x \leq t_N] = \frac{t_N}{E_v^\tau [H_x]} + O \left( t_N^2 2^{2(\varepsilon-1)N} \right) + O \left( 2^{(\varepsilon-1)N} \right) \leq ct_N 2^{(\varepsilon-1)N}.$$

The second lemma then gives the expected contribution of a single  $\ell_{t,x}^k$  to  $\sum_{k=1}^K U_N^k(t)$ .

**Lemma 5.4** *For every fixed  $t > 0$ ,  $k = 1, \dots, K$  and  $\varepsilon > 0$ ,  $\mathbb{P}$ -a.s. for  $N$  large enough, for all  $x \in \mathcal{D}_N$ ,*

$$E_v^\tau \left[ \ell_{t,x}^k \right] = \frac{tR_N}{K E_v^\tau [H_x]} E_x^\tau \left[ \ell_{T_{\text{mix}}}(x)^\alpha \right] + O \left( 2^{(2\gamma+3\varepsilon-2\varepsilon_0-2)N} \right).$$

With Lemma 5.4 it is easy to compute the expectation (5.3). Using that  $|\mathcal{D}_N| \leq c2^{(1-\gamma')N}$  by (2.8), and the definition (2.10) of  $R_N$ , for every  $\varepsilon > 0$ ,  $\mathbb{P}$ -a.s. for  $N$  large enough,

$$\begin{aligned} E_v^\tau \left[ \sum_{k=1}^K U_N^k(t) \right] &= 2^{(\gamma'-\gamma)N} \sum_{x \in \mathcal{D}_N} \sum_{k=1}^K \left( \frac{tR_N}{K E_v^\tau [H_x]} E_x^\tau \left[ \ell_{T_{\text{mix}}}(x)^\alpha \right] \right. \\ &\quad \left. + O \left( 2^{(2\gamma+3\varepsilon-2\varepsilon_0-2)N} \right) \right) \\ &= t + O \left( 2^{(\gamma'-\gamma)N} 2^{(1-\gamma')N} 2^{(2\gamma-2+3\varepsilon-\varepsilon_0)N} \right) \\ &= t + O \left( 2^{(\gamma-1+3\varepsilon-\varepsilon_0)N} \right). \end{aligned}$$

Choosing  $\varepsilon < \varepsilon_0/3$  and recalling (5.1) imply (5.3).

Next, we estimate the variance (5.4). Since  $\nu$  is the stationary measure for  $Y$ , the random variables  $U_N^k(t), k = 1, \dots, K$ , are identically distributed under  $P_\nu^\tau$ . Hence

$$\text{Var}_\nu^\tau \left( \sum_{k=1}^K U_N^k(t) \right) = K \text{Var}_\nu^\tau \left( U_N^1(t) \right) + 2 \sum_{1 \leq k < j \leq K} \text{Cov}_\nu^\tau \left( U_N^k(t), U_N^j(t) \right). \tag{5.5}$$

The covariances can be neglected easily. Indeed, since by definition  $U_N^k(t)$  depends on the trajectory of  $Y$  between times  $t_{k-1}$  and  $t_k - N^2 m_N$  only, we can use the Markov property at the later time to write

$$\begin{aligned} & \text{Cov}_\nu^\tau \left( U_N^k(t), U_N^j(t) \right) \\ &= E_\nu^\tau \left[ \left( U_N^k(t) - E_\nu^\tau U_N^k(t) \right) E^\tau \left[ U_N^j(t) - E_\nu^\tau U_N^j(t) \mid Y_{t_k - N^2 m_N} \right] \right]. \end{aligned} \tag{5.6}$$

By Lemma 3.4,  $|P^\tau[Y_{t_k} = y | Y_{t_k - N^2 m_N}] - \nu_y| \leq e^{-cN^2}$ . Using in addition that  $U_N^j \leq e^{c'N}$  for some sufficiently large  $c'$ , we see that the inner expectation satisfies

$$\left| E^\tau [U_N^j(t) - E_\nu^\tau U_N^j(t) \mid Y_{t_k - N^2 m_N}] \right| \leq e^{-cN^2/2}.$$

Inserting this inequality back to (5.6) and summing over  $k < j$  then implies that the second term in (5.5) is  $O(e^{-cN^2})$  and thus can be neglected when proving (5.4).

To control the variance of  $U_N^1(t)$  in (5.5), it is enough to bound its second moment, which is

$$E_\nu^\tau \left[ U_N^1(t)^2 \right] = 2^{2(\gamma' - \gamma)N} \left( \sum_{x \in \mathcal{D}_N} E_\nu^\tau \left[ (\ell_{t,x}^1)^2 \right] + \sum_{x \neq y \in \mathcal{D}_N} E_\nu^\tau [\ell_{t,x}^1 \ell_{t,y}^1] \right).$$

Since, by definition,  $\ell_{t,x}^1 \leq (N^2 m_N)^\alpha$  and  $\ell_{t,x}^1 \neq 0$  implies  $H_x \leq t R_N / K$ ,

$$\begin{aligned} E_\nu^\tau \left[ U_N^1(t)^2 \right] &\leq 2^{2(\gamma' - \gamma)N} N^{4\alpha} m_N^{2\alpha} \left( \sum_{x \in \mathcal{D}_N} P_\nu^\tau \left[ H_x \leq \frac{t R_N}{K} \right] \right. \\ &\quad \left. + \sum_{x \neq y \in \mathcal{D}_N} P_\nu^\tau \left[ H_x, H_y \leq \frac{t R_N}{K} \right] \right). \end{aligned} \tag{5.7}$$

By Lemmas 5.3 and 4.4, for every  $\varepsilon > 0$ ,  $\mathbb{P}$ -a.s. as  $N \rightarrow \infty$ ,

$$P_\nu^\tau \left[ H_x \leq \frac{t R_N}{K} \right] \leq c 2^{(\gamma - 1 + \varepsilon - \varepsilon_0)N}. \tag{5.8}$$



Moreover, by (2.8),  $|\mathcal{D}_N| \leq c2^{(1-\gamma')N}$ , and by (3.7),  $N^{4\alpha}m_N^{2\alpha} \leq 2^{\varepsilon N}$ , for every  $\varepsilon > 0$  and  $N$  large enough. It follows that the contribution of the first sum in (5.7) to the variance, including the prefactor  $K = 2^{\varepsilon_0 N}$  from (5.5), can be bounded by

$$c2^{(2(\gamma'-\gamma)+1-\gamma'+\gamma-1+2\varepsilon)N} = c2^{(\gamma'-\gamma+2\varepsilon)N}.$$

By (5.2),  $\gamma' - \gamma + 2\varepsilon \leq -\varepsilon_0 + 2\varepsilon < -\frac{1}{2}\varepsilon_0$  for  $\varepsilon < \varepsilon_0/4$ , and hence this contribution is smaller than  $c2^{-\frac{1}{2}\varepsilon_0 N}$  as required for (5.4).

For the second summation in (5.7) we write

$$P_v^\tau \left[ H_x, H_y \leq \frac{tR_N}{K} \right] \leq P_v^\tau \left[ H_x < H_y \leq \frac{tR_N}{K} \right] + P_v^\tau \left[ H_y < H_x \leq \frac{tR_N}{K} \right].$$

By the Markov property, each of these two probabilities can be bounded by

$$\begin{aligned} P_v^\tau \left[ H_x < H_y \leq \frac{tR_N}{K} \right] &= \int_0^{\frac{tR_N}{K}} P_v^\tau [H_x \in du] P_x^\tau \left[ H_y < \frac{tR_N}{K} - u \right] \\ &\leq \int_0^{\frac{tR_N}{K}} P_v^\tau [H_x \in du] \left( P_x^\tau [H_y \leq T_{\text{mix}}] + P_v^\tau \left[ H_y \leq \frac{tR_N}{K} \right] \right) \\ &\leq P_v^\tau \left[ H_x \leq \frac{tR_N}{K} \right] \left( P_x^\tau [H_y \leq T_{\text{mix}}] + P_v^\tau \left[ H_y \leq \frac{tR_N}{K} \right] \right). \end{aligned}$$

Using (5.8) and (2.8) again, the second sum in (5.7) is bounded by

$$c2^{(\gamma-1+\varepsilon-\varepsilon_0)N} \left( 2^{2(1-\gamma')N} 2^{(\gamma-1+\varepsilon-\varepsilon_0)N} + \sum_{x \neq y \in \mathcal{D}_N} P_x^\tau [H_y \leq T_{\text{mix}}] \right). \tag{5.9}$$

The first term in the parentheses of (5.9) together with the prefactors  $K$  from (5.5) and  $2^{2(\gamma'-\gamma)N} N^{4\alpha} m_N^{2\alpha} \leq 2^{(2(\gamma'-\gamma)+\varepsilon)N}$  from (5.7), contributes to the variance by at most

$$c2^{(\varepsilon_0+2(\gamma'-\gamma)+\varepsilon+2(1-\gamma')+2(\gamma-1+\varepsilon-\varepsilon_0))N} = c2^{(3\varepsilon-\varepsilon_0)N} \leq c2^{-\frac{1}{2}\varepsilon_0 N}$$

if  $\varepsilon$  is small enough, as required by (5.4).

For the second term in the parentheses of (5.9) we need the following lemma whose proof is again postponed.

**Lemma 5.5** *Let  $\mathcal{W}_t^x = \sum_{y \in \mathcal{D}_N, y \neq x} \mathbf{1}_{\{H_y \leq t\}}$ . Then for every  $\varepsilon > 0$ ,  $\mathbb{P}$ -a.s. for  $N$  large enough, for every  $x \in \mathcal{D}_N$ ,*

$$E_x^\tau [\mathcal{W}_{T_{\text{mix}}}^x] \leq 2^{\varepsilon N}.$$

Using Lemma 5.5, and including all the prefactors as before, the contribution of the second term in (5.9) to the variance (5.5) is bounded by

$$c2^{(\varepsilon_0+2(\gamma'-\gamma)+\varepsilon+1-\gamma'+\gamma-1+\varepsilon-\varepsilon_0+\varepsilon)N} = c2^{(\gamma'-\gamma+3\varepsilon)N} \leq 2^{-\frac{1}{2}\varepsilon_0 N},$$

where for the last inequality we used (5.2) again, and choose  $\varepsilon$  small enough. This completes the proof of (5.4) and thus of the proposition.  $\square$

We proceed by proving the lemmas used in the above proof.

*Proof of Lemma 5.3* By [1, Theorem 1] the hitting time  $H_x$  is approximately exponential in the sense that

$$\left| P_v^\tau[H_x > t] - e^{-\frac{t}{E_v^\tau[H_x]}} \right| \leq \frac{1}{\lambda_Y E_v^\tau[H_x]}.$$

Hence, using Propositions 3.1 and 4.1 to bound  $\lambda_Y$  and  $E_v^\tau[H_x]$  respectively, we have for every  $\varepsilon > 0$ ,  $\mathbb{P}$ -a.s. for  $N$  large enough,

$$\begin{aligned} P_v^\tau[H_x \leq t_N] &= \left( 1 - e^{-\frac{t_N}{E_v^\tau[H_x]}} \right) + O\left(2^{(\varepsilon-1)N}\right) \\ &= \frac{t_N}{E_v^\tau[H_x]} + O\left(t_N^2 2^{2(\varepsilon-1)N}\right) + O\left(2^{(\varepsilon-1)N}\right). \end{aligned}$$

Finally, if  $1 \leq t_N \leq 2^N$  this is bounded by  $ct_N 2^{(\varepsilon-1)N}$ , which proves the lemma.  $\square$

*Proof of Lemma 5.4* By the strong Markov property and the definition of  $\ell_{t,x}^k$ ,

$$\begin{aligned} E_v^\tau \left[ \ell_{t,x}^k \right] &\geq P_v^\tau \left[ H_x \in [t_{k-1}, t_k - 2N^2 m_N] \right] E_x^\tau \left[ \ell_{N^2 m_N}(x)^\alpha \right], \\ E_v^\tau \left[ \ell_{t,x}^k \right] &\leq P_v^\tau \left[ H_x \in [t_{k-1}, t_k - N^2 m_N] \right] E_x^\tau \left[ \ell_{N^2 m_N}(x)^\alpha \right]. \end{aligned} \tag{5.10}$$

We will now give approximations of the expressions appearing in (5.10).

Observe that for every  $s, t > 0$ ,

$$\ell_t(x)^\alpha \leq \ell_s(x)^\alpha + (\ell_t(x) - \ell_s(x))^\alpha.$$

Using this inequality with  $t = N^2 m_N$  and  $s = T_{\text{mix}}$  and applying the strong Markov property at  $T_{\text{mix}}$ , observing that  $Y_{T_{\text{mix}}}$  is  $\nu$ -distributed,

$$E_x^\tau \left[ \ell_{N^2 m_N}(x)^\alpha \right] \leq E_x^\tau \left[ \ell_{T_{\text{mix}}}(x)^\alpha \right] + E_v^\tau \left[ \ell_{N^2 m_N}(x)^\alpha \right].$$

By Lemma 5.3, using also that by (3.7),  $\ell_{N^2 m_N}(x)^\alpha \leq N^{2\alpha} m_N^\alpha \leq 2^{\varepsilon N}$  for every  $\varepsilon > 0$  and  $N$  large enough,

$$E_v^\tau \left[ \ell_{N^2 m_N}(x)^\alpha \right] \leq P_v^\tau \left[ H_x \leq N^2 m_N \right] 2^{\varepsilon N} \leq c 2^{(3\varepsilon-1)N}.$$

Hence we obtain the upper bound

$$E_x^\tau \left[ \ell_{N^2 m_N}(x)^\alpha \right] \leq E_x^\tau \left[ \ell_{T_{\text{mix}}}(x)^\alpha \right] + c 2^{(3\varepsilon-1)N}. \tag{5.11}$$

For a matching lower bound, note that

$$E_x^\tau [\ell_{N^2 m_N}(x)^\alpha] \geq E_x^\tau [\ell_{T_{\text{mix}}}(x)^\alpha \mathbf{1}_{\{T_{\text{mix}} \leq N^2 m_N\}}].$$

But from Proposition 3.3 it follows that

$$E_x^\tau [\ell_{T_{\text{mix}}}(x)^\alpha \mathbf{1}_{\{T_{\text{mix}} > N^2 m_N\}}] \leq E_x^\tau [T_{\text{mix}}^\alpha \mathbf{1}_{\{T_{\text{mix}} > N^2 m_N\}}] \leq \sum_{k=N^2}^\infty (km_N)^\alpha e^{-k} \leq ce^{-c'N^2},$$

so that

$$E_x^\tau [\ell_{N^2 m_N}(x)^\alpha] \geq E_x^\tau [\ell_{T_{\text{mix}}}(x)^\alpha] - ce^{-cN^2}. \tag{5.12}$$

Combining (5.11) and (5.12), we obtain

$$E_x^\tau [\ell_{N^2 m_N}(x)^\alpha] = E_x^\tau [\ell_{T_{\text{mix}}}(x)^\alpha] + O(2^{(3\epsilon-1)N}). \tag{5.13}$$

Note also that by (3.7), for every  $\epsilon > 0$  and  $N$  large enough,

$$E_x^\tau [\ell_{T_{\text{mix}}}(x)^\alpha] \leq E_x^\tau [T_{\text{mix}}^\alpha] \leq cm_N^\alpha \leq 2^{\epsilon N}. \tag{5.14}$$

To approximate the probabilities in (5.10), we apply Lemma 5.3 for  $t_N = t_{k-1}$  and  $t_N = t_k - iN^2 m_N$ , for a fixed  $t > 0$  and  $i = 1, 2$ . Using Lemma 4.4 to bound  $R_N$  and Proposition 4.1 to bound  $E_v^\tau [H_x]$ , for every  $\epsilon > 0$ ,  $\mathbb{P}$ -a.s. for  $N$  large enough, for both  $i = 1, 2$ ,

$$\begin{aligned} &P_v^\tau [H_x \in [t_{k-1}, t_k - iN^2 m_N]] \\ &= \frac{tR_N}{K E_v^\tau [H_x]} + O(2^{2(\gamma+\epsilon-\epsilon_0-1)N}) = O(2^{(\gamma+\epsilon-\epsilon_0-1)N}). \end{aligned} \tag{5.15}$$

Inserting both (5.15) and (5.13) in (5.10), and using (5.14), for every  $\epsilon > 0$ ,  $\mathbb{P}$ -a.s. for  $N$  large enough,

$$E_v^\tau [\ell_{t,x}^k] = \frac{tR_N}{K E_v^\tau [H_x]} E_x^\tau [\ell_{T_{\text{mix}}}(x)^\alpha] + O(2^{(2\gamma+3\epsilon-2\epsilon_0-2)N}).$$

This proves the lemma. □

*Proof of Lemma 5.2* Note first that

$$\left\{ L_N(t) \neq \sum_{k=1}^K U_N^k(t) \right\} \subseteq \left\{ \exists x \in \mathcal{D}_N : \ell_{tR_N}(x)^\alpha \neq \sum_{k=1}^K \ell_{t,x}^k \right\}.$$

To control the probability of this event, we introduce some more notation. Set  $H_x^{(0)} = 0$ ,  $H_x^{(1)} = H_x$ , and for  $k \geq 2$  define the time of the ‘ $k$ -th visit after mixing’ inductively

as

$$H_x^{(k)} = \inf\{t > T_{\text{mix}} \circ \theta_{H_x^{(k-1)}} + H_x^{(k-1)} : Y_t = x\}.$$

Let  $\mathcal{N}_t^x = \min\{k \geq 0, H_x^{(k)} \leq t\}$  be the number of ‘visits after mixing’ to  $x$  before time  $t$ . Finally, let  $I_k = [t_k - 2N^2m_N, t_k]$ . Then

$$\begin{aligned} P_\nu^\tau \left[ \exists x \in \mathcal{D}_N : \ell_{tR_N}(x)^\alpha \neq \sum_{k=1}^K \ell_{t,x}^k \right] &\leq P_\nu^\tau \left[ Y_s \in \mathcal{D}_N \text{ for some } s \in \bigcup_{k=1}^K I_k \right] \\ &\quad + P_\nu^\tau \left[ \exists x \in \mathcal{D}_N : \mathcal{N}_{tR_N}^x \geq 2 \right] \\ &\quad + P_\nu^\tau \left[ \exists x \in \mathcal{D}_N : T_{\text{mix}} \circ \theta_{H_x} > N^2m_N \right]. \end{aligned} \tag{5.16}$$

We show that each of the three terms on the right-hand side is smaller than  $c2^{-\frac{1}{2}\varepsilon_0N}$ , which will prove the lemma.

For the first term in (5.16), using the stationarity of  $\nu$  and the Markov property

$$P_\nu^\tau \left[ Y_s \in \mathcal{D}_N \text{ for some } s \in \bigcup_{k=1}^K I_k \right] \leq K \sum_{x \in \mathcal{D}_N} P_\nu^\tau \left[ H_x \leq 2N^2m_N \right]. \tag{5.17}$$

By Lemma 5.3,  $\mathbb{P}$ -a.s. for all  $x \in \mathcal{D}_N$ , for  $\varepsilon > 0$  small and  $N$  large enough,

$$P_\nu^\tau [H_x \leq 2N^2m_N] \leq 2^{(\varepsilon-1)N}.$$

Since  $|\mathcal{D}_N| \leq c2^{(1-\gamma')N}$  by (2.8), the right hand side of (5.17) is bounded by  $c2^{\varepsilon_0N}2^{(\varepsilon-\gamma')N}$ . Since  $\gamma' > 1/2$  and by definition  $\varepsilon_0 \leq 1/4$ , when  $\varepsilon$  is small enough this is smaller than  $c2^{-\frac{1}{2}\varepsilon_0N}$  as required.

For the second term in (5.16), by Lemma 5.3 and the strong Markov property at  $T_{\text{mix}}$ , for every  $\varepsilon > 0$ ,  $\mathbb{P}$ -a.s. for  $N$  large enough,

$$P_\nu^\tau [H_x^{(2)} \leq tR_N] \leq P_\nu^\tau [H_x \leq tR_N]^2 \leq c2^{2(\gamma-1+\varepsilon)N}.$$

Together with (2.8) to bound  $|\mathcal{D}_N|$ , and using (5.1) and (5.2),  $\mathbb{P}$ -a.s. for  $N$  large enough,

$$\begin{aligned} P_\nu^\tau \left[ \exists x \in \mathcal{D}_N : \mathcal{N}_{tR_N}^x \geq 2 \right] &\leq c2^{(1-\gamma')N}2^{2(\gamma-1+\varepsilon)N} \\ &= c2^{(\gamma-\gamma')N+(\gamma-1)N+\varepsilon N} \\ &\leq c2^{(-\varepsilon_0+\varepsilon)N} \leq c2^{-\frac{1}{2}\varepsilon_0N} \end{aligned}$$

as required.

Finally we give a bound on the third term in (5.16). By Proposition 3.3,  $P_x^\tau [T_{\text{mix}} > N^2 m_N] \leq e^{-cN^2}$ . Thus, with (2.8) to bound  $|\mathcal{D}_N|$ ,  $\mathbb{P}$ -a.s. for  $N$  large enough,

$$P_v^\tau \left[ \exists x \in \mathcal{D}_N : T_{\text{mix}} \circ \theta_{H_x} > N^2 m_N \right] \leq c2^{(1-\gamma')N} P_x^\tau [T_{\text{mix}} > N^2 m_N] \leq c'2^{-\frac{1}{2}\varepsilon_0 N}.$$

Together with the previous estimates, this implies that the right-hand side of (5.16) is bounded by  $c2^{-\frac{1}{2}\varepsilon_0 N}$ , and concludes the proof of the lemma.  $\square$

*Proof of Lemma 5.5* Let  $\mathcal{H}_0 = 0$  and define recursively for  $i \geq 1$

$$\mathcal{H}_i = \inf \{t \geq \mathcal{H}_{i-1} : Y_t \in \mathcal{D}_N \setminus \{Y_{\mathcal{H}_{i-1}}\}\}.$$

By (2.9),  $\mathbb{P}$ -a.s. for  $N$  large enough, the vertices in  $\mathcal{D}_N$  are at least distance  $\delta N$  from each other. In particular the balls  $A_x = B(x, \rho_x)$ ,  $x \in \mathcal{D}_N$ , defined in (4.2) are disjoint. Hence, when on  $y \in \mathcal{D}_N$ , the random walk  $Y$  should first leave  $A_y$  in order to visit  $\mathcal{D}_N \setminus \{y\}$ . The strong Markov property and Corollary 4.3 then imply that  $\mathcal{H}_i$  stochastically dominates a Gamma random variable with parameters  $i$  and  $\mu := c2^{N^{1-\delta}}$ .

If  $\mathcal{W}_i^x \geq i$ , then  $\mathcal{H}_i \leq t$ . Hence, for  $t \geq \mu$ ,

$$E_x^\tau [\mathcal{W}_t^x] = \sum_{i \geq 1} P_x^\tau [\mathcal{W}_t^x \geq i] \leq \sum_{i \geq 1} P_x^\tau [\mathcal{H}_i \leq t] \leq \sum_{i \geq 1} \int_0^t \mu^i u^{i-1} e^{-\mu u} \Gamma(i)^{-1} du = \mu t.$$

Using the trivial estimate  $E_x^\tau [\mathcal{W}_{T_{\text{mix}}}^x \mathbf{1}_{\{T_{\text{mix}} \geq N^2 m_N\}}] \leq |\mathcal{D}_N| P_x^\tau [T_{\text{mix}} \geq N^2 m_N]$ , it follows that

$$E_x^\tau [\mathcal{W}_{T_{\text{mix}}}^x] \leq E_x^\tau [\mathcal{W}_{N^2 m_N}^x] + |\mathcal{D}_N| P_x^\tau [T_{\text{mix}} \geq N^2 m_N] \leq \mu N^2 m_N + c2^{(\gamma'-1)N} e^{-cN^2} \leq 2^{\varepsilon N}$$

by (2.8), (3.7) and Proposition 3.3. This completes the proof.  $\square$

For later applications, we state two further consequences of the proof of Lemma 5.2.

**Lemma 5.6**  $\mathbb{P}$ -a.s. for  $N$  large enough,

$$P_v^\tau \left[ \exists x \in \mathcal{D}_N : \ell_{tR_N}(x) > N^2 m_N \right] \leq c2^{-\frac{1}{2}\varepsilon_0 N},$$

and

$$P_v^\tau \left[ |\{x \in \mathcal{D}_N : H_x \leq tR_N\}| \geq 2^{\frac{3}{2}\varepsilon_0 N} \right] \leq c2^{-\frac{1}{4}\varepsilon_0 N}.$$

*Proof* The first claim follows directly from the bounds on the second and third term on the right hand side of (5.16) in the proof of Lemma 5.2, since the local time in a vertex that is only ‘visited once after mixing’ is bounded by  $T_{\text{mix}} \circ \theta_{H_x}$ .

The second assertion can be seen in the following way. Using Lemma 5.3 to bound the probability of a single vertex  $x \in \mathcal{D}_N$  to be visited before time  $tR_N$  and (2.8) to bound the size of  $\mathcal{D}_N$ , for every  $\varepsilon > 0$ ,  $\mathbb{P}$ -a.s. for  $N$  large enough,

$$E_v^\tau [|\{x \in \mathcal{D}_N : H_x \leq tR_N\}|] \leq c2^{(1-\gamma')N}2^{(\gamma-1+\varepsilon)N} \leq c2^{(\gamma-\gamma'+\varepsilon)N}.$$

By (5.2) this is equal to  $c2^{(\varepsilon_0+\varepsilon)N}$ , so choosing  $\varepsilon < \varepsilon_0/4$  this is smaller than  $c2^{\frac{5}{4}\varepsilon_0N}$ . Then by the Markov inequality the probability that there are more than  $2^{\frac{3}{2}\varepsilon_0N}$  vertices visited is smaller than  $c2^{-\frac{1}{4}\varepsilon_0N}$ .  $\square$

### 6 Clock process of the deep traps

This section contains the main steps leading to the proof of Theorem 1.1. Recall from (1.8) that the ‘clock process of deep traps’  $S_{\mathcal{D}}$  is given by

$$S_{\mathcal{D}}(t) = \int_0^t (1 \vee \tau_{Y_s})\mathbf{1}_{\{Y_s \in \mathcal{D}_N\}}ds = \int_0^t \tau_{Y_s}\mathbf{1}_{\{Y_s \in \mathcal{D}_N\}}ds.$$

We now show that  $S_{\mathcal{D}}$  converges to a stable process.

**Proposition 6.1** *Under the assumptions of Theorem 1.1, the rescaled clock processes of the deep traps  $g_N^{-1}S_{\mathcal{D}}(tR_N)$  converge in  $\mathbb{P}$ -probability as  $N \rightarrow \infty$ , in  $P_v^\tau$  distribution on the space  $D([0, T], \mathbb{R})$  equipped with the Skorokhod  $M_1$ -topology, to an  $\alpha$ -stable subordinator  $V_\alpha$ .*

The proof of Proposition 6.1 consists of three steps. In a first step, we show convergence in distribution of one-dimensional marginals by showing that the Laplace transform of one-dimensional marginals converges. This step contains, to some extent, the principal insight of this paper and is split in two parts: We first show the quasi-annealed convergence mentioned in the introduction, which is then strengthened to convergence in probability with respect to the environment. The second and third step of the proof of Proposition 6.1 are rather standard and deal with the joint convergence of increments and the tightness.

#### 6.1 Quasi-annealed convergence

We establish here the connection between the Laplace transform of the clock process of deep traps and the local time functional  $L_N$  studied in Sect. 5. The key observation is that the depths of the deep traps are in some sense independent of the fast chain  $Y$ , and can be thus averaged out easily.

To formalize this, we introduce a two-step procedure to sample the environment  $\tau$ . Let  $\xi = (\xi_x)_{x \in \mathbb{H}_N}$  be i.i.d. Bernoulli random variables such that, cf. (2.7),

$$\mathbb{P}[\xi_x = 1] = 1 - \mathbb{P}[\xi_k = 0] = \mathbb{P}[x \in \mathcal{D}_N] = 2^{-\gamma'N}(1 + o(1)).$$

Further, let  $\bar{E} = (\bar{E}_x)_{x \in \mathbb{H}_N}$  be i.i.d. standard Gaussian random variables conditioned to be larger than  $\frac{1}{\beta\sqrt{N}} \log g'_N$ , and  $\underline{E} = (\underline{E}_x)_{x \in \mathbb{H}_N}$  i.i.d. standard Gaussian random variables conditioned to be smaller than  $\frac{1}{\beta\sqrt{N}} \log g'_N$ . The collections  $\xi$ ,  $\bar{E}$  and  $\underline{E}$  are mutually independent. The Hamiltonian of the REM can be obtained by setting

$$E_x = \bar{E}_x \mathbf{1}_{\{\xi_x=1\}} + \underline{E}_x \mathbf{1}_{\{\xi_x=0\}}. \tag{6.1}$$

From now on, we always assume that  $E_x$  are given by (6.1). Observe that in this procedure the set  $\mathcal{D}_N$  coincides with the set  $\{x \in \mathbb{H}_N : \xi_x = 1\}$ .

We use  $\mathcal{G} = \sigma(\xi, \underline{E})$  to denote the  $\sigma$ -algebra generated by the  $\xi$ 's and  $\underline{E}$ 's. In particular, the number and positions of deep traps and all the  $\tau_y$ ,  $y \notin \mathcal{D}_N$ , are  $\mathcal{G}$ -measurable. The depths of deep traps are however independent of  $\mathcal{G}$ .

In the next lemma we compute the quasi-annealed Laplace transform of  $S_{\mathcal{D}}$ . The term 'quasi-annealed' refers to the fact that we average over the energies of the deep traps  $\bar{E}_x$  (and over the law of the process), but we keep quenched the positions of the deep traps  $\xi_x$  and the energies of remaining traps  $\underline{E}_x$ .

**Lemma 6.2** *There is a constant  $\mathcal{K} \in (0, \infty)$  such that for every  $\lambda > 0$  and  $t \geq 0$ ,*

$$\mathbb{E} \left[ E_v^\tau \left[ e^{-\frac{\lambda}{gN} S_{\mathcal{D}}(tR_N)} \right] \middle| \mathcal{G} \right] \xrightarrow{N \rightarrow \infty} e^{-\mathcal{K}\lambda^\alpha t}, \quad \mathbb{P}\text{-a.s.}$$

*Proof* Recall the separation event  $\mathcal{S}$  defined in (2.9). This event depends only on  $\xi$  and is therefore  $\mathcal{G}$ -measurable, and by Lemma 2.1 it occurs  $\mathbb{P}$ -a.s. for  $N$  large enough. On  $\mathcal{S}$ , no deep traps  $x \in \mathcal{D}_N$  are neighbors. Since moreover  $\tau_x \geq 1$  for  $x \in \mathcal{D}_N$ , all the transition rates

$$q_{xy} \mathbf{1}_{\mathcal{S}} = \frac{\tau_x \wedge \tau_y}{1 \wedge \tau_x} \mathbf{1}_{\mathcal{S}}, \quad x, y \in \mathbb{H}_N,$$

are  $\mathcal{G}$ -measurable. That is, on the event  $\mathcal{S}$ , the law of the chain  $Y$  is in fact  $\mathcal{G}$ -measurable. Therefore, on  $\mathcal{S}$ , the order of taking expectations over the depth of the deep traps and the chain  $Y$  can be exchanged. Namely, denoting by  $\bar{\mathbb{E}}$  the expectation over the random variables  $\bar{E}_x$ , on  $\mathcal{S}$ ,

$$\begin{aligned} \mathbb{E} \left[ E_v^\tau \left[ e^{-\frac{\lambda}{gN} S_{\mathcal{D}}(tR_N)} \right] \middle| \mathcal{G} \right] &= E_v^\tau \left[ \bar{\mathbb{E}} \left[ e^{-\frac{\lambda}{gN} S_{\mathcal{D}}(tR_N)} \right] \right] \\ &= E_v^\tau \left[ \bar{\mathbb{E}} \left[ \exp \left\{ -\frac{\lambda}{gN} \int_0^{tR_N} \tau_{Y_s} \mathbf{1}_{\{Y_s \in \mathcal{D}_N\}} ds \right\} \right] \right] \\ &= E_v^\tau \left[ \bar{\mathbb{E}} \left[ \exp \left\{ -\frac{\lambda}{gN} \sum_{x \in \mathcal{D}_N} \ell_{tR_N}(x) \tau_x \right\} \right] \right]. \end{aligned} \tag{6.2}$$

We next approximate the inner expectation on the right-hand side of (6.2). Since its argument is bounded by one, it will be sufficient to control it on an event of  $P_v^\tau$ -probability tending to 1 as  $N \rightarrow \infty$ . Define the event

$$\mathcal{A} = \left\{ \text{for all } x \in \mathcal{D}_N, \ell_{tR_N}(x) \leq N^2 m_N \right\} \cap \left\{ |L_N(t) - t| \leq 2^{-\frac{1}{5}\varepsilon_0 N} \right\}. \tag{6.3}$$

By Proposition 5.1 and Lemma 5.6,  $\mathbb{P}$ -a.s. for  $N$  large enough,  $P_v^\tau[\mathcal{A}^c] \leq e^{-cN}$ .

When performing the inner expectation of (6.2), the local times  $\ell_{tR_N}(x)$  of  $Y$  as well as  $\mathcal{D}_N$  are fixed, the expectation is taken only over the energies of the deep traps. By independence of the  $\bar{E}_x$  it follows that

$$\begin{aligned} \bar{\mathbb{E}} \left[ e^{-\frac{\lambda}{g_N} S_{\mathcal{D}(tR_N)}} \right] &= \prod_{x \in \mathcal{D}_N} \bar{\mathbb{E}} \left[ \exp \left\{ -\frac{\lambda}{g_N} \ell_{tR_N}(x) e^{\beta\sqrt{N} \bar{E}_x} \right\} \right] \\ &= \exp \left\{ \sum_{x \in \mathcal{D}_N} \log \bar{\mathbb{E}} \left[ \exp \left\{ -\frac{\lambda}{g_N} \ell_{tR_N}(x) e^{\beta\sqrt{N} \bar{E}_x} \right\} \right] \right\}. \end{aligned} \tag{6.4}$$

For  $u \in [0, N^2 m_N]$ , let

$$\vartheta(u) = 1 - \bar{\mathbb{E}} \left[ \exp \left\{ -\frac{\lambda}{g_N} u e^{\beta\sqrt{N} \bar{E}_x} \right\} \right].$$

Since  $(\bar{E}_x)$  has standard Gaussian distribution conditioned on being larger than  $\frac{1}{\beta\sqrt{N}} \log g'_N$ , using that by (2.7),

$$\mathbb{P} \left[ E_x > \frac{1}{\beta\sqrt{N}} \log g'_N \right] = \mathbb{P}[x \in \mathcal{D}_N] = 2^{-\gamma'N} (1 + o(1)),$$

it follows that

$$\vartheta(u) = \frac{2^{\gamma'N}}{\sqrt{2\pi}} (1 + o(1)) \int_{\frac{1}{\beta\sqrt{N}} \log g'_N}^{\infty} e^{-\frac{s^2}{2}} \left( 1 - e^{-\frac{\lambda u}{g_N} e^{\beta\sqrt{N}s}} \right) ds.$$

We use the substitution  $s = \frac{1}{\beta\sqrt{N}} (\beta z + \log g_N - \log \lambda - \log u)$ . The lower limit of the integral then becomes

$$\frac{1}{\beta} (\log g'_N - \log g_N + \log \lambda + \log u) =: \omega(N).$$



For  $u \leq N^2 m_N$ ,  $\omega_N$  is asymptotically dominated by  $\log g'_N - \log g_N \leq -cN$ , and thus  $\lim_{N \rightarrow \infty} \omega(N) = -\infty$ . After the substitution,

$$\vartheta(u) = \frac{2\gamma'N}{\sqrt{2\pi}}(1 + o(1)) \int_{\omega(N)}^\infty e^{-\frac{1}{2\beta^2N}(\beta z + \log g_N - \log \lambda - \log u)^2} (1 - e^{-e^{\beta z}}) \frac{1}{\sqrt{N}} dz. \tag{6.5}$$

For  $u \in [0, N^2 m_N]$ , using the definition (1.6) of  $g_N$ , the exponent of the first exponential satisfies

$$\begin{aligned} &-\frac{1}{2\beta^2N}(\beta z + \log g_N - \log \lambda - \log u)^2 \\ &= -\frac{1}{2\beta^2N} \left( \beta z + \alpha\beta^2N - \frac{1}{\alpha} \log(\alpha\beta\sqrt{2\pi N}) - \log \lambda - \log u \right)^2 \\ &= -\frac{\alpha^2\beta^2}{2}N + \alpha \log \lambda + \alpha \log u + \log(\alpha\beta\sqrt{2\pi N}) - \alpha\beta z + \text{err}(z) + o(1). \end{aligned} \tag{6.6}$$

Here,  $o(1)$  is an error independent of the variable  $z$ . Note that for the  $\log^2 u$  part to be  $o(1)$  it is important that  $m_N$  defined in (3.7) is not too large, see also Remark 6.4. The second error term is

$$\text{err}(z) = -\frac{1}{2N}z^2 + \frac{1}{\beta N}z \left( \frac{1}{\alpha} \log(\alpha\beta\sqrt{2\pi N}) + \log \lambda + \log u \right).$$

Observe that  $\lim_{N \rightarrow \infty} \text{err}(z) = 0$  for every  $z \in \mathbb{R}$ , and that for every  $\varepsilon$  there is  $N_0$  large enough, so that for  $N \geq N_0$  and all  $z \in \mathbb{R}$

$$\text{err}(z) \leq \varepsilon|z|. \tag{6.7}$$

Inserting the results of the computation (6.6) back into (6.5), using that  $\alpha^2\beta^2/2 = \gamma \log 2$ , we obtain

$$\vartheta(u) = \alpha\beta 2^{(\gamma' - \gamma)N} \lambda^\alpha u^\alpha \int_{\omega(N)}^\infty e^{-\alpha\beta z + \text{err}(z)} (1 - e^{-e^{\beta z}}) dz (1 + o(1)). \tag{6.8}$$

We now claim that

$$\int_{\omega(N)}^\infty e^{-\alpha\beta z + \text{err}(z)} (1 - e^{-e^{\beta z}}) dz \xrightarrow{N \rightarrow \infty} \int_{\mathbb{R}} e^{-\alpha\beta z} (1 - e^{-e^{\beta z}}) dz =: C. \tag{6.9}$$

Indeed, the integrand converges point-wise on  $\mathbb{R}$  to  $e^{-\alpha\beta z} (1 - e^{-e^{\beta z}})$  which is integrable if  $\alpha < 1$ . Moreover, by (6.7), the integrand is bounded by  $e^{-\alpha\beta z + \varepsilon|z|} (1 - e^{-e^{\beta z}})$ , which is integrable if we choose  $\varepsilon < \beta(1 - \alpha) \wedge \alpha\beta$ . The claim (6.9) follows by the dominated convergence theorem.

We now come back to (6.4). Since on  $\mathcal{A}$ ,  $\ell_{IR_N}(x) \leq N^2 m_N$  for all  $x \in \mathcal{D}_N$ , and  $\gamma' < \gamma$ , we see that  $\vartheta(\ell_{IR_N}(x)) = o(1)$  uniformly in  $x \in \mathcal{D}_N$  on  $\mathcal{A}$ . With

$\log(1 - x) = -x(1 + O(x))$  as  $x \rightarrow 0$  this yields

$$\begin{aligned} \overline{\mathbb{E}} \left[ e^{-\frac{\lambda}{g_N} S_{\mathcal{D}}(tR_N)} \right] &= \exp \left\{ \sum_{x \in \mathcal{D}_N} \log(1 - \vartheta(\ell_{tR_N}(x))) \right\} \\ &= \exp \left\{ - \sum_{x \in \mathcal{D}_N} \vartheta(\ell_{tR_N}(x)) (1 + o(1)) \right\}. \end{aligned}$$

The inner sum can be easily computed from (6.8). Recalling that on  $\mathcal{A}$  the local time functional  $L_N(t)$  converges, denoting  $\mathcal{K} = \alpha\beta C$ , we obtain on  $\mathcal{A}$ ,

$$\begin{aligned} \sum_{x \in \mathcal{D}_N} \vartheta(\ell_{tR_N}(x)) &= \alpha\beta C \lambda^\alpha 2^{(\gamma' - \gamma)N} \sum_{x \in \mathcal{D}_N} \ell_{tR_N}(x)^\alpha (1 + o(1)) \\ &= \alpha\beta C \lambda^\alpha L_N(t) (1 + o(1)) \\ &= \mathcal{K} \lambda^\alpha t + o(1) \quad \text{as } N \rightarrow \infty. \end{aligned} \tag{6.10}$$

It follows that on  $\mathcal{A}$

$$\overline{\mathbb{E}} \left[ e^{-\frac{\lambda}{g_N} S_{\mathcal{D}}(tR_N)} \right] = e^{-\mathcal{K}t\lambda^\alpha(1+o(1))} = e^{-\mathcal{K}t\lambda^\alpha} + o(1) \quad \text{as } N \rightarrow \infty.$$

Inserting this into (6.2), using that  $P_v^\tau[\mathcal{A}^c] = O(e^{-cN})$ , we conclude that, on  $\mathcal{S}$ ,  $\mathbb{P}$ -a.s. as  $N \rightarrow \infty$ ,

$$\begin{aligned} \mathbb{E} \left[ E_v^\tau \left[ e^{-\frac{\lambda}{g_N} S_{\mathcal{D}}(tR_N)} \right] \middle| \mathcal{G} \right] &= E_v^\tau \left[ \overline{\mathbb{E}} \left[ e^{-\frac{\lambda}{g_N} S_{\mathcal{D}}(tR_N)} \right] \mathbf{1}_{\mathcal{A}} \right] + O(e^{-cN}) \\ &= e^{-\mathcal{K}t\lambda^\alpha} + o(1). \end{aligned}$$

Since  $\mathcal{S}$  occurs  $\mathbb{P}$ -a.s. for  $N$  large enough, this completes the proof. □

### 6.2 Quenched convergence

We strengthen the convergence in Lemma 6.2 in the following way.

**Lemma 6.3** *The one-dimensional marginals of the rescaled clock processes  $g_N^{-1} S_{\mathcal{D}}(tR_N)$  converge in  $\mathbb{P}$ -probability as  $N \rightarrow \infty$ , in  $P_v^\tau$ -distribution to an  $\alpha$ -stable law, that is for every  $t > 0$  and  $\lambda > 0$ ,*

$$E_v^\tau \left[ e^{-\frac{\lambda}{g_N} S_{\mathcal{D}}(tR_N)} \right] \xrightarrow{N \rightarrow \infty} e^{-\mathcal{K}\lambda^\alpha t} \quad \text{in } \mathbb{P}\text{-probability.}$$

*Proof* It will be enough to show that  $\mathbb{P}$ -a.s. for  $N$  large enough,

$$\mathbb{E} \left[ E_v^\tau \left[ e^{-\frac{\lambda}{g_N} S_{\mathcal{D}}(tR_N)} \right]^2 \middle| \mathcal{G} \right] = e^{-2K\lambda^\alpha t} + o(1). \tag{6.11}$$

Indeed, if (6.11) holds, then the conditional variance

$$\text{Var} \left[ E_v^\tau \left[ e^{-\frac{\lambda}{g_N} S_{\mathcal{D}}(tR_N)} \right] \middle| \mathcal{G} \right] \xrightarrow{N \rightarrow \infty} 0, \quad \mathbb{P}\text{-a.s.},$$

and the claim follows by an application of the Chebyshev inequality and Lemma 6.2.

To show (6.11), we rewrite

$$\mathbb{E} \left[ E_v^\tau \left[ e^{-\frac{\lambda}{g_N} S_{\mathcal{D}}(tR_N)} \right]^2 \middle| \mathcal{G} \right] = \mathbb{E} \left[ \hat{E}_v^\tau \left[ e^{-\frac{\lambda}{g_N} \sum_{x \in \mathcal{D}_N} (\ell_{tR_N}^{(1)}(x) + \ell_{tR_N}^{(2)}(x)) \tau_x} \right] \middle| \mathcal{G} \right],$$

where  $\ell^{(1)}$  and  $\ell^{(2)}$  are the local times of two independent Markov chains  $Y^{(1)}$  and  $Y^{(2)}$ , both having law  $P_v^\tau$ , and  $\hat{E}_v^\tau$  is the expectation with respect to the joint law  $\hat{P}_v^\tau$  of these chains. Again  $\mathbb{P}$ -a.s. for  $N$  large enough the separation event  $\mathcal{S}$  holds, and on this event the law  $\hat{P}_v^\tau$  is  $\mathcal{G}$ -measurable. Therefore we can exchange the expectations similarly as before. As in Lemma 6.2, it will be enough to control the expression on an event of  $\hat{P}_v^\tau$ -probability tending to 1 as  $N \rightarrow \infty$ . We thus set  $\hat{\mathcal{A}} = \mathcal{A}^{(1)} \cap \mathcal{A}^{(2)}$  where  $\mathcal{A}^{(i)}$  are defined for both chains  $Y^{(i)}$  as in (6.3). Applying Proposition 5.1 and Lemma 5.6 for both independent chains, we have that  $\mathbb{P}$ -a.s. as  $N \rightarrow \infty$ ,  $\hat{P}_v^\tau[\hat{\mathcal{A}}^c] = O(e^{-cN})$ .

Let  $\mathcal{C}$  be the event that  $Y^{(1)}$  and  $Y^{(2)}$  visit disjoint sets of deep traps,

$$\mathcal{C} = \left\{ \{x \in \mathcal{D}_N : \ell_{tR_N}^{(1)}(x) > 0\} \cap \{x \in \mathcal{D}_N : \ell_{tR_N}^{(2)}(x) > 0\} = \emptyset \right\}.$$

We claim that  $\hat{P}_v^\tau[\mathcal{C}^c] = O(e^{-cN})$ ,  $\mathbb{P}$ -a.s. as  $N \rightarrow \infty$ . Indeed, by Lemma 5.6, with probability larger than  $1 - c2^{-\frac{1}{4}\varepsilon_0 N}$ , the chain  $Y^{(1)}$  visits at most  $2^{\frac{3}{2}\varepsilon_0 N}$  different vertices in  $\mathcal{D}_N$ . By Lemma 5.3, each of those vertices has probability smaller than  $c2^{(\gamma-1+\varepsilon)N}$  of being hit by  $Y^{(2)}$ , for every  $\varepsilon > 0$ ,  $\mathbb{P}$ -a.s. for  $N$  large enough. Therefore by the choice (5.1) of  $\varepsilon_0$ ,  $\mathbb{P}$ -a.s. for  $N$  large enough,

$$\hat{P}_v^\tau[\mathcal{C}^c] \leq c2^{-\frac{1}{4}\varepsilon_0 N} + 2^{\frac{3}{2}\varepsilon_0 N} c'2^{(\gamma-1+\varepsilon)N} \leq c2^{-\frac{1}{4}\varepsilon_0 N} + c'2^{-\frac{1}{2}\varepsilon_0 N + \varepsilon N},$$

which decays exponentially if  $\varepsilon < \varepsilon_0/2$ .

Since on  $\mathcal{C}$  the  $\tau_x$  of the vertices  $x \in \mathcal{D}_N$  visited by  $Y^{(1)}$  and  $Y^{(2)}$  are independent, and since the integrand is bounded by 1, we have on the separation event  $\mathcal{S}$ ,

$$\begin{aligned} & \mathbb{E} \left[ E_v^\tau \left[ e^{-\frac{\lambda}{g_N} S_{\mathcal{D}}(tR_N)} \right]^2 \middle| \mathcal{G} \right] \\ &= \hat{E}_v^\tau \left[ \mathbb{E} \left[ e^{-\frac{\lambda}{g_N} \sum_{x \in \mathcal{D}_N} (\ell_{tR_N}^{(1)}(x) + \ell_{tR_N}^{(2)}(x)) \tau_x} \right] \right] \end{aligned}$$

$$\begin{aligned}
 &= \hat{E}_v^\tau \left[ \overline{\mathbb{E}} \left[ e^{-\frac{\lambda}{g_N} \sum_{x \in \mathcal{D}_N} (\ell_{tR_N}^{(1)}(x) + \ell_{tR_N}^{(2)}(x)) \tau_x} \right] \mathbf{1}_{\hat{\mathcal{A}} \cap \mathcal{C}} \right] + O(e^{-cN}) \\
 &= \hat{E}_v^\tau \left[ \overline{\mathbb{E}} \left[ e^{-\frac{\lambda}{g_N} \sum_{x \in \mathcal{D}_N} \ell_{tR_N}^{(1)}(x) \tau_x} \right] \overline{\mathbb{E}} \left[ e^{-\frac{\lambda}{g_N} \sum_{x \in \mathcal{D}_N} \ell_{tR_N}^{(2)}(x) \tau_x} \right] \mathbf{1}_{\hat{\mathcal{A}} \cap \mathcal{C}} \right] + O(e^{-cN}).
 \end{aligned}$$

Using the same procedure as in the proof of Lemma 6.2, on the event  $\hat{\mathcal{A}}$ , the two inner expectations,  $x \in \mathcal{D}_N$ , both converge to

$$\exp \left\{ -\mathcal{K} \lambda^\alpha 2^{(\gamma' - \gamma)N} \sum_{x \in \mathcal{D}_N} \ell_{tR_N}^{(i)}(x)^\alpha \right\} = \exp \left\{ -\mathcal{K} \lambda^\alpha L_N^{(i)}(t) \right\}, \quad i = 1, 2.$$

Moreover, on  $\hat{\mathcal{A}}$ , the local time functionals  $L_N^{(i)}(t)$  concentrate on  $t$  simultaneously. It follows that on  $\mathcal{S}$ ,  $\mathbb{P}$ -a.s. as  $N \rightarrow \infty$ ,

$$\mathbb{E} \left[ E_v^\tau \left[ e^{-\frac{\lambda}{g_N} S_{\mathcal{D}}(tR_N)} \right]^2 \middle| \mathcal{G} \right] = e^{-2\mathcal{K} \lambda^\alpha t} + o(1).$$

Noting again that  $\mathcal{S}$  occurs  $\mathbb{P}$ -a.s. for  $N$  large enough, this shows (6.11), and hence the lemma. □

*Remark 6.4* (a) Inspecting the last proof carefully, it follows that Lemma 6.3 can be slightly strengthened. Namely, the stated convergence holds a.s. with respect to  $\xi$  and  $\underline{E}$ , and in probability only with respect to  $\bar{E}$ . The same remark then applies to Theorem 1.1.

- (b) A closer analysis of the errors made in the computation of the quasi-annealed Laplace transform, in particular in (6.6), shows that the error in Lemma 6.2 and (6.11) is of order  $O(N^{-1} \log^2 N)$ , where the logarithmic part comes from the  $\log^2 u$  part in (6.6),  $u$  being bounded by  $N^2 m_N$ , and  $m_N$  being polynomial in  $N$ . Therefore the variance decay is not enough to apply the Borel–Cantelli lemma and obtain  $\mathbb{P}$ -a.s. convergence.
- (c) Note also that the previous proof, more precisely bounding the  $\log^2 u$  part of (6.6), requires that  $\log(N^2 m_N) \ll N^{1/2}$ . This is where our improved techniques to estimate the spectral gap in Proposition 3.1 are necessary. As we already remarked, the techniques of [23] show roughly that  $m_N \leq e^{\sqrt{N \log N}}$  only, which is not sufficient.

### 6.3 Joint convergence of increments

In the next step, we extend the convergence to joint convergence of increments.

**Lemma 6.5** *The increments of the rescaled clock processes  $g_N^{-1} S_{\mathcal{D}}(tR_N)$  converge jointly in  $\mathbb{P}$ -probability in  $P_v^\tau$ -distribution to the increments of an  $\alpha$ -stable subordinator.*

*Proof* Fix  $k \geq 1$  and  $0 = t_0 < t_1 < \dots < t_k$ . We will show that for every  $\lambda_1, \dots, \lambda_k \in (0, \infty)$  and  $\mathbb{P}$ -a.e. environment  $\tau$ ,

$$\lim_{N \rightarrow \infty} E_v^\tau \left[ e^{-\frac{1}{gN} \sum_{i=1}^k \lambda_i (S_{\mathcal{D}}(t_i R_N) - S_{\mathcal{D}}(t_{i-1} R_N))} \right] = \lim_{N \rightarrow \infty} \prod_{i=1}^k E_v^\tau \left[ e^{-\frac{\lambda_i}{gN} S_{\mathcal{D}}((t_i - t_{i-1}) R_N)} \right]. \tag{6.12}$$

Then the lemma follows by using the above proved convergence in  $\mathbb{P}$ -probability in  $P_v^\tau$ -distribution of the one-dimensional marginals.

Let  $I^i = [t_i R_N - N^2 m_N, t_i R_N]$ . For a set  $I \subset [0, \infty)$ , let  $\mathcal{V}(I)$  be the event

$$\mathcal{V}(I) = \{Y_s \notin \mathcal{D}_N \text{ for all } s \in I\}.$$

On the event  $\mathcal{V}(\cup_{i=1}^k I^i)$ , for every  $i \leq k$ ,

$$S_{\mathcal{D}}(t_i R_N) - S_{\mathcal{D}}(t_{i-1} R_N) = S_{\mathcal{D}}(t_i R_N - N^2 m_N) - S_{\mathcal{D}}(t_{i-1} R_N). \tag{6.13}$$

Moreover, by Lemma 5.3,  $\mathbb{P}$ -a.s. for all  $x \in \mathcal{D}_N$ , for  $\varepsilon > 0$  small and  $N$  large enough,

$$P_v^\tau [H_x \leq N^2 m_N] \leq 2^{(\varepsilon-1)N}.$$

By (2.8),  $|\mathcal{D}_N| \leq c2^{(1-\gamma')N}$ , hence the expected number of vertices  $x \in \mathcal{D}_N$  visited in a time-interval of length  $N^2 m_N$  is smaller than  $c2^{(\varepsilon-\gamma')N}$ ,  $\mathbb{P}$ -a.s. for  $N$  large enough. This still holds for a finite union of intervals of length  $N^2 m_N$ , and so we conclude that by the Markov inequality,  $P_v^\tau [\mathcal{V}(\cup_{i=1}^k I^i)] \rightarrow 1$ ,  $\mathbb{P}$ -a.s. as  $N \rightarrow \infty$ .

The reason to shorten the time intervals as above is to give the Markov chain  $Y$  the time it needs to mix. Define the event

$$\mathcal{M} = \{T_{\text{mix}} \circ \theta_{t_i R_N - N^2 m_N} \leq N^2 m_N \forall i = 1, \dots, k\}.$$

It is easy to see using Proposition 3.3 that  $P_v^\tau [\mathcal{M}] \rightarrow 1$ ,  $\mathbb{P}$ -a.s. as  $N \rightarrow \infty$ . On the event  $\mathcal{M}$  the Markov chain  $Y$  always mixes between  $t_i R_N - N^2 m_N$  and  $t_i R_N$  and thus, by Lemma 3.4, for every  $i = 1, \dots, k$  and  $y \in \mathbb{H}_N$ ,

$$P_v^\tau [Y_{t_i R_N} = y \mid \mathcal{M}] = \nu_y.$$

Therefore, on  $\mathcal{M}$ ,

$$\begin{aligned} & \left( S_{\mathcal{D}}(t_i R_N - N^2 m_N) - S_{\mathcal{D}}(t_{i-1} R_N) \right)_{i=1, \dots, k} \\ & \stackrel{d}{=} \left( S_{\mathcal{D}}^{(i)}((t_i - t_{i-1}) R_N - N^2 m_N) \right)_{i=1, \dots, k}, \end{aligned} \tag{6.14}$$

where the  $S_{\mathcal{D}}^{(i)}$  are the clock processes of the deep traps of independent stationary started processes  $Y^{(i)}$  having the same law as  $Y$ .

Combining observations (6.13) and (6.14), with the estimates on the probabilities of  $\mathcal{V}(\cup_{i=1}^k I^i)$  and  $\mathcal{M}$ , since the integrand is bounded by 1, we obtain that  $\mathbb{P}$ -a.s. as  $N \rightarrow \infty$ ,

$$\begin{aligned} & E_v^\tau \left[ e^{-\frac{1}{g_N} \sum_{i=1}^k \lambda_i (S_{\mathcal{D}}(t_i R_N) - S_{\mathcal{D}}(t_{i-1} R_N))} \right] \\ &= E_v^\tau \left[ e^{-\frac{1}{g_N} \sum_{i=1}^k \lambda_i (S_{\mathcal{D}}(t_i R_N - N^2 m_N) - S_{\mathcal{D}}(t_{i-1} R_N))} \mathbf{1}_{\mathcal{V}(\cup_{i=1}^k I^i) \cap \mathcal{M}} \right] + o(1) \\ &= E_v^\tau \left[ \prod_{i=1}^k E_v^\tau \left[ e^{-\frac{\lambda_i}{g_N} S_{\mathcal{D}}^{(i)}((t_i - t_{i-1}) R_N - N^2 m_N)} \right] \mathbf{1}_{\mathcal{V}(\cup_{i=1}^k I^i) \cap \mathcal{M}} \right] + o(1) \\ &= \prod_{i=1}^k E_v^\tau \left[ e^{-\frac{\lambda_i}{g_N} S_{\mathcal{D}}^{(i)}((t_i - t_{i-1}) R_N - N^2 m_N)} \right] + o(1). \end{aligned}$$

Using analogous arguments it can be shown that for every  $i = 1, \dots, k$ ,  $\mathbb{P}$ -a.s. as  $N \rightarrow \infty$ ,

$$E_v^\tau \left[ e^{-\frac{\lambda_i}{g_N} S_{\mathcal{D}}^{(i)}((t_i - t_{i-1}) R_N - N^2 m_N)} \right] = E_v^\tau \left[ e^{-\frac{\lambda_i}{g_N} S_{\mathcal{D}}^{(i)}((t_i - t_{i-1}) R_N)} \right] + o(1).$$

Combining the last two equations proves (6.12) and hence the lemma. □

### 6.4 Tightness in the Skorokhod topology

The last step in the proof of Proposition 6.1 is to show tightness.

**Lemma 6.6** *The sequence of probability measures  $P_N^\tau [g_N^{-1} S_{\mathcal{D}}(t R_N) \in \cdot]$  is  $\mathbb{P}$ -a.s. tight with respect to the Skorokhod  $M_1$ -topology on  $D([0, T], \mathbb{R})$ .*

*Proof* The proof is standard but we include it for the sake of completeness. By [33, Theorem 12.12.3], the tightness in the Skorokhod  $M_1$ -topology on  $D([0, T], \mathbb{R})$  is characterized in the following way: For  $f \in D([0, T], \mathbb{R})$ ,  $\delta > 0$ ,  $t \in [0, T]$ , let

$$\begin{aligned} w_f(\delta) &= \sup \left\{ \inf_{\alpha \in [0, 1]} |f(t) - (\alpha f(t_1) + (1 - \alpha) f(t_2))| : t_1 \leq t \leq t_2 \leq T, t_2 - t_1 \leq \delta \right\}, \\ v_f(t, \delta) &= \sup \{|f(t_1) - f(t_2)| : t_1, t_2 \in [0, T] \cap (t - \delta, t + \delta)\}. \end{aligned}$$

The sequence of probability measures  $P_N = P_N^\tau [g_N^{-1} S_{\mathcal{D}}(t R_N) \in \cdot]$  on  $D([0, T], \mathbb{R})$  is tight in the  $M_1$ -topology, if

- (i) For every  $\varepsilon > 0$  there is  $c$  such that

$$P_N[f : \|f\|_\infty > c] \leq \varepsilon, \quad N \geq 1. \tag{6.15}$$

(ii) For every  $\varepsilon > 0$  and  $\eta > 0$ , there exist  $\delta \in (0, T)$  and  $N_0$  such that

$$P_N[f : w_f(\delta) \geq \eta] \leq \varepsilon, \quad N \geq N_0, \tag{6.16}$$

and

$$P_N[f : v_f(0, \delta) \geq \eta] \leq \varepsilon \quad \text{and} \quad P_N[f : v_f(T, \delta) \geq \eta] \leq \varepsilon, \quad N \geq N_0. \tag{6.17}$$

Since the clock processes are increasing, (6.15) is equivalent to convergence of the distribution of  $g_N^{-1} S_{\mathcal{D}}(T R_N)$ , which follows from the convergence of the Laplace transform of the marginal at time  $T$ . (6.16) is immediate from the fact that the oscillating function  $w_f(\delta)$  is always zero since the processes  $g_N^{-1} S_{\mathcal{D}}(t R_N)$  are increasing. To check (6.17), again by the monotonicity of the  $g_N^{-1} S_{\mathcal{D}}(t R_N)$  it is enough to check that for  $\delta$  small enough and  $N \geq N_0$ ,  $P_v^\tau [g_N^{-1} S_{\mathcal{D}}(\delta R_N) \geq \eta] \leq \varepsilon$ . By the convergence of the marginal at time  $\delta$ , we may take  $\delta$  such that  $\mathbb{P}[V_\alpha(\delta) \geq \eta] \leq \frac{\varepsilon}{2}$  and  $N_0$  such that for  $N \geq N_0$ ,

$$\left| P_v^\tau \left[ \frac{1}{g_N} S_{\mathcal{D}}(\delta R_N) \geq \eta \right] - \mathbb{P}[V_\alpha(\delta) \geq \eta] \right| \leq \frac{\varepsilon}{2}.$$

The reasoning for  $v_f(T, \delta)$  is similar. □

### 7 Shallow traps

In this section we show that the convergence of the clock process of the deep traps shown in Sect. 6 is enough for convergence of the clock process itself.

**Proposition 7.1** *Under the assumptions of Theorem 1.1, the clock process of the deep traps approximates the clock process, namely, for every  $t \geq 0$ ,*

$$\frac{1}{g_N} (S(t R_N) - S_{\mathcal{D}}(t R_N)) \xrightarrow{N \rightarrow \infty} 0 \quad \mathbb{P}\text{-a.s. in } P_v^\tau\text{-probability.}$$

*Proof* We will split the set of shallow traps  $\mathcal{S}_N := \mathbb{H}_N \setminus \mathcal{D}_N$  into two parts and separately deal with the corresponding contributions to the clock process.

We start with ‘very shallow traps’. Let  $\delta > 0$  be a small constant which will be fixed later and  $h_N = e^{\delta \alpha \beta^2 N}$ . Define the set of very shallow traps as

$$\bar{\mathcal{S}}_N = \{x \in \mathbb{H}_N : \tau_x \leq h_N\}.$$

The contribution of this set to the clock process can easily be neglected as follows. Write

$$E_v^\tau \left[ \frac{1}{g_N} \int_0^{t R_N} (1 \vee \tau_{Y_s}) \mathbf{1}_{\{Y_s \in \bar{\mathcal{S}}_N\}} ds \right] = \frac{1}{g_N} \sum_{x \in \bar{\mathcal{S}}_N} (1 \vee \tau_x) E_v^\tau [\ell_{t R_N}(x)]$$

Note that  $E_v^\tau[\ell_{tR_N}(x)] = v_x t R_N = Z_N^{-1}(1 \wedge \tau_x) t R_N$ , and  $(1 \vee \tau_x)(1 \wedge \tau_x) = \tau_x \leq h_N$  on  $\bar{\mathcal{S}}_N$ . With (2.3) for  $Z_N$ , and Lemma 4.4 for  $R_N$ , for every  $\epsilon > 0$ ,  $\mathbb{P}$ -a.s. for  $N$  large enough, the right-hand side of the last equation can be bounded from above by

$$g_N^{-1} 2^N h_N Z_N^{-1} t R_N \leq c g_N^{-1} e^{\delta \alpha \beta^2 N} 2^{(\gamma + \epsilon)N}.$$

To obtain exponential decay of this expression, it is enough to take account of the exponential part of  $g_N$ , which is  $e^{\alpha \beta^2 N}$ . Then, up to sub-exponential factors, using that  $\gamma = \frac{\alpha^2 \beta^2}{2 \log 2}$ , the above is bounded by

$$\exp \left\{ ((\delta - 1)\alpha \beta^2 + \frac{1}{2}\alpha^2 \beta^2 + \epsilon \log 2)N \right\}.$$

Since  $\alpha < 1$ , by choosing  $\epsilon$  and  $\delta$  small enough this can be made smaller than  $e^{-cN}$  for some  $c > 0$ . Applying the Markov inequality and the Borel–Cantelli lemma,

$$\frac{1}{g_N} \int_0^{tR_N} (1 \vee \tau_{Y_s}) \mathbf{1}_{\{Y_s \in \bar{\mathcal{S}}_N\}} ds \xrightarrow{N \rightarrow \infty} 0 \quad \mathbb{P}\text{-a.s. in } P_v^\tau\text{-probability.} \tag{7.1}$$

To control the contribution of the remaining shallow traps  $\mathcal{S}_N \setminus \bar{\mathcal{S}}_N$ , we first split this set into slices  $\mathcal{S}_N^i$  as follows. Set

$$I_N = \left\lceil \frac{1}{\log 2} (\log g'_N - \log h_N) \right\rceil.$$

Note that by definition of  $g'_N$  and  $h_N$ , for  $\delta$  small as fixed above,  $I_N = cN + O(1)$  for some  $c > 0$ . For  $i = 1, \dots, I_N$ , let

$$\mathcal{S}_N^i = \left\{ x \in \mathcal{S}_N \setminus \bar{\mathcal{S}}_N : \tau_x \in [2^{-i} g'_N, 2^{-i+1} g'_N) \right\},$$

so that  $\mathcal{S}_N \setminus \bar{\mathcal{S}}_N = \cup_{i=1}^{I_N} \mathcal{S}_N^i$ .

We next control the sizes of the slices  $\mathcal{S}_N^i$ . By the tail approximation (2.2), for all  $i = 1, \dots, I_N$ ,

$$\begin{aligned} \mathbb{P}[y \in \mathcal{S}_N^i] &\leq \mathbb{P} \left[ E_x > \frac{1}{\beta \sqrt{N}} (\log g'_N - i \log 2) \right] \\ &= f_{N,i}^{(1)} \exp \left\{ -\frac{1}{2} \alpha^2 \beta^2 N + \alpha' i \log 2 - f_{N,i}^{(2)} - o(1) \right\} (1 + o(1)). \end{aligned} \tag{7.2}$$



We separately control the two expressions  $f_{N,i}^{(1)}$  and  $f_{N,i}^{(2)}$ . The first one equals

$$f_{N,i}^{(1)} = \frac{\alpha' \beta \sqrt{2\pi N}}{\frac{\sqrt{2\pi}}{\beta \sqrt{N}} (\log g'_N - i \log 2)}.$$

To control this, note that by definition of  $I_N$ , for all  $i = 1, \dots, I_N$ ,

$$\log g'_N - i \log 2 \geq \log h_N - \log 2 = \delta \alpha \beta^2 N - \log 2.$$

It follows that, for all  $i = 1, \dots, I_N$ ,  $f_{N,i}^{(1)}$  is bounded by some constant  $c > 0$ , which can be chosen to be independent of  $i$ . The second expression to control in (7.2) is

$$f_{N,i}^{(2)} = \frac{i^2 \log^2 2}{2\beta^2 N} + \frac{i \log 2}{\alpha' \beta^2 N} \log(\alpha' \beta \sqrt{2\pi N}).$$

This is strictly positive, so it can be omitted in (7.2) in order to obtain an upper bound. Using the obtained control on  $f_{N,i}^{(1)}$  and  $f_{N,i}^{(2)}$  in (7.2), as well as the fact that  $\gamma' = \frac{\alpha'^2 \beta^2}{2 \log 2}$ , we conclude that for all  $i = 1, \dots, I_N$ ,

$$\mathbb{P}[y \in \mathcal{S}_N^i] \leq c 2^{-\gamma' N} 2^{\alpha' i}.$$

In particular, the size  $|\mathcal{S}_N^i|$  of the  $i$ -th slice is dominated by a binomial random variable with parameters  $n = 2^N$  and  $p = c 2^{\alpha' i} 2^{-\gamma' N}$ . Then it follows by the Markov inequality that for every  $\epsilon > 0$ ,

$$\mathbb{P}\left[|\mathcal{S}_N^i| > 2^{\epsilon N} c 2^{\alpha' i} 2^{(1-\gamma')N}\right] \leq 2^{-\epsilon N}.$$

Since  $I_N = cN + O(1)$ , a union bound and the Borel–Cantelli lemma imply that for every  $\epsilon > 0$ ,  $\mathbb{P}$ -a.s. for  $N$  large enough,

$$|\mathcal{S}_N^i| \leq 2^{\epsilon N} c 2^{\alpha' i} 2^{(1-\gamma')N}, \quad \text{for all } i = 1, \dots, I_N. \tag{7.3}$$

Coming back to the contribution of the intermediate traps  $\mathcal{S}_N \setminus \bar{\mathcal{S}}_N$  to the clock process, we use as before that  $E_v^\tau[\ell_{tR_N}(y)] = v_y t R_N = \frac{1 \wedge \tau_y}{Z_N} t R_N$ , and  $(1 \vee \tau_y)(1 \wedge \tau_y) = \tau_y \leq 2^{-i+1} g'_N$  on  $\mathcal{S}_N^i$ . With (2.3) for  $Z_N$ , Lemma 4.4 for  $R_N$ , and (7.3) for the size of  $\mathcal{S}_N^i$ , we obtain that for every  $\epsilon > 0$ ,  $\mathbb{P}$ -a.s. for  $N$  large enough, for all  $i = 1, \dots, I_N$ ,

$$\begin{aligned} E_v^\tau \left[ \frac{1}{g_N} \int_0^{tR_N} (1 \vee \tau_{Y_s}) \mathbf{1}_{\{Y_s \in \mathcal{S}_N^i\}} ds \right] &= \frac{1}{g_N} \sum_{y \in \mathcal{S}_N^i} (1 \vee \tau_y) E_v^\tau[\ell_{tR_N}(y)] \\ &\leq g_N^{-1} |\mathcal{S}_N^i| 2^{-i+1} g'_N Z_N^{-1} t R_N \end{aligned}$$

$$\leq c \frac{g'_N}{g_N} 2^{(\alpha-1)i} 2^{(\gamma-\gamma'+2\varepsilon)N}.$$

Summing over  $i = 1, \dots, I_N$ ,  $\mathbb{P}$ -a.s. for  $N$  large enough,

$$E_v^\tau \left[ \frac{1}{g_N} \int_0^{tR_N} (1 \vee \tau_{Y_s}) \mathbf{1}_{\{Y_s \in \bigcup_{i=1}^{I_N} S_N^i\}} ds \right] \leq c' \frac{g'_N}{g_N} 2^{(\gamma-\gamma'+2\varepsilon)N}. \tag{7.4}$$

We claim that the right hand side of (7.4) decays exponentially in  $N$  for  $\varepsilon > 0$  small enough. To this end, as before, it is enough to take account of the exponential parts in both  $g_N$  and  $g'_N$ , which contribute to the right hand side of (7.4) by

$$e^{(\alpha'-\alpha)\beta^2 N} = 2^{(\sqrt{\gamma'}-\sqrt{\gamma})\frac{2\beta}{\beta_c} N}.$$

Hence, to show the exponential decay on the right hand side of (7.4), it is sufficient to prove that we can choose  $\varepsilon > 0$  small enough, such that

$$(\sqrt{\gamma'} - \sqrt{\gamma}) \frac{2\beta}{\beta_c} + \gamma - \gamma' + 2\varepsilon < 0. \tag{7.5}$$

With a first order approximation of the concave function  $\sqrt{x}$  at  $\gamma$ ,

$$\frac{1}{2\sqrt{\gamma}}(\gamma - \gamma') < \sqrt{\gamma} - \sqrt{\gamma'}.$$

Since,  $\frac{1}{2\sqrt{\gamma}} = \frac{\beta_c}{2\alpha\beta} > \frac{\beta_c}{2\beta}$  and  $\alpha < 1$ , this implies

$$\frac{\beta_c}{2\beta}(\gamma - \gamma') < \sqrt{\gamma} - \sqrt{\gamma'},$$

and (7.5) thus holds for  $\varepsilon > 0$  small enough. The right hand side of (7.4) then decays exponentially, and with Markov inequality we conclude that

$$\frac{1}{g_N} \int_0^{tR_N} (1 \vee \tau_{Y_s}) \mathbf{1}_{\{Y_s \in \bigcup_{i=1}^{I_N} S_N^i\}} ds \xrightarrow{N \rightarrow \infty} 0 \quad \mathbb{P}\text{-a.s. in } P_v^\tau\text{-probability.}$$

This together with (7.1) finishes the proof of the proposition. □

### 8 Conclusions

We first complete the proof of Theorem 1.1 by observing that it follows directly from Propositions 6.1, 7.1 and Lemma 4.4.

In the remaining part of this section we discuss how our main result, Theorem 1.1, and the techniques leading to its proof can be extended to show more usual aging statements in terms of two-point functions or in terms of the age process. The purpose

of this discussion is not to give the best possible results, but more to show what is missing with our techniques, and also to explain what is missing to obtain better results.

Let us start by pointing out what are the main obstacles impeding us from working with the standard objects considered in the literature previously. As example, consider first the usual two-point function

$$\Pi(t, \theta) = P_v^\tau[X_t = X_{\theta t}], \quad t > 0, \theta > 1, \tag{8.1}$$

where  $X$  is the Metropolis chain defined in (1.2). For this two-point function the desired statement would be, say,

$$\lim_{N \rightarrow \infty} \Pi(tg_N, \theta) = \text{Asl}_\alpha(\theta), \quad \text{in } \mathbb{P}\text{-probability, for every } t > 0, \tag{8.2}$$

where  $\text{Asl}_\alpha(\theta)$  is the probability, given by the arc-sine law, that the range of an  $\alpha$ -stable Lévy process does not intersect the interval  $[1, \theta]$ .

Statements like (8.2) usually follow from a clock-process convergence analogous to Theorem 1.1, if one can show that

“After hitting a deep trap  $x \in \mathcal{D}_N$ , the process  $Y$  mixes well before hitting another deep trap, that is before hitting  $\mathcal{D}_N \setminus \{x\}$ .” (8.3)

However, the strongest information we have in this direction is contained in Lemma 5.5:

“After hitting a deep trap  $x \in \mathcal{D}_N$ ,  $Y$  hits on average less than  $2^{\varepsilon N}$  other deep traps before mixing.”

That means that we cannot exclude the (unlikely) possibility that  $X$  is trapped in a cluster of deep traps, making the verification of (8.2) non-trivial.

An alternative way how to show aging, put forward in [22,31], is to consider the age process  $A_t = \tau_{X_t}$ . This process does not depend on the auxiliary chain  $Y$ , which makes it more physically relevant than the clock process. Aging in the language of age processes manifests itself by the convergence

$$b_N^{-1} A_{g_N t} \xrightarrow{N \rightarrow \infty} \mathcal{Z}_t, \tag{8.4}$$

in a sufficiently weak topology on the space of càdlàg processes, with a well chosen scaling factor  $b_N$ , and with  $\mathcal{Z}_t$  being a particular process, related to the limit behavior of the clock process, whose definition we quote from [31]: Let  $\Upsilon$  be an  $\alpha$ -stable Lévy process, and let

$$\mathcal{V}_t = \int_0^t T_s \, d\Upsilon_s,$$

where  $(T_t)_{t \geq 0}$  is an independent family of independent, mean one, exponential random variables. Let  $\mathcal{W} = \mathcal{V}^{-1}$  be the inverse of  $\mathcal{V}$ . Finally, define

$$\mathcal{Z}_t = \Upsilon_{\mathcal{W}_t} - \Upsilon_{\mathcal{W}_t-}.$$

In order to prove a statement like (8.4) in our context, the first difficulty comes in fixing a right scaling factor  $b_N$ . In [31], this factor consist of two ingredients: (a) the observation scale for the processes  $X$  and  $A$  (i.e.  $g_N$  here), and (b) the diagonal Green function of the suitably killed process  $Y$ , evaluated at the deep traps. As the consequence of the weak asymmetry assumption of [31], this Green function can be computed explicitly, and, more importantly, it can be shown to be essentially deterministic (see (2.5) and Theorems 10.1, 11.1 in [31]).

In our case, we do not have access to such Green function (actually one of the main points of this paper is to show that one *does not need it* to show the clock-process convergence). As a consequence, we even cannot guess the right scale  $b_N$ . Moreover, we actually doubt that a statement like (8.4) holds without modifications. The reason for this is the fact that the Green function (when suitably defined) will be random in our setting, and will take different values on different deep traps. This additional randomness should be incorporated into (8.4), by either modifying the age process  $A_t$  or the limit process  $\mathcal{Z}_t$ .

One possible idea would be to replace  $A_t$  with

$$\tilde{A}_t = \tilde{\tau}_{X_t}, \quad \text{where } \tilde{\tau}_x = E_x^\tau[S(T_{\text{mix}})]. \tag{8.5}$$

While  $\tau_x$  is directly related to the energy  $E_x$  of the state  $x$ ,  $\tilde{\tau}_x$  is a more dynamical quantity. It gives the mean holding time in  $x$  and ‘its neighborhood’ before mixing, which could, possibly, be related to the ‘activation energy’ needed to leave  $x$ . We expect that

$$g_N^{-1} \tilde{A}_{g_N t} \xrightarrow{N \rightarrow \infty} \mathcal{Z}_t,$$

but we cannot show it easily; the non-availability of (8.3), again, hinders proving that  $\tilde{\tau}_x^{-1} S(T_{\text{mix}})$  under  $P_x^\tau$  has an exponential distribution in the limit  $N \rightarrow \infty$ .

We now state our complementary aging results. To this end, set  $H_1 = \inf\{t \geq 0 : Y_t \in \mathcal{D}_N\}$ , and define inductively for  $i \geq 1$

$$\begin{aligned} M_i &= H_i + T_{\text{mix}} \circ \theta_{H_i}, \\ N_i &= M_i + T_{\text{mix}} \circ \theta_{M_i}, \\ H_{i+1} &= \inf\{t \geq N_i : Y_t \in \mathcal{D}_N\}. \end{aligned}$$

Let further

$$\mathcal{H}_i = S(H_i),$$

where  $S$  is the clock process as usual. Finally, let

$$\Pi'(t, \theta) = P_\nu^\tau [\{\mathcal{H}_i : i \in \mathbb{N}\} \cap [t, \theta t] = \emptyset]. \tag{8.6}$$

Heuristically,  $\mathcal{H}_i$  corresponds to times when the Metropolis chain  $X$  enters a ‘new’ deep trap, or possibly a cluster of deep traps, after mixing twice previously.  $\Pi'$  then gives the probability that no such entrance occurs between  $t$  and  $\theta t$ .  $\Pi'$  can also be viewed as an approximation to the probability that  $X$  remains in one (cluster of) deep trap(s) during the observation period, which is close in spirit to (8.1).

Observe, however, that the times  $\mathcal{H}_i$  are defined in terms of mixing of  $Y$  and not of  $X$ , which makes the two-point function  $\Pi'$  to some extent artificial. Optimally, two-point functions should only depend on the physically relevant chain  $X$ , and not on the auxiliary chain  $Y$ . Remark also that various modified two-point functions like  $\Pi'$  were considered in the literature previously, see e.g. [5, 9]; all of them are however expressed in terms of  $X$  only.

The function  $\Pi'$  exhibits aging given by the arc-sine law:

**Theorem 8.1** *Let  $\alpha, \beta$  and  $g_N$  be as in Theorem 1.1. Then for every  $t > 0$ ,*

$$\lim_{N \rightarrow \infty} \Pi'(tg_N, \theta) = \text{Asl}_\alpha(\theta), \quad \text{in } \mathbb{P}\text{-probability.}$$

*Remark 8.2* Observe that the random scale  $R_N$  does not enter the statement of Theorem 8.1. This partially confirms the heuristic arguments given in the introduction.

*Proof of Theorem 8.1* The result is far from being optimal and its proof is rather standard, so we only briefly describe its main ingredients.

We decompose  $S(H_i)$  into three parts

$$\begin{aligned} \mathcal{H}_i &= S(H_i) = \int_0^{H_i} (1 \vee \tau_{Y_s}) \, ds \\ &= \int_0^{H_i} (1 \vee \tau_{Y_s}) \mathbf{1}_{\{Y_s \notin \mathcal{D}_s\}} \, ds + \int_{I_i} \tau_{Y_s} \mathbf{1}_{\{Y_s \in \mathcal{D}_s\}} \, ds + \int_{J_i} \tau_{Y_s} \mathbf{1}_{\{Y_s \in \mathcal{D}_s\}} \, ds, \end{aligned} \tag{8.7}$$

where  $I_i = \cup_{j=1}^{i-1} [H_j, M_j]$  and  $J_i = \cup_{j=1}^{i-1} [M_j, N_j]$ . The first part gives the contribution of the shallow traps, the second and the third part then give the time spent in the deep traps in the corresponding time slots; observe that no deep trap is visited by  $Y$  in  $\cup_i [N_i, H_{i+1})$ .

Using Proposition 7.1, the first part can be neglected by showing

$$\sup_{i: S(H_i) \in [tg_N, \theta tg_N]} \frac{1}{g_N} \int_0^{H_i} (1 \vee \tau_{Y_s}) \mathbf{1}_{\{Y_s \notin \mathcal{D}_s\}} \, ds = 0, \quad \text{in } P_v^\tau\text{-probability, } \mathbb{P}\text{-a.s.}$$

The third part can be neglected as well, it is non-zero with very small probability. Indeed, using the same arguments as in (5.17) and below, we observe that, by the strong Markov property, the sequence  $\int_{M_j}^{N_j} \tau_{Y_s} \mathbf{1}_{\{Y_s \in \mathcal{D}_s\}} \, ds, j \geq 1$ , is an i.i.d. sequence. Since  $Y_{M_j}$  is  $\nu$ -distributed, we have for every  $j \geq 1$

$$P_v^\tau \left[ \int_{M_j}^{N_j} \tau_{Y_s} \mathbf{1}_{\{Y_s \in \mathcal{D}_s\}} \, ds \neq 0 \right] = P_v^\tau [H_{\mathcal{D}_N} \leq T_{\text{mix}}] \leq 2^{(\varepsilon - \gamma')N},$$

by Lemma 5.3 and Proposition 4.1. Moreover, by the second part of Lemma 5.6, we need to consider only  $j \leq 2^{\frac{3}{2}\varepsilon_0 N + \varepsilon N}$  to compute  $S(H_i)$  at the time scales we are interested in. It follows that the probability that the third term is not-zero is bounded by  $2^{(2\varepsilon - \gamma' + \frac{3}{2}\varepsilon_0)N}$  which is negligible since  $\gamma' > 1/2$  and  $\varepsilon_0 < 1/4$ .

The dominating contribution comes from the second part which is again a sum of an i.i.d. sequence, namely of

$$U_j = \int_{H_j}^{M_j} \tau_{Y_s} \mathbf{1}_{\{Y_s \in \mathcal{D}_s\}} ds, \quad j \geq 1. \tag{8.8}$$

We may now define the process

$$\tilde{S}(t) = \sum_{H_j \leq t} U_j,$$

and observe that  $g_N^{-1} \tilde{S}(tR_N)$  is close in the  $M_1$ -distance to  $g_N^{-1} S(tR_N)$ , by previous arguments. It follows that  $g_N^{-1} \tilde{S}(tR_N)$  converges to a Lévy process in the  $M_1$ -topology, and thus, since the jumps of  $\tilde{S}$  are given by an i.i.d. sum, also in  $J_1$ -topology.

We may now conclude using the usual continuity arguments,

$$\Pi'(tg_N, \theta) = P_n^\tau[\{g_N^{-1} \tilde{S}(t) : t \geq 0\} \cap [t, t\theta] = \emptyset] + o(1) = \text{Asl}_\alpha(\theta) + o(1),$$

as  $N \rightarrow \infty$ , as required. □

*Remark 8.3* It might be not obvious why we require  $Y$  to mix twice in the definition of  $H_i$ . The reason for this is the fact that while  $Y_{M_i}$  is  $\nu$ -distributed, it is not independent of  $U_i$ . We need thus an additional mixing to introduce an i.i.d. structure. Without this mixing,  $U_i$  would be a two-dependent sequence.

Finally, let us state an age-process type result. Using the notation (8.8) introduced in the proof of Theorem 8.1, let

$$\bar{A}_t = U_i \quad \text{whenever } t \in [\mathcal{H}_i, \mathcal{H}_{i+1}),$$

with  $\mathcal{H}_0 := 0$  and  $U_0 := 0$ . Roughly,  $A_t$  gives the total holding time in the (cluster of) deep trap(s) where  $X$  is located at time  $t$ . The next theorem is an aging statement for  $\bar{A}$ .

**Theorem 8.4** *Let  $\alpha, \beta, g_N$  be as in Theorem 1.1. Then for every  $T > 0$ , the rescaled age process*

$$\bar{A}_N(t) = g_N^{-1} \bar{A}(g_N t), \quad t \geq 0,$$

*converge in  $\mathbb{P}$ -probability as  $N \rightarrow \infty$ , in  $P_\nu^\tau$ -distribution on the space  $D([0, T], \mathbb{R})$  equipped with the Skorokhod  $J_1$ -topology, to the process  $\bar{Z}_t$  given by  $\bar{Z}_t = \Upsilon_{\mathcal{W}_t} - \Upsilon_{\mathcal{W}_t^-}$ , where  $\Upsilon$  is an  $\alpha$ -stable Lévy process and  $\mathcal{W} = \Upsilon^{-1}$  its inverse.*

Comparing this theorem with the ‘desired’ statement (8.5),  $\tilde{\tau}_x$ ’s of  $\tilde{A}$  are replaced by  $U$ ’s in  $\bar{A}$ , that means the expectation  $E_x^\tau$  is not taken. This is reflected in the change made in the limiting process: the definition of the process  $\tilde{Z}$  does not contain the additional exponential family  $T_t$ . The reason for this is the already mentioned fact that we cannot show easily that the holding times are exponential after rescaling them by their mean.

*Proof of Theorem 8.4* In the proof of Theorem 8.1 we showed that  $\mathcal{H}_i$  is well approximated by  $\sum_{j=1}^{i-1} U_j$ . As  $U_j$  is an i.i.d. sequence converging after rescaling to an  $\alpha$ -stable Lévy process, the claim of the Theorem can be proved by standard arguments similarly to [31]. Observe also that the stronger  $J_1$ -topology can be used here, because  $\bar{A}$  is considerably tamer than  $A$ . □

### Appendix: Extremal characterization of mean hitting time

In this appendix we give the proof of the formula (4.1) which gives a lower bound on the mean hitting time of a set when starting from stationarity. This formula is a continuous-time version of (a half of) Proposition 3.2 from [16]. This proposition, as well as the underlying result [2, Proposition 3.41], is stated for a continuous-time Markov chain whose waiting times are mean-one exponential random variables. We were not able to find analogous statements for general continuous-time Markov chains in the literature, so we provide short proofs here, for the sake of completeness.

We start by introducing some notation. Let  $Y$  be a reversible continuous-time Markov chain on a finite state space  $\mathcal{S}$  with transition rates  $q_{xy}$  and invariant probability measure  $\nu_x$ , denote by  $P_\nu$  and  $P_x$  the laws of  $Y$  started stationary and from  $x$  respectively, and by  $E_\nu, E_x$  the corresponding expectations. Define the conductances as  $c_{xy} = \nu_x q_{xy} = \nu_y q_{yx}$ . Let  $q_x = \sum_y q_{xy}$  and  $c_x = \sum_y c_{xy}$ . The transition probability from  $x$  to  $y$  is  $p_{xy} = \frac{q_{xy}}{q_x} = \frac{c_{xy}}{c_x}$ . In the same way as in Sect. 2, we define the hitting time  $H_x$  and the return time  $H_x^+$  to  $x$  by  $Y$ , and similarly  $H_A$  and  $H_A^+$  for sets  $A \subset \mathcal{S}$ .

A function  $g$  on  $\mathcal{S}$  is called harmonic in  $x$ , if  $\sum_y g(y)p_{xy} = g(x)$ . For  $x \in \mathcal{S}$  and  $B \subset \mathcal{S} \setminus \{x\}$ , the equilibrium potential  $g_{x,B}^*$  is defined as the unique function on  $\mathcal{S}$  that is harmonic on  $(x \cup B)^c$ , 1 on  $x$  and 0 on  $B$ . It is well known that

$$g_{x,B}^*(y) = P_y[H_x \leq H_B].$$

For a function  $g : \mathcal{S} \rightarrow \mathbb{R}$ , the Dirichlet form is defined as

$$D(g, g) = \frac{1}{2} \sum_{z \in \mathcal{S}} \sum_{y \sim z} \nu_z q_{zy} (g(z) - g(y))^2, \tag{8.9}$$

where  $y \sim z$  means that  $y$  and  $z$  are neighbors in the sense that  $q_{zy} > 0$ .

The following proposition is the required generalization of Proposition 3.2 of [16].

**Proposition 8.5** For every  $x \in \mathcal{S}$  and  $B \subset \mathcal{S} \setminus \{x\}$

$$\frac{1}{E_v[H_x]} \leq D(g_{x,B}^*, g_{x,B}^*)v(B)^{-2} = c_x P_x[H_x^+ > H_B]v(B)^{-2}. \tag{8.10}$$

To prove this proposition we will need a lemma which is a generalization of [2, Proposition 3.41] giving the extremal characterization of the mean hitting time.

**Lemma 8.6** For every  $x \in \mathcal{S}$ ,

$$\frac{1}{E_v[H_x]} = \inf \left\{ D(g, g) : g : \mathcal{S} \rightarrow \mathbb{R}, g(x) = 1, \sum_{y \in \mathcal{S}} v_y g(y) = 0 \right\}. \tag{8.11}$$

*Proof* The proof follows the lines of [2] with some minor changes to fit into the setting of general continuous-time chains.

We first show that there is a minimizing function  $g$  that equals  $g(y) = \frac{Z_{yx}}{Z_{xx}}$ , where

$$Z_{yx} = \int_0^\infty (P_y[Y_t = x] - v_x) dt.$$

To this end, we introduce the Lagrange multiplier  $\gamma$  and consider  $g$  as the minimizer of  $D(g, g) + \gamma \sum_z v_z g(z)$  with  $g(x) = 1$ . The contribution to this of  $g(y)$  for  $y \neq x$  is

$$\sum_{z \sim y} v_y q_{yz} (g(y) - g(z))^2 + \gamma v_y g(y),$$

which is minimized if

$$2 \sum_{z \sim y} v_y q_{yz} (g(y) - g(z)) + \gamma v_y = 0.$$

From this we get for all  $y \in \mathcal{S}$ , by introducing the term including the parameter  $\beta$  for the case  $y = x$ , that

$$g(y) = \sum_{z \sim y} \frac{q_{yz}}{q_y} g(z) - \frac{\gamma}{2} \frac{1}{q_y} + \frac{\beta}{q_y} \mathbf{1}_{\{y=x\}}.$$

Multiplying by  $q_y$  and  $v_y$ , and summing over all  $y \in \mathcal{S}$ ,

$$\sum_y \sum_{z \sim y} v_y q_{yz} g(y) = \sum_y \sum_{z \sim y} v_y q_{yz} g(z) - \frac{\gamma}{2} + \beta v_x.$$

By reversibility  $v_y q_{yz} = v_z q_{zy}$ , so the term on the left and the first term on the right are identical, which gives  $\frac{\gamma}{2} = \beta v_x$ . Thus there is a minimizing  $g$  such that

$$g(y) = \frac{\beta}{q_y} (\mathbf{1}_{\{y=x\}} - v_x) + \sum_{z \sim y} \frac{q_{yz}}{q_y} g(z). \tag{8.12}$$



We now show that up to the factor  $\beta$  the function  $y \mapsto Z_{yx}$  satisfies the same relation. Indeed, by the strong Markov property at the time  $J_1$  of the first jump of  $Y$ , which under  $P_y$  is an exponential random variable with mean  $\frac{1}{q_y}$ ,

$$\begin{aligned} Z_{yx} &= \int_0^\infty \left( \int_0^{J_1} (\mathbf{1}_{\{y=x\}} - \nu_x) dt + \sum_{z \sim y} \frac{q_{yz}}{q_y} \int_0^\infty (P_z[Y_t = x] - \nu_x) dt \right) P_y(dJ_1) \\ &= \frac{1}{q_y} (\mathbf{1}_{\{y=x\}} - \nu_x) + \sum_{z \sim y} \frac{q_{yz}}{q_y} Z_{zx}. \end{aligned}$$

The function  $g(y) = \frac{Z_{yx}}{Z_{xx}}$  thus satisfies the constraints of the variational problem in (8.11) and fulfills (8.12) with  $\beta = 1/Z_{xx}$ . It is thus the minimizer of this variational problem.

Moreover, by [2, Lemmas 2.11 and 2.12], we have  $Z_{xx} = E_\nu[H_x]\nu_x$  and  $\nu_x E_y[H_x] = Z_{xx} - Z_{yx}$ . Denoting  $h(y) = E_y[H_x]$  and using these equalities, we obtain

$$D(g, g) = \frac{1}{E_\nu[H_x]^2} D(h, h) = \frac{1}{E_\nu[H_x]},$$

where for the last equality we used  $D(h, h) = E_\nu[H_x]$ , by e.g. [1, Lemma 6]. This completes the proof. □

With this lemma the proof of Proposition 8.5 follows the lines of [16].

*Proof of Proposition 8.5* To prove the inequality in (8.10), it is sufficient to modify the function  $g_{x,B}^*$  so that it becomes admissible for the variational problem in Lemma 8.6. Write  $g^*$  for  $g_{x,B}^*$  and define  $\tilde{g}$  on  $\mathcal{S}$  as

$$\tilde{g}(z) = \frac{g^*(z) - \sum_{y \in \mathcal{S}} \nu_y g^*(y)}{1 - \sum_{y \in \mathcal{S}} \nu_y g^*(y)}.$$

Then  $\tilde{g}$  equals 1 on  $x$  and  $\sum_{z \in \mathcal{S}} \nu_z \tilde{g}(z) = 0$ . Hence, by Lemma 8.6,

$$\frac{1}{E_\nu[H_x]} \leq D(\tilde{g}, \tilde{g}) = D(g^*, g^*) \left( 1 - \sum_{y \in \mathcal{S}} \nu_y g^*(y) \right)^{-2}.$$

But  $g^*$  is non-negative, bounded by 1 and non-zero only on  $B^c$ , therefore  $\sum_{y \in \mathcal{S}} \nu_y g^*(y) \leq \nu(B^c)$ , the first part of Proposition 8.5 follows.

To prove the equality in (8.10), we show that

$$D(g_{x,B}^*, g_{x,B}^*) = P_x[H_x^+ > H_B]c_x. \tag{8.13}$$

Indeed, let again  $g^* = g_{x,B}^*$ . If  $g^*$  is harmonic in  $z$ , the second sum in the Dirichlet form (8.9) is

$$\sum_{y \sim z} c_{zy} (g^*(z) - g^*(y))^2 = \sum_{y \sim z} c_{zy} (g^*(y)^2 - g^*(z)^2).$$

This shows that the contribution to the Dirichlet form of every edge that connects two vertices in which  $g^*$  is harmonic or zero vanishes. Therefore  $D(g^*, g^*)$  reduces to

$$\begin{aligned} D(g^*, g^*) &= \frac{1}{2} \left( \sum_{y \sim x} c_{xy} (1 - g^*(y))^2 + \sum_{y \sim x} c_{xy} (1 - g^*(y)^2) \right) \\ &= \sum_{y \sim x} c_{xy} (1 - g^*(y)) \\ &= c_x \sum_{y \sim x} p_{xy} P_y [H_x > H_B] \\ &= c_x P_x [H_x^+ > H_B]. \end{aligned}$$

This proves (8.13) and thus the proposition.  $\square$

## References

1. Aldous, D.J., Brown, M.: Inequalities for rare events in time-reversible Markov chains. I, Stochastic inequalities (Seattle, WA, 1991). IMS Lecture Notes Monogr. Ser., vol. 22, Inst. Math. Statist., pp. 1–16. Hayward, CA (1992)
2. Aldous, D.J., Fill, J.A.: Reversible markov chains and random walks on graphs. Unfinished monograph, recomplied 2014. (2002). <http://www.stat.berkeley.edu/~aldous/RWG/book.html>
3. Barlow, M.T., Černý, J.: Convergence to fractional kinetics for random walks associated with unbounded conductances. *Probab. Theory Relat. Fields* **149**(3–4), 639–673 (2011)
4. Barlow, M.T., Deuschel, J.-D.: Invariance principle for the random conductance model with unbounded conductances. *Ann. Probab.* **38**(1), 234–276 (2010)
5. Ben Arous, G., Bovier, A., Černý, J.: Universality of the REM for dynamics of mean-field spin glasses. *Commun. Math. Phys.* **282**(3), 663–695 (2008)
6. Ben Arous, G., Bovier, A., Gaynard, V.: Aging in the random energy model. *Phys. Rev. Lett.* **88**(8), 087201 (2002)
7. Ben Arous, G., Bovier, A., Gaynard, V.: Glauber dynamics of the random energy model. I. Metastable motion on the extreme states. *Commun. Math. Phys.* **235**(3), 379–425 (2003)
8. Ben Arous, G., Bovier, A., Gaynard, V.: Glauber dynamics of the random energy model. II. Aging below the critical temperature. *Commun. Math. Phys.* **236**(1), 1–54 (2003)
9. Ben Arous, G., Černý, J.: The arcsine law as a universal aging scheme for trap models. *Commun. Pure Appl. Math.* **61**(3), 289–329 (2008)
10. Ben Arous, G., Gün, O.: Universality and extremal aging for dynamics of spin glasses on subexponential time scales. *Commun. Pure Appl. Math.* **65**(1), 77–127 (2012)
11. Bouchaud, J.P.: Weak ergodicity breaking and aging in disordered systems. *J. Phys. I Fr.* **2**(9), 1705–1713 (1992)
12. Bouchaud, J.-P., Cugliandolo, L.F., Kurchan, J., Mézard, M.: Out of Equilibrium Dynamics in Spin-Glasses and Other Glassy Systems. *Spin Glasses and Random Fields*. World Scientific, Singapore (1997)
13. Bouchaud, J.P., Dean, D.S.: Aging on Parisi's tree. *J. Phys. I Fr.* **5**(3), 265–286 (1995)
14. Černý, J.: On two-dimensional random walk among heavy-tailed conductances. *Electron. J. Probab.* **16**, 293–313 (2011)

15. Černý, J., Gayrard, V.: Hitting time of large subsets of the hypercube. *Random Struct. Algorithms* **33**(2), 252–267 (2008)
16. Černý, J., Teixeira, A., Windisch, D.: Giant vacant component left by a random walk in a random  $d$ -regular graph. *Ann. Inst. Henri Poincaré Probab. Stat.* **47**(4), 929–968 (2011)
17. Černý, J., Wassmer, T.: Randomly trapped random walks on  $\mathbb{Z}^d$ . *Stoch. Process. Appl.* **125**(3), 1032–1057 (2015)
18. Chamberlin, R.V.: Time decay of the thermoremanent magnetization in spin-glasses as a function of the time spent in the field-cooled state. *Phys. Rev. B* **30**(9), 5393–5395 (1984)
19. Derrida, B.: Random-energy model: limit of a family of disordered models. *Phys. Rev. Lett.* **45**(2), 79–82 (1980)
20. Derrida, B.: Random-energy model: an exactly solvable model of disordered systems. *Phys. Rev. B* (3) **24**(5), 2613–2626 (1981)
21. Diaconis, P., Stroock, D.: Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.* **1**(1), 36–61 (1991)
22. Fontes, L.R.G., Mathieu, P.: On the dynamics of trap models in  $\mathbb{Z}^d$ . *Proc. Lond. Math. Soc. (3)* **108**(6), 1562–1592 (2014)
23. Fontes, L.R.G., Isopi, M., Kohayakawa, Y., Picco, P.: The spectral gap of the REM under Metropolis dynamics. *Ann. Appl. Probab.* **8**(3), 917–943 (1998)
24. Gayrard, V.: Aging in reversible dynamics of disordered systems. II. Emergence of the arcsine law in the random hopping time dynamics of the REM (2010). [arXiv:1008.3849](https://arxiv.org/abs/1008.3849)
25. Gayrard, V.: Convergence of clock process in random environments and aging in Bouchaud’s asymmetric trap model on the complete graph. *Electron. J. Probab.* **17**(58), 22 (2012)
26. Gayrard, V.: Convergence of clock processes and aging in Metropolis dynamics of a truncated REM. To appear in *Annales Henri Poincaré* (2014). [arXiv:1402.0388](https://arxiv.org/abs/1402.0388)
27. Gayrard, V., Švejda, A.: Convergence of clock processes on infinite graphs and aging in Bouchaud’s asymmetric trap model on  $\mathbb{Z}^d$ . *ALEA Lat. Am. J. Probab. Math. Stat.* **11**(1), 781–822 (2014)
28. Hastad, J., Leighton, T., Newman, M.: Reconfiguring a hypercube in the presence of faults. In: *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing (New York, NY, USA), STOC ’87*, ACM, pp. 274–284 (1987)
29. Lundgren, L., Svedlindh, P., Nordblad, P., Beckman, O.: Dynamics of the relaxation-time spectrum in a CuMn spin-glass. *Phys. Rev. Lett.* **51**(10), 911–914 (1983)
30. Lyons, R., Peres, Y.: *Probability on trees and networks*. Cambridge University Press. Current version available at <http://mypage.iu.edu/~rdlyons/> (2014) (in preparation)
31. Mathieu, P., Mourrat, J.-C.: Aging of asymmetric dynamics on the random energy model. *Probab. Theory Relat. Fields* **161**(1–2), 351–427 (2015)
32. Mourrat, J.-C.: Scaling limit of the random walk among random traps on  $\mathbb{Z}^d$ . *Ann. Inst. Henri Poincaré Probab. Stat.* **47**(3), 813–849 (2011)
33. Whitt, W.: *Stochastic-Process Limits*. Springer Series in Operations Research. An introduction to stochastic-process limits and their application to queues. Springer-Verlag, New York (2002)