Probability Theory and Related Fields (2022) 182:1–109
https://doi.org/10.1007/s00440-021-01029-5

# The number of solutions for random regular NAE-SAT

**Allan Sly**[1]· **Nike Sun**[2]· **Yumeng Zhang**[3]

Received: 27 March 2019 / Revised: 2 February 2021 / Accepted: 6 February 2021 /
Published online: 20 November 2021


**Abstract**
Recent work has made substantial progress in understanding the transitions of random constraint satisfaction problems. In particular, for several of these models, the exact satisfiability threshold has been rigorously determined, confirming predictions of statistical physics. Here we revisit one of these models, random regular $k$-NAE- SAT: knowing the satisfiability threshold, it is natural to study, in the satisfiable regime, the number of solutions in a typical instance. We prove here that these solutions have a well-defined free energy (limiting exponential growth rate), with explicit value matching the one-step replica symmetry breaking prediction. The proof develops new techniques for analyzing a certain "survey propagation model" associated to this problem. We believe that these methods may be applicable in a wide class of related problems.

**Mathematics Subject Classification** 60K35 · 82B44

## 1 Introduction

In a **random constraint satisfaction problem** (CSP), we have $n$ variables taking values in a (finite) alphabet $\mathcal{X}$, subject to a random set of constraints. In previous works on models of this kind, it has emerged that the space of solutions—a random subset of $\mathcal{X}^n$—can have a complicated structure, posing obstacles to mathematical analysis. Major advances in intuition were achieved by statistical physicists, who developed powerful analytic heuristics to shed light on the behavior of random CSPs ([31] and references therein). Their insights and methods are fundamental to the current understanding of random CSPs.

Research supported in part by Allan Sly: NSF DMS-1208338, DMS-1352013, Sloan Fellowship, and Nike Sun: NSFMSPRF.

✉ Allan Sly
allansly@math.princeton.edu

1 Princeton University, Princeton, NJ, USA

2 Massachusetts Institute of Technology, Cambridge, MA, USA

3 Stanford University, Stanford, CA, USA

One prominent application of the physics heuristic is in giving explicit (non-rigorous) predictions for the locations of satisfiability thresholds in a large class of random CSPs ([36] and others). Recent works have given rigorous proofs for some of these thresholds: in the random regular NAE- SAT model [22,25], in the random $k$-SAT model [23], and in the independent set model on random regular graphs [24]. However, the satisfiability threshold is only one aspect of the rich picture that physicists have developed. There are deep conjectures for the behavior of these models inside the satisfiable regime, and it remains an outstanding mathematical challenge to prove them. In this paper we address one part of this challenge, concerning the **total number of solutions** for a typical instance in the satisfiable regime.

### 1.1 Main result

Given a CNF boolean formula, a **not-all-equal-SAT (NAE-SAT)** solution is an assignment $\underline{\boldsymbol{x}}$ of literals to variables such that both $\underline{\boldsymbol{x}}$ and its negation $\neg\underline{\boldsymbol{x}}$ evaluate to TRUE—equivalently, such that no clause gives the same evaluation to all its variables. A $k$-NAE- SAT problem is one in which each clause has exactly $k$ literals; it is termed **$d$-regular** if each variable appears in exactly $d$ clauses. Sampling such a formula in a uniformly random manner gives rise to the **random $d$-regular $k$-NAE-SAT model**. (The formal definition is given in Sect. 2.) See [3] for important early work on the closely related model of random (Erdős–Rényi) NAE- SAT. The appeal of this model is that it has certain symmetries making the analysis particularly tractable, yet it is expected to share most of the interesting qualitative phenomena exhibited by other commonly studied problems, including random $k$-SAT and random graph colorings.

Following convention, we fix $k$ and then parametrize the model by its clause-to-variable ratio, $\alpha \equiv d/k$. The **partition function**, denoted $Z \equiv Z_n$, is the number of valid NAE- SAT assignments for an instance on $n$ variables. It is conjectured that for each $k \geq 3$, the model has an exact satisfiability threshold $\alpha_{\text{sat}}(k)$: for $\alpha < \alpha_{\text{sat}}$ it is satisfiable ($Z > 0$) with high probability, but for $\alpha > \alpha_{\text{sat}}$ it is unsatisfiable ($Z = 0$) with high probability. This has been proved [25, Thm. 1] for all $k$ exceeding an absolute constant $k_0$, together with an exact formula for $\alpha_{\text{sat}}$ which matches the physics prediction. It can be approximated as

$$\alpha_{\text{sat}} = \left(2^{k-1} - \frac{1}{2} - \frac{1}{4\ln 2}\right)\ln 2 + \epsilon_k \tag{1}$$

where $\epsilon_k$ denotes an error tending to zero as $k \to \infty$.

We say the model has **free energy** $\mathsf{f}(\alpha)$ if $Z^{1/n}$ converges to $\mathsf{f}(\alpha)$ in probability as $n \to \infty$. *A priori*, the limit may not be well-defined. If it exists, however, Markov's inequality and Jensen's inequality imply that it must be upper bounded by the **replica symmetric free energy**

$$\mathsf{f}^{\text{RS}}(\alpha) \equiv (\mathbb{E}Z)^{1/n} = 2\left(1 - \frac{2}{2^k}\right)^{\alpha}. \tag{2}$$

(In this model and in other random regular models, the replica symmetry free energy is the same as the annealed free energy.) One of the intriguing predictions from the physics analysis [38,44] is that there is a critical value $\alpha_{\text{cond}}$ strictly below $\alpha_{\text{sat}}$, such that $\mathsf{f}(\alpha)$ and $\mathsf{f}^{\text{RS}}(\alpha)$ agree up to $\alpha = \alpha_{\text{cond}}$ and diverge thereafter. In particular, this implies that the function $\mathsf{f}(\alpha)$ must be non-analytic at $\alpha = \alpha_{\text{cond}}$. This is the **condensation transition** (or *Kauzmann transition*), and will be further described below in Sect. 1.2. For all $0 \leq \alpha < \alpha_{\text{sat}}$, the free energy is predicted to be given by a formula

$$\mathsf{f}(\alpha) = \mathsf{f}^{\text{1RSB}}(\alpha) \begin{cases} = \mathsf{f}^{\text{RS}}(\alpha) & \text{for } 0 \leq \alpha \leq \alpha_{\text{cond}}, \\ < \mathsf{f}^{\text{RS}}(\alpha) & \text{for } \alpha > \alpha_{\text{cond}}. \end{cases}$$

The function $\mathsf{f}^{\text{1RSB}}(\alpha)$ is quite explicit, although not extremely simple to state; it is formally presented below in Definition 1.3. The formula for $\mathsf{f}^{\text{1RSB}}(\alpha)$ is derived via the **one-step replica symmetry breaking** (1RSB) heuristic, discussed further below. Our main result is to prove this prediction for large $k$:

**Theorem 1** *In random regular $k$-*NAE- SAT *with $k \geq k_0$, for all $\alpha < \alpha_{sat}(k)$ the free energy $\mathsf{f}(\alpha)$ exists and equals the predicted value $\mathsf{f}^{\text{1RSB}}(\alpha)$.*

***Remark 1.1*** We allow for $k_0$ to be adjusted as long as it remains an absolute constant (so it need not equal the $k_0$ from [25]). It is assumed throughout the paper that $k \geq k_0$, even when not explicitly stated. The following considerations restrict the range of $\alpha = d/k$ that we must consider:

– A convenient upper bound on the satisfiable regime is given by

$$\alpha_{\text{sat}} \leq \alpha_{\text{RS}} \equiv \frac{\ln 2}{-\ln(1 - 2/2^k)} < 2^{k-1} \ln 2 \equiv \alpha_{\text{ubd}} .$$

This bound is certainly implied by the estimate (1) from [25], but it follows much more easily and directly from the first moment calculation (2). Indeed, we see from (2) that the function $\mathsf{f}^{\text{RS}}(\alpha)$ is decreasing in $\alpha$ and satisfies $\mathsf{f}^{\text{RS}}(\alpha_{\text{RS}}) = 1$, so $(\mathbb{E}Z)^{1/n} < 1$ for all $\alpha > \alpha_{\text{RS}}$. Thus, by Markov's inequality, we have that $\mathbb{P}(Z > 0) \leq \mathbb{E}Z$ tends to zero as $n \to \infty$, i.e., the random problem instance is unsatisfiable with high probability.

– For $\alpha > \alpha_{\text{sat}}$ we must have $\mathsf{f}(\alpha) = 0$. On the other hand, we can see by comparing (1) and (2) that $\alpha_{\text{sat}}$ is strictly smaller than $\alpha_{\text{RS}}$, and $\mathsf{f}^{\text{RS}}(\alpha_{\text{sat}})$ is strictly positive. This suggests that $\alpha_{\text{cond}}$ occurs strictly before $\alpha_{\text{sat}}$, since $\alpha_{\text{cond}} = \alpha_{\text{sat}}$ would mean that $\mathsf{f}(\alpha) = \mathsf{f}^{\text{RS}}(\alpha)$ up to $\alpha_{\text{sat}}$, and in this case we would expect to have $\alpha_{\text{sat}} = \alpha_{\text{RS}}$. Formally, it requires further argument to confirm that $\alpha_{\text{cond}} < \alpha_{\text{sat}}$ in random regular NAE- SAT, and we obtain this as a consequence of results in the present paper. However, the phenomenon of $\alpha_{\text{cond}} < \alpha_{\text{sat}}$ was previously confirmed by [20] and [8] for random hypergraph bicoloring and random regular SAT, both of which are very similar to random regular NAE- SAT. As for the value of $\mathsf{f}(\alpha)$ at the threshold $\alpha = \alpha_{\text{sat}}$, we point out that $\alpha = \alpha_{\text{sat}}$ makes sense in the setting of this paper only if $d_{\text{sat}}(k) \equiv k\alpha_{\text{sat}}(k)$ is integer-valued for some $k$. We have no reason

to think that this ever occurs; however, if it does, then the probability for $Z > 0$ is bounded away from both zero and one [25, Thm. 1]. In this case, $Z^{1/n}$ does not concentrate around a single value but rather on two values,

$$\left\{0, \lim_{\alpha \uparrow \alpha_{\text{sat}}} \mathsf{f}^{\text{1RSB}}(\alpha)\right\}.$$

– In [25, Propn. 1.1] it is shown that for $0 \le \alpha \le \alpha_{\text{lbd}} \equiv (2^{k-1} - 2)\ln 2$ and $n$ large enough,

$$\frac{\mathbb{E}(Z^2)}{(\mathbb{E}Z)^2} \le C \equiv C(k, \alpha) < \infty$$

where $Z \equiv Z_n$ and $C(k, \alpha)$ does not depend on $n$. Thus, for any fixed $0 < \epsilon < 1$ and $n$ large enough,

$$\mathbb{P}(Z \ge \epsilon \mathbb{E}Z) \overset{\odot}{\ge} \frac{\mathbb{E}(Z\mathbf{1}\{Z \ge \epsilon \mathbb{E}Z\})^2}{\mathbb{E}(Z^2)} \ge \frac{(1-\epsilon)^2(\mathbb{E}Z)^2}{\mathbb{E}(Z^2)} \ge \frac{(1-\epsilon)^2}{C} \equiv \delta\,.$$

where the step marked $\odot$ is by the Cauchy–Schwarz inequality. The results of [25, Sec. 6] imply the stronger statement that for any $0 \le \alpha \le \alpha_{\text{lbd}}$,

$$\lim_{\epsilon \downarrow 0} \liminf_{n \to \infty} \mathbb{P}(Z \ge \epsilon \mathbb{E}Z) = 1\,.$$

On the other hand we already noted in (2) that $\mathbb{E}(Z^{1/n}) \le (\mathbb{E}Z)^{1/n} = \mathsf{f}^{\text{RS}}(\alpha)$ for all $\alpha \ge 0$ and $n \ge 1$. It follows by combining these facts that $Z^{1/n}$ converges in probability to $\mathsf{f}^{\text{RS}}(\alpha)$ in probability for any $0 \le \alpha \le \alpha_{\text{lbd}}$. That is to say, the result of Theorem 1 is already proved for $\alpha \le \alpha_{\text{lbd}}$, with $\mathsf{f}(\alpha) = \mathsf{f}^{\text{RS}}(\alpha)$.
This also implies that the condensation transition $\alpha_{\text{cond}}$ must occur above $\alpha_{\text{lbd}}$.

In summary, we have $\alpha_{\text{lbd}} < \alpha_{\text{sat}} < \alpha_{\text{RS}} < \alpha_{\text{ubd}}$, and it remains to prove Theorem 1 for $\alpha \in (\alpha_{\text{lbd}}, \alpha_{\text{sat}})$. Thus, we can assume for the remainder of the paper that

$$(2^{k-1} - 2)\ln 2 = \alpha_{\text{lbd}} \le \alpha \le \alpha_{\text{ubd}} = 2^{k-1}\ln 2\,. \tag{3}$$

In the course of proving Theorem 1 we will also identify the condensation threshold $\alpha_{\text{cond}} \in (\alpha_{\text{lbd}}, \alpha_{\text{sat}})$ (characterized in Proposition 1.4 below).

The 1RSB heuristic, along with its implications for the condensation and satisfiability thresholds, has been studied in numerous recent works, which we briefly survey here. The existence of a condensation transition was first shown in random hypergraph bicoloring [20], which as we mentioned above is a model very similar to random NAE-SAT. We also point out [17] which is the first work to successfully analyze solution clusters within the condensation regime, leading to a very good lower bound on satisfiability threshold. This was an important precursor to subsequent works [23–25] on

exact satisfiability thresholds in random regular NAE- SAT, random SAT, and independent sets. Condensation has been demonstrated to occur even at positive temperature in hypergraph bicoloring (which is very similar to NAE- SAT) [11]. However, determining the precise location of $\alpha_{\text{cond}}$ is challenging, and was first achieved for the random graph coloring model [10] by an impressive and technically challenging analysis. A related paper pinpoints $\alpha_{\text{cond}}$ for random regular $k$-SAT (which again is very similar to NAE- SAT) [8]. Subsequent work [15] characterizes the condensation threshold in a more general family of models, and shows a correspondence with information-theoretic thresholds in statistical inference problems. The main contribution of this paper is to determine for the first time the free energy throughout the condensation regime $(\alpha_{\text{cond}}, \alpha_{\text{sat}})$.

### 1.2 Statistical physics predictions

According to the heuristic analysis by statistical physics methods, the random regular NAE- SAT model has a single level of replica symmetry breaking (1RSB). We summarize here some of the key phenomena that are predicted from the 1RSB framework [31,38, 44], referring the reader to [33, Ch. 19] for a full expository account. While much of the following description remains conjectural, the implications at the free energy level are rigorously established by the present paper. Throughout the following we write $\doteq$ to indicate equality up to subexponential factors ($\exp\{o(n)\}$).

Recall that we consider NAE- SAT with $k$ fixed, parametrized by the clause density $\alpha \equiv d/k$. Abbreviate $0 \equiv$ TRUE, $1 \equiv$ FALSE. For small $\alpha$, almost all of the solutions lie in a single well-connected subset of $\{0,1\}^n$. This holds until a **clustering transition** (or *dynamical transition*) $\alpha_{\text{clust}}$, above which the solution space becomes broken up into many well-separated pieces, or **clusters** (see [1,2,6,35]). Informally speaking, clusters are subsets of solutions which are characterized by the property that within-cluster distances are very small relative to between-cluster distances. Conjecturally, $\alpha_{\text{clust}}$ also coincides with the **reconstruction threshold** [28,31,39], and is small relative to $\alpha_{\text{sat}}$ when $k$ is large, with $\alpha_{\text{clust}}/\alpha_{\text{sat}} \asymp (\ln k)/k$.

For $\alpha$ above $\alpha_{\text{clust}}$ it is expected that the number of clusters of size $\exp\{ns\}$ has mean value $\exp\{n\Sigma(s;\alpha)\}$, and is **concentrated** about this mean. The function $\Sigma(s) \equiv \Sigma(s;\alpha)$ is referred to as the "cluster complexity." The 1RSB framework of statistical physics gives an **explicit** conjecture for $\Sigma$, discussed below in Sect. 1.3. Then, summing over cluster sizes $0 \le s \le \ln 2$ gives that the total number $Z$ of NAE- SAT solutions has mean

$$\mathbb{E}Z \doteq \sum_s \exp\{n[s + \Sigma(s)]\} \doteq \exp\{n[s_1 + \Sigma(s_1)]\}, \tag{4}$$

where $s_1 = \arg\max[s + \Sigma(s)]$. It is expected that $\Sigma$ is continuous and strictly concave in $s$, and also that $s + \Sigma(s)$ has a unique maximizer $s_1$ with $\Sigma'(s_1) = -1$. For NAE- SAT and related models, this explicit calculation reveals a critical value $\alpha_{\text{cond}} \in (\alpha_{\text{clust}}, \alpha_{\text{sat}})$, characterized as

$$\alpha_{\text{cond}} = \sup\{\alpha \ge \alpha_{\text{clust}} : \Sigma(s_1(\alpha);\alpha) \ge 0\}.$$

By contrast, the satisfiability threshold can be characterized as

$$\alpha_{\text{sat}} = \sup\{\alpha \geq \alpha_{\text{clust}} : \max_s \Sigma(s;\alpha) \geq 0\}.$$

For all $\alpha \geq \alpha_{\text{clust}}$, the expected partition function $\mathbb{E}Z$ is dominated by clusters of size $\exp\{ns_1\}$. However, for $\alpha > \alpha_{\text{cond}}$, we have $\Sigma(s_1) < 0$, so the expected number of clusters of this size is very small: $\exp\{n\Sigma(s_1)\}$ tends to zero exponentially fast as $n \to \infty$. This means that clusters of size $\exp\{ns_1\}$ are highly unlikely to appear in a typical realization of the model. Instead, in a typical realization we only expect to see clusters of size $\exp\{ns\}$ with $\Sigma(s) \geq 0$. As a result the solution space should be dominated (with high probability) by clusters of size $s_{\max}$ where

$$s_{\max} \equiv s_{\max}(\alpha) \equiv \arg\max\{s + \Sigma(s) : \Sigma(s) \geq 0\}.$$

Since $\Sigma$ is continuous, $s_{\max}$ is the largest root of $\Sigma$, and for $\alpha \in (\alpha_{\text{cond}}, \alpha_{\text{sat}})$ we should have

$$Z \doteq \exp\{ns_{\max}\} \ll \mathbb{E}Z = \exp\{n[s_1 + \Sigma(s_1)]\}$$

(where the approximation for $Z$ holds with high probability). The **1RSB free energy**, formally given by Definition 1.3 below, should be interpreted as an expression for the function $\mathsf{f}^{\text{1RSB}}(\alpha) = s_{\max}(\alpha)$.

### 1.3 The tilted cluster partition function

From the discussion of Sect. 1.2 we see that once the function $\Sigma(s;\alpha)$ is determined, it is possible to derive $\alpha_{\text{cond}}$, $\alpha_{\text{sat}}$, and $\mathsf{f}(\alpha)$. However, previous works have not taken the approach of actually computing $\Sigma$. Indeed, $\alpha_{\text{sat}}$ was determined [25] by an analysis involving only $\max_s \Sigma(s;\alpha)$, which contains less information than the full curve $\Sigma$. Related work on the exact determination (in a range of models) of $\alpha_{\text{cond}}$ [8,10,15] also avoids computing $\Sigma$, reasoning instead via the so-called "planted model."

In order to compute the free energy, however, we cannot avoid computing (some version of) the function $\Sigma$, which we will do by a physics-inspired approach. First consider the $\lambda$-tilted partition function

$$\bar{\boldsymbol{Z}}_\lambda \equiv \sum_{\boldsymbol{\gamma} \in \mathsf{CL}(\mathscr{G})} |\boldsymbol{\gamma}|^\lambda, \tag{5}$$

where $\mathsf{CL}(\mathscr{G})$ denotes the set of solution clusters of $\mathscr{G}$, and $|\boldsymbol{\gamma}|$ denotes the number of satisfying assignments inside the cluster $\boldsymbol{\gamma}$. According to the conjectural picture described above, we should have

$$\mathbb{E}\bar{\boldsymbol{Z}}_\lambda \doteq \sum_s (\exp\{ns\})^\lambda \exp\{n\Sigma(s)\} \doteq \exp\{n\mathfrak{F}(\lambda)\}$$

where $\mathfrak{F}$ is the Legendre dual of $-\Sigma$:

$$\mathfrak{F}(\lambda) \equiv (-\Sigma)^{\star}(\lambda) \equiv \max_{s} \left\{ \lambda s + \Sigma(s) \right\} = \lambda s_{\lambda} + \Sigma(s_{\lambda}) \,, \tag{6}$$

where $s_{\lambda} \equiv \arg\max_{s}[\lambda s + \Sigma(s)]$. Moreover, if $\Sigma(s_{\lambda}) \geq 0$, then $\boldsymbol{Z}_{\lambda}$ should concentrate near $\mathbb{E}\boldsymbol{Z}_{\lambda}$, and in this regime physicists have an exact prediction for $\mathfrak{F}(\lambda)$, which will be further discussed below in Sect. 1.5. In short, the physics approach to computing $\Sigma$ is to first compute $\mathfrak{F}(\lambda)$ (in the regime where $\Sigma(s_{\lambda}) \geq 0$), and then set $\Sigma = -\mathfrak{F}^{\star}$. Note that by differentiating $\mathfrak{F}(\lambda) = n^{-1} \ln \mathbb{E}\bar{\boldsymbol{Z}}_{\lambda}$ we find that $\mathfrak{F}$ is convex in $\lambda$, so the resulting $\Sigma$ will indeed be concave.

At first glance the reduction to computing $\mathfrak{F}(\lambda)$ may not seem to improve matters. It is not immediately clear how "clusters" should be defined. It turns out that in the regime we are studying, a reasonable definition is that two NAE- SAT solutions are connected if they differ by a single bit, and the **clusters** are the connected components of the solution space. A typical NAE- SAT solution will have a positive density of variables which are **free**, meaning their value can be changed without violating any clause; any such solution must belong in a cluster of exponential size. Each cluster may be a complicated subset of $\{0,1\}^n$—changing the value at one free variable may affect whether its neighbors are free, so a cluster need not be a simple subcube of $\{0,1\}^n$. Nevertheless, we wish to sum over the cluster sizes raised to non-integer powers. This computation is made tractable by constructing more explicit combinatorial models for the NAE- SAT solution clusters, as we next describe.

### 1.4 Modeling solution clusters

In our regime of interest (i.e., $k \geq k_0$ and $\alpha_{\mathrm{lbd}} \leq \alpha \leq \alpha_{\mathrm{ubd}}$; see Remark 1.1), the analysis of NAE- SAT solution clusters is greatly simplified by the fact that in a typical satisfying assignment the vast majority of variables are **frozen** rather than free. The result of this, roughly speaking, is that a cluster $\boldsymbol{\gamma} \in \mathsf{CL}(\mathscr{G})$ can be encoded by a configuration $\underline{x} \in \{0,1,\mathtt{f}\}^n$ (representing its circumscribed subcube, so $x_v = \mathtt{f}$ indicates a free variable) with no essential loss of information. This is formalized by a combinatorial model of "frozen configurations" representing the clusters (Definition 2.2). These frozen configurations can be viewed as the solutions of a certain CSP lifted from the original NAE- SAT problem — so the physics heuristics can be applied again to (the randomized version of) the lifted model. Variations on this idea appear in several places in the physics literature; in the specific context of random CSPs we refer to [13,34,42]. Analyzing the number of frozen configurations, corresponding to (5) with $\lambda = 0$, yields the satisfiability threshold for this model [25].

Analyzing (5) for general $\lambda$ requires deeper investigation of the arrangement of free variables in a typical frozen configuration $\underline{x}$. For this purpose it is convenient to view an NAE- SAT instance as a (bipartite) graph $\mathscr{G}$, where the vertices are given by variables and clauses, and the edges indicate which variables participate in which clauses (the formal description appears in Sect. 2). A key piece of intuition is that if we consider the subgraph of $\mathscr{G}$ induced by the free variables, together with the clauses through which

they interact, then this subgraph is predominantly comprised of disjoint components $\boldsymbol{T}$ of bounded size. (In fact, the majority of free variables are simply isolated vertices; a smaller fraction occur in linked pairs; a yet smaller fraction occur in components of size three or more.) Each free component $\boldsymbol{T}$ is surrounded by frozen variables, and we let $z(\boldsymbol{T})$ be the number of NAE- SAT assignments on $\boldsymbol{T}$ which are consistent with the frozen boundary. Since disjoint components $\boldsymbol{T}, \boldsymbol{T}'$ do not interact, the size of the cluster represented by $\underline{x}$ is simply the product of $z(\boldsymbol{T})$ over all $\boldsymbol{T}$.

Another key observation is that the random NAE- SAT graph has few short cycles, so almost all of the free components will be **trees**. As a result, their weights $z(\boldsymbol{T})$ can be evaluated recursively by **belief propagation** (BP), a well-known dynamic programming method (see e.g. [33, Ch. 14]). In the RSB heuristic framework, a cluster is represented by a vector $\underline{\mathtt{m}}$ of "messages," indexed by the directed edges of the NAE- SAT graph $\mathscr{G}$. Informally, for a given cluster, and for any variable $v$ adjacent to any clause $a$,

$$\begin{aligned} &\mathtt{m}_{v\to a} \text{ represents the "within-cluster law of } \boldsymbol{x}_v \text{ in absence of } a''; \\ &\mathtt{m}_{a\to v} \text{ represents the "within-cluster law of } \boldsymbol{x}_v \text{ in absence of } \partial v \backslash a'', \end{aligned} \tag{7}$$

where $\partial v$ denotes the neighboring clauses of $v$. Each message is a probability measure on $\{0, 1\}$, and the messages are related to one another via local consistency equations, which are known as the BP equations. A configuration $\underline{\mathtt{m}}$ which satisfies all the local consistency equations is a **BP solution**. Thus a cluster $\boldsymbol{\gamma}$ can be encoded either by a frozen configuration $\underline{x}$ or by a BP solution $\underline{\mathtt{m}}$; the latter has the key advantage that **the size of $\boldsymbol{\gamma}$ can be easily read off from $\underline{\mathtt{m}}$, as a certain product of local functions.** For the cluster size raised to power $\lambda$, simply raise each local function to power $\lambda$. Thus the configurations $\underline{\mathtt{m}}$ with $\lambda$-tilted weights form a **spin system** (Markov random field), whose partition function is the quantity of interest (5). This new spin system is sometimes termed the "auxiliary model" (e.g. [33, Ch. 19]).

### 1.5 One-step replica symmetry breaking

In Sect. 1.4 we described informally how a solution cluster $\boldsymbol{\gamma}$ can be encoded by a frozen configuration $\underline{x}$, or a BP solution $\underline{\mathtt{m}}$. An important caveat is that the converse need not hold. In the NAE- SAT model, for any value of $\alpha$, a trivial BP solution is always given by the "replica symmetric fixed point" (also called the "factorized fixed point"), where every $\mathtt{m}_{v\to a}$ is the uniform measure on $\{0, 1\}$. However, above $\alpha_{\mathrm{cond}}$, this is a spurious solution. One way to see this is via the heuristic "cavity calculation" of $\mathsf{f}^{\mathrm{RS}}(\alpha)$, which we now describe to motivate the more complicated expression for $\mathsf{f}^{\mathrm{1RSB}}(\alpha)$.

Given a random regular NAE- SAT instance $\mathscr{G}$ on $n$ variables, choose $k$ uniformly random variables $v_1, \ldots, v_k$, and assume for simplicity that no two of these share a clause. Remove the $k$ variables along with their $kd$ incident clauses, producing the "cavity graph" $\mathscr{G}''$. Then add $d(k-1)$ new clauses to $\mathscr{G}''$, producing the graph $\mathscr{G}'$. Under this operation (cf. [7]), $\mathscr{G}'$ is distributed as a random regular NAE- SAT instance on $n-k$ variables. If the free energy $\mathsf{f}(\alpha) = \lim_{n\to\infty} Z^{1/n}$ exists, then we would expect it to agree asymptotically with

$$\left(\frac{Z(\mathscr{G})}{Z(\mathscr{G}')}\right)^{1/k} = \left(\frac{Z(\mathscr{G})}{Z(\mathscr{G}'')}\right)^{1/k} \Big/ \left(\frac{Z(\mathscr{G}')}{Z(\mathscr{G}'')}\right)^{1/k}. \tag{8}$$

Let $U$ denote the set of "cavity neighbors": the variables in $\mathscr{G}''$ of degree $d-1$, which neighbored the clauses that were deleted from $\mathscr{G}$. Then $\mathscr{G}$ and $\mathscr{G}'$ differ from $\mathscr{G}''$ only in the addition of a few small subgraphs which are attached to $U$. Computing $Z(\mathscr{G})/Z(\mathscr{G}'')$ or $Z(\mathscr{G}')/Z(\mathscr{G}'')$ reduces to understanding the joint law of the spins $(\boldsymbol{x}_u)_{u\in U}$ under the NAE- SAT model defined by $\mathscr{G}''$. Since $\mathscr{G}$ is unlikely to have many cycles, the vertices of $U$ are typically far apart from one another in $\mathscr{G}''$. Therefore, one plausible scenario is that their spins are approximately independent under the NAE- SAT model on $\mathscr{G}''$, with $\boldsymbol{x}_u$ marginally distributed according to $\mathtt{m}_{u\to a}$ where $a$ is the deleted clause that neighbored $u$ in $\mathscr{G}$. If this is the case, then each $\mathtt{m}_{u\to a}$ must be uniform over $\{0,1\}$, by the negation symmetry of NAE- SAT. Under this assumption, we can calculate

$$\left(\frac{Z(\mathscr{G})}{Z(\mathscr{G}'')}\right)^{1/k} = 2(1-2/2^k)^d, \quad \left(\frac{Z(\mathscr{G}')}{Z(\mathscr{G}'')}\right)^{1/k} = (1-2/2^k)^{\alpha(k-1)}, \tag{9}$$

Substituting into (8) gives the replica symmetric free energy prediction $\mathsf{f}(\alpha) \doteq \mathsf{f}^{\mathrm{RS}}(\alpha)$, which we know to be false for large $\alpha$ (in particular, it is inconsistent with the known satisfiability threshold). Thus the replica symmetric fixed point, $\mathtt{m}_{u\to a} = \mathrm{unif}(\{0,1\})$ for all $u \to a$, is a spurious BP solution. In reality the $\boldsymbol{x}_u$ are **not** approximately independent in $\mathscr{G}''$, even though the $u$'s are far apart. This phenomenon of **non-negligible long-range dependence** may be taken as a definition of replica symmetry breaking (RSB) in this setting, and occurs precisely for $\alpha$ larger than $\alpha_{\mathrm{cond}}$.

Since above $\alpha_{\mathrm{cond}}$ the partition function cannot be estimated by (9) due to replica symmetry breaking, a different approach is needed. To this end, the **one-step RSB** (1RSB) heuristic posits that even when the original NAE- SAT model exhibits RSB, the (seemingly more complicated) "auxiliary model" of $\lambda$-weighted BP solutions $\underline{\mathtt{m}}$ **is replica symmetric, for $\lambda$ small enough**: conjecturally, as long as $\Sigma(s_\lambda) \geq 0$ for $s_\lambda \equiv \arg\max_s\{\lambda s + \Sigma(s)\}$ (cf. the discussion below (6)). That is, for such $\lambda$, the auxiliary model is predicted to have correlation decay, in contrast with the long-range correlations of the original model. This would mean that in the auxiliary model of the cavity graph $\mathscr{G}''$, the spins $(\mathtt{m}_{u\to a})_{u\in U}$ are approximately independent, with each $\mathtt{m}_{u\to a}$ marginally distributed according to some law $\dot{q}_{u\to a}$. The model has a replica symmetric fixed point, $\dot{q}_{u\to a} = \dot{q}_\lambda$ for all $u \to a$ (the analogue of $\mathtt{m}_{u\to a} = \mathrm{unif}(\{0,1\})$ for all $u \to a$). If we substitute this assumption into the cavity calculation (the analogues of (8) and (9)), we obtain the replica symmetric prediction for the auxiliary model free energy $\mathfrak{F}(\lambda)$, expressed as a function of $\dot{q}_\lambda$. As explained above, from $\mathfrak{F}(\lambda)$ we can derive the complexity function $\Sigma(s)$ and the 1RSB NAE- SAT free energy $\mathsf{f}^{\mathrm{1RSB}}(\alpha)$.

### 1.6 The 1RSB free energy prediction

Having described the heuristic reasoning, we now proceed to formally state the 1RSB free energy prediction. We first describe $\dot{q}_\lambda$ as a certain discrete probability measure over $\mathtt{m}$. Since $\mathtt{m}$ is a probability measure over $\{0,1\}$, we encode it by $x \equiv \mathtt{m}(1) \in [0,1]$.

A measure $q$ on $\mathtt{m}$ can thus be encoded by an element $\mu \in \mathscr{P}$ where $\mathscr{P}$ denotes the set of discrete probability measures on $[0, 1]$. For measurable $B \subseteq [0, 1]$, define

$$\hat{\mathscr{R}}_\lambda \mu(B) \equiv \frac{1}{\hat{\mathscr{Z}}(\mu)} \int \left(2 - \prod_{i=1}^{k-1} x_i - \prod_{i=1}^{k-1}(1-x_i)\right)^\lambda \mathbf{1}\left\{\frac{1 - \prod_{i=1}^{k-1} x_i}{2 - \prod_{i=1}^{k-1} x_i - \prod_{i=1}^{k-1}(1-x_i)} \in B\right\} \prod_{i=1}^{k-1} \mu(dx_i)\,,$$
$$\dot{\mathscr{R}}_\lambda \mu(B) \equiv \frac{1}{\dot{\mathscr{Z}}(\mu)} \int \left(\prod_{i=1}^{d-1} y_i + \prod_{i=1}^{d-1}(1-y_i)\right)^\lambda \mathbf{1}\left\{\frac{\prod_{i=1}^{d-1} y_i}{\prod_{i=1}^{d-1} y_i + \prod_{i=1}^{d-1}(1-y_i)} \in B\right\} \prod_{i=1}^{d-1} \mu(dy_i)\,, \quad (10)$$

where $\hat{\mathscr{Z}}(\mu)$ and $\dot{\mathscr{Z}}(\mu)$ are the normalizing constants such that $\hat{\mathscr{R}}_\lambda \mu$ and $\dot{\mathscr{R}}_\lambda \mu$ are also probability measures on $[0, 1]$. (In the context of $\lambda = 0$ we take the convention that $0^0 = 0$.) Denote $\mathscr{R}_\lambda \equiv \dot{\mathscr{R}}_\lambda \circ \hat{\mathscr{R}}_\lambda$. The map $\mathscr{R}_\lambda : \mathscr{P} \to \mathscr{P}$ represents the BP recursion for the auxiliary model. The following presents a solution for $\alpha$ in the interval $(\alpha_{\mathrm{lbd}}, \alpha_{\mathrm{ubd}})$ which we recall (Remark 1.1) is a superset of $(\alpha_{\mathrm{cond}}, \alpha_{\mathrm{sat}})$.

**Proposition 1.2** (proved in "Appendix B") *For $\lambda \in [0, 1]$, let $\dot{\mu}_{\lambda,0} \equiv \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1 \in \mathscr{P}$, and define recursively $\dot{\mu}_{\lambda,l+1} = \mathscr{R}_\lambda \dot{\mu}_{\lambda,l} \in \mathscr{P}$ for all $l \geq 0$. Define $S_l \equiv (\mathrm{supp}\, \dot{\mu}_{\lambda,l}) \backslash (\mathrm{supp}(\dot{\mu}_{\lambda,0} + \ldots + \dot{\mu}_{\lambda,l-1}))$; this is a finite subset of $[0, 1]$. Regard $\dot{\mu}_{\lambda,l}$ as an infinite sequence indexed by the elements of $S_1$ in increasing order, followed by the elements of $S_2$ in increasing order, and so on. For $k \geq k_0$ and $\alpha_{lbd} \leq \alpha \leq \alpha_{ubd}$, in the limit $l \to \infty$, $\dot{\mu}_{\lambda,l}$ converges in the $\ell^1$ sequence space to a limit $\dot{\mu}_\lambda \in \mathscr{P}$ satisfying the fixed point equation $\dot{\mu}_\lambda = \mathscr{R}_\lambda \dot{\mu}_\lambda$, as well as the estimates $\dot{\mu}_\lambda((0, 1)) \leq 7/2^k$ and $\dot{\mu}_\lambda(dx) = \dot{\mu}_\lambda(d(1-x))$.*

The limit $\dot{\mu}_\lambda$ of Proposition 1.2 encodes the desired replica symmetric solution $\dot{q}_\lambda$ for the auxiliary model. We can then express $\mathfrak{F}(\lambda)$ in terms of $\dot{\mu}_\lambda$ as follows. Writing $\hat{\mu}_\lambda \equiv \hat{\mathscr{R}}_\lambda \dot{\mu}_\lambda$, let $\dot{w}_\lambda, \hat{w}_\lambda, \bar{w}_\lambda \in \mathscr{P}$ be defined by

$$\dot{w}_\lambda(B) = (\dot{\mathfrak{Z}}_\lambda)^{-1} \int \left(\prod_{i=1}^{d} y_i + \prod_{i=1}^{d}(1-y_i)\right)^\lambda \mathbf{1}\left\{\prod_{i=1}^{d} y_i + \prod_{i=1}^{d}(1-y_i) \in B\right\} \prod_{i=1}^{d} \hat{\mu}_\lambda(dy_i)\,,$$
$$\hat{w}_\lambda(B) = (\hat{\mathfrak{Z}}_\lambda)^{-1} \int \left(1 - \prod_{i=1}^{k} x_i - \prod_{i=1}^{k}(1-x_i)\right)^\lambda \mathbf{1}\left\{1 - \prod_{i=1}^{k} x_i - \prod_{i=1}^{k}(1-x_i) \in B\right\} \prod_{i=1}^{k} \dot{\mu}_\lambda(dx_i)\,,$$
$$\bar{w}_\lambda(B) = (\bar{\mathfrak{Z}}_\lambda)^{-1} \iint \left(xy + (1-x)(1-y)\right)^\lambda \mathbf{1}\left\{xy + (1-x)(1-y) \in B\right\} \dot{\mu}_\lambda(dx) \hat{\mu}_\lambda(dy)\,, \quad (11)$$

with $\dot{\mathfrak{Z}}_\lambda, \hat{\mathfrak{Z}}_\lambda, \bar{\mathfrak{Z}}_\lambda$ the normalizing constants. The analogue of (9) for this model is

$$\left(\frac{\bar{\mathbf{Z}}_\lambda(\mathscr{G})}{\bar{\mathbf{Z}}_\lambda(\mathscr{G}'')}\right)^{1/k} = \dot{\mathfrak{Z}}_\lambda(\hat{\mathfrak{Z}}_\lambda/\bar{\mathfrak{Z}}_\lambda)^d, \quad \left(\frac{\bar{\mathbf{Z}}_\lambda(\mathscr{G}')}{\bar{\mathbf{Z}}_\lambda(\mathscr{G}'')}\right)^{1/k} = (\hat{\mathfrak{Z}}_\lambda)^{\alpha(k-1)},$$

and substituting into (8) gives the 1RSB prediction $\bar{\mathbf{Z}}_\lambda \doteq \exp\{\mathfrak{F}(\lambda)\}$ where

$$\mathfrak{F}(\lambda) \equiv \mathfrak{F}(\lambda; \alpha) \equiv \ln \dot{\mathfrak{Z}}_\lambda + \alpha \ln \hat{\mathfrak{Z}}_\lambda - k\alpha \ln \bar{\mathfrak{Z}}_\lambda\,. \quad (12)$$

Further, the maximizer of $s \mapsto (\lambda s + \Sigma(s))$ is predicted to be given by

$$s_\lambda \equiv s_\lambda(\alpha) \equiv \int \ln(x)\dot{w}_\lambda(dx) + \alpha \int \ln(x)\hat{w}_\lambda(dx) - k\alpha \int \ln(x)\bar{w}_\lambda(dx)\,. \quad (13)$$

If $s = s_\lambda$ for $\lambda \in [0, 1]$ then we define

$$\Sigma(s) \equiv \Sigma(s; \alpha) \equiv \mathfrak{F}(\lambda; \alpha) - \lambda s_\lambda(\alpha)\,. \quad (14)$$

We then use (14) to define the thresholds

$$\alpha_{\text{cond}} \equiv \sup\{\alpha : \Sigma(s_1; \alpha) > 0\}\,,$$
$$\alpha_{\text{sat}} \equiv \sup\{\alpha : \Sigma(s_0; \alpha) > 0\}\,.$$

We can now formally state the predicted free energy of the original NAE- SAT model:

**Definition 1.3** For $\alpha \in k^{-1}\mathbb{Z}$, 1RSB free energy prediction $\mathsf{f}^{\text{1RSB}}(\alpha)$ is defined as

$$\mathsf{f}^{\text{1RSB}}(\alpha) = \begin{cases} \mathsf{f}^{\text{RS}}(\alpha) = 2(1 - 2/2^k)^\alpha & \text{for } \alpha \le \alpha_{\text{cond}}, \\ \exp[\sup\{s : \Sigma(s) \ge 0\}] & \text{for } \alpha_{\text{cond}} \le \alpha < \alpha_{\text{sat}} \\ 0 & \text{for } \alpha > \alpha_{\text{sat}}. \end{cases} \quad (15)$$

(In regular $k$-NAE- SAT we must have integer $d = k\alpha$, so we need not consider $\alpha \notin k^{-1}\mathbb{Z}$.)

**Proposition 1.4** (proved in "Appendix B" ) *Assume $k \ge k_0$ and write $A \equiv [\alpha_{lbd}, \alpha_{ubd}] \cap (k^{-1}\mathbb{Z})$.*

a. *For each $\alpha \in A$, the function $s \mapsto \Sigma(s; \alpha)$ is well-defined, continuous, and strictly decreasing in $s$.*
b. *For each $0 \le \lambda \le 1$, the function $\alpha \mapsto \Sigma(s_\lambda; \alpha) = \mathfrak{F}(\lambda) - \lambda s_\lambda$ is strictly decreasing with respect to $\alpha \in A$. There is a unique $\alpha_\lambda \in A$ such that $\Sigma(s_\lambda; \alpha)$ is nonnegative for all $\alpha \le \alpha_\lambda$, and is negative for all $\alpha > \alpha_\lambda$. Taking $\lambda = 0$ we recover the estimate (1); and taking $\lambda = 1$ we obtain in addition*

$$\alpha_{cond} = \alpha_1 = (2^{k-1} - 1)\ln 2 + \epsilon_k\,. \quad (16)$$

*(The main purpose of this proposition is to show that $\Sigma(s_1) < 0$ for all $\alpha \in (\alpha_{cond}, \alpha_{sat})$, i.e., that the "condensation regime" is a contiguous range of values of $\alpha$. The expansion of $\alpha_{cond}$ matches an earlier result of [20], which was obtained for a slightly different but closely related model.)*

### 1.7 Proof approach

Since $\mathsf{f} = \mathsf{f}(\alpha)$ is *a priori* not well-defined, the statement $\mathsf{f} \le \mathsf{g}$ means formally that for all $\epsilon > 0$, $\mathbb{P}(Z^{1/n} \ge \mathsf{g} + \epsilon)$ tends to zero as $n \to \infty$. With this notation, we

will prove separately the upper bound $\mathsf{f}(\alpha) \le \mathsf{f}^{\mathrm{1RSB}}(\alpha)$ and the matching lower bound $\mathsf{f}(\alpha) \ge \mathsf{f}^{\mathrm{1RSB}}(\alpha)$. This implies the main result Theorem 1: the free energy $\mathsf{f}(\alpha)$ is indeed well-defined, and equals $\mathsf{f}^{\mathrm{1RSB}}(\alpha)$.

The upper bound is proved by an interpolation argument, which we defer to "Appendix E". This argument builds on similar bounds for spin glasses on Erdős–Rényi graphs [26,43], together with ideas from [12,27] for interpolation in random regular models. Let $Z_n(\beta)$ denote the partition function of NAE- SAT at inverse temperature $\beta > 0$. The interpolation method yields an upper bound on $\mathbb{E} \ln Z_n(\beta)$ which is expressed as the infimum of a certain function $\mathcal{P}(\mu; \beta)$, with $\mu$ ranging over probability measures on $[0, 1]$. We then choose $\mu$ according to Proposition 1.2, and take $\beta \to \infty$ to obtain the desired bound $\mathsf{f}(\alpha) \le \mathsf{f}^{\mathrm{1RSB}}(\alpha)$.

Most of the paper is devoted to establishing the matching lower bound. The proof strategy is inspired by the physics picture described above, and at a high level proceeds as follows. Take any $\lambda$ such that $\Sigma(s_\lambda)$ (as defined by (13) and (14)) is nonnegative, and let $\boldsymbol{Y}_\lambda$ be the number of clusters of size roughly $\exp\{n s_\lambda\}$. (As discussed in §1.3, we shall think of a cluster as a connected component of the solution space.) The informal statement of what we show is that

$$\boldsymbol{Y}_\lambda \doteq \exp\{n[\lambda s_\lambda + \Sigma(s_\lambda)]\}\,. \tag{17}$$

Adjusting $\lambda$ as indicated by (15) then proves the desired bound $\mathsf{f}(\alpha) \ge \mathsf{f}^{\mathrm{1RSB}}(\alpha)$.

Proving a formalized version of (17) occupies a significant part of the present paper. We introduce a slightly modified version of the messages $\mathtt{m}$ which record the topologies of the free trees $\boldsymbol{T}$. We then restrict to cluster encodings in which every free tree has fewer than $T$ variables, which limits the distance that information can propagate between free variables. We prove a version of (17) for every fixed $T$, and show that this yields the sharp lower bound in the limit $T \to \infty$. The proof of (17) for fixed $T$ is via the moment method for the auxiliary model, which boils down to a complicated optimization problem over many dimensions. It is known (see e.g. [25, Lem. 3.6]) that stationary points of the optimization problem correspond to "generalized" BP fixed points—these are measures $Q_{v\to a}(\mathtt{m}_{v\to a}, \mathtt{m}_{a\to v})$, rather than the simpler "one-sided" measures $q_{v\to a}(\mathtt{m}_{v\to a})$ considered in the 1RSB heuristic.

The one-sided property is a crucial simplification in physics calculations (cf. [33, Proposition 19.4]), but is challenging to prove in general. One contribution of this work that we wish to highlight is a novel resampling argument which yields a reduction to one-sided messages, and allows us to solve the moment optimization problem. (We are helped here by the truncation on the sizes of free trees.) Furthermore, the approach allows us to bring in methods from large deviations theory. With these we can show that the objective function has negative-definite Hessian at the optimizer, which is necessary for the second moment method. This resampling approach is quite general and should apply in a broad range of models.

### 1.8 Open problems

Beyond the free energy, it remains a challenge to establish the full picture predicted by statistical physicists for $\alpha \le \alpha_{\rm sat}$. We refer the reader to several recent works targeted at a broad class of models in the regime $\alpha \le \alpha_{\rm cond}$ [9,16,19], and to work on the location on $\alpha_{\rm cond}$ in a general family of models [14]. In the condensation regime $(\alpha_{\rm cond}, \alpha_{\rm sat})$, an initial step would be to show that most solutions lie within a bounded number of clusters. A much more refined prediction is that the mass distribution among the largest clusters forms a Poisson–Dirichlet process. Another question is to show that on a typical problem instance over $n$ variables, if $\underline{\boldsymbol{x}}^1, \underline{\boldsymbol{x}}^2$ are sampled independently and uniformly at random from the solutions of that instance, then the normalized overlap $R_{1,2} \equiv n^{-1}\{v : \boldsymbol{x}_v^1 = \boldsymbol{x}_v^2\}$ concentrates on two values (corresponding roughly to the two cases that $\underline{\boldsymbol{x}}^1, \underline{\boldsymbol{x}}^2$ come from the same cluster, or from different clusters)—this criterion is sometimes taken as the precise definition of 1RSB. During the final revision stage of this paper, some of the above questions were addressed by a new preprint [40].

Beyond the immediate context of random CSPs, understanding the condensation transition may deepen our understanding of the stochastic block model, a model for random networks with underlying community structure. Here again ideas from statistical physics have played an important role [21]. A great deal is now known rigorously for the case of two blocks [32,37], where there is no condensation regime. For models with more than two blocks, however, it is predicted that the condensation can occur, and may define a regime where detection is information-theoretically possible but computationally intractable. A condensation threshold has been established for the anti-ferromagnetic Potts model, corresponding to the disassortative regime of the stochastic block model. An analogous transition is expected in the ferromagnetic (assortative) case, and this remains open.

## 2 Combinatorial model

In this section we formalize a combinatorial model of clusters, which allows us to rigorously lower bound the tilted cluster partition function (5). We begin by reviewing the (standard) graphical depiction of NAE- SAT. A **not-all-equal-SAT (NAE-SAT)** problem instance is naturally represented by a **bipartite factor graph** $\mathscr{G}$ with signed edges, as follows. The vertex set of $\mathscr{G}$ is partitioned into a set $V = \{v_1, \dots, v_n\}$ of variables and a set $F = \{a_1, \dots, a_m\}$ of clauses; we then have a set $E$ of edges joining variables to clauses. For each edge $e \in E$ we write $v(e)$ for the incident variable, and $a(e)$ for the incident clause; and we assign an edge literal $\mathtt{L}_e \in \{\mathtt{0}, \mathtt{1}\}$ to indicate whether $v(e)$ participates affirmatively ($\mathtt{L}_e = \mathtt{0}$) or negatively ($\mathtt{L}_e = \mathtt{1}$) in $a(e)$. We define all edges to have length one-half, so two variables $v \ne v'$ lie at unit distance if and only if they appear in the same clause. Throughout this paper we denote $\mathcal{G} \equiv (V, F, E)$ for the bipartite graph without edge literals, and $\mathscr{G} \equiv (V, F, E, \underline{\mathtt{L}}) \equiv (\mathcal{G}, \underline{\mathtt{L}})$ for the NAE- SAT instance.

Formally we regard the edges $E$ as a permutation, as follows. Each variable $v \in V$ has incident half-edges $\delta v$, while each clause $a \in F$ has incident half-edges $\delta a$. Write $\delta V$ for the labelled set of all variable-incident half-edges, and $\delta F$ for the labelled set of

all clause-incident half-edges; we require that $|\delta V| = |\delta F| = \ell$. Then any permutation $\mathfrak{m}$ of $[\ell] \equiv \{1, \ldots, \ell\}$ defines $E$ by defining a matching between $\delta V$ and $\delta F$. Note that any permutation of $[\ell]$ is permitted, so multi-edges can occur. In this paper we assume that the graph is $(d, k)$-regular: each variable has $d$ incident edges, and each clause has $k$ incident edges, so $|E| = nd = mk$. A **random $k$-NAE-SAT instance** is given by $\mathscr{G} = (V, F, E, \underline{\mathtt{L}})$ where $|V| = n$, $|F| = m$, $E$ is given by a uniformly random permutation $\mathfrak{m}$ of $[nd]$, and $\underline{\mathtt{L}}$ is a uniformly random sample from $\{0, 1\}^E$. We write $\mathbb{P}$ and $\mathbb{E}$ for probability and expectation over the law of $\mathscr{G}$.

**Definition 2.1** *(solutions and clusters)* A **solution** of the NAE- SAT problem instance $\mathscr{G} = (V, F, E, \underline{\mathtt{L}})$ is any assignment $\underline{\boldsymbol{x}} \in \{0, 1\}^V$ such that for all $a \in F$, $(\mathtt{L}_e \oplus \boldsymbol{x}_{v(e)})_{e \in \delta a}$ is neither identically $0$ nor identically $1$. Let $\mathsf{SOL}(\mathscr{G}) \subseteq \{0, 1\}^V$ denote the set of all solutions of $\mathscr{G}$, and define a graph on $\mathsf{SOL}(\mathscr{G})$ by connecting any pair of solutions at unit Hamming distance. The (maximal) connected components of the $\mathsf{SOL}(\mathscr{G})$ graph are the **solution clusters**, hereafter denoted $\mathsf{CL}(\mathscr{G})$.

The aim of this section is to establish that (under a certain restriction) the NAE-SAT solution clusters can be represented by a combinatorial model of what we will term "colorings." We will describe the correspondence in a few stages. Informally, the progression is given by

$$\begin{aligned}&\text{NAE- SAT solution clusters } \boldsymbol{\gamma} \leftrightarrow \text{frozen configurations } \underline{x} \leftrightarrow \\ &\leftrightarrow \text{warning configurations } \underline{y} \leftrightarrow \text{message configurations } \underline{\tau} \leftrightarrow \text{colorings } \underline{\sigma}. \end{aligned} \tag{18}$$

Each step of (18) is formalized below. As mentioned previously, the key feature of the last model is that the size of a cluster $\boldsymbol{\gamma}$ can be easily read off from its corresponding coloring $\underline{\sigma}$, as a product of local functions. Some steps of the correspondence (18) appear in existing literature (see [13,25,33,34,42]) but we present them here in detail for completeness.

### 2.1 Frozen and warning configurations

We introduce a new value $\mathtt{f}$ (free), and adopt the convention $0 \oplus \mathtt{f} \equiv \mathtt{f} \equiv 1 \oplus \mathtt{f}$. For $l \geq 1$ and $\underline{x} \in \{0, 1, \mathtt{f}\}^l$ let $I^{\mathrm{NAE}}(\underline{x})$ be the indicator that $\underline{x}$ is neither identically $0$ nor identically $1$. Given an NAE- SAT instance $\mathscr{G} = (V, F, E, \underline{\mathtt{L}})$ and an assignment $\underline{x} \in \{0, 1, \mathtt{f}\}^V$, denote

$$I^{\mathrm{NAE}}(\underline{x}; \mathscr{G}) \equiv \prod_{a \in F} I^{\mathrm{NAE}}((\mathtt{L}_e \oplus x_{v(e)})_{e \in \delta a}) \,.$$

By Definition 2.1, an NAE- SAT solution is an assignment $\underline{\boldsymbol{x}} \in \{0, 1\}^V$ satisfying $I^{\mathrm{NAE}}(\underline{\boldsymbol{x}}; \mathscr{G}) = 1$.

**Definition 2.2** *(frozen configurations)* Given an NAE- SAT instance $\mathscr{G} = (V, F, E, \underline{\mathtt{L}})$, for any $e \in E$ let $\mathscr{G} \oplus 1_e$ denote the instance obtained by flipping the edge label $\mathtt{L}_e$ to $\mathtt{L}_e \oplus 1$. We say that $\underline{x} \in \{0, 1, \mathtt{f}\}^V$ is a valid **frozen configuration** on $\mathscr{G}$ if (i) no

NAE- SAT constraint is violated, meaning $I^{\mathrm{NAE}}(\underline{x}; \mathscr{G}) = 1$; and (ii) for all $v \in V$, $x_v$ takes a value in $\{0, 1\}$ only when **forced** to do so, meaning there is some $e \in \delta v$ such that

$$I^{\mathrm{NAE}}(\underline{x}; \mathscr{G} \oplus 1_e) = 0\,. \tag{19}$$

If no such $e \in \delta v$ exists then $x_v = \mathtt{f}$.

It is well known that on any given $\mathscr{G}$, every NAE- SAT solution $\underline{\boldsymbol{x}}$ can be mapped to a frozen configuration $\underline{x} = \underline{x}(\underline{\boldsymbol{x}})$ via a "coarsening" or "whitening" procedure [42], as follows. Initialize $\underline{x} = \underline{\boldsymbol{x}}$. Then, whenever $x_v \in \{0, 1\}$ but there exists no $e \in \delta v$ such that (19) holds, update $x_v$ to $\mathtt{f}$. Iterate until no further updates can be made; the result is then a valid frozen configuration. Two NAE- SAT solutions $\underline{\boldsymbol{x}}$, $\underline{\boldsymbol{x}}'$ map to the same frozen configuration $\underline{x}$ if and only if they lie in the same cluster $\boldsymbol{\gamma} \in \mathsf{CL}(\mathscr{G})$. Thus, for any given $\mathscr{G}$, we have a well-defined mapping from clusters $\boldsymbol{\gamma}$ to frozen configurations $\underline{x}$. This map is one-to-one but not necessarily onto: for instance, the all-free assignment $\underline{x} \equiv \mathtt{f}$ is always trivially a valid frozen configuration, but on many instances $\mathscr{G}$ there is no solution cluster $\boldsymbol{\gamma} \in \mathsf{CL}(\mathscr{G})$ whose coarsening is $\underline{x} \equiv \mathtt{f}$. Since the aim is to lower bound the clusters, the lack of surjectivity must be addressed. We will do so momentarily (Definition 2.4 below), but first we review an useful alternative representation of frozen configurations:

**Definition 2.3** *(warning configurations)* For the integers $l \geq 1$, define functions $\dot{Y} : \{0, 1, \mathtt{f}\}^l \to \{0, 1, \mathtt{f}, \mathtt{z}\}$ and $\hat{Y} : \{0, 1, \mathtt{f}\}^l \to \{0, 1, \mathtt{f}\}$ by

$$\dot{Y}(\underline{\hat{y}}) = \begin{cases} 0 & \text{if } 0 \in \{\hat{y}_i\} \subseteq \{0, \mathtt{f}\}, \\ 1 & \text{if } 1 \in \{\hat{y}_i\} \subseteq \{1, \mathtt{f}\}, \\ \mathtt{f} & \text{if } \{\hat{y}_i\} = \mathtt{f}, \\ \mathtt{z} & \text{otherwise;} \end{cases} \qquad \hat{Y}(\underline{\dot{y}}) = \begin{cases} 0 & \text{if } \{\dot{y}_i\} = \{1\}, \\ 1 & \text{if } \{\dot{y}_i\} = \{0\}; \\ \mathtt{f} & \text{otherwise.} \end{cases}$$

Denote $M \equiv \{0, 1, \mathtt{f}\}^2$. We write $\underline{y} \in M^E$ if $\underline{y} = (y_e)_{e \in E}$ where $y_e \equiv (\dot{y}_e, \hat{y}_e) \in M$. If edge $e$ joins variable $v$ to clause $a$, then $\dot{y}_e$ represents a "warning" along $e$ from $v$ to $a$, while $\hat{y}_e$ represents a "warning" along $e$ from $a$ to $v$. We say that $\underline{y} \in M^E$ is a valid **warning configuration** on $\mathscr{G}$ if it satisfies the local equations

$$y_e = (\dot{y}_e, \hat{y}_e) = \Big(\dot{Y}(\underline{\hat{y}}_{\delta v(e) \setminus e}), \mathtt{L}_e \oplus \hat{Y}((\underline{\mathtt{L}} \oplus \underline{\dot{y}})_{\delta a(e) \setminus e})\Big) \tag{20}$$

for all $e \in E$ (with no $\dot{y}_e = \mathtt{z}$).

It is well known that on any given $\mathscr{G}$ there is a natural bijection

$$\left\{ \begin{array}{c} \text{frozen configurations} \\ \underline{x} \in \{0, 1, \mathtt{f}\}^V \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{c} \text{warning configurations} \\ \underline{y} \in M^E \end{array} \right\}. \tag{21}$$

In the forward direction, given a (valid) frozen configuration $\underline{x}$, for any variable $v$ and any edge $e \in \delta v$ such that (19) holds, set $\hat{y}_e = x_v \in \{0, 1\}$; then in all other cases set $\hat{y}_e = \mathtt{f}$. Then, having defined all the $\hat{y}_e$, the $\dot{y}_e$ can only be defined by the local Eq.

(20). One can check that the resulting assignment $\underline{y} \in M^E$ is a warning configuration. Conversely, given a warning configuration $\underline{y}$, a frozen configuration $\underline{x}$ can be obtained by setting $x_v = \dot{Y}(\hat{y}_{\delta v})$ for all $v$.

### 2.2 Message configurations

We return to the question of surjectivity: does a given frozen configuration $\underline{x}$ encode a (nonempty) solution cluster $\boldsymbol{\gamma} \in \mathsf{CL}(\mathscr{G})$? We will now state an easy sufficient condition for this to hold. The condition is not in general necessary, but we will show that it captures enough of the solution space to deliver a sharp lower bound on the free energy.

**Definition 2.4** *(free cycles)* Let $\underline{x} \in \{0, 1, \mathtt{f}\}^V$ be a valid frozen configuration on $\mathscr{G} = (V, F, E, \underline{\mathtt{L}})$. We say that a clause $a \in F$ is **separating** (with respect to $\underline{x}$) if there exist $e', e'' \in \delta a$ such that $\mathtt{L}_{e'} \oplus x_{v(e')} = 0$ while $\mathtt{L}_{e''} \oplus x_{v(e'')} = 1$. For instance, a forcing clause is also separating. A cycle in $\mathscr{G}$ is a sequence of edges

$$e_1 e_2 \dots e_{2\ell-1} e_{2\ell} e_1,$$

where, taking indices modulo $2\ell$, it holds for each integer $i$ that $e_{2i-1}$ and $e_{2i}$ are distinct but share a clause, while $e_{2i}$ and $e_{2i+1}$ are distinct but share a variable. (In particular, if $v$ is joined to $a$ by two edges $e' \neq e''$, then $e'e''$ forms a cycle.) We say the cycle in $\mathscr{G}$ is **free** (with respect to $\underline{x}$) if all its variables are free and all its clauses are non-separating.

**Definition 2.5** *(free trees)* Let $\underline{x}$ be a frozen configuration on $\mathscr{G} = (V, F, E, \underline{\mathtt{L}})$ that has no free cyces. Let $H$ be the subgraph of $\mathscr{G}$ induced by the free variables and non-separating clauses of $\underline{x}$. Since $\underline{x}$ has no free cycles, $H$ must be a disjoint union of tree components $\boldsymbol{t}$, which we term the **free trees** of $\underline{x}$. For each $\boldsymbol{t}$, let $\boldsymbol{T} \equiv \boldsymbol{T}(\boldsymbol{t})$ be the subgraph of $\mathscr{G}$ induced by $\boldsymbol{t}$ together with its incident variables. The subgraphs $\boldsymbol{T}$ (which can contain cycles) will be termed the **free pieces** of $\underline{x}$. Each free variable is covered by exactly one free piece. In the simplest case, a free piece consists of a single free variable surrounded by $d$ separating clauses.

Let us say that $\underline{\boldsymbol{x}} \in \{0, 1\}^V$ **extends** $\underline{x} \in \{0, 1, \mathtt{f}\}^V$ if $\boldsymbol{x}_v = x_v$ for all $v$ such that $x_v \in \{0, 1\}$. If $\underline{x}$ is a frozen configuration on $\mathscr{G}$ with no free cycles, it is easy to extend $\underline{x}$ to valid NAE- SAT solutions $\underline{\boldsymbol{x}} \in \{0, 1\}^V$—we simply extend $\underline{x}$ on each free tree $\boldsymbol{t}$, since NAE- SAT on a tree is always solvable; the different free trees do not interact. Let $\boldsymbol{\gamma}$ denote the set of all valid NAE- SAT solutions on $\mathscr{G}$ that extend $\underline{x}$, and denote $\mathsf{size}(\underline{\mathrm{x}}) \equiv |\boldsymbol{\gamma}|$. Meanwhile, let $\mathfrak{T}(\underline{x})$ denote the set of all free pieces of $\underline{x}$. For each $\boldsymbol{T} \in \mathfrak{T}(\underline{x})$, let $\mathsf{size}(\underline{\mathrm{x}}; \boldsymbol{T})$ denote the number of valid NAE- SAT solutions on $\boldsymbol{T}$ that extend $\underline{x}|_{\boldsymbol{T}}$. It follows from our discussion that $\boldsymbol{\gamma} \in \mathsf{CL}(\mathscr{G})$ with

$$|\boldsymbol{\gamma}| = \mathsf{size}(\underline{\mathrm{x}}) = \prod_{\boldsymbol{T} \in \mathfrak{T}(\underline{\mathrm{x}})} \mathsf{size}(\underline{\mathrm{x}}; \boldsymbol{T})\,. \tag{22}$$

That is to say, the absence of free cycles is an easy sufficient condition for a frozen configuration to encode a nonempty cluster; and it further ensures that the cluster has a

relatively simple product structure (22). As noted previously, the structure within each free piece $\boldsymbol{T}$ can be understood by dynamic programming (BP). This is a well-known calculation (see e.g. [33, Ch. 14]) but we will review the details for our setting. To this end, we first introduce a combinatorial model of "message configurations" which will map directly to the natural BP variables.

Recall from Definition 2.3 that a warning configuration is denoted $\underline{y} \in M^E$ where each $y_e \equiv (\dot{y}_e, \hat{y}_e) \in M$. We denote a message configuration by $\underline{\tau} \in \mathscr{M}^E$ where each $\tau_e = (\dot{\tau}_e, \hat{\tau}_e) \in \mathscr{M}$ (for $\mathscr{M}$ to be defined below). It will be convenient to let E indicate a directed edge, pointing from tail vertex $t(\mathrm{E})$ to head vertex $h(\mathrm{E})$. If $e$ is the undirected version of E, then we denote

$$(y_{\mathrm{E}}, \tau_{\mathrm{E}}) = \begin{cases} (\dot{y}_e, \dot{\tau}_e) & \text{if } t(\mathrm{E}) \text{ is a variable,} \\ (\hat{y}_e, \hat{\tau}_e) & \text{if } t(\mathrm{E}) \text{ is a clause.} \end{cases}$$

We will make a definition such that $\tau_{\mathrm{E}}$ either takes the value "$\star$" or is a bipartite factor tree. The tree is **unlabelled** except that one vertex is distinguished as the root, and some edges are assigned $\mathtt{0}$ or $\mathtt{1}$ values as explained below. The root of $\tau_{\mathrm{E}}$ is required to have degree one, and should be thought of as corresponding to the head vertex $h(\mathrm{E})$.

In the context of message configurations $\underline{\tau}$, we use "$\mathtt{0}$" or "$\mathtt{1}$" to stand for the tree consisting of a single edge which is labelled $\mathtt{0}$ or $\mathtt{1}$ and rooted at the endpoint corresponding to the head vertex—the root is the incident clause in the case of $\dot{\tau}$, the incident variable in the case of $\hat{\tau}$. We use $\mathtt{s}$ to stand for the tree consisting of a single unlabelled edge, rooted at the incident variable; this will be related to the situation of separating clauses from Definition 2.4. Given a collection of rooted trees $t_1, \ldots, t_\ell$ whose roots $o_1, \ldots, o_\ell$ are all of the same type (either all variable or all clauses), we define $t = \mathsf{join}(\mathrm{t}_1, \ldots, \mathrm{t}_\cdot)$ by identifying all the $o_i$ as a single vertex $o$, then adding an edge which joins $o$ to a new vertex $o'$. The vertex $o$ has the same type (variable or clause) as the $o_i$; and the vertex $o'$ is assigned the opposite type and becomes the root of $t$.

**Definition 2.6** *(message configurations)* Start with $\dot{\mathscr{M}}_0 \equiv \{\mathtt{0}, \mathtt{1}, \star\}$ and $\hat{\mathscr{M}}_0 \equiv \varnothing$, and suppose inductively that $\dot{\mathscr{M}}_t, \hat{\mathscr{M}}_t$ have been defined. For $\underline{\hat{\tau}} \in (\hat{\mathscr{M}}_t)^{d-1}$ and $\underline{\dot{\tau}} \in (\dot{\mathscr{M}}_t)^{k-1}$, let us abbreviate $\{\hat{\tau}_i\} \equiv \{\hat{\tau}_1, \ldots, \hat{\tau}_{k-1}\}$ and $\{\dot{\tau}_i\} \equiv \{\dot{\tau}_1, \ldots, \dot{\tau}_{d-1}\}$. Define

$$\dot{T}(\underline{\hat{\tau}}) = \begin{cases} \mathtt{0} & \text{if } \mathtt{0} \in \{\hat{\tau}_i\} \subseteq \hat{\mathscr{M}}_t \backslash \{\mathtt{1}\}, \\ \mathtt{1} & \text{if } \mathtt{1} \in \{\hat{\tau}_i\} \subseteq \hat{\mathscr{M}}_t \backslash \{\mathtt{0}\}, \\ \mathtt{z} & \text{if } \{\mathtt{0}, \mathtt{1}\} \subseteq \{\hat{\tau}_i\}, \\ \star & \text{if } \star \in \{\hat{\tau}_i\} \subseteq \hat{\mathscr{M}}_t \backslash \{\mathtt{0}, \mathtt{1}\}, \\ \mathsf{join}\{\hat{\tau}_i\} & \text{if } \{\hat{\tau}_i\} \subseteq \hat{\mathscr{M}}_t \backslash \{\mathtt{0}, \mathtt{1}, \star\}; \end{cases} \qquad \hat{T}(\underline{\dot{\tau}}) = \begin{cases} \mathtt{0} & \text{if } \{\dot{\tau}_i\} = \{\mathtt{1}\}, \\ \mathtt{1} & \text{if } \{\dot{\tau}_i\} = \{\mathtt{0}\}, \\ \mathtt{s} & \text{if } \{\mathtt{0}, \mathtt{1}\} \subseteq \{\dot{\tau}_i\}, \\ \star & \text{if } \{\mathtt{0}, \mathtt{1}\} \not\subseteq \{\dot{\tau}_i\} \text{ and } \star \in \{\dot{\tau}_i\}, \\ \mathsf{join}\{\tau_i\} & \text{otherwise.} \end{cases}$$

Then, for $t \geq 0$, define recursively the sets

$$\hat{\mathscr{M}}_{t+1} \equiv \hat{\mathscr{M}}_t \cup \hat{T}[(\dot{\mathscr{M}}_t)^{k-1}],$$
$$\dot{\mathscr{M}}_{t+1} \equiv \dot{\mathscr{M}}_t \cup \dot{T}[(\hat{\mathscr{M}}_{t+1})^{d-1}] \backslash \{\mathtt{z}\}$$

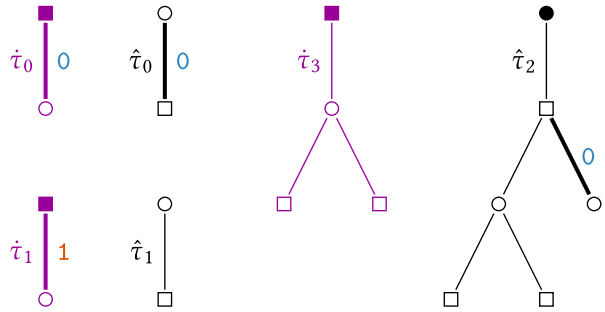



**Fig. 1** Examples of messages (Definition 2.6). Variables are indicated by circle nodes, clauses by square nodes, and edges by lines. For simplicity we assume that all edges depicted have literals $\mathtt{L}_e = \mathtt{0}$. Each message is shown with its root as a filled node at the top. The variable-to-clause messages $\dot\tau$ are rooted at clauses, while the clause-to-variable messages $\hat\tau$ are rooted at variables. The heavy lines indicates edges inside the message that are labelled $\mathtt{0}$ or $\mathtt{1}$. In our notation we have $\dot\tau_0 = \mathtt{0}$ and $\dot\tau_1 = \mathtt{1}$. Next $\hat\tau_0 = \hat T(\dot\tau_1, \dot\tau_1) = \mathtt{0}$, while $\hat\tau_1 = \hat T(\dot\tau_0, \dot\tau_1) = \mathtt{s}$. Finally $\dot\tau_3 = \dot T(\hat\tau_1, \hat\tau_1) = \mathsf{join}(\mathtt{0}, \mathtt{0}_1)$ and $\hat\tau_2 = \hat T(\dot\tau_3, \dot\tau_0) = \mathsf{join}(\mathtt{1}_3, \mathtt{1}_0)$

We then let $\dot{\mathscr{M}}$ be the union of all the $\dot{\mathscr{M}}_t$, let $\hat{\mathscr{M}}$ be the union of all the $\hat{\mathscr{M}}_t$, and let $\mathscr{M} = \dot{\mathscr{M}} \times \hat{\mathscr{M}}$. On $\mathscr{G} = (V, F, E, \underline{\mathtt{L}})$, the assignment $\underline{\tau} \in \mathscr{M}^E$ is a valid **message configuration** if (i) it satisfies the local equations

$$\tau_e = (\dot\tau_e, \hat\tau_e) = \left(\dot T(\underline{\hat\tau}_{\delta v(e)\setminus e}), \mathtt{L}_e \oplus \hat T((\underline{\mathtt{L}} \oplus \underline{\dot\tau})_{\delta a(e)\setminus e})\right) \tag{23}$$

for all $e \in E$ (with no $\dot\tau_e = \mathtt{z}$), and (ii) if one element of $\{\dot\tau_e, \hat\tau_e\}$ equals $\star$ then the other element is in $\{\mathtt{0}, \mathtt{1}\}$. In (23), we take the convention that $\mathtt{L}_e \oplus \mathtt{f} = \mathtt{f}$ and $\mathtt{L}_e \oplus \star = \star$, and if $\tau$ is a tree with labels then $\mathtt{L}_e \oplus \tau$ is defined by applying $\mathtt{L}_e \oplus \cdot$ entrywise to all labels of $\tau$. See Fig. 1.

Suppose $\underline{x}$ is a frozen configuration on $\mathscr{G}$, and let $\underline{y}$ be its corresponding warning configuration from (21). Given $\underline{y}$, we define $\underline{\tau}$ in four stages:

1. If $\dot y_e \in \{\mathtt{0}, \mathtt{1}\}$ then set $\dot\tau_e = \dot y_e$; likewise if $\hat y_e \in \{\mathtt{0}, \mathtt{1}\}$ then set $\hat\tau_e = \hat y_e$.
2. If $(\underline{\mathtt{L}} \oplus \underline{\dot y})_{\delta a(e)\setminus e}$ has both $\mathtt{0}$ and $\mathtt{1}$ entries, then set $\hat\tau_e = \mathtt{s}$.
3. Apply the local Eq. (23) recursively to define $\dot\tau_e$, $\hat\tau_e$ wherever possible.
4. Lastly, if any $\dot\tau_e$ or $\hat\tau_e$ remains undefined, then set it to $\star$.

An example with $\star$ messages is given in Fig. 2.

**Lemma 2.7** *The mapping described above defines a bijection*

$$\left\{\begin{array}{c}\textit{frozen configurations } \underline{x} \in \{0, 1, \mathtt{f}\}^V \\ \textit{without free cycles}\end{array}\right\} \longleftrightarrow \left\{\begin{array}{c}\textit{message configurations} \\ \underline{\tau} \in \mathscr{M}^E\end{array}\right\}.$$

***Proof*** Let $\underline{x} \in \{\mathtt{0}, \mathtt{1}, \mathtt{f}\}^V$ be a frozen configuration on $\mathscr{G} = (V, F, E, \underline{\mathtt{L}})$ without free cycles, and let $\underline{y} \in M^E$ be the warning configuration which corresponds to $\underline{x}$ via (21). We first check that the mapping $\underline{y} \mapsto \underline{\tau}$, as described above, gives a message configuration which is valid, i.e., satisfies conditions (i) and (ii) of Definition 2.6. In the first stage, the mapping procedure sets $\dot\tau_e = \dot y_e$ whenever $\dot y_e \in \{\mathtt{0}, \mathtt{1}\}$, and $\hat\tau_e = \hat y_e$

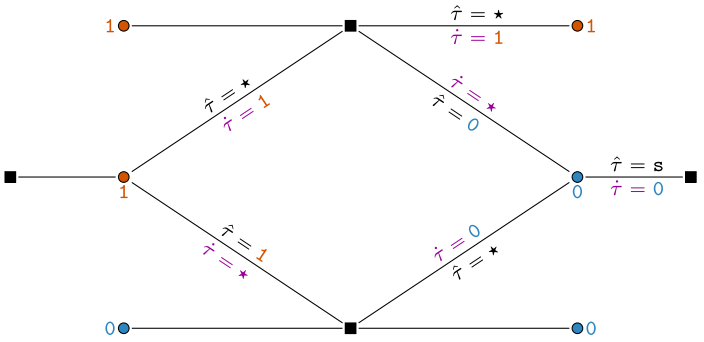



**Fig. 2** Example of how $\star$ messages can arise in the mapping from $\underline{y}$ to $\underline{\tau}$ (§2.2). The figure shows a subgraph of $\mathscr{G}$ with variables indicated by circle nodes, clauses by square nodes, and edges by lines. All edges in the figure are assumed to have label $\mathtt{L}_e = \mathtt{0}$. All variables shown are frozen to $\mathtt{0}$ or $\mathtt{1}$, and all clauses shown are separating. To avoid clutter we did not label the edges with the warnings $y_e$; instead, each variable $v$ is labeled with its frozen configuration spin $x_v$, according to the $\underline{x} \leftrightarrow \underline{y}$ bijection (21). The clauses force the variables along the cycle in the clockwise direction, resulting in $\star$ values in the final $\underline{\tau}$ in the counterclockwise direction of the cycle. (Note also that $\underline{y}$ can be recovered from $\underline{\tau}$ by changing $\star$ to $\mathtt{f}$; cf. Lemma 2.7)

whenever $\hat{y}_e \in \{\mathtt{0}, \mathtt{1}\}$. One can argue by induction that the rest of the procedure does not create any additional $\mathtt{0}$ or $\mathtt{1}$ messages, so that in the final configuration the $\{\mathtt{0}, \mathtt{1}\}$ values of $\underline{\tau}$ will match those of $\underline{y}$. The second and third stages of the procedure are clearly consistent with the local Eq. (23). Note in particular that the third stage does not produce any $\dot\tau_e = \mathtt{z}$ message, because it would contradict the assumption that $\underline{y}$ is a valid warning configuration; it also does not produce any $\star$ message. All $\star$ messages are created in the fourth stage, and this is clearly consistent with the mapping of $\star$ messages under $\dot{T}$ and $\hat{T}$. This concludes the proof that $\underline{\tau}$ satisfies condition (i) of Definition 2.6. To check condition (ii), suppose $\tau_{\mathrm{E}} = \star$, and let $\mathrm{F}$ denote the reversal of $\mathrm{E}$. From the above construction, it must be that $y_{\mathrm{E}} = \mathtt{f}$ and $\tau_{\mathrm{E}'} = \star$ for some $\mathrm{E}'$ that points to the tail vertex $t(\mathrm{E})$ but does not equal $\mathrm{F}$. Consequently $\mathrm{E}$ must belong to a directed cycle $\mathrm{E}_1\mathrm{E}_2 \ldots \mathrm{E}_{2k}\mathrm{E}_1$ with all the $\tau_{\mathrm{E}_i}$ equal to $\star$. Whenever $\mathrm{E}$ points from a separating clause $a$ to free variable $v$, we must have $\tau_{\mathrm{E}} = \mathtt{s}$. As a result, if all the variables along the cycle are free, then none of the clauses can be separating, contradicting the assumption that $\underline{x}$ has no free cycles. Therefore some variable $v$ on the cycle must take value $x_v \in \{\mathtt{0}, \mathtt{1}\}$, and by relabelling we may assume $v = t(\mathrm{E}_1)$. Let $\mathrm{F}_i$ denote the reversal of $\mathrm{E}_i$: since $x_v \neq \mathtt{f}$ but $y_{\mathrm{E}_1} = \mathtt{f}$, it must be that $y_{\mathrm{F}_1} = x_v$. This means that the clause $a = h(\mathrm{E}_1) = t(\mathrm{F}_1)$ is forcing to $v$, so in particular $y_{\mathrm{F}_2} \in \{\mathtt{0}, \mathtt{1}\}$. Continuing in this way we see that $y_{\mathrm{F}_i} \in \{\mathtt{0}, \mathtt{1}\}$ for all $i$, and it follows that $\underline{\tau}$ satisfies condition (ii), and so is a valid message configuration.

The mapping from $\underline{y}$ to $\underline{\tau}$ is clearly injective. To see that it is surjective, let $\underline{\tau}$ be any valid message configuration. Projecting $\dot{\mathscr{M}}\backslash\{\mathtt{0}, \mathtt{1}\} \mapsto \mathtt{f}$ and $\hat{\mathscr{M}}\backslash\{\mathtt{0}, \mathtt{1}\} \mapsto \mathtt{f}$ yields a valid warning configuration $\underline{y}$, which in turn maps to a valid frozen configuration $\underline{x}$. It remains then to check that $\underline{x}$ has no free cycles. Indeed, along a free cycle, all the warnings (in either direction) must be $\mathtt{f}$. This means none of the messages can be in $\{\mathtt{0}, \mathtt{1}\}$, and as a result none of the messages can be $\star$, by condition (ii) of Definition 2.6. This means all the messages must be in $\dot{\mathscr{M}}\backslash\{\mathtt{0}, \mathtt{1}, \star\}$ or $\hat{\mathscr{M}}\backslash\{\mathtt{0}, \mathtt{1}, \star\}$. Suppose in one direction of the cycle we have the directed edges $\mathrm{E}_1\mathrm{E}_2 \ldots \mathrm{E}_{2k}\mathrm{E}_1$. By

definition of $\dot{T}$ and $\hat{T}$, $\tau_{\mathrm{E}_i}$ is a proper subtree of $\tau_{\mathrm{E}_{i+1}}$ for all $i$, with indices modulo $2k$. Going around the cycle we find that $\tau_{\mathrm{E}_1}$ is a proper subtree of $\tau_{\mathrm{E}_{2k+1}} = \tau_{\mathrm{E}_1}$, which gives the contradiction. □

### 2.3 Bethe formula

We now describe the dynamic programming (BP) calculation which will ultimately take a message configuration $\underline{\tau}$ and evaluate a product of local functions to compute the size of its associated cluster. The first step is to define the dynamic programming variables; these will formalize the measures $\mathtt{m}$ which were introduced previously in (7). Recall that for $l \geq 1$ and $\underline{x} \in \{0, 1, \mathtt{f}\}^l$, we write $I^{\mathrm{NAE}}(\underline{x})$ for the indicator that the entries of $\underline{x}$ are not identically $0$ or identically $1$.

**Definition 2.8** Recall that message configuration spins belong to the space $\mathscr{M} = \dot{\mathscr{M}} \times \hat{\mathscr{M}}$ (Definition 2.6). Let $\mathscr{P}(\{0, 1\})$ denote the space of probability measures on $\{0, 1\}$. Define the mappings $\dot{\mathtt{m}} : \dot{\mathscr{M}} \to \mathscr{P}(\{0, 1\})$ and $\hat{\mathtt{m}} : \hat{\mathscr{M}} \to \mathscr{P}(\{0, 1\})$ as follows. For $\dot{\tau} \in \{0, 1\}$ let $\dot{\mathtt{m}}(\dot{\tau})$ be the unit measure supported on $\dot{\tau}$. Likewise, for $\hat{\tau} \in \{0, 1\}$ let $\hat{\mathtt{m}}(\hat{\tau})$ be the unit measure supported on $\hat{\tau}$. For $\dot{\tau} \in \dot{\mathscr{M}} \backslash \{0, 1, \star\}$ or $\hat{\tau} \in \hat{\mathscr{M}} \backslash \{0, 1, \star\}$ we let $\dot{\mathtt{m}}(\dot{\tau})$ and $\hat{\mathtt{m}}(\hat{\tau})$ be recursively defined: if $\dot{\tau} = \dot{T}(\hat{\tau}_1, \ldots, \hat{\tau}_{d-1})$ where no $\hat{\tau}_j = \star$, define

$$\dot{z}(\dot{\tau}) \equiv \sum_{\boldsymbol{x} \in \{0,1\}} \prod_{i=1}^{d-1} [\hat{\mathtt{m}}(\hat{\tau}_i)](\boldsymbol{x}) \,, \quad [\dot{\mathtt{m}}(\dot{\tau})](\boldsymbol{x}) \equiv \frac{1}{\dot{z}(\dot{\tau})} \prod_{i=1}^{d-1} [\hat{\mathtt{m}}(\hat{\tau}_i)](\boldsymbol{x}) \,. \tag{24}$$

Note that $\hat{\tau}_1, \ldots, \hat{\tau}_{d-1}$ can be recovered from $\dot{\tau}$ modulo permutation of the indices, so these quantities are well-defined. We see inductively that for $\dot{\tau} \in \dot{\mathscr{M}} \backslash \{0, 1, \star\}$, the normalizing factor $\dot{z}(\dot{\tau})$ is positive, and $\dot{\mathtt{m}}(\dot{\tau})$ is a nondegenerate probability measure on $\{0, 1\}$. Similarly, if $\hat{\tau} \in \hat{\mathscr{M}} \backslash \{0, 1, \star\}$ equals $\hat{T}(\dot{\tau}_1, \ldots, \dot{\tau}_{k-1})$ where none of the $\dot{\tau}_i$ are $\star$, then set

$$\begin{aligned} \hat{z}(\hat{\tau}) &\equiv \sum_{\boldsymbol{x} \in \{0,1\}} \sum_{\underline{\dot{\boldsymbol{x}}} \in \{0,1\}^{k-1}} I^{\mathrm{NAE}}(\boldsymbol{x}, \underline{\dot{\boldsymbol{x}}}) \prod_{i=1}^{k-1} [\dot{\mathtt{m}}(\dot{\tau}_i)](\dot{\boldsymbol{x}}_i) = 2 - \prod_{i=1}^{k-1} [\dot{\mathtt{m}}(\dot{\tau}_i)](0) - \prod_{i=1}^{k-1} [\dot{\mathtt{m}}(\dot{\tau}_i)](1) \,, \\ [\hat{\mathtt{m}}(\hat{\tau})](\boldsymbol{x}) &\equiv \frac{1}{\hat{z}(\hat{\tau})} \sum_{\underline{\dot{\boldsymbol{x}}} \in \{0,1\}^{k-1}} I^{\mathrm{NAE}}(\boldsymbol{x}, \underline{\dot{\boldsymbol{x}}}) \prod_{i=1}^{k-1} [\dot{\mathtt{m}}(\dot{\tau}_i)](\dot{\boldsymbol{x}}_i) = \frac{1}{\hat{z}(\hat{\tau})} \left( 1 - \prod_{i=1}^{k-1} [\dot{\mathtt{m}}(\dot{\tau}_i)](\boldsymbol{x}) \right) . \end{aligned} \tag{25}$$

Again, we see inductively that for $\hat{\tau} \in \hat{\mathscr{M}} \backslash \{0, 1, \star\}$, the normalizing factor $\hat{z}(\hat{\tau})$ is positive, and $\hat{\mathtt{m}}(\hat{\tau})$ is a nondegenerate probability measure on $\{0, 1\}$. Finally, we will see below that for our purposes we can take $\dot{\mathtt{m}}(\star)$ and $\hat{\mathtt{m}}(\star)$ to be arbitrary nondegenerate probability measures on $\{0, 1\}$; we therefore define them both to equal the uniform measure on $\{0, 1\}$.

Given a valid message configuration $\underline{\tau}$ on $\mathscr{G}$, define $\underline{\mathtt{m}} = (\mathtt{m}_e)_{e \in E}$ where $\mathtt{m}_e \equiv (\dot{\mathtt{m}}_e, \hat{\mathtt{m}}_e)$ with $\dot{\mathtt{m}}_e \equiv \dot{\mathtt{m}}(\dot{\tau}_e)$ and $\hat{\mathtt{m}}_e \equiv \hat{\mathtt{m}}(\hat{\tau}_e)$. It follows from Definition 2.8 that $\underline{\mathtt{m}}$

satisfies the following local consistency equations, which are inherited from the Eq. (23) satisfied by $\underline{\tau}$, in combination with the above definitions (24) and (25). If $\dot{\tau}_e \neq \star$, then $\dot{\mathrm{m}}_e$ is given by the equation

$$\dot{\mathrm{m}}_e(\boldsymbol{x}) = \frac{1}{\dot{z}(\dot{\tau}_e)} \prod_{e' \in \delta v(e) \setminus e} \hat{\mathrm{m}}_{e'}(\boldsymbol{x}) \tag{26}$$

for $\boldsymbol{x} \in \{0, 1\}$. Likewise, if $\hat{\tau}_e \neq \star$, then $\hat{\mathrm{m}}_e$ is given by the equation

$$\hat{\mathrm{m}}_e(\boldsymbol{x}) = \frac{1}{\hat{z}(\hat{\tau}_e)} \sum_{\underline{\dot{\boldsymbol{x}}}_{\delta a(e)} \in \{0,1\}^d} \mathbf{1}\{\dot{\boldsymbol{x}}_e = \boldsymbol{x}\} I^{\mathrm{NAE}}((\underline{\dot{\boldsymbol{x}}} \oplus \underline{L})_{\delta a(e)}) \prod_{e' \in \delta a(e) \setminus e} \dot{\mathrm{m}}_{e'}(\boldsymbol{x}_{e'}) \tag{27}$$

for $\boldsymbol{x} \in \{0, 1\}$. The Eqs. (26) and (27) are known as the **BP equations**. We now proceed to the calculation of the cluster size (22). To this end, we define the local functions

$$\begin{aligned}
\bar{\varphi}(\dot{\tau}, \hat{\tau}) &\equiv \left\{ \sum_{\boldsymbol{x} \in \{0,1\}} \dot{\mathrm{m}}[\dot{\tau}](\boldsymbol{x}) \cdot \hat{\mathrm{m}}[\hat{\tau}](\boldsymbol{x}) \right\}^{-1}, \\
\hat{\varphi}^{\mathrm{lit}}(\dot{\tau}_1, \ldots, \dot{\tau}_k) &\equiv \sum_{\underline{\boldsymbol{x}} \in \{0,1\}^k} I^{\mathrm{NAE}}(\underline{\boldsymbol{x}}) \prod_{i=1}^{k} \dot{\mathrm{m}}[\dot{\tau}_i](\boldsymbol{x}_i) = 1 \\
&\quad - \sum_{\boldsymbol{x} \in \{0,1\}} \prod_{i=1}^{k} \dot{\mathrm{m}}[\dot{\tau}_i](\boldsymbol{x}) = \frac{\hat{z}(\hat{T}((\dot{\tau}_j)_{j \neq i}))}{\bar{\varphi}(\dot{\tau}_i, \hat{T}((\dot{\tau}_j)_{j \neq i}))}, \\
\dot{\varphi}(\hat{\tau}_1, \ldots, \hat{\tau}_d) &\equiv \sum_{\boldsymbol{x} \in \{0,1\}} \prod_{i=1}^{d} \hat{\mathrm{m}}[\hat{\tau}_i](\boldsymbol{x}) = \frac{\dot{z}(\dot{T}((\hat{\tau}_j)_{j \neq i}))}{\bar{\varphi}(\hat{\tau}_i, \dot{T}((\hat{\tau}_j)_{j \neq i}))},
\end{aligned} \tag{28}$$

where the last identity in the last two lines holds for any choice of $i$. The BP calculation is summarized by the following:

**Lemma 2.9** *Suppose on $\mathscr{G} = (V, F, E, \underline{L})$ that $\underline{x}$ is a frozen configuration with no free cycles, and let $\underline{\tau}$ be its corresponding message configuration from* Lemma 2.7. *Let $\boldsymbol{T} \in \mathfrak{T}(\underline{x})$ be a free piece of $\underline{x}$, and let $\boldsymbol{t}$ be the free tree inside it. Then the number of* NAE- SAT *extensions of $\underline{x}|_{\boldsymbol{T}}$ on $\boldsymbol{T}$ is given by*

$$\mathit{size}(\underline{x}; \boldsymbol{T}) = \prod_{v \in V(\boldsymbol{t})} \left\{ \dot{\varphi}(\underline{\tau}_{\delta v}) \prod_{e \in \delta v} \bar{\varphi}(\tau_e) \right\} \prod_{a \in F(\boldsymbol{t})} \hat{\varphi}^{lit}((\underline{\tau} \oplus \underline{L})_{\delta a}) \tag{29}$$

*where $V(\boldsymbol{t})$ and $F(\boldsymbol{t})$ denote respectively the variables and clauses in $\boldsymbol{t}$. (An example calculation is worked out in* Fig. 3.*)*

***Proof*** As we have mentioned before, this calculation is well known (see e.g. [33, Ch. 14]) but we will review it here, beginning with a minor technical point. As noted

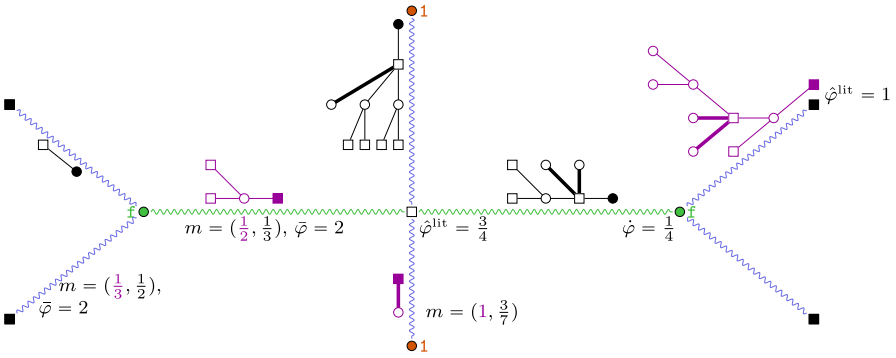



**Fig. 3** Example of correspondence (Lemma 2.7) between frozen and message configurations. Variables are indicated by circle nodes, clauses by square nodes, and edges by lines. The graph formed by the wavy lines is a free piece $\boldsymbol{T}$, with the free tree $\boldsymbol{t}$ in green and $\boldsymbol{T}\backslash\boldsymbol{t}$ in blue (Definition 2.5). Each variable $v$ is labelled with its frozen configuration spin value $x_v \in \{0, 1, N\}$. The four separating clauses are indicated by filled black squares. The message configuration is only partially shown, with the remaining values given by the obvious symmetries. The clause-to-variable messages $\hat{\tau}$ are shown in black, and the variable-to-clause messages $\dot{\tau}$ are shown in purple. Each message is a tree, with root vertex shown as a filled node. The heavy black and purple lines indicate edges inside the messages that are labeled 1. For instance, the message coming up out of the bottom variable is a tree consisting of a single edge, labelled 1 (indicated in the figure by a heavy purple line), rooted at its incident clause. We then calculate on each edge the values $m = (\dot{\mathrm{m}}[\dot{\tau}](1), \hat{\mathrm{m}}[\hat{\tau}](1))$, and use this to determine the factors $\dot{\varphi}$, $\hat{\varphi}^{\mathrm{lit}}$, $\bar{\varphi}$ from (28). In this example, $\boldsymbol{t}$ has two free variables each with $\dot{\varphi} = 1/4$, four separating clauses each with $\hat{\varphi}^{\mathrm{lit}} = 1$, and one non-separating clause with $\hat{\varphi}^{\mathrm{lit}} = 3/4$. There are six edges incident to $\boldsymbol{t}$, each with $\bar{\varphi} = 2$. Multiplying all these factors together (Lemma 2.9) gives $\mathsf{size}(\underline{x}; \boldsymbol{T}) = 3$. Indeed, in this small example it is easy to see that there are exactly three NAE- SAT assignments extending the frozen configuration $\underline{x}$ on $\boldsymbol{T}$, since the two free variables cannot both take value 1, but the remaining three possibilities give valid NAE- SAT assignments (color figure online)

in Definition 2.5, $\boldsymbol{t}$ is a tree but $\boldsymbol{T}$ has a cycle wherever a variable $v \in \boldsymbol{T}\backslash\boldsymbol{t}$ is joined by more than one edge to $\boldsymbol{t}$. However, since $\underline{x}|_{\boldsymbol{T}\backslash\boldsymbol{t}}$ is $\{0, 1\}$-valued, these cycles play no role in the question of extending $\underline{x}|_{\boldsymbol{T}}$ to a valid NAE- SAT assignment on $\boldsymbol{T}$—one can simply duplicate variables in $\boldsymbol{T}\backslash\boldsymbol{t}$ so that each one joins to $\boldsymbol{t}$ by exactly one edge. We may therefore assume for the rest of the proof that all the free pieces $\boldsymbol{T} \in \mathfrak{T}(\underline{x})$ are acyclic.

For any $\boldsymbol{T} \in \mathfrak{T}(\underline{x})$ and any edge $e \in \boldsymbol{T}$, delete from $\boldsymbol{T}$ the edges $\delta a(e)\backslash e$, and let $\dot{\boldsymbol{T}}_e$ denote the component containing $e$ in what remains, rooted at $a(e)$. Likewise, delete from $\boldsymbol{T}$ the edges $\delta v(e)\backslash e$, and let $\hat{\boldsymbol{T}}_e$ denote the component containing $e$ in what remains, rooted at $v(e)$. For each variable $w \in \dot{\boldsymbol{T}}_e\backslash\boldsymbol{t}$, let $\acute{x}_w \in \{0, 1\}$ be the boolean sum of $x_w$ together with all the edge literals L on the path joining $w$ to $a(e)$ in $\dot{\boldsymbol{T}}_e$. Note then that $\dot{\tau}_e$ encodes the isomorphism class of $\dot{\boldsymbol{T}}_e$, labelled with boundary data $\acute{x}_w$ (for all the variables $w \in \dot{\boldsymbol{T}}_e\backslash\boldsymbol{t}$). A similar relation holds between $\hat{\tau}_e$ and $\hat{\boldsymbol{T}}_e$. For each $e \in \boldsymbol{T}$, let $\dot{\mathsf{s}}_e(\boldsymbol{x}; \underline{x})$ count the number of valid NAE- SAT assignments that extend $\underline{x}|_{\dot{\boldsymbol{T}}_e}$ on $\dot{\boldsymbol{T}}_e$ and take value $\boldsymbol{x}$ on $v(e)$. Let $\hat{\mathsf{s}}_e(\boldsymbol{x}; \underline{x})$ count the number of valid NAE- SAT assignments that extend $\underline{x}|_{\hat{\boldsymbol{T}}_e}$ on $\hat{\boldsymbol{T}}_e$ and take value $\boldsymbol{x}$ on $v(e)$. Denote

$$\dot{\mathsf{s}}_e(\underline{x}) \equiv \sum_{\boldsymbol{x}\in\{0,1\}} \dot{\mathsf{s}}_e(\boldsymbol{x}; \underline{x}) , \quad \hat{\mathsf{s}}_e(\underline{x}) \equiv \sum_{\boldsymbol{x}\in\{0,1\}} \hat{\mathsf{s}}_e(\boldsymbol{x}; \underline{x}) .$$

There are two boundary cases: if edge $e$ joins a free variable in $\boldsymbol{t}$ to a separating clause in $\boldsymbol{T}\backslash\boldsymbol{t}$, then we have $\hat{\mathsf{s}}_e(0;\underline{x}) = \hat{\mathsf{s}}_e(1;\underline{x}) = 1$. If edge $e$ instead joins a non-separating clause in $\boldsymbol{t}$ to a frozen variable in $\boldsymbol{T}\backslash\boldsymbol{t}$, then we have $\dot{\mathsf{s}}_e(\boldsymbol{x};\underline{x}) = \mathbf{1}\{\boldsymbol{x} = x_{v(e)}\}$. By induction started from these boundary cases we find that for all $e \in \boldsymbol{T}$,

$$\dot{\mathrm{m}}_e(\boldsymbol{x}) = \frac{\dot{\mathsf{s}}_e(\boldsymbol{x};\underline{x})}{\dot{\mathsf{s}}_e(\underline{x})}\,, \quad \hat{\mathrm{m}}_e(\boldsymbol{x}) = \frac{\hat{\mathsf{s}}_e(\boldsymbol{x};\underline{x})}{\hat{\mathsf{s}}_e(\underline{x})}\,.$$

It follows that for any variable $v \in \boldsymbol{t}$, any clause $a \in \boldsymbol{t}$, and any edge $e \in \boldsymbol{T}$, we have the identities

$$\mathsf{size}(\underline{x};\boldsymbol{T}) = \dot{\varphi}(\underline{\hat{\tau}}_{\delta v}) \prod_{e'\in\delta v} \hat{\mathsf{s}}_{e'}(\underline{x}) = \hat{\varphi}^{\mathrm{lit}}(\underline{\dot{\tau}}_{\delta a}) \prod_{e'\in\delta a} \dot{\mathsf{s}}_{e'}(\underline{x}) = \frac{\dot{\mathsf{s}}_e(\underline{x})\hat{\mathsf{s}}_e(\underline{x})}{\bar{\varphi}(\tau_e)}\,.$$

Combining the identities and rearranging gives (writing $E(\boldsymbol{t})$ for the edges of $\boldsymbol{t}$)

$$\prod_{v\in V(\boldsymbol{t})} \left\{\dot{\varphi}(\underline{\hat{\tau}}_{\delta v}) \prod_{e\in\delta v} \bar{\varphi}(\tau_e)\right\} \prod_{a\in F(\boldsymbol{t})} \hat{\varphi}^{\mathrm{lit}}((\underline{\dot{\tau}} \oplus \underline{\mathrm{L}})_{\delta a})$$
$$= \frac{\mathsf{size}(\underline{\mathrm{x}};\boldsymbol{T})^{|\mathrm{V}(\boldsymbol{t})|+|\mathrm{F}(\boldsymbol{t})|}}{\mathsf{size}(\underline{\mathrm{x}};\boldsymbol{T})^{|\mathrm{E}(\boldsymbol{t})|}} \cdot \left\{\prod_{v\in V(\boldsymbol{t})} \prod_{e\in\delta v\backslash\boldsymbol{t}} \frac{\dot{\mathsf{s}}_e(\underline{x})}{\mathsf{size}(\underline{\mathrm{x}};\boldsymbol{T})}\right\} \Big/ \left\{\prod_{a\in F(\boldsymbol{t})} \prod_{e\in\delta a\backslash\boldsymbol{t}} \dot{\mathsf{s}}_e(\underline{x})\right\}.$$

For $a \in F(\boldsymbol{t})$ and $e \in \delta a\backslash\boldsymbol{t}$, the variable $v(e)$ is frozen and so we have $\dot{\mathsf{s}}_e(\underline{x}) = 1$. For any $v \in V(\boldsymbol{t})$ and $e \in \delta v\backslash\boldsymbol{t}$, we have $\dot{\mathsf{s}}_e(\underline{x}) = \mathsf{size}(\underline{\mathrm{x}};\boldsymbol{T})$. The tree $\boldsymbol{t}$ has Euler characteristic one. The right-hand side of the above equation then simplifies to $\mathsf{size}(\underline{\mathrm{x}};\boldsymbol{T})$, thereby proving the claim. □

**Corollary 2.10** *Suppose on $\mathscr{G} = (V, F, E, \underline{L})$ that $\underline{x}$ is a frozen configuration with no free cycles, and let $\underline{\tau}$ be its corresponding message configuration from* Lemma 2.7. *Then the number of valid* NAE- SAT *extensions of $\underline{x}$ is given by the product formula*

$$\mathit{size}(\underline{x}) = \prod_{v\in V} \dot{\varphi}(\underline{\hat{\tau}}_{\delta v}) \prod_{a\in F} \hat{\varphi}^{\mathit{lit}}((\underline{\dot{\tau}} \oplus \underline{L})_{\delta a}) \prod_{e\in E} \bar{\varphi}(\tau_e)\,.$$

*This identity holds as long as $\hat{m}(\star)$ and $\hat{m}(\star)$ are fixed nondegenerate probability measures on $\{0, 1\}$.*

***Proof*** Let $V'$ denote the set of free variables, and let $E'$ denote the set of all edges incident to $V'$. Let $F'$ the set of non-separating clauses. From (22) and Lemma 2.9 we have

$$\mathsf{size}(\underline{x}) = \prod_{\boldsymbol{T}\in\mathfrak{T}(\underline{x})} \mathsf{size}(\boldsymbol{x};\boldsymbol{T}) = \prod_{v\in V'} \dot{\varphi}(\underline{\hat{\tau}}_{\delta v}) \prod_{a\in F'} \hat{\varphi}^{\mathrm{lit}}((\underline{\dot{\tau}} \oplus \underline{\mathrm{L}})_{\delta a}) \prod_{e\in E'} \bar{\varphi}(\tau_e)\,. \tag{30}$$

For any edge $e \in E\backslash E'$, the incident variable $v(e)$ must lie in $V\backslash V'$, meaning $x_{v(e)} \in \{0, 1\}$. We now partition $E\backslash E'$ into the disjoint union of $E_{\mathrm{r}}$ and $E_{\mathrm{b}}$, as follows. Let $E_{\mathrm{r}}$

be the set of edges $e \in E \backslash E'$ such that $\hat{\mathrm{m}}_e$ is fully supported on $x_{v(e)}$. Let $E_{\mathtt{b}}$ be the set of edges $e \in E \backslash E'$ such that $\hat{\mathrm{m}}_e$ is a nondegenerate measure on $\{0, 1\}$; note that $\dot{\mathrm{m}}_e$ must then be fully supported on $x_{v(e)}$. Consider a clause $a \in F \backslash F'$. If $a$ is non-forcing, then $\delta a \cap E_{\mathtt{r}} = \varnothing$ and $\hat{\varphi}^{\mathrm{lit}}((\underline{\dot{\tau}} \oplus \underline{\mathtt{L}})_{\delta a}) = 1$. Otherwise, $a$ is forcing in the direction of some edge $e \in \delta a$, in which case $\delta a \cap E_{\mathtt{r}} = \{e\}$ and $\hat{\varphi}^{\mathrm{lit}}((\underline{\dot{\tau}} \oplus \underline{\mathtt{L}})_{\delta a}) = \dot{\mathrm{m}}_e(x_{v(e)}) = 1/\bar{\varphi}(\tau_e)$. We conclude for all $a \in F \backslash F'$ that

$$\hat{\varphi}^{\mathrm{lit}}((\underline{\dot{\tau}} \oplus \underline{\mathtt{L}})_{\delta a}) \prod_{e \in \delta a \cap E_{\mathtt{r}}} \bar{\varphi}(\tau_e) = 1 \,. \tag{31}$$

For $v \in V \backslash V'$, for all $e \in \delta v \cap E_{\mathtt{b}}$ we have $\dot{\mathrm{m}}_e(x_v) = 1$ and so $\bar{\varphi}(\tau_e) = 1/\hat{\mathrm{m}}(x_v)$. Thus, for all $v \in V \backslash V'$,

$$\dot{\varphi}(\underline{\hat{\tau}}_{\delta v}) \prod_{e \in \delta v \cap E_{\mathtt{b}}} \bar{\varphi}(\tau_e) = 1 \,. \tag{32}$$

The identities (31) and (32) remain valid even for vertices incident to $\star$ messages, as long as $\dot{\mathrm{m}}(\star)$ and $\hat{\mathrm{m}}(\star)$ are fixed nondegenerate probability measures on $\{0, 1\}$. Combining the identities with (30) proves the claim. □

### 2.4 Colorings

We conclude this section by defining the coloring model, building on an encoding introduced by [18]. It is a simplification of the message configuration model (Definition 2.6) that takes advantage of some of the cancellations ((31) and (32)) seen above. In short, following the notation of Corollary 2.10, for edges in $E \backslash E'$ it is not necessary to keep all the information of $\tau_e$; instead, it suffices to keep track only of whether $e$ belongs to $E_{\mathtt{r}}$ or $E_{\mathtt{b}}$, along with the value of $x_{v(e)} \in \{0, 1\}$. The colorings encode precisely this information. The resulting bijection between colorings and message configurations is the last step of (18).

Recall messages take values $\tau \equiv (\dot{\tau}, \hat{\tau}) \in \mathscr{M} \equiv \dot{\mathscr{M}} \times \hat{\mathscr{M}}$ (Definition 2.6), and let $\{\mathtt{f}\} \subseteq \mathscr{M}$ denote the subset of values $\tau \in \mathscr{M}$ where we have both $\dot{\tau} \in \dot{\mathscr{M}} \backslash \{0, 1, \star\}$ and $\hat{\tau} \in \hat{\mathscr{M}} \backslash \{0, 1, \star\}$. Denote $\Omega \equiv \{\mathtt{r}_0, \mathtt{r}_1, \mathtt{b}_0, \mathtt{b}_1\} \cup \{\mathtt{f}\}$. Define a projection $\mathsf{S} : \mathscr{M} \to \Omega$ by

$$\mathsf{S}(\tau) = \begin{cases} \mathtt{r}_0 & \text{if } \hat{\tau} = 0, \\ \mathtt{r}_1 & \text{if } \hat{\tau} = 1, \\ \mathtt{b}_0 & \text{if } \hat{\tau} \neq 0 \text{ and } \dot{\tau} = 0, \\ \mathtt{b}_1 & \text{if } \hat{\tau} \neq 1 \text{ and } \dot{\tau} = 1, \\ \tau & \text{otherwise (meaning that } \tau \in \{\mathtt{f}\}). \end{cases} \tag{33}$$

(Note that $\mathsf{S}(\tau) \in \{\mathtt{r}_0, \mathtt{r}_1\}$ includes the case $\dot{\tau} = \star$, and $\mathsf{S}(\tau) \in \{\mathtt{b}_0, \mathtt{b}_1\}$ includes the case $\hat{\tau} = \star$.) We define a partial inverse to $\mathsf{S}$ as follows. If $\sigma \in \{\mathtt{f}\}$ then define $\tau \equiv \tau(\sigma) \equiv \sigma \equiv (\dot{\sigma}, \hat{\sigma})$. If $\sigma = \mathtt{r}_{\boldsymbol{x}}$ for $\boldsymbol{x} \in \{0, 1\}$ then define $\hat{\tau} \equiv \hat{\tau}(\sigma) \equiv \boldsymbol{x}$ and leave $\dot{\tau}(\sigma)$ undefined. If $\sigma = \mathtt{b}_{\boldsymbol{x}}$ for $\boldsymbol{x} \in \{0, 1\}$ then define $\dot{\tau} \equiv \dot{\tau}(\sigma) \equiv \boldsymbol{x}$ and leave $\hat{\tau}(\sigma)$ undefined. For $\sigma \in \{\mathtt{r}_0, \mathtt{r}_1, \mathtt{b}_0, \mathtt{b}_1\}$ we denote $(\dot{\sigma}, \hat{\sigma}) \equiv (\sigma, \sigma)$. The

coloring model is the image of the message configuration model under the projection $\mathsf{S}$, formally given by the following:

**Definition 2.11** *(colorings)* For $\underline{\sigma} \in \Omega^d$, abbreviate $\{\sigma_i\} \equiv \{\sigma_1, \ldots, \sigma_d\}$, and define

$$\dot{I}(\underline{\sigma}) \equiv \begin{cases} 1 & \text{if } \mathtt{r}_0 \in \{\sigma_i\} \subseteq \{\mathtt{r}_0, \mathtt{b}_0\}, \\ 1 & \text{if } \mathtt{r}_1 \in \{\sigma_i\} \subseteq \{\mathtt{r}_1, \mathtt{b}_1\}, \\ 1 & \{\sigma_i\} \subseteq \{\mathtt{f}\}, \text{ and } \dot{\sigma}_i = \dot{T}((\hat{\sigma}_j)_{j \neq i}) \text{ for all } i, \\ 0 & \text{otherwise.} \end{cases}$$

For $\underline{\sigma} \in \Omega^k$, abbreviate $\{\sigma_i\} \equiv \{\sigma_1, \ldots, \sigma_k\}$, and define

$$\hat{I}^{\mathrm{lit}}(\underline{\sigma}) = \begin{cases} 1 & \text{if } \exists i : \sigma_i = \mathtt{r}_0 \text{ and } \{\sigma_j\}_{j \neq i} = \{\mathtt{b}_1\}, \\ 1 & \text{if } \exists i : \sigma_i = \mathtt{r}_1 \text{ and } \{\sigma_j\}_{j \neq i} = \{\mathtt{b}_0\}, \\ 1 & \text{if } \{\mathtt{b}_0, \mathtt{b}_1\} \subseteq \{\sigma_i\} \subseteq \{\mathtt{b}_0, \mathtt{b}_1\} \cup \{\sigma \in \{\mathtt{f}\} : \hat{\sigma} = \mathtt{s}\}, \\ 1 & \text{if } \{\sigma_i\} \subseteq \{\mathtt{b}_0\} \cup \{\mathtt{f}\}, |\{\sigma_i\} \cap \{\mathtt{f}\}| \geq 2, \text{ and } \hat{\sigma}_i = \hat{T}((\dot{\tau}(\sigma_j))_{j \neq i}) \\ & \text{for all } i \text{ where } \sigma_i \neq \mathtt{b}_0; \\ 1 & \text{if } \{\sigma_i\} \subseteq \{\mathtt{b}_1\} \cup \{\mathtt{f}\}, |\{\sigma_i\} \cap \{\mathtt{f}\}| \geq 2, \text{ and } \hat{\sigma}_i = \hat{T}((\dot{\tau}(\sigma_j))_{j \neq i}) \\ & \text{for all } i \text{ where } \sigma_i \neq \mathtt{b}_1; \\ 0 & \text{otherwise.} \end{cases}$$

(In the definition of $\hat{I}^{\mathrm{lit}}(\underline{\sigma})$ we used that if $\{\sigma_i\} \subseteq \{\mathtt{b}_0, \mathtt{b}_1\} \cup \{\mathtt{f}\}$, then $\dot{\tau}(\sigma_i)$ is defined for all $i$.) On an NAE- SAT instance $\mathscr{G} = (V, F, E, \underline{\mathtt{L}})$, a configuration $\underline{\sigma} \in \Omega^E$ is a valid **coloring** if $\dot{I}(\underline{\sigma}_{\delta v}) = 1$ for all $v \in V$, and $\hat{I}^{\mathrm{lit}}((\underline{\sigma} \oplus \underline{\mathtt{L}})_{\delta a}) = 1$ for all $a \in F$.

**Lemma 2.12** *On any given* NAE- SAT *instance* $\mathscr{G} = (V, F, E, \underline{\mathtt{L}})$*, we have a bijection*

$$\left\{ \begin{array}{c} \textit{message configurations} \\ \underline{\tau} \in \mathscr{M}^E \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{c} \textit{colorings} \\ \underline{\sigma} \in \Omega^E \end{array} \right\}.$$

***Proof*** Given a valid message configuration $\underline{\tau}$, a valid coloring $\underline{\sigma}$ is obtained by coordinatewise application of the projection map $\mathsf{S}$ from (33). In the other direction, given a valid coloring $\underline{\sigma}$, let $x_v = \mathtt{0}$ if $\underline{\sigma}_{\delta v}$ has any $\mathtt{r}_0$ entries, $x_v = \mathtt{1}$ if $\underline{\sigma}_{\delta v}$ has any $\mathtt{r}_1$ entries, and $x_v = \mathtt{f}$ otherwise. The resulting $\underline{x} \in \{\mathtt{0}, \mathtt{1}, \mathtt{f}\}^V$ is a valid frozen configuration, and the argument of Lemma 2.7 implies that it has no free cycles. It then maps by Lemma 2.7 to a valid message configuration $\underline{\tau}$, which completes the correspondence. □

Recall the definitions (28) of $\bar{\varphi}$, $\hat{\varphi}^{\mathrm{lit}}$, and $\dot{\varphi}$. For $\underline{\sigma} \in \Omega^d$, let

$$\dot{\Phi}(\underline{\sigma}) = \begin{cases} \dot{\varphi}(\hat{\underline{\sigma}}) & \text{if } \dot{I}(\underline{\sigma}) = 1 \text{ and } \{\sigma_i\} \subseteq \{\mathtt{f}\}; \\ 1 & \text{if } \dot{I}(\underline{\sigma}) = 1 \text{ and } \{\sigma_i\} \subseteq \{\mathtt{r}_0, \mathtt{r}_1, \mathtt{b}_0, \mathtt{b}_1\}; \\ 0 & \text{otherwise (meaning that } \dot{I}(\underline{\sigma}) = 0). \end{cases}$$

(Note if $\{\sigma_i\} \subseteq \{\mathtt{f}\}$ then $\hat{\underline{\sigma}} = \hat{\underline{\tau}}$ and $\dot{\varphi}(\hat{\underline{\sigma}})$ is well-defined.) For $\underline{\sigma} \in \Omega^k$, let

$$\hat{\Phi}^{\text{lit}}(\underline{\sigma}) = \begin{cases} \hat{\varphi}^{\text{lit}}((\dot{\tau}(\sigma_i))_i) & \text{if } \hat{I}^{\text{lit}}(\underline{\sigma}) = 1 \text{ and } \{\sigma_i\} \subseteq \{\mathtt{b}_0, \mathtt{b}_1\} \cup \{\mathtt{f}\}; \\ 1 & \text{if } \hat{I}^{\text{lit}}(\underline{\sigma}) = 1 \text{ and } \{\sigma_i\} \cap \{\mathtt{r}_0, \mathtt{r}_1\} \neq \varnothing; \\ 0 & \text{otherwise (meaning that } \hat{I}^{\text{lit}}(\underline{\sigma}) = 0). \end{cases} \tag{34}$$

(Note if $\{\sigma_i\} \subseteq \{\mathtt{b}_1, \mathtt{b}_1\} \cup \{\mathtt{f}\}$ then $\dot{\tau}(\sigma_i)$ is well-defined for all $i$.) Finally, let

$$\bar{\Phi}(\sigma) = \begin{cases} \bar{\varphi}(\sigma) & \text{if } \sigma \in \{\mathtt{f}\}, \\ 1 & \text{if } \sigma \in \{\mathtt{r}_0, \mathtt{r}_1, \mathtt{b}_0, \mathtt{b}_1\}. \end{cases}$$

The following is a straightforward consequence of Lemma 2.9:

**Lemma 2.13** *Suppose on $\mathscr{G} = (V, F, E, \underline{L})$ that $\underline{x}$ is a frozen configuration with no free cycles. Let $\underline{\sigma}$ be the coloring that corresponds to $\underline{x}$ by* Lemmas 2.7 and 2.12. *Then the number of valid* NAE- SAT *extensions of $\underline{x}$ is given by $\mathsf{size}(\underline{\mathrm{x}}) = \boldsymbol{w}^{lit}_{\mathscr{G}}(\underline{\mathrm{o}})$ where we define*

$$\boldsymbol{w}^{lit}_{\mathscr{G}}(\underline{\sigma}) \equiv \prod_{v \in V} \dot{\Phi}(\underline{\sigma}_{\delta v}) \prod_{a \in F} \hat{\Phi}^{lit}((\underline{\sigma} \oplus \underline{L})_{\delta a}) \prod_{e \in E} \bar{\Phi}(\sigma_e)\,. \tag{35}$$

***Proof*** This is a rewriting of (30). □

**Definition 2.14** *(T-colorings)* If $\sigma \in \{\mathtt{f}\}$, then $\dot{\sigma}$ is a tree rooted at a clause $a$ incident to a single edge $e(a)$, while $\hat{\sigma}$ is a tree rooted at a variable $v$ incident to a single edge $e(v)$. Glue $\dot{\sigma}$ and $\hat{\sigma}$ together by identifying $e(a)$ with $e(v)$, and let $|\sigma|$ count the number of free variables in the resulting tree. (Note that $|\sigma|$ must be finite because we only consider colorings of finite NAE- SAT instances $\mathscr{G}$.) Thus $|\sigma| = |\dot{\sigma}| + |\hat{\sigma}| - 1$ where $|\dot{\sigma}|$ is the number of free variables in the tree $\dot{\sigma}$, and $|\hat{\sigma}|$ is the number of free variables in the tree $\hat{\sigma}$. If $\sigma \in \Omega \backslash \{\mathtt{f}\} = \{\mathtt{r}_0, \mathtt{r}_1, \mathtt{b}_0, \mathtt{b}_1\}$ then define $|\sigma| \equiv 0$. If $\underline{\sigma}$ is a valid coloring on $\mathscr{G} = (V, F, E, \underline{L})$, then $|\sigma_e|$ must be finite on every edge $e \in E$, by Definition 2.11. For $0 \leq T \leq \infty$ define $\Omega_T \equiv \{\sigma \in \Omega : |\sigma| \leq T\}$; we then call $\underline{\sigma}$ a **$T$-coloring** if $\sigma_e \in \Omega_T$ for all $e \in E$. Define $\boldsymbol{Z}_{\lambda,T}$ to be the partition function of $\lambda$-tilted $T$-colorings,

$$\boldsymbol{Z}_{\lambda,T} \equiv \boldsymbol{Z}_{\lambda,T}(\mathscr{G}) \equiv \sum_{\underline{\sigma} \in (\Omega_T)^E} \boldsymbol{w}^{\text{lit}}_{\mathscr{G}}(\underline{\sigma})^{\lambda}\,. \tag{36}$$

Denote $\boldsymbol{Z}_{\lambda} \equiv \boldsymbol{Z}_{\lambda,\infty}$ and note that as $T \uparrow \infty$ we have $\boldsymbol{Z}_{\lambda,T} \uparrow \boldsymbol{Z}_{\lambda,\infty} \equiv \boldsymbol{Z}_{\lambda}$.

**Proposition 2.15** *On an* NAE- SAT *instance $\mathscr{G} = (V, F, E, \underline{L})$, recall $\mathsf{CL}(\mathscr{G})$ denotes the set of solution clusters (connected components of $\mathsf{SOL}(\mathscr{G})$), and define*

$$\bar{\boldsymbol{Z}}_{\lambda} \equiv \bar{\boldsymbol{Z}}_{\lambda}(\mathscr{G}) \equiv \sum_{\boldsymbol{\gamma} \in \mathsf{CL}(\mathscr{G})} |\boldsymbol{\gamma}|^{\lambda}\,. \tag{37}$$

*Then $\bar{\boldsymbol{Z}}_{\lambda}$ is lower bounded by $\boldsymbol{Z}_{\lambda}$, where $\boldsymbol{Z}_{\lambda}$ is the increasing limit of $\boldsymbol{Z}_{\lambda,T}$ as defined by* (36).

***Proof*** On $\mathscr{G}$, the colorings $\underline{\sigma}$ are in bijection (Lemma 2.12) with the message configurations $\underline{\tau}$, which in turn are in bijection (Lemma 2.7) with the frozen configurations $\underline{x}$ that do not have free cycles. Each such frozen configuration defines a distinct cluster $\boldsymbol{\gamma} \in \mathsf{CL}(\mathscr{G})$, of size $|\boldsymbol{\gamma}| = \mathsf{size}(\underline{x}) = \boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G}}(\underline{\alpha})$. The claimed inequality directly follows.□

To summarize what we have obtained so far, note that the quantity $\bar{\boldsymbol{Z}}_\lambda$ of (37) is a formal definition of the "λ-tilted cluster partition function" introduced in (5). In a sequence (18) of combinatorial mappings, we have produced in (36) a mathematically well-defined quantity $\boldsymbol{Z}_{\lambda,T}$ which lower bounds $\bar{\boldsymbol{Z}}_\lambda$ (Proposition 2.15), and will be much more tractable thanks to the product formula for cluster sizes (Lemma 2.13). The lower bound of Theorem 1 is based on the second moment method applied to $\boldsymbol{Z}_{\lambda,T}$. In preparation for the moment calculation, we conclude the current section by discussing some simplifications obtained by averaging over the literals of the NAE- SAT instance.

### 2.5 Averaging over edge literals

Our eventual purpose is to calculate $\mathbb{E}\boldsymbol{Z}_{\lambda,T}$ and $\mathbb{E}[(\boldsymbol{Z}_{\lambda,T})^2]$, where $\mathbb{E}$ is expectation over the NAE- SAT instance $\mathscr{G}$. Recall that $\mathscr{G} = (\mathcal{G}, \underline{\mathtt{L}})$ where $\mathcal{G} = (V, F, E)$ is the graph without the edge literals $\underline{\mathtt{L}}$. Then $\mathbb{E}\boldsymbol{Z}_{\lambda,T} = \mathbb{E}(\mathbb{E}(\boldsymbol{Z}_{\lambda,T} \mid \mathcal{G}))$ where

$$\mathbb{E}(\boldsymbol{Z}_{\lambda,T} \mid \mathcal{G}) = \sum_{\underline{\sigma} \in (\Omega_T)^E} \mathbb{E}(\boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G}}(\underline{\sigma})^\lambda \mid \mathcal{G})\,.$$

For any $l \geq 1$ and any function $g : \{0,1\}^l \to \mathbb{R}$, let $\mathbb{E}^{\mathrm{lit}} g$ denote the average value of $g(\underline{\mathtt{L}})$ over all $\underline{\mathtt{L}} \in \{0,1\}^l$. For any $\underline{\sigma} \in \Omega^E$, we have $\mathbb{E}(\boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G}}(\underline{\sigma})^\lambda \mid \mathcal{G}) = \boldsymbol{w}_{\mathcal{G}}(\underline{\sigma})^\lambda$ where (compare (35))

$$\boldsymbol{w}_{\mathcal{G}}(\underline{\sigma}) \equiv \prod_{v \in V} \dot{\Phi}(\underline{\sigma}_{\delta v}) \prod_{a \in F} \hat{\Phi}(\underline{\sigma}_{\delta a}) \prod_{e \in E} \bar{\Phi}(\sigma_e)\,, \quad \hat{\Phi}(\underline{\sigma}) \equiv \left(\mathbb{E}^{\mathrm{lit}}[\hat{\Phi}^{\mathrm{lit}}(\underline{\sigma} \oplus \underline{\mathtt{L}})^\lambda]\right)^{1/\lambda} \tag{38}$$

— that is to say, even after averaging over $\underline{\mathtt{L}}$, the contribution of each $\underline{\sigma} \in \Omega^E$ is still given by a product formula. This means that $\mathbb{E}(\boldsymbol{Z}_{\lambda,T} \mid \mathcal{G})$ is the partition function of a "factor model":

**Definition 2.16** *(factor model)* On a bipartite graph $\mathcal{G} = (V, F, E)$, the **factor model** specified by $g \equiv (\dot{g}, \hat{g}, \bar{g})$ is the probability measure $\nu_{\mathcal{G}}$ on configurations $\underline{\xi} \in \mathscr{X}^E$ defined by

$$\nu_{\mathcal{G}}(\underline{\xi}) = \frac{1}{Z} \prod_{v \in V} \dot{g}(\underline{\xi}_{\delta v}) \prod_{a \in F} \hat{g}(\underline{\xi}_{\delta a}) \prod_{e \in E} \bar{g}(\xi_e),$$

with $Z$ the normalizing constant.

A further observation is that for $\mathscr{G} = (\mathcal{G}, \underline{\mathtt{L}})$ and $\underline{\sigma} \in \Omega^E$, as we go over all possibilities of $\underline{\mathtt{L}}$ while keeping $\mathcal{G}$ fixed, the weight $\boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G}}(\underline{\sigma})$ (the size of the cluster encoded by $\underline{\sigma}$ on $\mathscr{G}$, as given by (35)) does not take more than one positive value. In

other words, we can extract the cluster size without referring to the edge literals. The precise statement is as follows:

**Lemma 2.17** *The function* $\hat{\Phi}^{lit}$ *of* (34) *can be factorized as* $\hat{\Phi}^{lit}(\underline{\sigma} \oplus \underline{L}) = \hat{I}^{lit}(\underline{\sigma} \oplus \underline{L})\hat{F}(\underline{\sigma})$ *for*

$$\hat{F}(\underline{\sigma}) \equiv \begin{cases} 1 & \text{if } \underline{\sigma} \in \{\mathtt{r}_0, \mathtt{r}_1, \mathtt{b}_0, \mathtt{b}_1\}^k, \\ \dfrac{\hat{z}(\hat{\sigma}_j)}{\bar{\varphi}(\sigma_j)} & \text{if } \underline{\sigma} \in \Omega^k \text{ with } \sigma_j \in \{\mathtt{f}\}. \end{cases}$$

*As a consequence, the function of* (38) *satisfies* $\hat{\Phi}(\underline{\sigma})^\lambda = \hat{v}(\underline{\sigma})\hat{F}(\underline{\sigma})^\lambda$ *where* $\hat{v}(\underline{\sigma}) \equiv \mathbb{E}^{\mathrm{lit}}[\hat{I}^{lit}(\underline{\sigma} \oplus \underline{L})]$.

***Proof*** For $\underline{\sigma} \in \Omega^k$ abbreviate $\{\sigma_i\} \equiv \{\sigma_1, \ldots, \sigma_k\}$. If $\{\sigma_i\} \subseteq \{\mathtt{r}_0, \mathtt{r}_1, \mathtt{b}_0, \mathtt{b}_1\}$, then the definition (34) implies $\hat{\Phi}^{\mathrm{lit}}(\underline{\sigma} \oplus \underline{L}) = \hat{I}^{\mathrm{lit}}(\underline{\sigma} \oplus \underline{L})$ for all $\underline{L}$, so the factorization holds with $\hat{F}(\underline{\sigma}) \equiv 1$. If $\{\sigma_i\}$ nontrivially intersects both $\{\mathtt{r}_0, \mathtt{r}_1\}$ and $\{\mathtt{f}\}$, then $\hat{\Phi}^{\mathrm{lit}}(\underline{\sigma} \oplus \underline{L}) = \hat{I}^{\mathrm{lit}}(\underline{\sigma} \oplus \underline{L}) = 0$ for all $\underline{L}$, so we can set $\hat{F}(\underline{\sigma})$ arbitrarily. It remains to consider the case where $\{\sigma_i\}$ nontrivially intersects $\{\mathtt{f}\}$ but does not intersect $\{\mathtt{r}_0, \mathtt{r}_1\}$. Recalling the discussion around (33), this means that $\dot{\tau}_i \equiv \dot{\tau}(\sigma_i) \in \dot{\mathscr{M}} \backslash \{\star\}$ is well-defined for all $i$—if $\sigma_i \in \{\mathtt{f}\}$ then $\dot{\tau}_i = \dot{\sigma}_i \in \dot{\mathscr{M}} \backslash \{0, 1, \star\}$, and if $\sigma_i = \mathtt{b}_{\boldsymbol{x}}$ then $\dot{\tau}_i = \boldsymbol{x} \in \{0, 1\}$. Following (23), given any $\underline{\mathtt{L}} \in \{0, 1\}^k$, let us define $\hat{\tau}_{\underline{\mathtt{L}},i} \equiv \mathtt{L}_i \oplus \hat{T}((\dot{\tau}_j \oplus \mathtt{L}_j)_{j \neq i})$. If $\hat{I}^{\mathrm{lit}}(\underline{\sigma} \oplus \underline{L}) = 1$, then it follows from (25), (28), and (34) that

$$\begin{aligned} \hat{\Phi}^{\mathrm{lit}}(\underline{\sigma} \oplus \underline{L}) = \hat{\varphi}^{\mathrm{lit}}(\underline{\dot{\tau}} \oplus \underline{L}) &= \sum_{\underline{\boldsymbol{x}} \in \{0,1\}^k} I^{\mathrm{NAE}}(\underline{\boldsymbol{x}} \oplus \underline{L}) \prod_{j=1}^{k} [\dot{\mathtt{m}}(\dot{\tau}_j)](\boldsymbol{x}_j) \\ &= \hat{z}(\hat{\tau}_{\underline{L},i}) \sum_{\boldsymbol{x}_i} [\dot{\mathtt{m}}(\dot{\tau}_i)](\boldsymbol{x}_i) \cdot [\hat{\mathtt{m}}(\hat{\tau}_{\underline{L},i})](\boldsymbol{x}_i) = \frac{\hat{z}(\hat{\tau}_{\underline{L},i})}{\bar{\varphi}(\dot{\tau}_i, \hat{\tau}_{\underline{L},i})}. \end{aligned} \tag{39}$$

We will have $\hat{I}^{\mathrm{lit}}(\underline{\sigma} \oplus \underline{L}) = 1$ if and only if it holds for all $1 \leq i \leq k$ that $\hat{\tau}_{\underline{L},i}$ is compatible with $\sigma_i$, in the sense that $\mathsf{S}(\dot{\tau}_i, \hat{\tau}_{\underline{L},i}) = \sigma_i$. In particular, if $\sigma_i \in \{\mathtt{f}\}$ (and we assumed $\underline{\sigma}$ has at least one such entry), we must have $(\dot{\tau}_i, \hat{\tau}_{\underline{L},i}) = (\dot{\sigma}_i, \hat{\sigma}_i)$. It follows that for any $\underline{\sigma} \in \Omega^k$ having at least one entry in $\{\mathtt{f}\}$, we can define $\hat{F}(\underline{\sigma}) \equiv \hat{z}(\hat{\sigma}_i)/\bar{\varphi}(\sigma_i)$ for any $i$ where $\sigma_i \in \{\mathtt{f}\}$. This completes the proof. □

**Corollary 2.18** *On a bipartite graph* $\mathcal{G} = (V, F, E)$, *suppose* $\underline{\sigma} \in \Omega^E$ *satisfies* $\dot{I}(\underline{\sigma}_{\delta v}) = 1$ *for all* $v \in V$. *Then, for* $\mathscr{G} = (\mathcal{G}, \underline{L})$, *it follows from* (35) *that*

$$\boldsymbol{w}^{lit}_{\mathscr{G}}(\underline{\sigma}) = \left\{ \prod_{a \in F} \hat{I}^{lit}((\underline{\sigma} \oplus \underline{L})_{\delta a}) \right\} \boldsymbol{W}_{\mathcal{G}}(\underline{\sigma}), \quad \boldsymbol{W}_{\mathcal{G}}(\underline{\sigma}) \equiv \prod_{v \in V} \dot{\Phi}(\underline{\sigma}_{\delta v}) \prod_{a \in F} \hat{F}(\underline{\sigma}_{\delta a}) \prod_{e \in E} \bar{\Phi}(\sigma_e).$$

*Combining with* (38) *gives, with* $\hat{v}$ *as defined by* Lemma 2.17,

$$\boldsymbol{w}_{\mathcal{G}}(\underline{\sigma})^\lambda = \boldsymbol{p}_{\mathcal{G}}(\underline{\sigma}) \boldsymbol{W}_{\mathcal{G}}(\underline{\sigma})^\lambda, \quad \boldsymbol{p}_{\mathcal{G}}(\underline{\sigma}) \equiv \mathbb{E}\left[ \prod_{a \in F} \hat{I}^{lit}((\underline{\sigma} \oplus \underline{L})_{\delta a}) \,\middle|\, \mathcal{G} \right] = \prod_{a \in F} \hat{v}(\underline{\sigma}_{\delta a}).$$

***Proof*** Immediate consequence of Lemma 2.17. □

In the notation of Definition 2.16, the conditional first moment $\mathbb{E}(\boldsymbol{Z}_{\lambda,T} \mid, \mathcal{G})$ is the partition function of the factor model with specification $(\dot{\Phi}, \hat{\Phi}, \bar{\Phi})^{\lambda}$ restricted to the alphabet $\Omega_T$. Similarly, the conditional second moment $\mathbb{E}[(\boldsymbol{Z}_{\lambda,T})^2 \mid \mathcal{G}]$ is the partition function of the factor model on the alphabet $(\Omega_T)^2$ with specification $(\dot{\Phi}_2, \hat{\Phi}_2, \bar{\Phi}_2)^{\lambda}$, where $\dot{\Phi}_2 \equiv \dot{\Phi} \otimes \dot{\Phi}$, $\bar{\Phi}_2 \equiv \bar{\Phi} \otimes \bar{\Phi}$, and for any $\underline{\sigma} \equiv (\underline{\sigma}^1, \underline{\sigma}^2) \in \Omega^{2k}$ we have

$$\hat{\Phi}_2(\underline{\sigma}) \equiv \left(\mathbb{E}^{\mathrm{lit}}\Big[\hat{\Phi}^{\mathrm{lit}}(\underline{\sigma}^1 \oplus \underline{\mathtt{L}})^{\lambda}\hat{\Phi}^{\mathrm{lit}}(\underline{\sigma}^2 \oplus \underline{\mathtt{L}})^{\lambda}\Big]\right)^{1/\lambda} = \hat{v}_2(\underline{\sigma})^{1/\lambda}(\hat{F} \otimes \hat{F})(\underline{\sigma}) \,,$$

for $\hat{v}_2(\underline{\sigma}) \equiv \mathbb{E}^{\mathrm{lit}}[\hat{I}^{\mathrm{lit}}(\underline{\sigma}^1 \oplus \underline{\mathtt{L}})\hat{I}^{\mathrm{lit}}(\underline{\sigma}^1 \oplus \underline{\mathtt{L}})]$ (by Corollary 2.18). We emphasize that $\hat{\Phi}$, $\hat{\Phi}_2$ both depend on $\lambda$, although we suppress it from the notation. Moreover, $\hat{\Phi}_2 \neq \hat{\Phi} \otimes \hat{\Phi}$ since $\underline{\sigma}^1$ and $\underline{\sigma}^2$ are coupled through their interaction with the same literals $\underline{\mathtt{L}} \in \{0, 1\}^k$. Lastly, we have written $\underline{\sigma}$ in the first moment and $\underline{\sigma} \equiv (\underline{\sigma}^1, \underline{\sigma}^2)$ in the second moment—this is a deliberate abuse of notation, which allows us to treat the two cases in a unified manner. To distinguish the cases we shall refer to the "first-moment" or "single-copy" model, versus the "second-moment" or "pair" model. We turn next to the analysis of these models.

## 3 Proof outline

Having formally set up our combinatorial model of NAE- SAT solution clusters (Sect. 2), we now give a more detailed outline for the (first and second) moment calculation that proves the lower bound of Theorem 1. (As we mentioned before, the upper bound of Theorem 1 is proved by an interpolation argument which builds on prior results in spin glass theory [12,26,43]. It does not involve the combinatorial model or the moment method, and is deferred to "Appendix E".)

### 3.1 Empirical measures and moments

We use standard multi-index notations in what follows—in particular, for any ordered sequence $z = (z_1, \ldots, z_l)$ of nonnegative integers summing to $n$, we denote

$$\binom{n}{z} \equiv n! \Big/ \prod_{i=1}^{l} z_i! \,.$$

If $\pi$ is any nonnegative measure on a discrete space, write $\mathcal{H}(\pi) = -\langle \pi, \ln \pi \rangle$ for its Shannon entropy. It follows from Stirling's formula that for any fixed $\pi$, in the limit $n \to \infty$ we have

$$\binom{n}{n\pi} \asymp \frac{\exp\{n\mathcal{H}(\pi)\}}{n^{(|\operatorname{supp}\pi|-1)/2}} \,.$$

On a bipartite graph $\mathcal{G}$, we will summarize colorings $\underline{\sigma}$ according to some "local statistics," as follow:

**Definition 3.1** *(empirical measures)* Given a bipartite graph $\mathcal{G} = (V, F, E)$ and $\underline{\sigma} \in \Omega^E$, define

$$\begin{aligned}
\dot{H}(\underline{\dot{\sigma}}) &= |\{v \in V : \underline{\sigma}_{\delta v} = \underline{\dot{\sigma}}\}|/|V| \qquad \text{for } \underline{\dot{\sigma}} \in \Omega^d, \\
\hat{H}(\underline{\hat{\sigma}}) &= |\{a \in F : \underline{\sigma}_{\delta a} = \underline{\hat{\sigma}}\}|/|F| \qquad \text{for } \underline{\hat{\sigma}} \in \Omega^k, \\
\bar{H}(\sigma) &= |\{e \in E : \sigma_e = \sigma\}|/|E| \qquad \text{for } \sigma \in \Omega;
\end{aligned}$$

The triple $H \equiv H(\mathcal{G}, \underline{\sigma}) \equiv (\dot{H}, \hat{H}, \bar{H})$ is the **empirical measure** of $\underline{\sigma}$ on $\mathcal{G}$.

Recall from (38) that $\boldsymbol{w}_{\mathcal{G}}(\underline{\sigma})^{\lambda}$ is the contribution to $\mathbb{E}(\boldsymbol{Z}_{\lambda} \mid \mathcal{G})$ from $\underline{\sigma} \in \Omega^E$. We saw in Corollary 2.18 that $\boldsymbol{w}_{\mathcal{G}}(\underline{\sigma})^{\lambda} = \boldsymbol{p}_{\mathcal{G}}(\underline{\sigma})\boldsymbol{W}_{\mathcal{G}}(\underline{\sigma})^{\lambda}$ where $\boldsymbol{p}_{\mathcal{G}}(\underline{\sigma})$ is the probability (conditional on $\mathcal{G}$) that $\underline{\sigma}$ is a valid coloring on $(\mathcal{G}, \underline{\mathbb{L}})$; and $\boldsymbol{W}_{\mathcal{G}}(\underline{\sigma})$ is the size of the cluster encoded if $\underline{\sigma}$ is valid. Now all these quantities can be expressed solely in terms of $H = H(\mathcal{G}, \underline{\sigma})$: we have $\boldsymbol{p}_{\mathcal{G}}(\underline{\sigma}) = \exp(n\boldsymbol{v}(H))$ and $\boldsymbol{W}_{\mathcal{G}}(\underline{\sigma}) = \exp(n\boldsymbol{s}(H))$ where

$$\begin{aligned}
\boldsymbol{v}(H) &\equiv (d/k)\langle \ln \hat{v}, \hat{H} \rangle = (d/k) \sum_{\underline{\sigma} \in \Omega^k} \hat{H}(\underline{\sigma}) \ln \hat{v}(\underline{\sigma}) \,, \\
\boldsymbol{s}(H) &\equiv \langle \ln \dot{\Phi}, \dot{H} \rangle + (d/k)\langle \ln \hat{F}, \hat{H} \rangle + d\langle \ln \bar{\Phi}, \bar{H} \rangle \,.
\end{aligned}$$

Given $\mathscr{G} = (\mathcal{G}, \underline{\mathbb{L}})$, let $\boldsymbol{Z}_{\lambda,T}(H)$ be the contribution to $\boldsymbol{Z}_{\lambda,T}$ from colorings $\underline{\sigma} \in (\Omega_T)^E$ such that $H(\mathcal{G}, \underline{\sigma}) = H$. In what follows we will often suppress the dependence on $\lambda$ and $T$, and write simply $\boldsymbol{Z} \equiv \boldsymbol{Z}_{\lambda,T}$.

**Definition 3.2** *(simplex)* For $d, k, T$ fixed, the **simplex of empirical measures** is the space $\boldsymbol{\Delta} \equiv \boldsymbol{\Delta}(T)$ of triples $H \equiv (\dot{H}, \hat{H}, \bar{H})$ satisfyng the following conditions: $\dot{H}$ is a probability measure supported within the set of $\underline{\sigma} \in (\Omega_T)^d$ such that $\dot{I}(\underline{\sigma}) = 1$; $\hat{H}$ is a probability measure supported within the set of $\underline{\sigma} \in (\Omega_T)^k$ such that $\hat{v}(\underline{\sigma})$ is positive; and both $\dot{H}$ and $\hat{H}$ must have marginal $\bar{H}$, that is,

$$\frac{1}{d} \sum_{\underline{\sigma} \in \Omega^d} \dot{H}(\underline{\sigma}) \sum_{i=1}^{d} \mathbf{1}\{\sigma_i = \sigma\} = \bar{H}(\sigma) = \frac{1}{k} \sum_{\underline{\sigma} \in \Omega^k} \hat{H}(\underline{\sigma}) \sum_{j=1}^{k} \mathbf{1}\{\sigma_j = \sigma\} \tag{40}$$

for all $\sigma \in \Omega$. It follows that $\bar{H}$ is a probability measure supported on $\Omega_T$.

It follows from Corollary 2.18 that if $\mathbb{E}$ is expectation over a $(d, k)$-regular NAE- SAT instance on $n$ variables, then $\mathbb{E}\boldsymbol{Z}(H)$ is positive if and only if $H \in \boldsymbol{\Delta}$ and $(n\dot{H}, m\hat{H})$ is integer-valued. For such $H$, it follows from the definition of the random regular NAE- SAT graph that

$$\mathbb{E}\boldsymbol{Z}(H) = \left[ \left\{ \binom{n}{n\dot{H}} \binom{m}{m\hat{H}} \Big/ \binom{nd}{nd\bar{H}} \right\} \exp\{n\boldsymbol{v}(H)\} \right] \cdot \exp\{n\lambda \boldsymbol{s}(H)\} \,. \tag{41}$$

In (41), the first factor (in square brackets) is the expected number $\mathbb{E}\boldsymbol{Z}_{\lambda=0,T}$ of valid colorings with empirical profile $H$. The remaining factor $\exp\{n\lambda \boldsymbol{s}(H)\}$ is explained by the fact that any such coloring encodes a cluster of size $\exp\{n\boldsymbol{s}(H)\}$. By Stirling's formula, in the limit $n \to \infty$ (with $T$ fixed),

$$\mathbb{E}\boldsymbol{Z}_{\lambda=0,T} \asymp \frac{\exp\{n[\mathcal{H}(\dot{H}) + (d/k)\mathcal{H}(\hat{H}) - d\mathcal{H}(\bar{H}) + \boldsymbol{v}(H)]\}}{n^{\wp(H)/2}} \equiv \frac{\exp\{n\boldsymbol{\Sigma}(H)\}}{n^{\wp(H)/2}}$$

where $\wp(H) \equiv |\operatorname{supp} \dot{H}| + |\operatorname{supp} \hat{H}| - |\operatorname{supp} \bar{H}| - 1$, and the exponential rate $\boldsymbol{\Sigma}(H)$ is a formal analogue of the "cluster complexity" function $\Sigma(s)$ appearing in (4). In analogy with (6) we let

$$\boldsymbol{F} \equiv \boldsymbol{F}_{\lambda,T} \equiv \boldsymbol{\Sigma}(H) + \lambda \boldsymbol{s}(H)\,. \tag{42}$$

Then altogether the first moment can be estimated as

$$\mathbb{E}\boldsymbol{Z}(H) \asymp \left(\frac{\exp\{n\boldsymbol{\Sigma}(H)\}}{n^{\wp(H)/2}}\right) \exp\{n\lambda \boldsymbol{s}(H)\} = \frac{\exp\{n\boldsymbol{F}(H)\}}{n^{\wp(H)/2}}\,. \tag{43}$$

Note that $\asymp$ hides a dependence on $T$, since we keep $T$ fixed throughout our moment analysis.

### 3.2 Outline of first moment

For any subset of empirical measures $\mathbf{H} \subseteq \boldsymbol{\Delta}$, we write $\underline{\sigma} \in \mathbf{H}$ to indicate that $H(\mathcal{G}, \underline{\sigma}) \in \mathbf{H}$, and write $\boldsymbol{Z}(\mathbf{H}) \equiv \boldsymbol{Z}_{\lambda,T}(\mathbf{H})$ for the contribution to $\boldsymbol{Z}_{\lambda,T}$ from colorings $\underline{\sigma} \in \mathbf{H}$. It then follows from (43) that

$$\mathbb{E}\boldsymbol{Z}(\mathbf{H}) = \sum_{H \in \mathbf{H}} \mathbb{E}\boldsymbol{Z}(H) = n^{O(1)} \exp\left\{n \max\{\boldsymbol{F}(H) : H \in \mathbf{H}\}\right\},$$

for $\boldsymbol{F}$ as in (42). Thus, calculating the first moment $\mathbb{E}\boldsymbol{Z}$ essentially reduces to the problem of maximizing $\boldsymbol{F}$ over $\boldsymbol{\Delta}$. The physics theory suggests that $\boldsymbol{F}$ is uniquely maximized at a point $H_\star \in \boldsymbol{\Delta}$ which is given explicitly in terms of a replica symmetric fixed point for the $\lambda$-tilted $T$-coloring model. (Recall from §1.5 that in the original NAE- SAT model, the replica symmetric fixed point was described by the measure $\mathtt{m} = \mathrm{unif}(\{0, 1\})$. In the coloring model, with spins $\sigma \equiv (\dot{\sigma}, \hat{\sigma}) \in \Omega_T$, the replica symmetric fixed point will be characterized by a measure $\dot{q}$ on the space $\Omega_T$ of possible values for $\dot{\sigma}$.)

There are several obstacles to the rigorous moment computation. From a physics perspective, the replica symmetric fixed point of the $\lambda$-tilted coloring model at $T = \infty$ is equivalent to the fixed point described by Proposition 1.2, which was used to define the 1RSB prediction (12). Mathematically, however, we work with $T$ finite so that $\boldsymbol{\Delta}$ has finite dimension and $\wp(H)$ is defined. Therefore we need to explicitly construct a replica symmetric fixed point at finite $T$, and use it to define $H_\star = H_{\lambda,T} \in \boldsymbol{\Delta}$

(Definition 5.6 below). We must then take $T \to \infty$ and show that the limit matches the fixed point of Proposition 1.2, so that $\boldsymbol{F}(H_\star) = \boldsymbol{F}_{\lambda,T}(H_{\lambda,T})$ converges as $T \to \infty$ the 1RSB prediction (12). The construction of the fixed point at finite $T$ is stated in Proposition 5.5 below, and proved in "Appendix A". The correspondence with (12) in the $T = \infty$ limit is stated in Proposition 3.13 below, and proved in "Appendix B".

A more difficult problem is to show that $\boldsymbol{F}$ is in fact maximized at $H_\star$. The function $\boldsymbol{F}$ is generally not convex, and must be optimized over a space $\boldsymbol{\Delta}$ whose dimension grows with $d, k, T$. Moreover, an analogous but even more difficult optimization must be solved to compute the second moment $\mathbb{E}(\boldsymbol{Z}^2)$. The main part of this analysis is carried out in Sects. 4 and 5. In the remainder of this section we make some preparatory calculations and explain how the pieces will be fit together to prove the main result Theorem 1. Recall from Remark 1.1 that we can restrict consideration to $\alpha$ satisfying (3). In this regime, we make *a priori* estimates to show that the optimal $H$ satisfy some basic restrictions. Abbreviate $\{\mathtt{r}\} \equiv \{\mathtt{r}_0, \mathtt{r}_1\}$, recall $\{\mathtt{f}\} \equiv \Omega\backslash\{\mathtt{r}_0, \mathtt{r}_1, \mathtt{b}_0, \mathtt{b}_1\}$, and let

$$\mathbf{N}_\circ \equiv \left\{ H \in \boldsymbol{\Delta} : \max\{\bar{H}(\mathtt{f}), \bar{H}(\mathtt{r})\} \le \frac{7}{2^k} \right\},$$
$$\mathbf{N} = \left\{ H \in \mathbf{N}_\circ : \|H - H_\star\| \le \frac{1}{n^{1/3}} \right\} \subseteq \mathbf{N}_\circ, \tag{44}$$

where $\|\cdot\|$ denotes the $\ell^1$ norm throughout this paper. We next show that empirical measures $H \notin \mathbf{N}_\circ$ give a negligible contribution to the first moment:

**Lemma 3.3** *For $k \ge k_0$, $\alpha \equiv d/k$ satisfying (3), and $0 \le \lambda \le 1$, $\mathbb{E}\boldsymbol{Z}(\boldsymbol{\Delta}\backslash\mathbf{N}_\circ)$ is exponentially small in $n$.*

***Proof*** Let $Z^{\mathtt{f}}$ count the NAE- SAT solutions $\underline{\boldsymbol{x}} \in \{0,1\}^V$ which map—via coarsening and the bijection (18)—to warning configurations $\underline{y}$ with more than $7/2^k$ fraction of edges $e$ such that $\dot{y}_e = \hat{y}_e = \mathtt{f}$. Similarly, let $Z^{\mathtt{r}}$ count NAE- SAT solutions $\underline{\boldsymbol{x}}$ mapping to warning configurations $\underline{y}$ with more than $7/2^k$ fraction of edges $e$ such that $\hat{y}_e \in \{0,1\}$. It follows from Proposition 2.15 that for any $0 \le \lambda \le 1$ we have $\boldsymbol{Z}(\boldsymbol{\Delta}\backslash\mathbf{N}_\circ) \le Z^{\mathtt{f}} + Z^{\mathtt{r}}$. For $\alpha$ satisfying (3), $\mathbb{E}Z^{\mathtt{f}}$ is exponentially small in $n$ by [25, Propn. 2.2]. As for $Z^{\mathtt{r}}$, let us say that an edge $e \in E$ is **blocked** under $\underline{\boldsymbol{x}} \in \{0,1\}^V$ if $\mathtt{L}_e \oplus x_{v(e)} = 1 \oplus \mathtt{L}_{e'} \oplus x_{v(e')}$ for all $e' \in \delta a(e)\backslash e$. Note that if $\underline{\boldsymbol{x}}$ maps to $\underline{y}$, the only possibility for $y_e \in \{\mathtt{r}_0, \mathtt{r}_1\}$ is that $e$ was blocked under $\underline{\boldsymbol{x}}$. (The converse need not hold.) If we condition on $\underline{\boldsymbol{x}}$ being a valid NAE- SAT solution, then each clause contains a blocking edge independently with chance $\theta = 2k/(2^k - 2)$; note also that a clause can contain at most one blocking edge. It follows that

$$\mathbb{E}Z^{\mathtt{r}} \le (\mathbb{E}Z)\mathbb{P}\Big(\mathrm{Bin}(m,\theta) \ge 7nd/2^k\Big).$$

This is exponentially small in $n$ by a Chernoff bound together with the trivial bound $\mathbb{E}Z \le 2^n$. □

We assume throughout what follows that $k \geq k_0$, $\alpha$ satisfies (3), and $0 \leq \lambda \leq 1$. In this regime, Lemma 3.3 tells us that $\max\{\boldsymbol{F}(H) : H \notin \mathbf{N}_\circ\}$ is negative. On the other hand, we shall assume that the global maximum of $\boldsymbol{F}$ is nonnegative, since otherwise $\mathbb{E}\boldsymbol{Z}$ is exponentially small in $n$ and there is nothing to prove. From this we have that any maximizer $H$ of $\boldsymbol{F}$ must lie in $\mathbf{N}_\circ$. In Sects. 4 and 5 we develop a more refined analysis to solve the optimization problem for $\boldsymbol{F}$ restricted to $\mathbf{N}_\circ$:

**Proposition 3.4** *(proved in Sect. 5) Assuming the global maximum of $\boldsymbol{F}$ is nonnegative, the unique maximizer of $\boldsymbol{F}$ is an explicitly characterized point $H_\star$ in the interior of $\mathbf{N}_\circ$. Moreover, there is a positive constant $\epsilon = \epsilon(k, \lambda, T)$ so that for all $\|H - H_\star\| \leq \epsilon$ we have $\boldsymbol{F}(H) \leq \boldsymbol{F}(H_\star) - \epsilon\|H - H_\star\|^2$.*

A consequence of the above is that we can compute the first moment of $\boldsymbol{Z}$ up to constant factors. In the following, let $\dot{\wp} \equiv \dot{\wp}(T)$ count the number of $d$-tuples $\underline{\sigma} \in (\Omega_T)^d$ for which $\dot{I}(\underline{\sigma}) > 0$. Let $\hat{\wp} \equiv \hat{\wp}(T)$ count the number of $k$-tuples $\underline{\sigma} \in (\Omega_T)^k$ for which $\hat{v}(\underline{\sigma}) > 0$. Let $\bar{\wp} \equiv |\Omega_T|$, and denote $\wp \equiv \dot{\wp} + \hat{\wp} - \bar{\wp} - 1$. Recall from (44) the definition of $\mathbf{N}$.

**Corollary 3.5** *The coloring partition function $\boldsymbol{Z} \equiv \boldsymbol{Z}_{\lambda,T}$ has first moment $\mathbb{E}\boldsymbol{Z} \asymp \exp\{n\boldsymbol{F}(H_\star)\}$. Moreover the expectation is dominated by $\mathbf{N}$ in the sense that $\mathbb{E}\boldsymbol{Z}(\mathbf{N}) = (1 - o(1))\mathbb{E}\boldsymbol{Z}$.*

***Proof*** Define the $\bar{\wp} \times \dot{\wp}$ matrix $\dot{M}$ with entries

$$\dot{M}(\sigma', \underline{\sigma}) \equiv \sum_{i=1}^{d} \mathbf{1}\{\sigma_i = \sigma'\}, \quad \sigma' \in \Omega_T \text{ and } \underline{\sigma} \in (\Omega_T)^d \cap (\operatorname{supp} \dot{I}).$$

Similarly define the $\bar{\wp} \times \hat{\wp}$ matrix $\hat{M}$ with entries

$$\hat{M}(\sigma', \underline{\sigma}) = \sum_{i=1}^{k} \mathbf{1}\{\sigma_i = \sigma'\}, \quad \sigma' \in \Omega_T \text{ and } \underline{\sigma} \in (\Omega_T)^k \cap (\operatorname{supp} \hat{v}).$$

Lastly define the $\bar{\wp} \times (\dot{\wp} + \hat{\wp})$ matrix $M \equiv \begin{pmatrix} \dot{M} & -\hat{M} \end{pmatrix}$. It follows from the discussion after Definition 3.2 that $\mathbb{E}\boldsymbol{Z}(H)$ is positive if and only if (i) $\dot{H}$ and $\hat{H}$ are nonnegative; (ii) $\langle \mathbf{1}, \dot{H} \rangle = 1$; (iii) $(k\dot{H}, d\hat{H})$ lies in the kernel of $M$; and (iv) $(n\dot{H}, m\hat{H})$ is integer-valued. Conditions (i)–(iii) are equivalent to $H \in \boldsymbol{\Delta}$. One can verify that the matrix $M$ is of full rank, from which it follows that the space of vectors $(\dot{H}, \hat{H})$ satisfying the conditions (ii) and (iii) has dimension $\wp$. In Lemma 4.4 we will show that $M$ satisfies a stronger condition, which implies that the space of $(\dot{H}, \hat{H})$ satisfying (ii)–(iv) is an affine transformation of $(n^{-1}\mathbb{Z})^\wp$, where the coefficients of the transformation are uniformly bounded. Then, by substituting the result of Proposition 3.4 in to (43), we conclude

$$\mathbb{E}\boldsymbol{Z} \asymp \sum_{z \in (n^{-1}\mathbb{Z})^\wp} \frac{\exp\{n[\boldsymbol{F}(H_\star) - \Theta(\|z\|^2)]\}}{n^{\wp/2}} \asymp \exp\{n\boldsymbol{F}(H_\star)\}$$

Empirical measures $H \notin \mathbf{N}$ correspond to vectors $z \in (n^{-1}\mathbb{Z})^{\wp}$ with norm $\|z\| \geq n^{-1/3}$. These give a negligible contribution to the above sum which proves the second claim $\mathbb{E}\boldsymbol{Z}(\mathbf{N}) = (1 - o(1))\mathbb{E}\boldsymbol{Z}$. □

### 3.3 Outline of second moment

By a similar calculation as above, calculating the second moment $\mathbb{E}(\boldsymbol{Z}^2)$ reduces to the problem of maximizing a function $\boldsymbol{F}_2 \equiv \boldsymbol{F}_{2,\lambda,T}$ over a space $\boldsymbol{\Delta}_2$ of **pair empirical measures**. In fact we will calculate the second moment not of $\boldsymbol{Z}$ itself, but rather of a more restricted random variable $\boldsymbol{S} \leq \boldsymbol{Z}$, defined below. This leads to a more tractable analysis, as we now explain.

Concretely, $\boldsymbol{\Delta}_2$ is the space of triples $H = (\dot{H}, \hat{H}, \bar{H})$ satisfying the following conditions: $\dot{H}$ is a probability measure on $(\Omega_T)^{2d} \cap (\operatorname{supp} \dot{I})^2$, $\hat{H}$ is a probability measure on $(\Omega_T)^{2d} \cap (\operatorname{supp} \hat{v}_2)$, and $\bar{H}$ is a probability measure on $(\Omega_T)^2$ which can be obtained from both $\dot{H}$ and $\hat{H}$ as the edge marginal (40). Repeating the derivation of (43) in the second moment setting, we have the following. For any $H \in \boldsymbol{\Delta}_2$, the expected number of valid coloring pairs $(\underline{\sigma}, \underline{\sigma}') \in (\Omega_T)^{2E}$ of type $H$ is

$$\mathbb{E}(\boldsymbol{Z}_{\lambda=0,T})^2(H) \asymp \frac{\exp\{n[\mathcal{H}(\dot{H}) + (d/k)\mathcal{H}(\hat{H}) - d\mathcal{H}(\bar{H}) + \boldsymbol{v}_2(H)]\}}{n^{\wp(H)/2}} \equiv \frac{\exp\{n\boldsymbol{\Sigma}_2(H)\}}{n^{\wp(H)/2}}$$

Given a pair empirical measure $H \in \boldsymbol{\Delta}_2$, take the marginal on the first element of the pair to define the **first-copy marginal** $H^1 \in \boldsymbol{\Delta}$; define likewise the **second-copy marginal** $H^2 \in \boldsymbol{\Delta}$. The contribution to $\boldsymbol{Z}^2$ from any valid pair is given by $\exp\{n\lambda \boldsymbol{s}_2(H)\}$ where $\boldsymbol{s}_2(H) \equiv \boldsymbol{s}(H^1) + \boldsymbol{s}(H^2)$. Thus $\boldsymbol{F}_2$ is given explicitly by

$$\mathbb{E}\boldsymbol{Z}^2(H) \asymp \frac{\exp\{n\boldsymbol{F}_2(H)\}}{n^{\wp(H)/2}} \equiv \frac{\exp\{n[\boldsymbol{\Sigma}_2(H) + \lambda \boldsymbol{s}_2(H)]\}}{n^{\wp(H)/2}} \tag{45}$$

(cf. (42) and (43)). In view of Corollary 3.5, it would suffice for our purposes to calculate the second moment of $\boldsymbol{Z}(\mathbf{N})$ rather than $\boldsymbol{Z}$, which amounts to maximizing $\boldsymbol{F}_2$ on the restricted set

$$\mathbf{N}_2 \equiv \left\{ H \in \boldsymbol{\Delta}_2 : H^1 \in \mathbf{N} \text{ and } H^2 \in \mathbf{N} \right\}.$$

Following [18] we can simplify the analysis by a further restriction, as follows:

**Definition 3.6** *(separability)* If $\underline{\sigma}, \underline{\sigma}'$ are valid colorings on $\mathscr{G}$, define their **separation** $\mathsf{sep}(\underline{\mathfrak{o}}, \underline{\mathfrak{o}}')$ to be the fraction of variables where their corresponding frozen configurations (from Lemmas 2.7 and 2.12) differ. Write $\underline{\sigma}' \succcurlyeq \underline{\sigma}$ if the frozen configuration of $\underline{\sigma}'$ has more free variables than that of $\underline{\sigma}$. We say that a coloring is **separable** if $\underline{\sigma} \in \mathbf{N}$ (recall this means $H(\mathcal{G}, \underline{\sigma}) \in \mathbf{N}$) and

$$|\{\underline{\sigma}' \in \mathbf{N} : \underline{\sigma}' \succcurlyeq \underline{\sigma} \text{ and } \mathsf{sep}(\underline{\mathfrak{o}}, \underline{\mathfrak{o}}') \notin \mathrm{I}_{\mathrm{se}}\}| \leq \exp\{(\ln n)^4\},$$

where $I_{\mathrm{se}} \equiv [(1 - k^4/2^{k/2})/2, (1 + k^4/2^{k/2})/2]$ and it is implicit that both $\underline{\sigma}, \underline{\sigma}'$ must both be valid on $\mathscr{G}$. Let $\boldsymbol{S} \equiv \boldsymbol{S}_{\lambda,T}$ be the contribution to $\boldsymbol{Z}(\mathbf{N})$ from separable colorings.

**Proposition 3.7** (proved in "Appendix *D*") *The first moment is dominated by separable colorings in the sense that* $\mathbb{E}\boldsymbol{S} = (1 - o(1))\mathbb{E}\boldsymbol{Z}(\mathbf{N})$.

We will apply the second moment method to lower bound $\boldsymbol{S}$; the result will follow since $\boldsymbol{S} \le \boldsymbol{Z}(\mathbf{N}) \le \boldsymbol{Z}$. For $H \in \boldsymbol{\Delta}_2$, all pairs $(\underline{\sigma}, \underline{\sigma}') \in H$ must have the same separation $\mathsf{sep}(\underline{\alpha}, \underline{\alpha}')$, which we can thus denote as $\mathsf{sep}(\mathrm{H})$. Partition $\mathbf{N}_2$ into $\mathbf{N}_{\mathrm{se}} \equiv \{H \in \mathbf{N}_2 : \mathsf{sep}(\mathrm{H}) \in \mathrm{I}_{\mathrm{se}}\}$ (the near-uncorrelated regime) and $\mathbf{N}_{\mathrm{ns}} \equiv \mathbf{N}_2 \backslash \mathbf{N}_{\mathrm{se}}$ (the correlated regime). Denote the corresponding contributions to $\boldsymbol{S}^2$ by $\boldsymbol{S}^2(\mathbf{N}_{\mathrm{se}})$ and $\boldsymbol{S}^2(\mathbf{N}_{\mathrm{ns}})$.

**Corollary 3.8** *For separable colorings, the second moment contribution from the correlated regime* $\mathbf{N}_{ns}$ *is bounded by* $\mathbb{E}[\boldsymbol{S}^2(\mathbf{N}_{ns})] \le \exp\{n\lambda \boldsymbol{s}(H_\star) + o(n)\}\,\mathbb{E}\boldsymbol{S}$.

***Proof*** By the symmetry between the roles of $\underline{\sigma}$ and $\underline{\sigma}'$, and the definition of separability, we have

$$\boldsymbol{S}^2(\mathbf{N}_{\mathrm{ns}}) \le 2 \sum_{\substack{(\underline{\sigma},\underline{\sigma}') \in \mathbf{N}_{\mathrm{ns}}, \\ \underline{\sigma} \text{ separable}}} \mathbf{1}\{\underline{\sigma}' \succcurlyeq \underline{\sigma}\} \boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G},T}(\underline{\sigma})^\lambda \boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G},T}(\underline{\sigma}')^\lambda \le \exp\{n\boldsymbol{s}(H_\star)\lambda + o(n)\}\boldsymbol{S}\,.$$

Taking expectations gives the claim. □

We then conclude the second moment calculation by computing $\mathbb{E}(\boldsymbol{Z}^2(\mathbf{N}_{\mathrm{se}}))$, which is an upper bound on $\mathbb{E}(\boldsymbol{S}^2(\mathbf{N}_{\mathrm{se}}))$. Therefore we must maximize the function $\boldsymbol{F}_2$ on $\mathbf{N}_{\mathrm{se}}$. As in the first moment, the physics theory suggests that the unique maximizer of $\boldsymbol{F}_2$ on $\mathbf{N}_{\mathrm{se}}$ is given by a specific pair empirical measure $H_\bullet$ which is defined in terms of $H_\star$. In the NAE- SAT model there is a small complication in this definition: we will have $\dot{H}_\bullet = \dot{H}_\star \otimes \dot{H}_\star$ and $\bar{H}_\bullet = \bar{H}_\star \otimes \bar{H}_\star$, but $\hat{H}_\bullet \ne \hat{H}_\star \otimes \hat{H}_\star$ because the first and second copies interact via the edge literals. It therefore requires a small calculation to argue that $\boldsymbol{F}_2(H_\bullet) = 2\boldsymbol{F}_2(H_\star)$. We address this by giving a simple sufficient condition for $H \in \boldsymbol{\Delta}_2$ to satisfy $\boldsymbol{F}_2(H) = \boldsymbol{F}(H^1) + \boldsymbol{F}(H^2)$ where $H^j \in \boldsymbol{\Delta}$ are its single-copy marginal.

**Lemma 3.9** *Consider* $H \in \boldsymbol{\Delta}_2$ *with single-copy marginals* $H^1, H^2 \in \boldsymbol{\Delta}$. *Suppose there are functions* $g^1, g^2$ *which are invariant to literals in the sense that* $g^j(\underline{\sigma}^j) = g^j(\underline{\sigma}^j \oplus \underline{L})$ *for all* $\underline{\sigma}^j \in (\Omega_T)^k$, $\underline{L} \in \{0,1\}^k$, *and*

$$\hat{H}^j(\underline{\sigma}^j) = \hat{v}(\underline{\sigma}^j) g^j(\underline{\sigma})\,, \quad \hat{H}(\underline{\sigma}^1, \underline{\sigma}^2) = \hat{v}_2(\underline{\sigma}^1, \underline{\sigma}^2) \prod_{j=1,2} g^1(\underline{\sigma}^1)\,.$$

*If in addition* $\dot{H} = \dot{H}^1 \otimes \dot{H}^2$ *and* $\bar{H} = \bar{H}^1 \otimes \bar{H}^2$, *then* $\boldsymbol{F}_2(H) = \boldsymbol{F}(H^1) + \boldsymbol{F}(H^2)$.

***Proof*** Let $K^j(\underline{\sigma}^j, \underline{\mathtt{L}}) \equiv \hat{I}^{\mathrm{lit}}(\underline{\sigma} \oplus \underline{\mathtt{L}}) g^j(\underline{\sigma})/2^k$. This defines a probability measure on $(\Omega_T)^k \times \{0,1\}^k$ where the marginal on $(\Omega_T)^k$ is $\hat{H}^j$, and the marginal on $\{0,1\}^k$ is uniform by the assumption on $g^j$. It follows that $K^j(\underline{\sigma}^j \,|\, \underline{\mathtt{L}}) = \hat{I}^{\mathrm{lit}}(\underline{\sigma} \oplus \underline{\mathtt{L}}) g^j(\underline{\sigma})$ and $\hat{H}^j(\underline{\sigma}^j) = \mathbb{E}^{\mathrm{lit}}[K^j(\underline{\sigma}^j \,|\, \underline{\mathtt{L}})]$. Let $(X^1, X^2, L)$ be a random variable with law

$$\mathbb{P}((X^1, X^2, L) = (\underline{\sigma}^1, \underline{\sigma}^2, \underline{\mathtt{L}})) = \frac{1}{2^k} \prod_{j=1,2} K^j(\underline{\sigma}^j \,|\, \underline{\mathtt{L}}) \,.$$

The marginal law of $L$ is uniform on $\{0,1\}^k$, and the $X^j$ are conditionally independent given $L$. The marginal law of $(X^1, X^2)$ is $\hat{H}$, and the marginal law of $X^j$ is $\hat{H}^j$. The law of $L$ conditional on $X^j$ is uniform over $2^k \hat{v}(X^j)$ possibilities, whereas the law of $L$ conditional on $(X^1, X^2)$ is uniform over $2^k \hat{v}_2(X^1, X^2)$ possibilities. It follows that

$$\begin{aligned}
\mathcal{H}(\hat{H}) = \mathcal{H}(X^1, X^2) &= \mathcal{H}(L) - \mathcal{H}(L|X^1, X^2) + \sum_{j=1,2} \mathcal{H}(X^j|L) \\
&= -\langle \hat{H}, \ln \hat{v}_2 \rangle + \sum_{j=1,2} \mathcal{H}(X^j|L) \\
&= -\langle \hat{H}, \ln \hat{v}_2 \rangle + \sum_{j=1,2} [\mathcal{H}(\hat{H}^j) + \langle \hat{H}^j, \ln \hat{v} \rangle] \,.
\end{aligned}$$

Rearranging gives $\mathbf{\Sigma}_2(H) = \mathbf{\Sigma}(H^1) + \mathbf{\Sigma}(H^2)$, and the result follows. □

It will be clear from the explicit definition that the measure $H_\star$ of Proposition 3.4 can be expressed as $H_\star(\underline{\sigma}) = \hat{v}(\underline{\sigma}) g_\star(\underline{\sigma})$ where $g_\star$ is invariant to literals. Let $H_\bullet = (\dot{H}_\bullet, \hat{H}_\bullet, \bar{H}_\bullet)$ where

$$\hat{H}_\bullet(\underline{\sigma}^1, \underline{\sigma}^2) \equiv \hat{v}_2(\underline{\sigma}^1, \underline{\sigma}^2) \prod_{j=1,2} g_\star(\underline{\sigma}^j) \,,$$

$\dot{H}_\bullet = \dot{H}_\star \otimes \dot{H}_\star$, and $\bar{H}_\bullet = \bar{H}_\star \otimes \bar{H}_\star$. The following is the second moment analogue of Proposition 3.4.

**Proposition 3.10** *(*proved in Section 5) The unique maximizer of $\boldsymbol{F}_2$ in $\mathbf{N}_{\mathrm{se}}$ is $H_\bullet$. Moreover, there is a positive constant $\epsilon = \epsilon(k, \lambda, T)$ so that for $\|H - H_\bullet\| \le \epsilon$ we have $\boldsymbol{F}_2(H) \le \boldsymbol{F}_2(H_\bullet) - \epsilon \|H - H_\bullet\|^2$.

**Corollary 3.11** *For the coloring model, the second moment contribution from the near-uncorrelated regime $\mathbf{N}_{se}$ is given by the estimate $\mathbb{E}[\boldsymbol{Z}^2(\mathbf{N}_{se})] \asymp \exp\{2n\boldsymbol{F}(H_\star)\}$.*

***Proof*** Recall from Corollary 3.5 the definition of $(\dot{\wp}, \hat{\wp}, \bar{\wp})$ for the single-copy model, and define $(\dot{\wp}_2, \hat{\wp}_2, \bar{\wp}_2)$ analogously for the pair model. Let $\wp_2 \equiv \dot{\wp}_2 + \hat{\wp}_2 - \bar{\wp}_2 - 1$. For any $H^1, H^2 \in \mathbf{N}$, let $\mathbf{N}_{\mathrm{se}, H^1, H^2}$ denote the set of $H \in \mathbf{N}_{\mathrm{se}}$ with single-copy marginals $H^1, H^2$. This is a space of dimension $\wp_2 - 2\wp$, and it follows from Proposition 3.10 and Lemma 4.4 that

$$\mathbb{E}[\boldsymbol{Z}^2(\mathbf{N}_{\mathrm{se}, H^1, H^2})] \asymp \frac{\exp\{n[\boldsymbol{F}_2(H_\bullet) - \Theta(\|(H^1, H^2) - (H_\star, H_\star)\|^2)]\}}{n^{\wp}} \,.$$

Summing over $H^1, H^2 \in \mathbf{N}$ then gives $\mathbb{E}[\boldsymbol{Z}^2(\mathbf{N}_{\rm se})] \asymp \exp\{n\boldsymbol{F}_2(H_\bullet)\}$, which in turn equals $\exp\{2n\boldsymbol{F}(H_\star)\}$ by applying Lemma 3.9. □

### 3.4 Conclusion of main result

We now explain that the main theorem follows. We continue to assume, as we have done throughout the section, that $k \geq k_0$, $\alpha$ satisfies (3), and $0 \leq \lambda \leq 1$. The measure $H_\star$ of Proposition 3.4 depends on $\lambda$ and $T$, and we now make this explicit by writing $H_\star \equiv H_{\lambda,T}$.

**Corollary 3.12** *For any $0 \leq \lambda \leq 1$ and $T$ finite such that $\boldsymbol{\Sigma}(H_{\lambda,T})$ is positive, the separable contribution to $\boldsymbol{Z}_{\lambda,T}$ is well-concentrated about its mean:*

$$\lim_{\epsilon \downarrow 0} \liminf_{n\to\infty} \mathbb{P}\left(\epsilon(\mathbb{E}\boldsymbol{S}) \leq \boldsymbol{S} \leq \frac{\mathbb{E}\boldsymbol{S}}{\epsilon}\right) = 1\,.$$

***Proof*** The upper bound follows trivially from Markov's inequality, so the task is to show the lower bound. In the first moment, we have $\mathbb{E}\boldsymbol{S} \asymp \exp\{n\boldsymbol{F}(H_\star)\}$ by Corollary 3.5 and Proposition 3.7. In the second moment, since $\boldsymbol{S} \leq \boldsymbol{Z}$, we have $\mathbb{E}(\boldsymbol{S}^2) \leq \mathbb{E}[\boldsymbol{S}^2(\mathbf{N}_{\rm ns})] + \mathbb{E}[\boldsymbol{Z}^2(\mathbf{N}_{\rm se})]$. Combining with Corollaries 3.8 and 3.11 gives

$$\frac{\mathbb{E}(\boldsymbol{S}^2)}{(\mathbb{E}\boldsymbol{S})^2} \leq \frac{\exp\{n\lambda \boldsymbol{s}(H_\star) + o(n)\}}{\mathbb{E}\boldsymbol{S}} + O(1) \asymp \frac{\exp\{o(n)\}}{\exp\{n\boldsymbol{\Sigma}(H_\star)\}} + O(1)\,,$$

which immediately implies $\mathbb{P}(\boldsymbol{S} \geq \delta(\mathbb{E}\boldsymbol{S})) \geq \delta$ for some positive constant $\delta$. This can be strengthened to the asserted concentration result by an easy adaptation of the method described in [25, Sec. 6]. □

**Proposition 3.13** (proved in "Appendix *B*") *For $0 \leq \lambda \leq 1$ and $H_{\lambda,T} = H_\star \in \boldsymbol{\Delta}$ as given by* Proposition 3.4*, the triple $(\boldsymbol{s}(H_{\lambda,T}), \boldsymbol{\Sigma}(H_{\lambda,T}), \boldsymbol{F}(H_{\lambda,T}))$ converges as $T \to \infty$ to $(s_\lambda, \Sigma(s_\lambda), \mathfrak{F}(\lambda))$ from* Definition 1.3.

***Proof of Theorem 1*** In "Appendix E" we prove the upper bound, $\mathsf{f}(\alpha) \leq \mathsf{f}^{\rm 1RSB}(\alpha)$ for all $0 \leq \alpha < \alpha_{\rm sat}$. For any $\lambda, T$ such that $\boldsymbol{\Sigma}(H_{\lambda,T})$ is positive, Corollary 3.12 gives

$$\liminf_{n\to\infty} (\boldsymbol{Z}_{\lambda,T}(\mathbf{N}))^{1/n} \geq \lim_{n\to\infty} (\boldsymbol{S}_{\lambda,T})^{1/n} = \exp\{\boldsymbol{F}(H_{\lambda,T})\} = \exp\{\boldsymbol{\Sigma}(H_{\lambda,T}) + \lambda \boldsymbol{s}(H_{\lambda,T})\}\,.$$

On the other hand, $\boldsymbol{Z}_{\lambda,T}(\mathbf{N})$ consists entirely of clusters of size $\exp\{n\boldsymbol{s}(H_{\lambda,T}) + o(n)\}$. Therefore, if $\boldsymbol{\Sigma}(H_{\lambda,T})$ is positive, it must be that $\mathsf{f}(\alpha) \geq \boldsymbol{s}(H_{\lambda,T})$. The lower bound $\mathsf{f}(\alpha) \geq \mathsf{f}^{\rm 1RSB}(\alpha)$ then follows by appealing to Proposition 3.13, so the theorem is proved. □

The next two sections are devoted to the optimization of $\boldsymbol{F}$ in $\mathbf{N}_\circ$, and of $\boldsymbol{F}_2$ in $\mathbf{N}_{\rm se}$. In Sect. 4 we show that the optimization of $\boldsymbol{F}$ and $\boldsymbol{F}_2$ over small regions can be reduced to an optimization problem on trees. In Sect. 5 we solve the tree optimization problem by connecting it to the analysis of the BP recursion for the coloring model. This allows us to prove Propositions 3.4 and 3.11, thereby completing the proof of the main result Theorem 1.

## 4 Reduction to tree optimization by local updates

In this section we prove the key reduction that ultimately allows us to compute the (first and second) moments of $\boldsymbol{Z} \equiv \boldsymbol{Z}_{\lambda,T}$ (Propositions 3.4 and 3.10). As we have already seen, the calculation reduces to the optimization of functions $\boldsymbol{F}$ and $\boldsymbol{F}_2$ from (42) and (45). These functions are generally not convex over the entirety of their domains $\boldsymbol{\Delta}$ and $\boldsymbol{\Delta}_2$, but we expect them to be convex in neighborhoods around their maximizers $H_\star$ and $H_\bullet$ (as given in Definition 5.6 below). With this in mind, we rely on other means (*a priori* estimates and separability) to restrict the domains—from $\boldsymbol{\Delta}$ to $\mathbf{N}_\circ$ in the first moment (Lemma 3.3), and from $\boldsymbol{\Delta}_2$ to $\mathbf{N}_{\mathrm{se}}$ in the second moment (Corollary 3.8). Within these restricted regions, we will show that $\boldsymbol{F}$ and $\boldsymbol{F}_2$ can be optimized by a **local** update procedure that reduces the (nonconvex) graph optimization to a (convex) tree optimization.

### 4.1 Local update

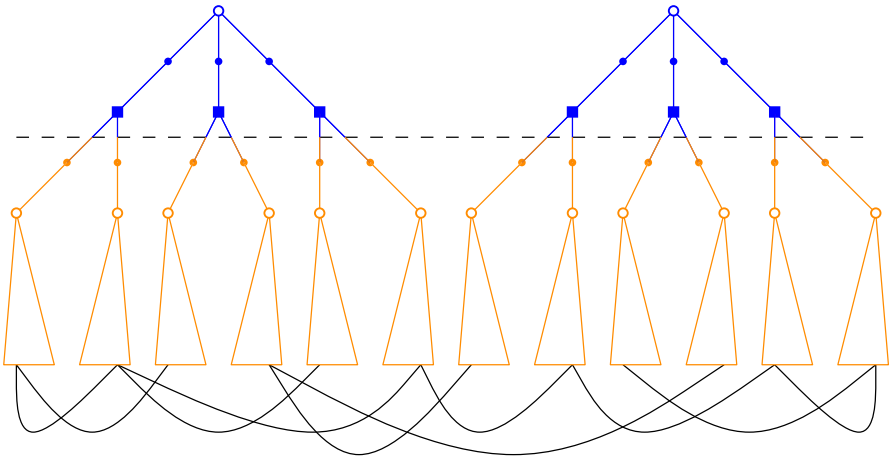
We begin with an overview. Throughout this section, we assume $1 \le T < \infty$. Suppose $\underline{\sigma}$ is a $T$-coloring on $\mathscr{G}$. Sample from $\mathscr{G}$ a subset of variables $Y$, and let $\mathscr{N} \equiv \mathscr{N}(Y)$ be the subgraph of $\mathscr{G}$ induced by $Y$, together with the clauses neighboring $Y$ and their incident half-edges. The half-edges at the boundary of $\mathscr{N}$ will be referred to as the **leaf edges of** $\mathscr{N}$.

Form a modified instance $\mathscr{G}'$ (see Fig. 4) by resampling the edge literals on $\mathscr{N}$ as well as the matching between $\mathscr{N}$ and $\mathscr{G}\backslash\mathscr{N}$. In both $\mathscr{G}$ and $\mathscr{G}'$, we will say **cut edges** to refer to the edges $e = (av)$ where $a$ is a clause in $\mathscr{N}$ and $v$ is a variable in the complement of $\mathscr{N}$; these are the edges cut by the dashed lines in Fig. 4. According to our terminology, the leaf edges of $\mathscr{N}$ are the half-edges that lie just above the dashed lines, so each leaf edge of $\mathscr{N}$ is half of a cut edge. The coloring is updated accordingly to produce $\underline{\eta}$, a $T$-coloring on $\mathscr{G}'$ which agrees as much as possible with $\underline{\sigma}$ on $\mathscr{G}\backslash\mathscr{N}$: in particular, $\underline{\sigma}$ and $\underline{\eta}$ will agree in the variable-to-clause colors on the cut edges. We will define the procedure so that it gives a Markov chain $\pi$ on triples $(\mathscr{G}, Y, \underline{\sigma})$ with reversing measure given by $\mu(\mathscr{G}, Y, \underline{\sigma}) = \mathbb{P}(\mathscr{G})\mathbb{P}(Y \mid \mathscr{G})\boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G}}(\underline{\sigma})^\lambda$. (Note that $\mu$ is not normalized to be a probability measure.)

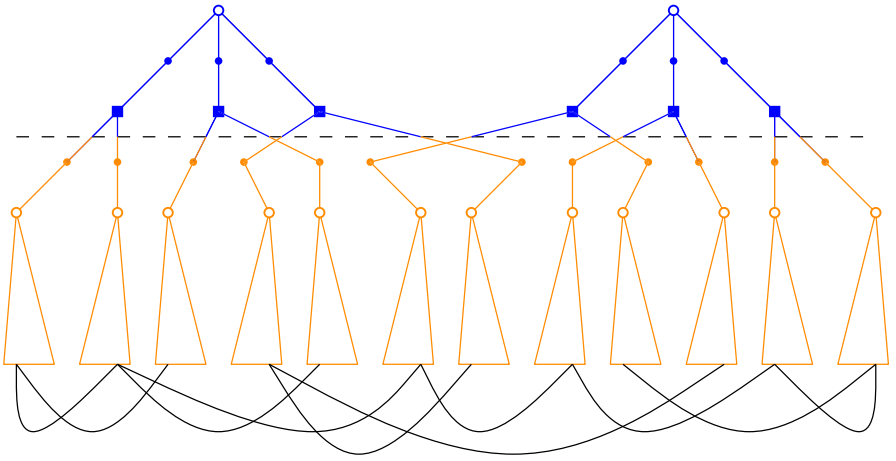
Reversibility implies that for any subset $A$ of the state space, if $B$ is the set of states reachable in one step from $A$, then

$$\begin{aligned}\mu(A) &= \sum_{\mathrm{A}\in A}\sum_{\mathrm{B}\in B}\mu(\mathrm{A})ß(\mathrm{A},\mathrm{B}) = \sum_{\mathrm{A}\in A}\sum_{\mathrm{B}\in B}\bar{}(\mathrm{B})ß(\mathrm{B},\mathrm{A}) = \sum_{\mathrm{B}\in B}\bar{}(\mathrm{B})\sum_{\mathrm{A}\in A}ß(\mathrm{B},\mathrm{A})\\ &= \sum_{\mathrm{B}\in B}\mu(\mathrm{B})ß(\mathrm{B},\mathrm{A}) \le \Big\{\sum_{\mathrm{B}\in B}\bar{}(\mathrm{B})\Big\}\Big\{\max_{\mathrm{B}\in B}ß(\mathrm{B},\mathrm{A})\Big\} = \bar{}(\mathrm{B})\max_{\mathrm{B}\in B}ß(\mathrm{B},\mathrm{A})\,. \quad (46)\end{aligned}$$

We will design the sampling procedure to ensure that (i) the vertices in $Y$ are far from one another and from any short cycles, and (ii) the empirical measure $H^{\mathrm{sm}}$ of $\underline{\sigma}$ on $\mathscr{N}$ is close to $H^{\mathrm{sy}}$, a certain symmetrization of the overall empirical measure $H(\mathcal{G}, \underline{\sigma})$. Then $\mathbb{E}\boldsymbol{Z}(H) \approx \mu(A)$ where $A$ is the set of states $(\mathscr{G}, Y, \underline{\sigma})$ with $H^{\mathrm{sm}} \approx H^{\mathrm{sy}}$. The update produces a state $(\mathscr{G}', Y, \underline{\eta}) \in B$ with possibly different $H^{\mathrm{sm}}$, but with the **same**



**(a)** $(\mathscr{G}, Y, \underline{\sigma})$



**(b)** $(\mathscr{G}', Y, \underline{\eta})$

**Fig. 4** One step of the local update procedure. In this figure, open circles indicate variable factors $\dot{\Phi}$, solid squares indicate clause factors $\hat{\Phi}^{\text{lit}}$, and each variable-clause edge is bisected by a small dot indicating the edge factor $\bar{\Phi}$. The initial state (top panel) is a triple $(\mathscr{G}, Y, \underline{\sigma})$ where $\mathscr{G}$ is an NAE- SAT instance, $Y$ is a subset of variables (blue circles), and $\underline{\sigma}$ is a coloring on $\mathscr{G}$ (not shown). Let $\mathscr{N} \equiv \mathscr{N}(Y)$ be the subgraph of $\mathscr{G}$ induced by the variables in $Y$, together with the clausees neighboring $Y$ and their incident half-edges (shown in blue, above the dashed line). The local update procedure resamples the edge literals on $\mathscr{N}$, as well as the matching between $\mathscr{N}$ and $\mathscr{G}\backslash\mathscr{N}$, to produce a modified instance $\mathscr{G}'$. The coloring is updated accordingly so that we arrive at the new state $(\mathscr{G}', Y, \underline{\eta})$ (bottom panel). In both $\mathscr{G}$ and $\mathscr{G}'$, the edges cut by the dashed lines will be referred to as **cut edges**. The half-edges just above the dashed lines will be referred to as the **leaf edges of** $\mathscr{N}$ (color figure online)

empirical measure $\dot{h}^{\text{tr}}(H^{\text{sm}})$ of variable-to-clause colors $\dot{\sigma}$ on the leaf edges of $\mathscr{N}$. Bounding $\pi(\text{B}, \text{A})$ reduces to calculating the weight of configurations on $\mathscr{N}$ with empirical measure $H^{\text{sm}} \approx H^{\text{sy}}$, relative to the weight of all configurations on $\mathscr{N}$ with empirical measure $\dot{h}^{\text{tr}}(H^{\text{sm}}) \approx \dot{h}^{\text{tr}}(H^{\text{sy}})$ on the leaf edges of $\mathscr{N}$. Because $\mathscr{N}$ is a disjoint union of **trees**, this reduces to a convex optimization problem which lends

itself much more readily to analysis. The purpose of the current section is to formalize this graphs-to-trees reduction. We begin with the precise definitions of $H^{\mathrm{sy}}$, $H^{\mathrm{sm}}$ and $\dot{h}^{\mathrm{tr}}(H^{\mathrm{sm}})$. Recall our notation $\sigma \equiv (\dot{\sigma}, \hat{\sigma}) \in \Omega$ from the discussion following (33). As $\sigma$ goes over all of $\Omega_T$, write $\dot{\Omega}_T$ for the possible values of $\dot{\sigma}$, and $\hat{\Omega}_T$ for the possible values of $\hat{\sigma}$. Let $\dot{\Omega} \equiv \dot{\Omega}_\infty$ and $\hat{\Omega} \equiv \hat{\Omega}_\infty$, so

$$\dot{\Omega} \equiv \{\mathtt{r}_0, \mathtt{r}_1, \mathtt{b}_0, \mathtt{b}_1\} \cup (\dot{\mathscr{M}} \backslash \{0, 1, \star\}),$$
$$\hat{\Omega} \equiv \{\mathtt{r}_0, \mathtt{r}_1, \mathtt{b}_0, \mathtt{b}_1\} \cup (\hat{\mathscr{M}} \backslash \{0, 1, \star\})$$

**Definition 4.1** *(sample empirical measures)* Given an NAE- SAT instance $\mathscr{G} \equiv (\mathcal{G}, \underline{\mathrm{L}})$, a $T$-coloring $\underline{\sigma}$ on $\mathscr{G}$, and a nonempty subset of variables $Y \subseteq V$, we record the local statistics of "$\underline{\sigma}$ around $Y$" as follows. Let $\dot{H}^{\mathrm{sm}}$ be the empirical measure of variable-incident colorings in $Y$: for $\underline{\eta} \in (\Omega_T)^d$,

$$\dot{H}^{\mathrm{sm}}(\underline{\eta}) \equiv \frac{1}{|Y|} \sum_{v \in Y} \mathbf{1}\{\underline{\sigma}_{\delta v} = \underline{\eta}\}.$$

Let $\bar{H}^{\mathrm{sm}}$ be the empirical measure of colors on the edges incident to $Y$: for $\eta \in \Omega_T$,

$$\bar{H}^{\mathrm{sm}}(\eta) \equiv \frac{1}{|Y|d} \sum_{v \in Y} \sum_{e \in \delta v} \mathbf{1}\{\sigma_e = \eta\}.$$

For $\underline{\eta} \in (\Omega_T)^k$ and $1 \le j \le k$ define the rotation $\underline{\eta}^{(j)} \equiv (\eta_j, \ldots, \eta_k, \eta_1, \ldots, \eta_{j-1})$. For any $v \in Y$ and $e \in \delta v$, let $j(e)$ be the index of $e$ in $\delta a(e)$. For $\underline{\eta} \in (\Omega_T)^k$ let

$$\hat{H}^{\mathrm{sm}}(\underline{\eta}) \equiv \frac{1}{|Y|d} \sum_{v \in Y} \sum_{e \in \delta v} \mathbf{1}\{(\underline{\sigma}_{\delta a(e)})^{j(e)} = \underline{\eta}\}.$$

Then $H^{\mathrm{sm}} \equiv (\dot{H}^{\mathrm{sm}}, \hat{H}^{\mathrm{sm}}, \bar{H}^{\mathrm{sm}})$ is the **sample empirical measure** for the state $(\mathscr{G}, Y, \underline{\sigma})$; we shall write this hereafter as $H^{\mathrm{sm}} = H^{\mathrm{sm}}(\mathcal{G}, Y, \underline{\sigma})$. Note that $H^{\mathrm{sm}}$ lies in the space $\mathbf{\Delta}^{\mathrm{sm}}$ which is defined similarly to $\mathbf{\Delta}$ but with condition (40) replaced by

$$\frac{1}{d} \sum_{\underline{\sigma} \in \Omega^d} \dot{H}^{\mathrm{sm}}(\underline{\sigma}) \sum_{i=1}^{d} \mathbf{1}\{\sigma_i = \tau\} = \bar{H}^{\mathrm{sm}}(\tau) = \sum_{\underline{\sigma} \in \Omega^k} \hat{H}^{\mathrm{sm}}(\underline{\sigma}) \mathbf{1}\{\sigma_1 = \tau\}. \tag{47}$$

We emphasize that (40) and (47) differ on the right-hand side. However, if we have $H = (\dot{H}, \hat{H}, \bar{H}) \in \mathbf{\Delta}$ such that $\hat{H}$ is invariant under rotation of the indices $1 \le j \le k$, then $H \in \mathbf{\Delta}^{\mathrm{sm}}$ as well. With this in mind, for $H \in \mathbf{\Delta}$ we define $H^{\mathrm{sy}} \equiv (\dot{H}, \hat{H}^{\mathrm{sy}}, \bar{H}) \in \mathbf{\Delta}^{\mathrm{sm}}$ where $\hat{H}^{\mathrm{sy}}$ is the average over all $k$ rotations of $\hat{H}$. Later we will sample $Y$ such that $H^{\mathrm{sm}}$ falls very close to $H^{\mathrm{sy}}$ with high probability. Lastly, for any $H^{\mathrm{sm}} \in \mathbf{\Delta}^{\mathrm{sm}}$ we

let $\dot{h}^{\mathrm{tr}}(H^{\mathrm{sm}})$ be the measure on $\dot{\Omega}_T$ given by

$$[\dot{h}^{\mathrm{tr}}(H^{\mathrm{sm}})](\dot{\eta}) \equiv \frac{1}{k-1} \sum_{\underline{\sigma} \in (\Omega_T)^k} \sum_{j=2}^{k} \mathbf{1}\{\dot{\sigma}_j = \dot{\eta}\} \hat{H}^{\mathrm{sm}}(\underline{\sigma}) \,.$$

Thus $\dot{h}^{\mathrm{tr}}(H^{\mathrm{sm}})$ represents the empirical measure of spins $\dot{\sigma}$ on the leaf edges of $\mathscr{N}$, i.e., the edges cut by the dashed lines in Fig. 4.

For $H \in \boldsymbol{\Delta}$, recall from (43) that $\boldsymbol{F}(H) = \boldsymbol{\Sigma}(H) + \lambda \boldsymbol{s}(H)$ where $\boldsymbol{\Sigma}(H)$ is the cluster complexity and $\boldsymbol{s}(H)$ is (the exponential rate of) the cluster size. The tree analogue of $\boldsymbol{\Sigma}(H)$ is $\boldsymbol{\Sigma}^{\mathrm{tr}}(H^{\mathrm{sm}})$ where

$$\boldsymbol{\Sigma}^{\mathrm{tr}}(H) \equiv \mathcal{H}(\dot{H}) + d\mathcal{H}(\hat{H}) - d\mathcal{H}(\bar{H}) + \boldsymbol{v}(H) \tag{48}$$

— the only difference being that $\boldsymbol{\Sigma}$ has coefficient $\alpha = d/k$ on the clause entropy term $\mathcal{H}(\hat{H})$, while $\boldsymbol{\Sigma}^{\mathrm{tr}}$ has coefficient $d$. As we see below, this occurs because the ratio of variables to clauses to edges is $1 : d : d$ for the disjoint union of trees $\mathscr{N}$, versus $1 : \alpha : d$ for the full graph $\mathscr{G}$. We will also see that $\boldsymbol{\Sigma}^{\mathrm{tr}}$ is always concave, though $\boldsymbol{\Sigma}$ need not be. Likewise, the tree analogue of $\boldsymbol{s}(H)$ is $\boldsymbol{s}^{\mathrm{tr}}(H^{\mathrm{sm}})$ where

$$\boldsymbol{s}^{\mathrm{tr}}(H) \equiv \langle \ln \dot{\Phi}, \dot{H} \rangle + d\langle \ln \hat{F}, \hat{H} \rangle + d\langle \ln \bar{\Phi}, \bar{H} \rangle \,.$$

The tree analogue of $\boldsymbol{F}(H)$ is $\boldsymbol{\Lambda}(H^{\mathrm{sm}})$ where

$$\boldsymbol{\Lambda}(H) \equiv \boldsymbol{\Sigma}^{\mathrm{tr}}(H) + \lambda \boldsymbol{s}^{\mathrm{tr}}(H) \,. \tag{49}$$

Recall Definition 4.1: given $(\mathscr{G}, Y, \underline{\sigma})$ with sample empirical measure $H^{\mathrm{sm}}$, the empirical measure of spins $\dot{\sigma}$ on the leaf edges of $\mathscr{N}(Y)$ is given by $\dot{h}^{\mathrm{tr}} = \dot{h}^{\mathrm{tr}}(H^{\mathrm{sm}})$. Then, for any probability measure $\dot{h}$ on $\dot{\Omega}_T$, we let

$$\boldsymbol{\Lambda}^{\mathrm{op}}(\dot{h}) \equiv \sup\{\boldsymbol{\Lambda}(H) : H \in \boldsymbol{\Delta}^{\mathrm{sm}} \text{ with } \dot{h}^{\mathrm{tr}}(H) = \dot{h}\} \,,$$

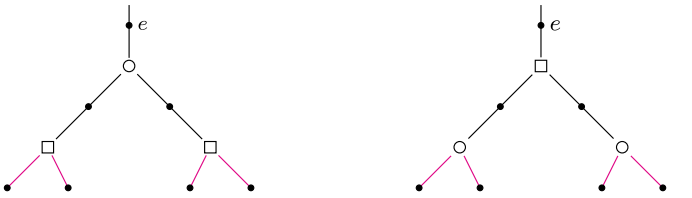
where we emphasize that the supremum is taken over $\boldsymbol{\Delta}^{\mathrm{sm}}$ rather than $\boldsymbol{\Delta}$. For $H \in \boldsymbol{\Delta}^{\mathrm{sm}}$ we define

$$\boldsymbol{\Xi}(H) \equiv \boldsymbol{\Xi}_{\lambda,T}(H) \equiv \boldsymbol{\Lambda}^{\mathrm{op}}(\dot{h}^{\mathrm{tr}}(H)) - \boldsymbol{\Lambda}(H) \,. \tag{50}$$

The interpretation of $\boldsymbol{\Xi}$, formalized below, is that for any $H \in \boldsymbol{\Delta}$, if $A$ is the set of states with $H^{\mathrm{sm}} \approx H^{\mathrm{sy}}$ and $B$ is the set of states reachable in one step of the chain from $A$, then $\max\{\pi(\mathrm{B}, \mathrm{A}) : \mathrm{B} \in \mathrm{B}\}$ is approximately $\exp\{-|Y|\, \boldsymbol{\Xi}(H^{\mathrm{sy}})\}$, where we note that $\boldsymbol{\Xi}(H^{\mathrm{sy}}) \geq 0$ since $\boldsymbol{\Xi}$ is nonnegative on all of $\boldsymbol{\Delta}^{\mathrm{sm}}$, and $H^{\mathrm{sy}} \in \boldsymbol{\Delta} \cap \boldsymbol{\Delta}^{\mathrm{sm}}$. Formally, we have the following bound:

**Theorem 4.2** *For $\epsilon$ small enough (depending only on $d, k, T$), it holds for all $H \in \boldsymbol{\Delta}$ that*

$$\boldsymbol{F}(H) \leq \max\left\{\boldsymbol{F}(H') : \|H' - H\| \leq \epsilon (dk)^{2T}\right\} - \epsilon\, \boldsymbol{\Xi}(H^{sy}) \,.$$




**(a)** variable-to-clause tree. **(b)** clause-to-variable tree.

**Fig. 5** Two types of directed tree $\mathscr{T}$. In each case the root edge $e$ is shown at the top, and the boundary $\delta\mathscr{T}$ is highlighted in purple. For $e \in \mathcal{W}$, the unit-radius neighborhood of $e$ in $\mathscr{G}\backslash\mathscr{N}$ typically looks like the left panel (color figure online)

*The analogous statement holds in the second moment with* $\boldsymbol{F}_2 = \boldsymbol{F}_{2,\lambda,T}$ *and* $\boldsymbol{\Xi}_2 = \boldsymbol{\Xi}_{2,\lambda,T}$.

For the sake of exposition, we will give the proof of Theorem 4.2 for $\boldsymbol{F}$ only; the assertion for $\boldsymbol{F}_2$ follows from the same argument with essentially no modifications. The first task is to define the Markov chain that was informally discussed above. There are a few issues to be addressed: how to sample $Y$ ensuring certain desirable properties; how to resample the matching between $\mathscr{N}$ and $\mathscr{G}\backslash\mathscr{N}$; and how to produce a valid coloring $\underline{\eta}$ on $\mathscr{G}'$ without changing the spins $\dot{\sigma}$ on the cut edges. We address the last issue next.

### 4.2 Tree updates

Recall that in the bipartite factor graph $\mathscr{G} = (V, F, E, \underline{\mathrm{L}})$, each edge joins a variable to a clause and is defined to have length one-half. For the discussion that follows, it is useful to bisect each edge $e \in E$ with an artificial vertex indicating the edge factor $\bar{\Phi}$; these are shown as small dots in Fig. 4. Thus an edge $e$ joining $a \in F$ to $v \in V$ becomes two quarter-length edges, $(ae)$ and $(ev)$, where $e$ now refers to the artificial vertex. Given a coloring $\underline{\sigma}$ on the original graph, we obtain a coloring on the new graph by simply duplicating the color on each edge, setting $\sigma_{ae} = \sigma_e = \sigma_{ev}$. We then define $\mathscr{N}(v)$ as the (5/8)-neighborhood of variable $v$, and define $\mathscr{N} = \mathscr{N}(Y)$ as the union of $\mathscr{N}(v)$ for all $v \in Y$: in the top panel of Fig. 4, $\mathscr{N}$ is the subgraph shown in blue, above the dashed line.

Directly below the same dashed line, the small solid orange dots correspond to the boundary edges, hereafter denoted $\mathcal{W}$, of the cavity graph $\mathscr{G}_\partial \equiv \mathscr{G}\backslash\mathscr{N}$. For $e \in \mathcal{W}$, let $\mathscr{T}$ be its neighborhood in $\mathscr{G}\backslash\mathscr{N}$ of some radius $\ell > T$ where $2\ell$ is a positive integer. Assuming $e$ is not close to a short cycle, $\mathscr{T}$ is what we will call a **directed tree** rooted at $e$. In this case we also call $\mathscr{T}$ a **variable-to-clause tree** since the root edge has no incident clause; a **clause-to-variable tree** is similarly defined. We always visualize a directed tree $\mathscr{T}$ as in Fig. 5, with the root edge $e$ at the top, so that paths leaving the root travel downwards. On an edge $e = (av)$, the **upward color** is $\dot{\sigma}_{av}$ if $a$ lies above $v$, and $\hat{\sigma}_{av}$ if $v$ lies above $a$. We let $\delta\mathscr{T}$ denote the boundary edges of $\mathscr{T}$, not including the root edge.

Suppose $\underline{\sigma}$ is a valid $T$-coloring of a directed tree $\mathscr{T}$ with root spin $\sigma_e = \sigma$, and consider a new root spin $\eta \in \Omega_T$. If $\sigma$ and $\eta$ agree on the upward color of the root edge, then there is a **unique valid coloring**

$$\underline{\eta} = \mathsf{update}(\underline{\sigma}, ; \mathscr{T}) \in (\ddot{}_{\mathrm{T}})^{\mathrm{E}(\mathscr{T})}$$

which has root spin $\eta$, and agrees with $\underline{\sigma}$ in all the upward colors. Indeed, the only possibility for $\sigma \neq \eta$ is that both $\sigma, \eta \in \{\mathtt{f}\}$. Then, recalling (23), the coloring $\mathsf{update}(\underline{\sigma}, ; \mathscr{T})$ is uniquely defined by recursively applying the mappings $\dot{T}$ and $\hat{T}$, starting from the root and continuing downwards. Since we assumed that $\underline{\sigma}$ was a valid $T$-coloring and $\eta \in \Omega_T$, it is easy to verify that the resulting $\underline{\eta}$ is also a valid $T$-coloring, so the update procedure respects the restriction to $\Omega_T$. From now on we assume all edge colors belong to $\Omega_T$.

**Lemma 4.3** *Suppose $\underline{\sigma}$ is a valid $T$-coloring of the directed tree $\mathscr{T}$ with root color $\sigma$, and $\eta \in \Omega_T$ agrees with $\sigma$ on the upward color of the root edge. If $\underline{\eta} = \mathsf{update}(\underline{\sigma}, ; \mathscr{T})$ agrees with $\underline{\sigma}$ on the boundary $\delta\mathscr{T}$, then $\boldsymbol{w}^{lit}_{\mathscr{T}}(\underline{\sigma}) = \boldsymbol{w}^{lit}_{\mathscr{T}}(\underline{\eta})$.*

***Proof*** It follows from the construction that on any edge $e$ in the tree, $\sigma_e$ and $\eta_e$ agree on the upward color; moreover, if $\sigma_e \neq \eta_e$ then we must have $\sigma_e, \eta_e \in \{\mathtt{f}\}$. For each vertex $x \in \mathscr{T}$, let $e(x)$ denote the parent edge of $x$, that is, the unique edge of $\mathscr{T}$ which lies above $x$. We then have

$$\boldsymbol{w}^{\mathrm{lit}}_{\mathscr{T}}(\underline{\sigma}) = \prod_{e \in \delta\mathscr{T}} \bar{\Phi}(\sigma_e) \prod_{v \in V(\mathscr{T})} \left\{ \dot{\Phi}(\underline{\sigma}_{\delta v}) \bar{\Phi}(\sigma_{e(v)}) \right\} \prod_{a \in F(\mathscr{T})} \left\{ \hat{\Phi}^{\mathrm{lit}}((\underline{\sigma} \oplus \underline{\mathtt{L}})_{\delta a}) \bar{\Phi}(\sigma_{e(a)}) \right\}.$$

For a clause $a$ in $\mathscr{T}$ with $e(a) = e$, if both $\sigma_e, \eta_e \in \{\mathtt{f}\}$ then it follows directly from (39) that

$$\hat{\Phi}^{\mathrm{lit}}((\underline{\sigma} \oplus \underline{\mathtt{L}})_{\delta a}) \bar{\Phi}(\sigma_e) = \hat{\varphi}^{\mathrm{lit}}((\underline{\dot{\tau}} \oplus \underline{\mathtt{L}})_{\delta a}) \bar{\varphi}(\sigma_e) = \hat{z}(\hat{\sigma}_e) = \hat{z}(\hat{\eta}_e) = \hat{\Phi}^{\mathrm{lit}}((\underline{\eta} \oplus \underline{\mathtt{L}})_{\delta a}) \bar{\Phi}(\eta_e)\,.$$

For a variable $v$ in $\mathscr{T}$ with $e(v) = e$, if $\sigma_e, \eta_e \in \{\mathtt{f}\}$, then a similar calculation as (39) gives

$$\dot{\Phi}(\underline{\sigma}_{\delta v}) \bar{\Phi}(\sigma_e) = \dot{\varphi}(\hat{\underline{\sigma}}_{\delta v}) \bar{\varphi}(\sigma_e) = \dot{z}(\dot{\sigma}_e) = \dot{z}(\dot{\eta}_e) = \dot{\Phi}(\underline{\eta}_{\delta v}) \bar{\Phi}(\eta_e)\,,$$

where we used the fact that necessarily we have $\underline{\sigma}_{\delta v} \in \{\mathtt{f}\}^d$. To conclude, we recall that $\underline{\sigma}$ and $\underline{\eta}$ agree on $\delta\mathscr{T}$ by assumption, so we have $\boldsymbol{w}^{\mathrm{lit}}_{\mathscr{T}}(\underline{\sigma}) = \boldsymbol{w}^{\mathrm{lit}}_{\mathscr{T}}(\underline{\eta})$ as claimed. □

We also use the directed tree as a device to prove the following lemma, which was used in the proofs of Corollaries 3.5 and 3.11.

**Lemma 4.4** *Let $\dot{M}, \hat{M}$ be as defined in* Corollary 3.5*, and let $\dot{M}_2, \hat{M}_2$ be their analogues in the pair model. For any $\sigma, \eta \in \Omega$ there exists an integer-valued vector $(\dot{H}, \hat{H})$ so that*

$$\langle \mathbf{1}, \dot{H} \rangle = 0 = \langle \mathbf{1}, \hat{H} \rangle \quad \text{and} \quad \dot{M}\dot{H} - \hat{M}\hat{H} = \mathbf{1}_\sigma - \mathbf{1}_\eta\,,$$

*where* $\mathbf{1}$ *denotes the all-ones vector, and* $\mathbf{1}_\sigma$ *denotes the vector which is one in the* $\sigma$ *coordinate and zero elsewhere. The analogous statement holds for* $(\dot{M}_2, \hat{M}_2)$.

***Proof*** We define a graph on $\Omega_T$ by putting an edge between $\sigma$ and $\eta$ if there exist valid colorings $\underline{\sigma}, \underline{\eta}$ on some directed tree $\mathscr{T}$ which take values $\sigma, \eta$ on the root edge, but agree on the boundary edges $\delta\mathscr{T}$. If $\sigma, \eta$ are connected in this way, then taking

$$\dot{H}(\underline{\rho}) = \sum_{v \in V(\mathscr{T})} \mathbf{1}\{\underline{\sigma}_{\delta v} = \underline{\rho}\} - \sum_{v \in V(\mathscr{T})} \mathbf{1}\{\underline{\eta}_{\delta v} = \underline{\rho}\}, \quad \underline{\rho} \in (\Omega_T)^d$$
$$\hat{H}(\underline{\rho}) = \sum_{a \in F(\mathscr{T})} \mathbf{1}\{\underline{\sigma}_{\delta a} = \underline{\rho}\} - \sum_{a \in F(\mathscr{T})} \mathbf{1}\{\underline{\eta}_{\delta a} = \underline{\rho}\}, \quad \underline{\rho} \in (\Omega_T)^k.$$

gives $\dot{M}\dot{H} - \hat{M}\hat{H} = \mathbf{1}_\sigma - \mathbf{1}_\eta$ as required. It therefore suffices to show that the graph we have defined on $\Omega_T$ is connected (hence complete). First, if $\dot{\sigma} = \dot{\eta}$, it is clear that $\sigma$ and $\eta$ can be connected by colorings $\underline{\sigma}, \underline{\eta}$ of some variable-to-clause tree $\mathscr{T}$, with $\underline{\eta} = \mathsf{update}(\underline{\alpha}, ; \mathscr{T})$. Similarly, if $\hat{\sigma} = \hat{\eta}$, then $\sigma$ and $\eta$ can be connected by a clause-to-variable tree. This implies that $\{\mathtt{f}\}$ is connected. Next, if $\sigma = \mathtt{r}_{\boldsymbol{x}}$ and $\eta = \mathtt{b}_{\boldsymbol{x}}$, then they can be connected by a variable-to-clause tree rooted at edge $e$, containing a single variable factor $v = v(e)$, with $\underline{\sigma}_{\delta v \setminus e}$ identically equal to $\mathtt{r}_{\boldsymbol{x}}$. If $\sigma = \mathtt{b}_{\boldsymbol{x}}$ and $\eta = (\dot{\tau}, \mathtt{s})$ for any $\dot{\tau} \in \dot{\Omega} \setminus \{\mathtt{r}, \mathtt{b}\}$, then they can be connected by a clause-to-variable tree rooted at edge $e$, containing a single clause factor $a = a(e)$, with any $\underline{\sigma}_{\delta a \setminus e}$ such that $(\underline{\sigma} \oplus \underline{\mathtt{L}})_{\delta a \setminus e}$ contains both $\{\mathtt{b}_0, \mathtt{b}_1\}$ entries. It follows that $\Omega_T$ is indeed connected, which proves the assertion concerning $(\dot{M}, \hat{M})$. The proof for $(\dot{M}_2, \hat{M}_2)$ is very similar and we omit the details. □

### 4.3 Markov chain

We now define a Markov chain on tuples $(\mathscr{G}, Y, \underline{\sigma})$ where $\mathscr{G}$ is an NAE- SAT instance, $\underline{\sigma}$ is a valid $T$-coloring on $\mathscr{G}$, and $Y \subseteq V$ is a subset of variables such that

$$\text{the subgraphs } B_{2T}(v), \text{ for } v \in Y, \text{ are mutually disjoint trees,} \tag{51}$$

where $B_{2T}(v)$ is to the $2T$-neighborhood of $v$ in $\mathscr{G}$. Recall that $\mathscr{N} = \mathscr{N}(Y)$ is the $\frac{5}{8}$-neighborhood of $Y$; we write $\mathscr{N} = (\mathcal{N}, \underline{\mathtt{L}}_{\mathcal{N}})$ where $\mathcal{N}$ is the graph without edge literals. Write $\delta\mathcal{N}$ for the boundary of $\mathcal{N}$, consisting of clause-incident edges that are not incident to $Y$ (just above the dashed line in Fig. 4). Write $\underline{\sigma}_{\mathcal{N}}$ for a $T$-coloring on $\mathcal{N}$ (including $\delta\mathcal{N}$), and let

$$\boldsymbol{w}^{\mathrm{lit}}_{\mathcal{N}}(\underline{\sigma}_{\mathcal{N}} | \underline{\mathtt{L}}_{\mathcal{N}}) \equiv \boldsymbol{w}^{\mathrm{lit}}_{\mathscr{N}}(\underline{\sigma}_{\mathcal{N}}) \equiv \prod_{v \in Y} \left\{ \dot{\Phi}(\underline{\sigma}_{\delta v}) \prod_{e \in \delta v} \left\{ \hat{\Phi}^{\mathrm{lit}}((\underline{\sigma} \oplus \underline{\mathtt{L}})_{\delta a(e)}) \bar{\Phi}(\sigma_e) \right\} \right\}. \tag{52}$$

On the other hand we have $\mathscr{G} \setminus \mathscr{N} \equiv \mathscr{G}_\partial \equiv (\mathcal{G}_\partial, \underline{\mathtt{L}}_\partial)$ where $\mathcal{G}_\partial \equiv (V_\partial, F_\partial, E_\partial)$, and $\mathcal{W}$ denotes the boundary of $\mathscr{G}_\partial$ (just below the dashed line in Fig. 4). Write $\underline{\sigma}_\partial$ for a

coloring on $\mathcal{G}_\partial$ (including $\mathcal{W}$), and let

$$\boldsymbol{w}^{\mathrm{lit}}_\partial(\underline{\sigma}_\partial) \equiv \prod_{v\in V_\partial} \dot{\Phi}(\underline{\sigma}_{\delta v}) \prod_{a\in F_\partial} \hat{\Phi}^{\mathrm{lit}}((\underline{\sigma}\oplus \underline{\mathtt{L}})_{\delta a}) \prod_{e\in E_\partial} \bar{\Phi}(\sigma_e)\,.$$

By matching $\delta\mathcal{N}$ to $\mathcal{W}$ (along the dashed line in Fig. 4), the graphs $\mathcal{G}_\partial$ and $\mathcal{N}$ combine to form the original instance $\mathcal{G}$. If $\underline{\sigma}$ is a valid coloring on $\mathcal{G}$, then $\underline{\sigma}_{\delta\mathcal{N}}$ and $\underline{\sigma}_{\mathcal{W}}$ must agree, and we have

$$\boldsymbol{w}^{\mathrm{lit}}_{\mathcal{G}}(\underline{\sigma}) = \boldsymbol{w}^{\mathrm{lit}}_\partial(\underline{\sigma}_\partial)\boldsymbol{w}^{\mathrm{lit}}_{\mathcal{N}}(\underline{\sigma}_{\mathcal{N}}|\underline{\mathtt{L}}_{\mathcal{N}})\,. \tag{53}$$

Let $\dot{h}^{\mathrm{tr}}(\underline{\sigma}_{\delta\mathcal{N}}) = \dot{h}^{\mathrm{tr}}$ be the empirical measure of the spins $(\dot{\sigma}_e)_{e\in\delta\mathcal{N}}$. Given initial state $(\mathcal{G}, Y, \underline{\sigma})$, we take one step of the Markov chain as follows:

1. Detach $\mathcal{N}$ from $\mathcal{G}$. On $\mathcal{N}$, sample a new assignment $(\underline{\mathtt{L}}'_{\mathcal{N}}, \underline{\eta}_{\mathcal{N}})$ from the probability measure

$$p((\underline{\mathtt{L}}'_{\mathcal{N}}, \underline{\eta}_{\mathcal{N}}) \mid (\underline{\mathtt{L}}_{\mathcal{N}}, \underline{\sigma}_{\mathcal{N}})) = \frac{\mathbf{1}\{\dot{h}^{\mathrm{tr}}(\underline{\eta}_{\delta\mathcal{N}}) = \dot{h}^{\mathrm{tr}}\}\boldsymbol{w}^{\mathrm{lit}}_{\mathcal{N}}(\underline{\eta}_{\mathcal{N}}|\underline{\mathtt{L}}'_{\mathcal{N}})^\lambda}{z(|Y|, \dot{h}^{\mathrm{tr}})} \tag{54}$$

   where the denominator is the normalizing constant obtained by summing over all possible $(\underline{\mathtt{L}}'_{\mathcal{N}}, \underline{\eta}_{\mathcal{N}})$.
2. Form the new graph $\mathcal{G}'$ by sampling a uniformly random matching of $\delta\mathcal{N}$ with $\mathcal{W}$, subject to the constraint that $e \in \mathcal{W}$ must be matched to $e' \in \delta\mathcal{N}$ with $\dot{\sigma}_e = \dot{\eta}_{e'}$. The number of such matchings depends only on $|Y|$ and $\dot{h}^{\mathrm{tr}}$, so we denote it as $\mathcal{M}(|Y|, \dot{h}^{\mathrm{tr}})$. For each matched pair $(e, e')$ where $\hat{\sigma}_e \neq \hat{\eta}_{e'}$, let $\mathscr{T} = \mathscr{T}(e)$ be the radius-$2T$ neighborhood of $e$ in the graph $\mathcal{G}_\partial$. Let

$$\underline{\eta}_{\mathscr{T}} \equiv \mathsf{update}(\underline{\mathfrak{o}}_{\mathscr{T}}, \mathfrak{e}; \mathscr{T})$$

   and note that, since $\underline{\sigma}$ is a valid $T$-coloring, $\underline{\eta}_{\mathscr{T}}$ and $\underline{\sigma}_{\mathscr{T}}$ must agree at the boundary of $\mathscr{T}$. Finally, on the rest of $\mathcal{G}_\partial$ outside the radius-$2T$ neighborhood of $\mathcal{W}$, we simply take $\underline{\eta}$ and $\underline{\sigma}$ to be the same.

The state of the Markov chain after one step is $(\mathcal{G}', Y, \underline{\eta})$ where $\underline{\eta}$ is a valid $T$-coloring on $\mathcal{G}'$.

**Lemma 4.5** *Suppose we have a sampling mechanism for a random subset of variables $Y$ in $\mathcal{G}$ such that, whenever $(\mathcal{G}, Y, \underline{\sigma})$ and $(\mathcal{G}', Y, \underline{\eta})$ appear in the same orbit of the Markov chain, we have*

$$\mathbb{P}(Y \mid \mathcal{G}) = \mathbb{P}(Y \mid \mathcal{G}')\,. \tag{55}$$

*A reversing measure for the Markov chain is then given by $\mu(\mathcal{G}, Y, \underline{\sigma}) = \mathbb{P}(\mathcal{G})\mathbb{P}(Y \mid \mathcal{G})\boldsymbol{w}^{lit}_{\mathcal{G}}(\underline{\sigma})^\lambda$.*

***Proof*** Given $\mathrm{A} = (\mathcal{G}, \mathrm{Y}, \underline{\mathfrak{o}})$, let $\mathrm{B} = (\mathcal{G}', \mathrm{Y}, )$ be any state reachable from A in a single step of the chain. By the factorization (53), together with assumption (55) and the fact that $\mathbb{P}(\mathcal{G}) = \mathbb{P}(\mathcal{G}')$,

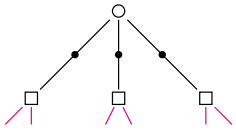


**Fig. 6** If $s \equiv |Y|$ and $\mathscr{N} = \mathscr{N}(Y) = (\mathcal{N}, \underline{\mathtt{L}})$, the graph $\mathcal{N}$ consists of $s$ disjoint copies of the tree $\mathcal{D}$ shown here. The boundary $\delta\mathcal{D}$ is highlighted in purple (color figure online)

$$\frac{\mu(\mathrm{A})}{\mu(\mathrm{B})} = \frac{\mathbb{P}(\mathscr{G})\mathbb{P}(Y \mid \mathscr{G})\boldsymbol{w}_{\partial}^{\mathrm{lit}}(\underline{\sigma}_{\partial})^{\lambda}\boldsymbol{w}_{\mathcal{N}}^{\mathrm{lit}}(\underline{\sigma}_{\mathcal{N}}|\underline{\mathtt{L}}_{\mathcal{N}})^{\lambda}}{\mathbb{P}(\mathscr{G}')\mathbb{P}(Y \mid \mathscr{G}')\boldsymbol{w}_{\partial}^{\mathrm{lit}}(\underline{\eta}_{\partial})^{\lambda}\boldsymbol{w}_{\mathcal{N}}^{\mathrm{lit}}(\underline{\eta}_{\mathcal{N}}|\underline{\mathtt{L}}'_{\mathcal{N}})^{\lambda}} = \frac{\boldsymbol{w}_{\partial}^{\mathrm{lit}}(\underline{\sigma}_{\partial})^{\lambda}\boldsymbol{w}_{\mathcal{N}}^{\mathrm{lit}}(\underline{\sigma}_{\mathcal{N}}|\underline{\mathtt{L}}_{\mathcal{N}})^{\lambda}}{\boldsymbol{w}_{\partial}^{\mathrm{lit}}(\underline{\eta}_{\partial})^{\lambda}\boldsymbol{w}_{\mathcal{N}}^{\mathrm{lit}}(\underline{\eta}_{\mathcal{N}}|\underline{\mathtt{L}}'_{\mathcal{N}})^{\lambda}} = \frac{\boldsymbol{w}_{\mathcal{N}}^{\mathrm{lit}}(\underline{\sigma}_{\mathcal{N}}|\underline{\mathtt{L}}_{\mathcal{N}})^{\lambda}}{\boldsymbol{w}_{\mathcal{N}}^{\mathrm{lit}}(\underline{\eta}_{\mathcal{N}}|\underline{\mathtt{L}}'_{\mathcal{N}})^{\lambda}},$$

where the last identity is by Lemma 4.3. On the other hand, with $\pi$ denoting the transition probabilities for the Markov chain, (54) implies

$$\frac{\pi(\mathrm{A}, \mathrm{B})}{\pi(\mathrm{B}, \mathrm{A})} = \frac{\boldsymbol{w}_{\mathcal{N}}^{\mathrm{lit}}(\underline{\eta}_{\mathcal{N}}|\underline{\mathtt{L}}'_{\mathcal{N}})^{\lambda}}{\mathcal{M}(|Y|, \dot{h}^{\mathrm{tr}})z(|Y|, \dot{h}^{\mathrm{tr}})} \frac{\mathcal{M}(|Y|, \dot{h}^{\mathrm{tr}})z(|Y|, \dot{h}^{\mathrm{tr}})}{\boldsymbol{w}_{\mathcal{N}}^{\mathrm{lit}}(\underline{\sigma}_{\mathcal{N}}|\underline{\mathtt{L}}_{\mathcal{N}})^{\lambda}} = \frac{\mu(\mathrm{B})}{\mu(\mathrm{A})}.$$

Rearranging proves reversibility, $\mu(\mathrm{A})ß(\mathrm{A}, \mathrm{B}) = \bar{}(\mathrm{B})ß(\mathrm{B}, \mathrm{A})$. (We remark that since the Markov chain breaks up into many disjoint orbits, the reversing measure $\mu$ is not unique.) □

### 4.4 From graph to tree optimizations

If $Y$ satisfies condition (51) and we define $\mathscr{N} = \mathscr{N}(Y) = (\mathcal{N}, \underline{\mathtt{L}})$ as before, then $\mathcal{N}$ consists of $|Y| \equiv s$ disjoint copies of the tree $\mathcal{D}$ shown in Fig. 6. Recall from Definition 4.1 the definition of $H^{\mathrm{sm}} = H^{\mathrm{sm}}(\mathcal{G}, Y, \underline{\sigma})$, and note $H^{\mathrm{sm}}$ depends only on $\underline{\sigma}_{\mathcal{N}}$. For any $H^{\mathrm{sm}} \in \boldsymbol{\Delta}^{\mathrm{sm}}$ we let $\boldsymbol{Z}(H^{\mathrm{sm}}; \mathscr{N})$ be the partition function of all colorings on $\mathscr{N}$ with empirical measure $H^{\mathrm{sm}}$—the only randomness comes from the literals $\underline{\mathtt{L}}_{\mathcal{N}}$. The expected number of valid colorings is

$$\mathbb{E}\boldsymbol{Z}_{\lambda=0,T}(H^{\mathrm{sm}}; \mathscr{N}) = \exp\{sd\langle \hat{H}^{\mathrm{sm}}, \hat{v}\rangle\}\binom{s}{s\dot{H}^{\mathrm{sm}}}\binom{ds}{ds\hat{H}^{\mathrm{sm}}}\Big/\binom{ds}{ds\bar{H}^{\mathrm{sm}}}$$

which by Stirling's formula is $s^{O(1)}\exp\{s\,\boldsymbol{\Sigma}^{\mathrm{tr}}(H^{\mathrm{sm}})\}$ (see (48)). Any valid coloring $\underline{\sigma}_{\mathcal{N}}$ with empirical measure $H^{\mathrm{sm}}$ contributes weight $\boldsymbol{w}_{\mathcal{N}}^{\mathrm{lit}}(\underline{\sigma}_{\mathcal{N}}|\underline{\mathtt{L}}_{\mathcal{N}})^{\lambda} = \exp\{s\lambda\boldsymbol{s}^{\mathrm{tr}}(H^{\mathrm{sm}})\}$, so altogether

$$\mathbb{E}\boldsymbol{Z}(H^{\mathrm{sm}}; \mathscr{N}) = s^{O(1)}\exp\{s[\boldsymbol{\Sigma}^{\mathrm{tr}}(H^{\mathrm{sm}}) + \lambda\boldsymbol{s}^{\mathrm{tr}}(H^{\mathrm{sm}})]\} = s^{O(1)}\exp\{s\,\boldsymbol{\Lambda}(H^{\mathrm{sm}})\}, \tag{56}$$

with $\boldsymbol{\Lambda}$ as in (49). (This calculation clarifies why we refer to $\boldsymbol{\Sigma}^{\mathrm{tr}}, \boldsymbol{\Lambda}$ as the "tree analogues" of $\boldsymbol{\Sigma}, \boldsymbol{F}$.)

Fix $H^{\mathrm{sm}} \in \boldsymbol{\Delta}^{\mathrm{sm}}$ and let $A(H^{\mathrm{sm}})$ be the set of all $(\mathscr{G}, Y, \underline{\sigma})$ with $H^{\mathrm{sm}}(\mathcal{G}, Y, \underline{\sigma}) = H^{\mathrm{sm}}$. Let $B(H^{\mathrm{sm}})$ be the set of states reachable in one step from $A(H^{\mathrm{sm}})$. Then, for all $\mathrm{B} \in \mathrm{B}(\mathrm{H}^{\mathrm{sm}})$,

$$\pi(\mathrm{B}, \mathrm{A}(\mathrm{H}^{\mathrm{sm}})) = \frac{\mathbb{E}\boldsymbol{Z}(\mathrm{H}^{\mathrm{sm}}; \mathscr{N})}{\sum_{\mathrm{H}' \in \boldsymbol{\Delta}^{\mathrm{sm}}} \mathbf{1}\{\mathrm{Ptr}(\mathrm{H}') = \mathrm{Ptr}(\mathrm{H}^{\mathrm{sm}})\} \mathbb{E}\boldsymbol{Z}(\mathrm{H}'; \mathscr{N})} = \mathrm{s}^{\mathrm{O}(1)} \exp\{-\mathrm{s}\, \boldsymbol{\Xi}(\mathrm{H}^{\mathrm{sm}})\}. \quad (57)$$

This is the key calculation for Theorem 4.2. To complete the proof, what remains is to produce a sampling mechanism $\mathbb{P}(Y \mid \mathscr{G})$ which satisfies our earlier conditions (51) and (55), together with some concentration bound to ensure that in most cases $Y$ is large and $H^{\mathrm{sm}} \approx H^{\mathrm{sy}}$. We formalize this as follows:

**Definition 4.6** *(sampling mechanism)* Let $\mathbb{P}(Y \mid \mathscr{G})$ be the probability of sampling the subset of variables $Y$ from the NAE- SAT instance $\mathscr{G}$. We call this a **good sampling mechanism** if the following holds: first, whenever $(\mathscr{G}, Y, \underline{\sigma})$ and $(\mathscr{G}', Y, \underline{\eta})$ appear in the same orbit of the Markov chain, we must have $\mathbb{P}(Y \mid \mathscr{G}) = \mathbb{P}(Y \mid \mathscr{G}')$ (used in Lemma 4.5 to show reversibility). Next we require that for every $\mathscr{G}$, and for every $Y$ with $\mathbb{P}(Y \mid \mathscr{G})$ positive, the neighborhoods $B_{2T}(v)$ for $v \in Y$ are mutually disjoint trees (condition (51), required for defining the Markov chain). Lastly we require that for all but an exceptional set $\mathscr{B}$ of "bad" NAE- SAT instances, with $\mathbb{P}(\mathscr{G} \in \mathscr{B}) \leq \exp\{-n(\ln n)^{1/2}\}$, we have

$$\sum_{Y : n\epsilon \leq |Y| \leq 4n\epsilon} \mathbb{P}(Y \mid \mathscr{G}) \mathbf{1}\left\{ \left\| H^{\mathrm{sm}}(\mathcal{G}, Y, \underline{\sigma}) - H^{\mathrm{sy}}(\mathcal{G}, \underline{\sigma}) \right\| \leq \frac{1}{(\ln \ln n)^{1/2}} \right\} \geq \frac{1}{2}. \quad (58)$$

for all colorings $\underline{\sigma}$ on $\mathscr{G}$.

**Proposition 4.7** *Assume the existence of a good sampling mechanism in the sense of* Definition 4.6. *Then, for $\epsilon$ small enough (depending only on $d, k, T$), it holds for all $H \in \boldsymbol{\Delta}$ that*

$$\frac{\mathbb{E}\boldsymbol{Z}(H)}{\mathbb{E}\boldsymbol{Z}(\mathbf{N}_{H,\epsilon})} \leq \frac{\exp\{o_n(1)\}}{\exp\{n\epsilon\, \boldsymbol{\Xi}(H^{sy})\}}$$

*where $\mathbf{N}_{H,\epsilon} \equiv \{H' \in \boldsymbol{\Delta} : \|H - H'\| \leq \epsilon(dk)^{2T}\}$ and $H^{sy}$ is the symmetrization of $H$ from* Definition 4.1.

***Proof*** Abbreviate $\delta \equiv 1/(\ln \ln n)^{1/2}$. Given $H \in \boldsymbol{\Delta}$ and its symmetrization $H^{\mathrm{sy}} \in \boldsymbol{\Delta} \cap \boldsymbol{\Delta}^{\mathrm{sm}}$, let

$$A = \left\{ (\mathscr{G}, Y, \underline{\sigma}) : |Y| \geq n\epsilon \text{ and } \left\| H^{\mathrm{sm}}(\mathcal{G}, Y, \underline{\sigma}) - H^{\mathrm{sy}} \right\| \leq 2\delta \right\}.$$

With $\mu$ the reversing measure from Lemma 4.5, we have

$$\mu(A) \geq \sum_{\mathscr{G} \notin \mathscr{B}} \mathbb{P}(\mathscr{G}) \sum_{\underline{\sigma}} \boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G}}(\underline{\sigma})^{\lambda} \sum_{Y : |Y| \geq n\epsilon} \mathbb{P}(Y \mid \mathscr{G}) \mathbf{1}\left\{ \begin{array}{c} \|H^{\mathrm{sm}}(\mathcal{G}, Y, \underline{\sigma}) - H^{\mathrm{sy}}(\mathcal{G}, \underline{\sigma})\| \leq \delta \\ \text{and } \|H(\mathcal{G}, \underline{\sigma}) - H\| \leq \delta \end{array} \right\},$$

using $\|H(\mathcal{G}, \underline{\sigma}) - H\| \leq \|H^{\mathrm{sy}}(\mathcal{G}, \underline{\sigma}) - H^{\mathrm{sy}}\|$ together with the triangle inequality. Applying (58) gives

$$\mu(A) \geq \frac{1}{2} \sum_{\mathscr{G} \notin \mathscr{B}} \mathbb{P}(\mathscr{G}) \sum_{\underline{\sigma}} \boldsymbol{w}_{\mathscr{G}}^{\mathrm{lit}}(\underline{\sigma})^{\lambda} \mathbf{1}\Big\{ \|H(\mathcal{G}, \underline{\sigma}) - H\| \leq \delta \Big\} \geq \frac{\mathbb{E}[\boldsymbol{Z}(H); \mathscr{G} \notin \mathscr{B}]}{2} \geq \frac{\mathbb{E}\boldsymbol{Z}(H)}{4},$$

where the last step follows from the bound on $\mathbb{P}(\mathscr{G} \in \mathscr{B})$. If $B$ is the set of states reachable from $A$ in one step of the Markov chain, then crudely $\mu(B) \leq \mathbb{E}\boldsymbol{Z}(\mathbf{N}_{H,\epsilon})$, and

$$\max_{\mathrm{B} \in \mathrm{B}} \pi(\mathrm{B}, \mathrm{A}) \leq \frac{\exp\{\mathrm{o_n}(1)\}}{\exp\{\mathrm{n}\, \Xi(\mathrm{H^{sy}})\}}$$

by our earlier calculation (57). The result follows by substituting into (46). □

### 4.5 Sampling mechanism

To complete the proof of Theorem 4.2, it remains for us to define a sampling mechanism satisfying the conditions of Definition 4.6. To this end, given a $(d, k)$-regular graph $\mathcal{G}$, let $V_t \subseteq V$ be the subset of variables $v \in V$ such that the $t$-neighborhood $B_t(v)$ around $v$ is a tree. Recall the following form of the Chernoff bound: if $X$ is a binomial random variable with mean $\mu$, then for all $t \geq 1$ we have $\mathbb{P}(X \geq t\mu) \leq \exp\{-t\mu \ln(t/e)\}$.

**Lemma 4.8** *Suppose $\mathcal{G}$ is sampled from the $(d, k)$-regular configuration model on $n$ vertices. For any fixed $t$ we have $\mathbb{P}(|V \setminus V_t| \geq n/(\ln \ln n)) \leq \exp\{-n(\ln n)^{1/2}\}$ for $n$ large enough (depending on $d, k, t$).*

***Proof*** Let $\gamma$ count the total number of cycles in $\mathcal{G}$ of length at most $2t$. If $v \notin V_t$ then $v$ must certainly lie within distance $t$ of one of these cycles, so crudely we have

$$|V \setminus V_t| \leq 2t(dk)^t \gamma\,. \tag{59}$$

Consider breadth-first search exploration in $\mathcal{G}$ started from an arbitrary variable, say $v = 1$. At each step of the exploration we reveal one edge, so the exploration takes $nd$ steps total. Conditioned on everything revealed in the first $t$ steps, the chance that the edge revealed at step $t + 1$ will form a new cycle of length $\leq 2t$ is upper bounded by$(dk)^{2t}/(nd - t)$. It follows that the total number of cycles revealed up to time $nd(1 - \delta)$ is stochastically dominated by a binomial random variable

$$\gamma' \sim \mathrm{Bin}\left(nd(1-\delta), \frac{(dk)^{2t}}{nd\delta}\right).$$

The final $nd\delta$ exploration steps form at most $nd\delta$ cycles, so $\gamma \leq \gamma' + nd\delta$. Applying the Chernoff bound (as stated above) with $\delta = 1/(\ln \ln n)^2$, we obtain

$$\mathbb{P}(\gamma \geq 2nd\delta) \leq \mathbb{P}(\gamma' \geq nd\delta) \leq \exp\left\{ - nd\delta \ln\left(\frac{nd\delta^2}{e(dk)^{2t}}\right)\right\} \leq \exp\{-n(\ln n)^{1/2}\}$$

for large enough $n$. Recalling (59) gives the claimed bound. □

Given an instance $\mathscr{G} = (\mathcal{G}, \underline{\mathrm{L}})$, let $V_t$ be as defined above and take $V' \equiv V_{4T}$. We then take i.i.d. random variables $I_v \sim \mathrm{Ber}(\epsilon')$ indexed by $v \in V'$ (for $\epsilon'$ a constant to be determined) and let

$$Y_v \equiv \mathbf{1}\{I_v = 1, \text{ and } I_u = 0 \text{ for all } u \in B_{4T}(v) \backslash \{v\}\}. \tag{60}$$

We then define $\mathbb{P}(Y \mid \mathscr{G})$ to be the law of the set $Y = \{v \in V' : Y_v = 1\}$. Note that the random variables $Y_v$, for $v \in V'$, all have the same expected value, so we can define $\epsilon \equiv (\mathbb{E}Y_v)/2$.

**Lemma 4.9** *Define $\mathscr{B}$ to be the set of all $\mathscr{G} = (\mathcal{G}, \underline{\mathrm{L}})$ with $|V \backslash V_{4T}| \le n/(\ln \ln n)$. For the sampling mechanism described above, condition (58) holds for any $\mathscr{G} \notin \mathscr{B}$ and any coloring $\underline{\sigma}$ on $\mathscr{G}$.*

***Proof*** Fix an instance $\mathscr{G} \notin \mathscr{B}$ and a coloring $\underline{\sigma}$ on $\mathscr{G}$. Recalling Definition 4.1, for each $v \in V$ denote

$$X_v \equiv (\dot{X}_v, \hat{X}_v, \bar{X}_v) \equiv H^{\mathrm{sm}}(\mathcal{G}, \{v\}, \underline{\sigma}).$$

Assume without loss that $V' \equiv V_{4T} = [n'] \equiv \{v_1, \ldots, v_{n'}\}$, and for $0 \le \ell \le n'$ let $\mathscr{F}_\ell$ denote the sigma-field generated by $Y_1, \ldots, Y_\ell$. Consider

$$S \equiv \sum_{v \le n'} A_v Y_v$$

where we can take different choices of $A_v$ to prove various different bounds:

- taking $A_v = 1$ gives $S = |Y|$ and $\mathbb{E}S = 2n'\epsilon$;
- taking $A_v = \dot{X}_v(\underline{\eta})$ gives $S = |Y|\dot{H}^{\mathrm{sm}}(\underline{\eta})$ and $|\mathbb{E}S - 2n'\epsilon\dot{H}(\underline{\eta})| \le n - n'$;
- taking $A_v = \hat{X}_v(\underline{\eta})$ gives $S = |Y|\hat{H}^{\mathrm{sm}}(\underline{\eta})$ and $|\mathbb{E}S - 2n'\epsilon\hat{H}(\underline{\eta})| \le n - n'$;
- taking $A_v = \bar{X}_v(\eta)$ gives $S = |Y|\bar{H}^{\mathrm{sm}}(\eta)$ and $|\mathbb{E}S - 2n'\epsilon\bar{H}(\eta)| \le n - n'$,

where we recall that $n - n' = |V \backslash V_{4T} \le n/(\ln \ln n)$. Consider the Doob martingale

$$M_\ell \equiv \mathbb{E}(S \mid \mathscr{F}_\ell) \equiv \sum_{v \le n'} A_v \, \mathbb{E}(Y_v \mid \mathscr{F}_\ell).$$

For $\ell \le n'$, if $v$ lies at distance greater than $8T$ from any variable in $[\ell] \equiv \{v_1, \ldots, v_\ell\}$, then

$$\mathbb{E}(Y_v \mid \mathscr{F}_\ell) = \mathbb{E}Y_v = 2\epsilon.$$

Thus, the only possibility for $\mathbb{E}(Y_v \mid \mathscr{F}_{\ell+1}) \ne \mathbb{E}(Y_v \mid \mathscr{F}_\ell)$ is that $v$ lies within distance $8T$ of vertex $\ell + 1$. The number of such $v$ is at most $(dk)^{8T}$, so we conclude

$$|M_{\ell+1} - M_\ell| \le (dk)^{8T} \|A\|_\infty \le (dk)^{8T}.$$

It follows by the Azuma–Hoeffding martingale inequality that

$$\mathbb{P}(|S - \mathbb{E}S| \geq x) \leq \exp\left\{ - \frac{x^2}{2n'(dk)^{16T}} \right\}.$$

The result follows by summing over the choices of $A$ listed above, combined with our above estimates on $\mathbb{E}S$ for each choice of $A$. □

***Proof of Theorem 4.2*** It follows from Lemmas 4.8 and 4.9 that the sampling mechanism described by (60) satisfies the conditions of Definition 4.6. The result then follows by taking $n \to \infty$ in Proposition 4.7. □

## 5 Solution of tree optimization

From Theorem 4.2 we see that if $H \in \mathbf{\Delta}$ is a local maximizer for the first moment exponent $\boldsymbol{F} = \boldsymbol{F}_{\lambda,T}$, then its symmetrization $H^{\mathrm{sy}}$ must be a zero of the function $\mathbf{\Xi} = \mathbf{\Xi}_{\lambda,T}$ defined by (50). The analogous statement holds in the second moment with $\boldsymbol{F}_2$ and $\mathbf{\Xi}_2$. The functions $\mathbf{\Xi}$, $\mathbf{\Xi}_2$ correspond to **tree** optimization problems, which we solve in this section by relating them to the BP recursions for the coloring model.

**Proposition 5.1** *For $0 \leq \lambda \leq 1$ and $1 \leq T < \infty$, let $H_\star \in \mathbf{\Delta}$ and $H_\bullet \in \mathbf{\Delta}_2$ be as in* Definition 5.6 *below.*

a. *On $\{H \in \mathbf{N}_\circ : H = H^{sy}\}$, $\mathbf{\Xi}$ is uniquely minimized at $H = H_\star$, with $\mathbf{\Xi}(H_\star) = 0$.*
b. *On $\{H \in \mathbf{N}_{se} : H = H^{sy}\}$, $\mathbf{\Xi}_2$ is uniquely minimized at $H = H_\bullet$, with $\mathbf{\Xi}_2(H_\bullet) = 0$.*

*Moreover there is a positive constant $\epsilon = \epsilon(d, k, T)$ such that*

1. *$\mathbf{\Xi}(H) \geq \epsilon \|H - H_\star\|^2$ for all $H \in \mathbf{\Delta}$ with $H = H^{sy}$ and $\|H - H_\star\| \leq \epsilon$, and*
2. *$\mathbf{\Xi}_2(H) \geq \epsilon \|H - H_\bullet\|^2$ for all $H \in \mathbf{\Delta}_2$ with $H = H^{sy}$ and $\|H - H_\bullet\| \leq \epsilon$.*

### 5.1 Tree optimization problem

Recall from the previous section that in the local update procedure, we sample a subset of variables $Y$ and consider its neighborhood $\mathscr{N} = (\mathcal{N}, \underline{\mathrm{L}}_{\mathcal{N}})$. Writing $s = |Y|$, the graph $\mathcal{N}$ is the disjoint union of $\mathcal{D}_1, \ldots, \mathcal{D}_s$ where each $\mathcal{D}_i$ is a copy of the tree $\mathcal{D}$ of Fig. 6. Let $\mathbf{\Pi}$ be the space of probability measures on colorings of $\mathcal{D}$. Any coloring $\underline{\sigma}_{\mathcal{N}}$ can be summarized by $\nu \in \mathbf{\Pi}$ where $\nu(\underline{\sigma}_{\mathcal{D}})$ is the fraction of copies $\mathcal{D}_i$ with $\underline{\sigma}_{\mathcal{D}_i} = \underline{\sigma}_{\mathcal{D}}$. The sample empirical measure $H^{\mathrm{sm}} = H^{\mathrm{sm}}(\mathcal{G}, Y, \underline{\sigma})$ can be obtained as a linear projection of $\nu$, and we hereafter denote this relation by $H^{\mathrm{sm}} = H^{\mathrm{tr}}(\nu)$. Recalling (52), we have $\mathbb{E}^{\mathrm{lit}}[(\boldsymbol{w}^{\mathrm{lit}}_{\mathscr{N}}(\underline{\sigma}_{\mathcal{N}}))^\lambda] = \boldsymbol{w}_{\mathcal{D}}(\underline{\sigma}_{\mathcal{D}_1})^\lambda \cdots \boldsymbol{w}_{\mathcal{D}}(\underline{\sigma}_{\mathcal{D}_s})^\lambda$ where

$$\boldsymbol{w}_{\mathcal{D}}(\underline{\sigma}_{\mathcal{D}}) = \dot{\Phi}(\underline{\sigma}_{\delta v}) \prod_{e \in \delta v} \left\{ \bar{\Phi}(\sigma_e) \hat{\Phi}(\underline{\sigma}_{\delta a(e)}) \right\}.$$

**Lemma 5.2** *The function* $\mathbf{\Lambda}$ *of* (49) *is concave on* $\mathbf{\Delta}^{sm}$*, and can be expressed as*

$$\mathbf{\Lambda}(H) = \sup\left\{\mathcal{H}(\nu) + \lambda\langle \ln \boldsymbol{w}_{\mathcal{D}}, \nu\rangle : \nu \in \mathbf{\Pi} \text{ with } H^{tr}(\nu) = H\right\}. \tag{61}$$

***Proof*** The function $\mathbf{\Lambda}(H)$ is the sum of $\mathbf{\Sigma}^{\mathrm{tr}}(H)$ and the linear function $\lambda \boldsymbol{s}^{\mathrm{tr}}(H)$, so it suffices to show that $\mathbf{\Sigma}^{\mathrm{tr}}$ is concave on $\mathbf{\Delta}^{\mathrm{sm}}$. For $H = (\dot{H}, \hat{H}, \bar{H}) \in \mathbf{\Delta}^{\mathrm{sm}}$, if $X \in \Omega^k$ is a random variable with law $\hat{H}$, then the first coordinate $X_1$ has marginal law $\bar{H}$ by (47). It follows that for any $H \in \mathbf{\Delta}^{\mathrm{sm}}$ we can express

$$\mathbf{\Sigma}^{\mathrm{tr}}(H) = \mathcal{H}(\dot{H}) + d\mathcal{H}(X) - d\mathcal{H}(X_1) + \boldsymbol{v}(H) = \mathcal{H}(\dot{H}) + d\mathcal{H}(X \mid X_1) + \boldsymbol{v}(H)\,.$$

The entropy function is concave and $\boldsymbol{v}$ is linear, so this proves that $\mathbf{\Sigma}^{\mathrm{tr}}$ (hence $\mathbf{\Lambda}$) is indeed concave on $\mathbf{\Delta}^{\mathrm{sm}}$. In fact this can be argued alternatively, as follows. Recalling (56), note that for $H \in \mathbf{\Pi}$ we have

$$s^{O(1)} \exp\{s\mathbf{\Lambda}(H)\} = \mathbb{E}\boldsymbol{Z}(H; \mathscr{N}) = \sum_{\nu \in \mathbf{\Pi}} \mathbf{1}\{H^{\mathrm{tr}}(\nu) = H\} \binom{s}{s\nu} (\boldsymbol{w}_{\mathcal{D}})^{\lambda\nu}\,.$$

Expanding with Stirling's formula gives the representation (61), which also implies concavity of $\mathbf{\Lambda}$. □

Thus, for $H \in \mathbf{\Delta}^{\mathrm{sm}}$, we have $\mathbf{\Xi}(H) = \mathbf{\Lambda}^{\mathrm{op}}(\dot{h}^{\mathrm{tr}}(H)) - \mathbf{\Lambda}(H)$ where $\mathbf{\Lambda}$ is given by (61), and

$$\mathbf{\Lambda}^{\mathrm{op}}(\dot{h}) = \sup\left\{\mathcal{H}(\nu) + \lambda\langle \ln \boldsymbol{w}_{\mathcal{D}}, \nu\rangle : \nu \in \mathbf{\Pi} \text{ with } \dot{h}^{\mathrm{tr}}(H^{\mathrm{tr}}(\nu)) = \dot{h}\right\}. \tag{62}$$

Both (61) and (62) fall in the general category of entropy maximization problems subject to linear constraints. In "Appendix C" we review basic calculations for problems of this type. The discussion there, in particular Remark C.7, implies that for any $\dot{h}$, there is a unique measure $\nu = \nu^{\mathrm{op}}(\dot{h})$ achieving the maximum in (62). Moreover, there exists a probability measure $\dot{q}$ on $\dot{\Omega}_T$—serving the role of Lagrange multipliers for the constrained maximization—such that $\nu^{\mathrm{op}}(\dot{h})$ can be expressed as

$$\nu(\underline{\sigma}_{\mathcal{D}}) = \nu_{\dot{q}}(\underline{\sigma}_{\mathcal{D}}) \equiv \frac{\boldsymbol{w}_{\mathcal{D}}(\underline{\sigma})^{\lambda}}{Z} \prod_{e \in \delta\mathcal{D}} \dot{q}(\dot{\sigma}_e)\,, \tag{63}$$

where $Z$ is the normalizing constant. The analogous statement holds for the second moment.

### 5.2 BP recursions

We now state the BP recursions for the $\lambda$-tilted $T$-coloring model. In the standard formulation (e.g. [33, Ch. 14]), this is a pair of relations for probability measures $\dot{\boldsymbol{q}}, \hat{\boldsymbol{q}}$ on $\Omega_T$:

$$\dot{\boldsymbol{q}}(\sigma) = [\dot{\boldsymbol{B}}_{\lambda,T}(\hat{\boldsymbol{q}})](\sigma) \cong \mathbf{1}\{\sigma \in \Omega_T\}\bar{\Phi}(\sigma)^{\lambda} \sum_{\underline{\sigma}\in(\Omega_T)^d} \mathbf{1}\{\sigma_1 = \sigma\}\dot{\Phi}(\underline{\sigma})^{\lambda} \prod_{i=2}^{d} \hat{\boldsymbol{q}}(\sigma_i)$$

$$\hat{\boldsymbol{q}}(\sigma) = [\hat{\boldsymbol{B}}_{\lambda,T}(\dot{\boldsymbol{q}})](\sigma) \cong \mathbf{1}\{\sigma \in \Omega_T\}\bar{\Phi}(\sigma)^{\lambda} \sum_{\underline{\sigma}\in(\Omega_T)^k} \mathbf{1}\{\sigma_1 = \sigma\}\hat{\Phi}(\underline{\sigma})^{\lambda} \prod_{i=2}^{k} \dot{\boldsymbol{q}}(\sigma_i)$$

where $\cong$ denotes equality up to normalization, so that the mapping always outputs a probability measure. Recall from Definition 4.1 that for $\dot{\sigma} \equiv (\dot{\sigma}, \hat{\sigma}) \in \Omega_T$ we have $\dot{\sigma} \in \dot{\Omega}_T$ and $\hat{\sigma} \in \hat{\Omega}_T$. For our purposes we can assume a **one-sided** dependence, meaning there are probability measures $\dot{q}$ on $\dot{\Omega}_T$ and $\hat{q}$ on $\hat{\Omega}_T$ such that $\dot{\boldsymbol{q}}(\sigma) \cong \dot{q}(\dot{\sigma})\mathbf{1}\{\sigma \in \Omega_T\}$ and $\hat{\boldsymbol{q}}(\sigma) \cong \hat{q}(\hat{\sigma})\mathbf{1}\{\sigma \in \Omega_T\}$. One can then check (e.g. [33, Ch. 19]) that the BP recursions preserve the one-sided property, so that $\dot{\boldsymbol{B}}_{\lambda,T}$ and $\hat{\boldsymbol{B}}_{\lambda,T}$ restrict to mappings

$$\dot{\mathtt{BP}} \equiv \dot{\mathtt{BP}}_{\lambda,T} : \mathscr{P}(\hat{\Omega}_T) \to \mathscr{P}(\dot{\Omega}_T), \quad \hat{\mathtt{BP}} \equiv \hat{\mathtt{BP}}_{\lambda,T} : \mathscr{P}(\dot{\Omega}_T) \to \mathscr{P}(\hat{\Omega}_T). \tag{64}$$

We also denote $\mathtt{BP} \equiv \mathtt{BP}_{\lambda,T} \equiv \dot{\mathtt{BP}} \circ \hat{\mathtt{BP}}$. Given any $\dot{q} \in \mathscr{P}(\dot{\Omega}_T)$, write $\hat{q} \equiv \mathtt{BP}\dot{q}$, and let $H \equiv H_{\dot{q}}$ be defined by

$$\dot{H}_{\dot{q}}(\underline{\sigma}) = \frac{\dot{\Phi}(\underline{\sigma})^{\lambda}}{\dot{\mathfrak{z}}} \prod_{i=1}^{d} \hat{q}(\hat{\sigma}_i), \quad \hat{H}_{\dot{q}}(\underline{\sigma}) = \frac{\hat{\Phi}(\underline{\sigma})^{\lambda}}{\hat{\mathfrak{z}}} \prod_{i=1}^{d} \dot{q}(\dot{\sigma}_i), \quad \bar{H}_{\dot{q}}(\sigma) = \frac{\bar{\Phi}(\sigma)^{-\lambda}}{\bar{\mathfrak{z}}} \dot{q}(\dot{\sigma})\hat{q}(\hat{\sigma}) \tag{65}$$

where $\dot{\mathfrak{z}}$, $\hat{\mathfrak{z}}$, and $\bar{\mathfrak{z}}$ are normalizing constants, all dependent on $\dot{q}$. Clearly, $H_{\dot{q}} = (H_{\dot{q}})^{\mathrm{sy}}$. If $\dot{q}$ is a fixed point of BP, then $H_{\dot{q}} \in \boldsymbol{\Delta}$. An entirely similar discussion applies to the pair (second moment) model, where the BP recursion reduces to a pair of mappings between $\mathscr{P}((\dot{\Omega}_T)^2)$ and $\mathscr{P}((\hat{\Omega}_T)^2)$. If $\dot{q} \in \mathscr{P}((\dot{\Omega}_T)^2)$ is a fixed point of BP, then (65) defines an element $H_{\dot{q}} \in \boldsymbol{\Delta}_2$.

**Lemma 5.3** *For $1 \le T < \infty$, if $\dot{q} \in \mathscr{P}(\dot{\Omega}_T)$ is any fixed point of $\mathtt{BP}_{\lambda,T}$ which has full support on $\dot{\Omega}_T$, then $\Xi(H_{\dot{q}}) = 0$. The analogous statement holds for the second moment.*

***Proof*** Consider the optimization problem (62) for $\boldsymbol{\Lambda}^{\mathrm{op}}(\dot{h})$ with $\dot{h} = \dot{h}^{\mathrm{tr}}(H_{\dot{q}})$. As noted above, $\nu^{\mathrm{op}}(\dot{h})$ can be written (63) as $\nu_{\tilde{q}}$ for some measure $\tilde{q} \in \mathscr{P}(\dot{\Omega}_T)$, which may not be unique if the constraint $\dot{h}^{\mathrm{tr}}(H^{\mathrm{tr}}(\nu)) = \dot{h}$ is rank-deficient. However, if $\dot{q}$ has full support on $\dot{\Omega}_T$, then $\dot{h} = \dot{h}^{\mathrm{tr}}(H_{\dot{q}})$ does also. In this case it is straightforward to check that the constraints are indeed of full rank, so $\tilde{q}$ is unique. Because $\dot{q}$ is a fixed point of BP, the measure $\nu_{\dot{q}}$ satisfies $H^{\mathrm{tr}}(\nu_{\dot{q}}) = H_{\dot{q}}$, so it also satisfies the weaker constraint $\dot{h}^{\mathrm{tr}}(H^{\mathrm{tr}}(\nu_{\dot{q}})) = \dot{h}$. It follows by the above uniqueness argument that $\dot{q} = \tilde{q}$. Therefore, $\nu_{\dot{q}}$ solves the optimization problem (62) for $\boldsymbol{\Lambda}^{\mathrm{op}}(\dot{h}^{\mathrm{tr}}(H_{\dot{q}}))$, as well as the optimization problem (61) for $\boldsymbol{\Lambda}(H_{\dot{q}})$, so we conclude $\Xi(H_{\dot{q}}) = 0$ as claimed. □

**Lemma 5.4** *For $0 \le \lambda \le 1$ and $1 \le T < \infty$, let $\Xi = \Xi_{\lambda,T}$, $\Xi_2 = \Xi_{2,\lambda,T}$, and $BP = BP_{\lambda,T}$.*

a. *If $H \in \mathbf{N}_\circ$ with $H = H^{sy}$ and $\Xi(H) = 0$, then $H = H_{\dot{q}}$ where $\dot{q} \in \mathscr{P}(\dot{\Omega}_T)$ is a fixed point of* BP.
b. *If $H \in \mathbf{N}_{se}$ with $H = H^{sy}$ and $\Xi_2(H) = 0$, then $H = H_{\dot{q}}$ where $\dot{q} \in \mathscr{P}((\dot{\Omega}_T)^2)$ is a fixed point of* BP.

***Proof*** Let $\mu = \nu^{\mathrm{op}}(H)$ denote the solution of the optimization problem (61) for $\mathbf{\Lambda}(H)$, and let $\nu = \nu^{\mathrm{op}}(\dot{h}^{\mathrm{tr}}(H))$ denote the solution of the optimization problem (62) for $\mathbf{\Lambda}^{\mathrm{op}}(\dot{h}^{\mathrm{tr}}(H))$. Since (62) has a unique optimizer, we have $\Xi(H) = 0$ if and only if $\mu = \nu$. This means $H^{\mathrm{tr}}(\nu) = H$, but also $\nu = \nu_{\dot{q}}$ from (63), which gives

$$\hat{H}(\underline{\sigma}) \cong \hat{\Phi}(\underline{\sigma})^{\lambda}((\mathtt{BP}\dot{q})(\dot{\sigma}_1)) \prod_{i=2}^{k} \dot{q}(\dot{\sigma}_i)\,. \tag{66}$$

We now claim that in order for $\hat{H} = \hat{H}^{\mathrm{sy}}$, we must have $\mathtt{BP}\dot{q} = \dot{q}$. Note that if $\hat{\Phi}$ were fully supported on $(\Omega_T)^k$, and both $\dot{q}$ and $\mathtt{BP}\dot{q}$ were fully supported on $\dot{\Omega}_T$, the claim would be obvious. Since $\hat{\Phi}$ is certainly not fully supported, and we also do not know *a priori* whether $\dot{q}$ and $\mathtt{BP}\dot{q}$ are fully supported, the claim requires some argument, which differs slightly between the first- and second-moment cases:

a. In the first moment, Lemma 3.3 implies that $\dot{q}(\dot{\sigma})$ is positive for at least one $\dot{\sigma} \in \{\mathtt{b}_0, \mathtt{b}_1\}$. Assume without loss that $\dot{q}(\mathtt{b}_0)$ is positive; it follows that $(\mathtt{BP}\dot{q})(\dot{\sigma})$ is positive for both $\dot{\sigma} = \mathtt{b}_0, \mathtt{b}_1$. For any $\dot{\sigma} \in \dot{\Omega}$, there exists $\hat{\sigma}$ such that

$$\hat{\Phi}((\dot{\sigma}, \hat{\sigma}), \mathtt{b}_0, \ldots, \mathtt{b}_0) > 0\,. \tag{67}$$

The symmetry of $\hat{H}$ then gives the relation

$$\frac{(\mathtt{BP}\dot{q})(\dot{\sigma})}{(\mathtt{BP}\dot{q})(\mathtt{b}_0)} = \frac{\dot{q}(\dot{\sigma})}{\dot{q}(\mathtt{b}_0)},$$

so it follows that $\mathtt{BP}\dot{q} = \dot{q}$ in the first moment.
b. In the second moment, since we restrict to $H \in \mathbf{N}_{\mathrm{se}}$, $\dot{q}(\dot{\sigma})$ is positive for at least one $\dot{\sigma} \in \{\mathtt{b}_0, \mathtt{b}_1\}^2$. Assume without loss that $\dot{q}(\mathtt{b}_0\mathtt{b}_0)$ is positive. For any $\dot{\sigma} \notin \{\mathtt{r}_0\mathtt{r}_1, \mathtt{r}_1\mathtt{r}_0\}$, there exists $\hat{\sigma}$ such that the second-moment analogue of (67) holds. The preceding argument gives

$$\frac{(\mathtt{BP}\dot{q})(\dot{\sigma})}{(\mathtt{BP}\dot{q})(\mathtt{b}_0\mathtt{b}_0)} = \frac{\dot{q}(\dot{\sigma})}{\dot{q}(\mathtt{b}_0\mathtt{b}_0)} \quad \text{for all } \dot{\sigma} \notin \{\mathtt{r}_0\mathtt{r}_1, \mathtt{r}_1\mathtt{r}_0\}\,.$$

Since $(\mathtt{BP}\dot{q})(\dot{\sigma})$ is positive for all $\dot{\sigma} \in \{\mathtt{b}_0, \mathtt{b}_1\}^2$, it follows that the same holds for $\dot{q}$, so

$$\frac{(\mathtt{BP}\dot{q})(\dot{\sigma})}{(\mathtt{BP}\dot{q})(\mathtt{b}_0\mathtt{b}_1)} = \frac{\dot{q}(\dot{\sigma})}{\dot{q}(\mathtt{b}_0\mathtt{b}_1)} \quad \text{for all } \dot{\sigma} \notin \{\mathtt{r}_0\mathtt{r}_0, \mathtt{r}_1\mathtt{r}_1\}\,.$$

Combining these, we have for $\dot\sigma \in \{\mathtt{r}_0\mathtt{r}_1, \mathtt{r}_1\mathtt{r}_0\}$ that

$$\frac{(\mathtt{BP}\dot q)(\dot\sigma)}{(\mathtt{BP}\dot q)(\mathtt{b}_0\mathtt{b}_0)} = \frac{(\mathtt{BP}\dot q)(\dot\sigma)}{(\mathtt{BP}\dot q)(\mathtt{b}_0\mathtt{b}_1)}\frac{(\mathtt{BP}\dot q)(\mathtt{b}_0\mathtt{b}_1)}{(\mathtt{BP}\dot q)(\mathtt{b}_0\mathtt{b}_0)} = \frac{\dot q(\dot\sigma)}{\dot q(\mathtt{b}_0\mathtt{b}_0)},$$

and this proves $\mathtt{BP}\dot q = \dot q$ in the second moment.

Altogether, the above proves in both the first- and second-moment settings that $\dot q$ is a BP fixed point. □

### 5.3 BP contraction and conclusion

The next step is to (explicitly) define a subset $\mathbf{\Gamma}$ of measures $\dot q$ on which we have a contraction estimate of the form $\|\mathtt{BP}\dot q - \dot q_\star\| \le c\|\dot q - \dot q_\star\|$ for a constant $c < 1$. A useful feature of NAE- SAT is that its BP recursions are self-averaging: if $\dot q$ is a measure on $\dot\Omega_T$, let

$$\dot q^{\mathrm{av}}(\dot\sigma) \equiv \frac{\dot q(\dot\sigma) + \dot q(\dot\sigma \oplus 1)}{2}.$$

Then $\hat{\mathtt{BP}}\dot q = \hat{\mathtt{BP}}\dot q^{\mathrm{av}}$, and consequently $\mathtt{BP}\dot q = \mathtt{BP}\dot q^{\mathrm{av}}$. The analogous statement holds in the second moment. It then suffices to prove contraction on the measures $\dot q = \dot q^{\mathrm{av}}$, since for general $\dot q$ it implies

$$\|\mathtt{BP}\dot q - \dot q_\star\| = \|\mathtt{BP}\dot q^{\mathrm{av}} - \dot q_\star\| \le c\|\dot q^{\mathrm{av}} - \dot q_\star\| \le c\|\dot q - \dot q_\star\|.$$

Abbreviate $\{\mathtt{r}\} \equiv \{\mathtt{r}_0, \mathtt{r}_1\}$ and $\{\mathtt{b}\} \equiv \{\mathtt{b}_0, \mathtt{b}_1\}$. In a mild abuse of notation we now write $\{\mathtt{f}\}$ for $(\dot\Omega \cup \hat\Omega)\backslash\{\mathtt{r}, \mathtt{b}\}$; so for instance $\dot q(\mathtt{f}) = \dot q(\dot\Omega\backslash\{\mathtt{r}, \mathtt{b}\}) = \dot q(\dot{\mathscr{M}}\backslash\{0, 1, \star\})$. For the first moment analysis, let $\mathbf{\Gamma}$ be the set of measures $\dot q \in \dot{\mathscr{P}}(\dot\Omega_T)$ satisfying $\dot q = \dot q^{\mathrm{av}}$, such that

$$\frac{\dot q(\mathtt{r}) + 2^k\dot q(\mathtt{f})}{C} \le \dot q(\mathtt{b}) \le \frac{\dot q(\mathtt{r})}{1 - C/2^k} \tag{68}$$

for $C$ a large constant (to be determined). For the second moment analysis, let $\mathbf{\Gamma}(c, \kappa)$ be the set of measures $\dot q \in \dot{\mathscr{P}}((\dot\Omega_T)^2)$ satisfying $\dot q = \dot q^{\mathrm{av}}$, such that

$$|\dot q(\mathtt{b}_0\mathtt{b}_0) - \dot q(\mathtt{b}_0\mathtt{b}_1)| \le (k^9/2^{ck})\dot q(\mathtt{bb}), \text{ and } \dot q(\mathtt{ff}) + \dot q(\{\mathtt{fr}, \mathtt{rf}\})/2^k + \dot q(\mathtt{rr})/4^k \le (C/2^k)\dot q(\mathtt{bb}); \tag{1$\mathbf{\Gamma}$}$$

$$\dot q(\{\mathtt{rf}, \mathtt{fr}\}) \le (C/2^{k\kappa})\dot q(\mathtt{bb}) \text{ and } \dot q(\mathtt{rr}) \le C2^{k(1-\kappa)}\dot q(\mathtt{bb}); \tag{2$\mathbf{\Gamma}$}$$

$$\dot q(\mathtt{r}_{\boldsymbol{x}}\dot\sigma) \ge [1 - C/2^k]\dot q(\mathtt{b}_{\boldsymbol{x}}\dot\sigma) \text{ and } \dot q(\dot\sigma\,\mathtt{r}_{\boldsymbol{x}}) \ge [1 - C/2^k]\dot q(\dot\sigma\,\mathtt{b}_{\boldsymbol{x}}) \text{ for all } \boldsymbol{x} \in \{0, 1\}, \dot\sigma \in \dot\Omega. \tag{3$\mathbf{\Gamma}$}$$

(To clarify the notation: since $\mathtt{b} \equiv \{\mathtt{b}_0, \mathtt{b}_1\}$, in (1$\mathbf{\Gamma}$) the expression $\dot q(\mathtt{bb})$ refers to $\dot q(\{\mathtt{b}_0, \mathtt{b}_1\}^2)$. Similarly, $\dot q(\mathtt{fr})$ refers to $\dot q(\{\mathtt{f}\} \times \{\mathtt{r}_0, \mathtt{r}_1\})$ where in this context $\{\mathtt{f}\} = (\dot\Omega \cup \hat\Omega)\backslash\{\mathtt{r}, \mathtt{b}\}$.)

**Proposition 5.5** (proved in "Appendix *A*") *Assume* $0 \le \lambda \le 1$. *In the first moment, we have:*

a. *For any* $1 \le T \le \infty$, *the map* $BP \equiv BP_{\lambda,T}$ *has a unique fixed point* $\dot{q}_\star \equiv \dot{q}_{\lambda,T} \in \mathbf{\Gamma}$. *For any* $\dot{q} \in \mathbf{\Gamma}$, *we have* $BP\dot{q} \in \mathbf{\Gamma}$ *also, with* $\|BP\dot{q} - \dot{q}_\star\| = O(k^2/2^k)\|\dot{q} - \dot{q}_\star\|$.
b. *In the limit* $T \to \infty$, $\|\dot{q}_{\lambda,T} - \dot{q}_{\lambda,\infty}\| \to 0$.

*In the second moment, for any* $1 \le T \le \infty$, *we have the following:*

A. *The map* $BP \equiv BP_{\lambda,T}$ *has a unique fixed point in* $\mathbf{\Gamma}(1,1)$, *given by* $\dot{q}_\star \otimes \dot{q}_\star$ *with* $\dot{q}_\star$ *as in part a. Moreover, for* $c \in (0,1]$ *and* $k$ *sufficiently large, there is no other fixed point of* $BP$ *in* $\mathbf{\Gamma}(c,1)$: *if* $\dot{q} \in \mathbf{\Gamma}(c,1)$ *then* $BP\dot{q} \in \mathbf{\Gamma}(1,1)$, *with* $\|BP\dot{q} - \dot{q}_\star\| = O(k^4/2^k)\|\dot{q} - \dot{q}_\star\|$.
B. *If for some* $c \in (0,1]$ *we have* $\dot{q} \in \mathbf{\Gamma}(c,0)$ *and* $\dot{q} = BP\dot{q}$, *then* $\dot{q} \in \mathbf{\Gamma}(c,1)$.

**Definition 5.6** *(optimal empirical measures)* For $0 \le \lambda \le 1$ and $1 \le T < \infty$, take the fixed point $\dot{q}_\star = \dot{q}_{\lambda,T}$ as given by Proposition 5.5a, and use (65) to define $H_\star = H_{\dot{q}_\star} \in \mathbf{\Delta}$ and $H_\bullet = H_{\dot{q}_\star \otimes \dot{q}_\star} \in \mathbf{\Delta}_2$. Note that this agrees with our earlier definition of $H_\bullet$, in the discussion below Lemma 3.9.

**Lemma 5.7** *Let* $\dot{q}$ *be any fixed point of* $BP$ *that arises from* Lemma 5.4.

a. *If* $H = H_{\dot{q}} \in \mathbf{N}_\circ$, *then* $\dot{q} = \dot{q}_\star$ *and so* $H = H_\star$.
b. *If* $H = H_{\dot{q}} \in \mathbf{N}_{se}$, *then* $\dot{q} = \dot{q}_\star \otimes \dot{q}_\star$ *and so* $H = H_\bullet$.

***Proof*** Since $\dot{q} = \mathrm{BP}\dot{q}$, we must have $\dot{q} = \dot{q}^{\mathrm{av}}$. Below we argue separately for the first and second moment. In each case we repeatedly take advantage of the fact that $H = H_{\dot{q}}$ is symmetric.

a. For the first moment, by Proposition 5.5a it suffices to show that $\dot{q}$ must lie in the set $\mathbf{\Gamma}$ defined by (68). It follows directly from the relation $\dot{q} = \mathrm{BP}\dot{q}$ that $\dot{q}(\mathtt{r}) \ge \dot{q}(\mathtt{b})$. By definition of $\mathbf{N}_\circ$ we must have $\bar{H}(\mathtt{r}) \le 7/2^k$ and $\bar{H}(\mathtt{f}) \le 7/2^k$, so the vast majority of clauses must have all incident colors in $\{\mathtt{b}\} = \{\mathtt{b}_0, \mathtt{b}_1\}$:

$$1 - \frac{14k}{2^k} \le \hat{H}(\mathtt{b}^k) = \frac{1}{\hat{\mathfrak{z}}} \sum_{\underline{\sigma} \in \mathtt{b}^k} \hat{\Phi}(\underline{\sigma})^\lambda \prod_{i=1}^{k} \dot{q}(\dot{\sigma}_i) \le \frac{\dot{q}(\mathtt{b})^k}{\hat{\mathfrak{z}}}.$$

Next, if $\underline{\sigma} \in \mathtt{rb}^{k-1}$, we have $\hat{\Phi}(\underline{\sigma})^\lambda = \mathbb{E}^{\mathrm{lit}}[\hat{\Phi}^{\mathrm{lit}}(\underline{\sigma} \oplus \underline{\mathtt{L}})^\lambda] \ge \mathbb{E}^{\mathrm{lit}}\hat{\Phi}^{\mathrm{lit}}(\underline{\sigma} \oplus \underline{\mathtt{L}}) = 2/2^k$, so

$$\frac{7}{2^k} \ge \bar{H}(\mathtt{r}) = \hat{H}(\mathtt{rb}^{k-1}) \ge \frac{\dot{q}(\mathtt{r})\dot{q}(\mathtt{b})^{k-1}}{2^{k-1}\hat{\mathfrak{z}}} \ge \frac{\dot{q}(\mathtt{r})}{\dot{q}(\mathtt{b})} \frac{2}{2^k}\left(1 - \frac{14k}{2^k}\right),$$

which gives $\dot{q}(\mathtt{r})/\dot{q}(\mathtt{b}) \le 4$ for large $k$. Similarly, if $\underline{\sigma} \in \mathtt{fb}^{k-1}$ with $\hat{\sigma}_1 = \mathtt{s}$ (indicating a separating clause), then $\hat{\Phi}(\underline{\sigma})^\lambda \ge \mathbb{E}^{\mathrm{lit}}\hat{\Phi}^{\mathrm{lit}}(\underline{\sigma} \oplus \underline{\mathtt{L}}) = 1 - 4/2^k$, so

$$\frac{7}{2^k} \ge \bar{H}(\mathtt{f}) \ge \hat{H}(\mathtt{fb}^{k-1}) \ge \left(1 - \frac{4}{2^k}\right)\frac{\dot{q}(\mathtt{f})\dot{q}(\mathtt{b})^{k-1}}{\hat{\mathfrak{z}}} \ge \frac{\dot{q}(\mathtt{f})}{\dot{q}(\mathtt{b})}\left(1 - \frac{4}{2^k}\right)\left(1 - \frac{14k}{2^k}\right),$$

which gives $\dot{q}(\mathtt{f})/\dot{q}(\mathtt{b}) \le 8/2^k$ for large $k$. Combining these estimates proves $\dot{q} \in \mathbf{\Gamma}$.

b. For the second moment, by Proposition 5.5A it suffices to verify $\dot{q} \in \mathbf{\Gamma}(1,1)$, as defined by (1$\mathbf{\Gamma}$)–(3$\mathbf{\Gamma}$). Condition (3$\mathbf{\Gamma}$) is immediate from the relation $\dot{q} = \mathtt{BP}\dot{q}$. Moreover, by Proposition 5.5B it suffices to show $\dot{q} \in \mathbf{\Gamma}(1,0)$, in which case condition (2$\mathbf{\Gamma}$) follows from (1$\mathbf{\Gamma}$). It remains to verify (1$\mathbf{\Gamma}$). For this end, we denote $\mathbb{B} \equiv \{\mathtt{b}_0, \mathtt{b}_1\}^2$, and partition this into $\mathbb{B}_= \equiv \{\mathtt{b}_0\mathtt{b}_0, \mathtt{b}_1\mathtt{b}_1\}$ and $\mathbb{B}_\ddagger \equiv \{\mathtt{b}_0\mathtt{b}_1, \mathtt{b}_1\mathtt{b}_0\}$. By definition, for any $H \in \mathbf{N}_{\mathrm{se}}$ the single-copy marginals $H^1, H^2$ lie in $\mathbf{N} \subseteq \mathbf{N}_\circ$, so the total density of $\{\mathtt{r}, \mathtt{f}\}$ edges in either copy is very small. As a result the vast majority of clauses have all incident colors in $\mathbb{B}$:

$$1 - \frac{28k}{2^k} \le \hat{H}(\mathbb{B}^k) \le \frac{\dot{q}(\mathbb{B})^k}{\hat{\mathfrak{z}}}.$$

For $H \in \mathbf{N}_{\mathrm{se}}$, we have $|\bar{H}(\mathbb{B}_=) - \bar{H}(\mathbb{B}_\ddagger)| \le k^4/2^{k/2} + O(2^{-k})$. The fraction of edges in not-all-$\mathbb{B}$ clauses is $O(k/2^k)$, and for $\underline{\sigma} \in \mathbb{B}^k$ we have $1 \ge \hat{\Phi}(\underline{\sigma})^\lambda \ge \mathbb{E}^{\mathrm{lit}}\hat{\Phi}^{\mathrm{lit}}(\underline{\sigma} \oplus \underline{\mathtt{L}}) = 1 - O(k/2^k)$, so

$$\bar{H}(\mathbb{B}_=) - \bar{H}(\mathbb{B}_\ddagger) - O(k/2^k) = \left\{\frac{\dot{q}(\mathbb{B}_=) - \dot{q}(\mathbb{B}_\ddagger)}{\dot{q}(\mathbb{B})} + O(k/2^k)\right\}\frac{\dot{q}(\mathbb{B})^k}{\hat{\mathfrak{z}}}.$$

Rearranging gives $|\dot{q}(\mathbb{B}_=) - \dot{q}(\mathbb{B}_\ddagger)|/\dot{q}(\mathbb{B}) \le 2k^4/2^{k/2}$, which proves the first part of (1$\mathbf{\Gamma}$) (with $c = 1$). It remains to show the second part of (1$\mathbf{\Gamma}$). If we denote $\mathbb{R}_= \equiv \{\mathtt{r}_0\mathtt{r}_0, \mathtt{r}_1\mathtt{r}_1\}$ and consider $\underline{\sigma} \in \mathbb{R}_=(\mathbb{B}_=)^{k-1}$, then (similarly as above) we have $\hat{\Phi}(\underline{\sigma})^\lambda \ge \mathbb{E}^{\mathrm{lit}}\hat{\Phi}^{\mathrm{lit}}(\underline{\sigma} \oplus \underline{\mathtt{L}}) = 2/2^k$, so

$$\frac{7}{2^k} \ge \bar{H}(\mathbb{R}_=) = \hat{H}(\mathbb{R}_=(\mathbb{B}_=)^{k-1}) \ge \frac{2}{2^k}\frac{\dot{q}(\mathbb{R}_=)\dot{q}(\mathbb{B}_=)^{k-1}}{\hat{\mathfrak{z}}} \ge \frac{2}{4^k}\frac{\dot{q}(\mathbb{R}_=)}{\dot{q}(\mathbb{B})}\left(1 - \frac{O(k^5)}{2^{k/2}}\right),$$

where the last inequality is by the preceding estimates on $\dot{q}(\mathbb{B})$ and $\dot{q}(\mathbb{B}_=)$. The same calculation bounds $\dot{q}(\mathbb{R}_\ddagger)$ for $\mathbb{R}_\ddagger \equiv \{\mathtt{r}_0\mathtt{r}_1, \mathtt{r}_1\mathtt{r}_0\}$. Next consider $\sigma = ((\dot{\sigma}, \mathtt{s}), \mathtt{r}) \in \{\mathtt{fr}\}$: if $\underline{\sigma} \in \Omega^k$ with $\sigma_1 = \sigma$ and the other $k-1$ entries in $\mathbb{B}$, then $\hat{\Phi}(\underline{\sigma})^\lambda \ge 4/2^k$ as long as the other entries are not all $\mathbb{B}_=$ or all $\mathbb{B}_\ddagger$. Therefore

$$\frac{7}{2^k} \ge \bar{H}(\mathtt{fr}) \ge \frac{4}{2^k}\frac{\dot{q}(\mathtt{fr})\dot{q}(\mathbb{B})^{k-1}}{\hat{\mathfrak{z}}}\left(1 - \frac{\dot{q}(\mathbb{B}_=)^{k-1}}{\dot{q}(\mathbb{B})^{k-1}} - \frac{\dot{q}(\mathbb{B}_\ddagger)^{k-1}}{\dot{q}(\mathbb{B})^{k-1}}\right) \ge \frac{\dot{q}(\mathtt{fr})}{\dot{q}(\mathbb{B})}\left(1 - \frac{O(k)}{2^k}\right),$$

and the same calculation bounds $\dot{q}(\mathtt{rf})$. Finally, for $\sigma = ((\dot{\sigma}^1, \mathtt{s}), (\dot{\sigma}^2, \mathtt{s})) \in \{\mathtt{ff}\}$, we can consider $\underline{\sigma} \in \Omega^k$ with $\sigma_1 = \sigma$ and the other $k-1$ entries in $\mathbb{B}$; therefore

$$\frac{7}{2^k} \ge \bar{H}(\mathtt{ff}) \ge \frac{\dot{q}(\mathtt{ff})\dot{q}(\mathbb{B})^{k-1}}{\hat{\mathfrak{z}}}\left(1 - \frac{4}{2^k}\right) \ge \frac{\dot{q}(\mathtt{ff})}{\dot{q}(\mathbb{B})}\left(1 - \frac{O(k)}{2^k}\right).$$

Combining these estimate verifies the second part of (1$\mathbf{\Gamma}$).

Altogether we have shown that if $H = H_{\dot{q}}$ lies in $\mathbf{N}_\circ$, then $\dot{q} \in \mathbf{\Gamma}$ and so $\dot{q} = \dot{q}_\star$; and if $H = H_{\dot{q}}$ lies in $\mathbf{N}_{\mathrm{se}}$ then $\dot{q} \in \mathbf{\Gamma}(1,1)$ and so $\dot{q} = \dot{q}_\star \otimes \dot{q}_\star$. This concludes the proof. □

***Proof of Proposition 5.1*** We will prove the claim in the first moment; the result for the second moment follows by the same argument. It follows from Lemmas 5.3, 5.4 and 5.7 that the unique minimizer of $\mathbf{\Xi}$ on the set $\{H \in \mathbf{N}_\circ : H = H^{\mathrm{sy}}\}$ is $H_\star$ (as given by Definition 5.6), with $\mathbf{\Xi}(H_\star) = 0$. It remains to establish that, for $H \in \mathbf{\Delta}$ with $H = H^{\mathrm{sy}}$ and $\|H - H_\star\| \le \epsilon$, we have $\mathbf{\Xi}(H) \ge \epsilon \|H - H_\star\|^2$. As in the proof of Lemma 5.4, let $\mu = \nu^{\mathrm{op}}(H)$ be the solution of the optimization problem (61) for $\mathbf{\Lambda}(H)$, and let $\nu = \nu^{\mathrm{op}}(\dot{h}^{\mathrm{tr}}(H))$ be the solution of the optimization problem (62) for $\mathbf{\Lambda}^{\mathrm{op}}(\dot{h}^{\mathrm{tr}}(H))$. We have from (63) that for some $\dot{q} \in \mathscr{P}(\dot{\Omega}_T)$,

$$\nu(\underline{\sigma}_{\mathcal{D}}) = \nu_{\dot{q}}(\underline{\sigma}_{\mathcal{D}}) = \frac{\boldsymbol{w}_{\mathcal{D}}(\underline{\sigma})^\lambda}{Z} \prod_{e \in \delta \mathcal{D}} \dot{q}(\dot{\sigma}_e) \,.$$

For $e \in \delta\mathcal{D}$, abbreviate $g_e(\underline{\sigma}_{\mathcal{D}}) \equiv \ln \dot{q}(\sigma_e)$. Then, for any probability measure $\varpi$ on colorings $\underline{\sigma}_{\mathcal{D}}$, the quantity $\Lambda(\omega) \equiv \mathcal{H}(\varpi) + \lambda \langle \ln \boldsymbol{w}_{\mathcal{D}}, \varpi \rangle$ can be expressed as

$$\begin{aligned} \Lambda(\omega) &= \mathcal{H}(\varpi) + \left\langle \ln \nu + \ln Z - \sum_{e \in \delta \mathcal{D}} g_e, \varpi \right\rangle \\ &= -\mathcal{D}_{\mathrm{KL}}(\omega | \nu) + \ln Z - |\delta \mathcal{D}| \langle \ln \dot{q}, \dot{h}^{\mathrm{tr}}(H^{\mathrm{tr}}(\varpi)) \rangle \,. \end{aligned}$$

We have $\dot{h}^{\mathrm{tr}}(H^{\mathrm{tr}}(\varpi)) = \dot{h}^{\mathrm{tr}}(H)$ for both $\varpi = \nu$ and $\varpi = \mu$, so $\mathbf{\Xi}(H) = \Lambda(\mu) - \Lambda(\nu) = \mathcal{D}_{\mathrm{KL}}(\mu|\nu)$. (For further discussion, see Proposition C.6.) It is well known that $\mathcal{D}_{\mathrm{KL}}(\mu|\nu) \gtrsim \|\mu - \nu\|^2$, so to conclude it remains for us to show that $\|\mu - \nu\| \gtrsim \|H - H_\star\|$. To this end, let $\nu_\star \equiv \nu_{\dot{q}_\star}$, and note that $H = H^{\mathrm{tr}}(\mu)$ while $H_\star = H^{\mathrm{tr}}(\nu_\star)$. Recall from the discussion preceding Lemma 5.2 that $\varpi \mapsto H^{\mathrm{tr}}(\varpi)$ is a linear projection, so

$$\|H - H_\star\| \lesssim \|\mu - \nu_\star\| \le \|\mu - \nu\| + \|\nu - \nu_\star\| \lesssim \|\mu - \nu\| + \|\dot{q} - \dot{q}_\star\| \,,$$

where the last bound holds since $\nu = \nu_{\dot{q}}$ and $\nu_\star = \nu_{\dot{q}_\star}$. Recall from Proposition 5.5a (or Proposition 5.5A for the second moment) that we have the contraction estimate $\|\mathtt{BP}\dot{q} - \dot{q}_\star\| \le c\|\dot{q} - \dot{q}_\star\|$ for $c \in (0,1)$, so

$$(1-c)\|\dot{q} - \dot{q}_\star\| \le \|\dot{q} - \dot{q}_\star\| - \|\mathtt{BP}\dot{q} - \dot{q}_\star\| \le \|\dot{q} - \mathtt{BP}\dot{q}\| \,.$$

Let $K \equiv (\dot{K}, \hat{K}, \bar{K}) \equiv H^{\mathrm{tr}}(\nu)$, and note that $\hat{K}$ need not be symmetric: if we let $\hat{K}'(\underline{\sigma}) \equiv \hat{K}(\sigma_2, \ldots, \sigma_k, \sigma_1)$ for $\underline{\sigma} \in (\Omega_T)^k$, then $\hat{K}$ and $\hat{K}'$ need not agree. On the other hand $H = H^{\mathrm{tr}}(\mu) = H^{\mathrm{sy}}$, so

$$\|\hat{K} - \hat{K}'\| \le \|\hat{H} - \hat{K}\| + \|\hat{H} - \hat{K}'\| = 2\|\hat{H} - \hat{K}\| \le 2\|H - K\| \lesssim \|\mu - \nu\| \,.$$

For any $k$-tuple $\underline{\dot{h}} \equiv (\dot{h}_1, \ldots, \dot{h}_k)$ of probability measures on $\dot{\Omega}_T$, consider

$$\hat{H}^{\mathrm{op}}(\underline{\dot{h}}) \equiv \arg\max_{\hat{\nu}} \left\{ \mathcal{H}(\hat{\nu}) + \lambda \langle \ln \hat{\Phi}, \hat{\nu} \rangle : \hat{\nu}(\dot{\sigma}_1 = \cdot) = \dot{h}_i \text{ for all } i \right\}$$

where $\hat{\nu}$ denotes any probability measure on $\operatorname{supp} \hat{\upsilon} \subseteq (\Omega_T)^k$. The unique optimizer $\hat{H}^{\mathrm{op}}(\underline{\dot{h}})$ can be described in terms of another $k$-tuple of probability measures on $\dot{\Omega}_T$, denoted $\underline{\dot{q}} \equiv (\dot{q}_1, \ldots, \dot{q}_k)$, which serve as Lagrange multipliers: $\hat{H}^{\mathrm{op}}(\underline{\dot{h}}) = \hat{H}(\underline{\dot{q}})$ where

$$[\hat{H}(\underline{\dot{q}})](\underline{\sigma}) \cong \hat{\Phi}(\underline{\sigma})^{\lambda} \prod_{i=1}^{k} \dot{q}_i(\dot{\sigma}_i) \,.$$

In particular, $\hat{H}^{\mathrm{op}}(\underline{\dot{h}}_\star) = \hat{H}(\underline{\dot{q}}_\star)$ for $\underline{\dot{h}}_\star \equiv (\dot{h}^{\mathrm{tr}}(H_\star), \ldots, \dot{h}^{\mathrm{tr}}(H_\star))$ and $\underline{\dot{q}}_\star \equiv (\dot{q}_\star, \ldots, \dot{q}_\star)$. For $\underline{\dot{h}}$ near $\underline{\dot{h}}_\star$, there is a unique $\underline{\dot{q}}$ satisfying $\hat{H}^{\mathrm{op}}(\underline{\dot{h}}) = \hat{H}(\underline{\dot{q}})$, and we can determine this $\underline{\dot{q}}$ as a smooth function of $\underline{\dot{h}}$. Thus

$$\|\hat{K} - \hat{K}'\| = \|\hat{H}(\mathtt{BP}\dot{q}, \dot{q}, \ldots, \dot{q}) - \hat{H}(\dot{q}, \mathtt{BP}\dot{q}, \dot{q}, \ldots, \dot{q})\| \gtrsim \|\dot{q} - \mathtt{BP}\dot{q}\| \,.$$

Combining the above inequalities gives $\|H - H_\star\| \lesssim \|\mu - \nu\|$ as desired. □

***Proof of Propositions 3.4 and 3.10*** Note that for $\bar{H}$ fixed, $\boldsymbol{F}(H) = \boldsymbol{F}(\dot{H}, \hat{H}, \bar{H})$ is a strictly concave function of $\dot{H}, \hat{H}$. It follows that for all $H \in \boldsymbol{\Delta}$ we have $\boldsymbol{F}(H) \leq \boldsymbol{F}(L_H) - \epsilon \|H - L_H\|^2$ for

$$L_H \equiv \arg\max_L \left\{ \boldsymbol{F}(L) : L \in \boldsymbol{\Delta} \text{ with } \bar{L} = \bar{H} \right\}.$$

Clearly $L_H = (L_H)^{\mathrm{sy}}$, so it follows from Theorem 4.2 and Proposition 5.1 that

$$\boldsymbol{F}(L_H) \leq \boldsymbol{F}(H_\star) - \epsilon \|L_H - H_\star\|^2.$$

Combining the inequalities (and adjusting $\epsilon$ as needed) gives $\boldsymbol{F}(H) \leq \boldsymbol{F}(H_\star) - \epsilon \|H - H_\star\|^2$. This concludes the proof of Proposition 3.4, and Proposition 3.10 follows by exactly the same argument. □

**Acknowledgements** We are grateful to Amir Dembo, Jian Ding, Andrea Montanari, and Lenka Zdeborová for helpful conversations. We thank the anonymous referee and Youngtak Sohn for pointing out errors and giving many helpful comments on drafts of the paper. We also gratefully acknowledge the hospitality of the Simons Institute at Berkeley, where part of this work was completed during a spring 2016 semester program.

**Data Availability** Data sharing is not applicable to this article as no datasets were generated or analyzed during this study.

## Appendix A: Contraction estimates

We now prove Proposition 5.5, on the contraction of the BP recursion for the coloring model. In Section A.1 we analyze the recursions for the first moment (single-copy) model and prove Proposition 5.5a. In Section A.2 we analyze the the recursions for the first moment (pair) model and prove the remainder of Proposition 5.5. We assume throughout the section that $0 \le \lambda \le 1$ and $1 \le T \le \infty$.

### A.1. Single-copy coloring recursions

Recall from Sect. 5.2 that the BP recursion is a pair (64) of mappings $\dot{\mathrm{BP}} : \mathscr{P}(\hat{\Omega}_T) \to \mathscr{P}(\dot{\Omega}_T)$ and $\hat{\mathrm{BP}} : \mathscr{P}(\dot{\Omega}_T) \to \mathscr{P}(\hat{\Omega}_T)$. Recall that for our purposes we can restrict attention to measures satisfying $\dot{q} = \dot{q}^{\mathrm{av}}$ and $\hat{q} = \hat{q}^{\mathrm{av}}$. Under this restriction, the BP recursion is quite explicit, as we now describe. Recall from Definition 2.8, equations (24) and (25), that for $\dot{\tau} \in \dot{\mathscr{M}}$ and $\hat{\tau} \in \hat{\mathscr{M}}$ we defined $\dot{\mathrm{m}}(\dot{\tau})$ and $\hat{\mathrm{m}}(\hat{\tau})$ as probability measures on $\{\mathtt{0}, \mathtt{1}\}$. For convenience, we also define

$$\dot{\mathrm{m}}(\mathtt{r}_1) = \hat{\mathrm{m}}(\mathtt{b}_1) = \delta_1, \quad \dot{\mathrm{m}}(\mathtt{r}_0) = \hat{\mathrm{m}}(\mathtt{b}_0) = \delta_0 \,. \tag{69}$$

In what follows we often represent a probability measure on $\{\mathtt{0}, \mathtt{1}\}$ by the probability assigned to $\mathtt{1}$, writing $\dot{m}(\dot{\tau}) \equiv \dot{\mathrm{m}}[\dot{\tau}](\mathtt{1})$ and $\hat{m}(\hat{\tau}) \equiv \hat{\mathrm{m}}[\hat{\tau}](\mathtt{1})$. Thus, equations (24), (25), and (69) together define mappings $\dot{m} : \dot{\Omega} \to [0, 1]$ and $\hat{m} : \hat{\Omega} \to [0, 1]$. Recall that we denote $\{\mathtt{r}\} \equiv \{\mathtt{r}_1, \mathtt{r}_0\}$, $\{\mathtt{b}\} \equiv \{\mathtt{b}_1, \mathtt{b}_0\}$, and $\{\mathtt{f}\} \equiv \Omega \backslash \{\mathtt{r}, \mathtt{b}\}$. We also write $\{\mathtt{f}\} \equiv (\dot{\Omega} \cup \hat{\Omega}) \backslash \{\mathtt{r}, \mathtt{b}\}$; the precise meaning of $\{\mathtt{f}\}$ will be unambiguous from context. Then, for $\boldsymbol{x} \in \{\mathtt{0}, \mathtt{1}\}$, let us abbreviate

$$\mathtt{g} \equiv \mathtt{b} \cup \mathtt{f}, \ \mathtt{g}_{\boldsymbol{x}} \equiv \mathtt{b}_{\boldsymbol{x}} \cup \mathtt{f}, \ \mathtt{y} \equiv \mathtt{r} \cup \mathtt{f}, \ \mathtt{p}_{\boldsymbol{x}} \equiv \mathtt{b}_{\boldsymbol{x}} \cup \mathtt{r}_{\boldsymbol{x}} \,.$$

The variable recursion $\dot{\mathrm{BP}} \equiv \dot{\mathrm{BP}}_{\lambda,T}$ is given by

$$(\dot{\mathrm{BP}}\hat{q})(\dot{\sigma}) \cong \begin{cases} \hat{q}(\mathtt{p}_1)^{d-1} & \text{if } \dot{\sigma} \in \{\mathtt{r}_0, \mathtt{r}_1\}, \\ \hat{q}(\mathtt{p}_1)^{d-1} - (\hat{q}(\mathtt{b}_1))^{d-1} & \text{if } \dot{\sigma} \in \{\mathtt{b}_0, \mathtt{b}_1\}, \\ \dot{z}(\dot{\sigma})^{\lambda} \displaystyle\sum_{\hat{\sigma}_2, \ldots, \hat{\sigma}_d} \mathbf{1}\Big\{\dot{\sigma} = \dot{T}\Big((\hat{\sigma}_i)_{i \ge 2}\Big)\Big\} \prod_{i=2}^{d} \hat{q}(\hat{\sigma}_i) & \text{if } \dot{\sigma} \in \dot{\Omega}_T \backslash \{\mathtt{r}, \mathtt{b}\}, \end{cases}$$

where $\cong$ indicates the normalization which makes $\dot{\mathrm{BP}}\hat{q}$ a probability measure on $\dot{\Omega}_T$. For the clause recursion, let us write $\underline{\dot{\sigma}} \sim \hat{\sigma}$ if $\underline{\dot{\sigma}} \equiv (\dot{\sigma}_2, \ldots, \dot{\sigma}_k) \in (\dot{\Omega}_T)^{k-1}$ is compatible with $\hat{\sigma}$, in the sense that

$$\Big\{\underline{\sigma} = ((\dot{\sigma}, \hat{\sigma}), (\dot{\sigma}_2, \hat{\sigma}_2), \ldots (\dot{\sigma}_k, \hat{\sigma}_k)) \in (\Omega_T)^k : \hat{I}^{\mathrm{lit}}(\underline{\sigma}) = 1\Big\} \neq \varnothing. \tag{70}$$

The clause recursion $\hat{\mathtt{BP}} \equiv \hat{\mathtt{BP}}_{\lambda,T}$ is given by

$$(\hat{\mathtt{BP}}\dot{q})(\hat{\sigma}) \cong \begin{cases} \dot{q}(\mathtt{b}_0)^{k-1} & \text{if } \hat{\sigma} \in \{\mathtt{r}_0, \mathtt{r}_1\}, \\ \hat{z}(\hat{\sigma})^{\lambda} \sum_{\dot{\sigma}_2,\ldots,\dot{\sigma}_k} \mathbf{1}\Big\{\hat{\sigma} = \hat{T}\Big((\dot{\sigma}_i)_{i\geq 2}\Big)\Big\} \prod_{i=2}^{k} \dot{q}(\dot{\sigma}_i) & \text{if } \hat{\sigma} \in \hat{\Omega}_T \backslash \{\mathtt{r}, \mathtt{b}\}, \\ \sum_{\dot{\sigma} \sim \mathtt{b}_1} \Big(1 - \prod_{i=2}^{k} \dot{m}(\dot{\sigma}_i)\Big)^{\lambda} \prod_{i=2}^{k} \dot{q}(\dot{\sigma}_i) & \text{if } \hat{\sigma} \in \{\mathtt{b}_0, \mathtt{b}_1\}, \end{cases}$$

where the last line uses the convention (69). Recall that $\mathtt{BP} \equiv \dot{\mathtt{BP}} \circ \hat{\mathtt{BP}} \equiv \mathtt{BP}_{\lambda,T}$. We will show the following contraction result (assuming, as always, $0 \leq \lambda \leq 1$ and $1 \leq T \leq \infty$).

**Proposition A.1** *If $\dot{q}_1, \dot{q}_2 \in \mathbf{\Gamma}$, then $\mathtt{BP}\dot{q}_1, \mathtt{BP}\dot{q}_2 \in \mathbf{\Gamma}$ and $\|\mathtt{BP}\dot{q}_1 - \mathtt{BP}\dot{q}_2\| = O(k^2/2^k)\|\dot{q}_1 - \dot{q}_2\|$.*

Before the proof of Proposition A.1 we deduce the following consequences:

***Proof of Proposition 5.5a*** Let $\dot{q}^{(0)}$ be the uniform measure on $\{\mathtt{b}_0, \mathtt{b}_1, \mathtt{r}_1, \mathtt{r}_0\}$, and let $\dot{q}^{(l)} \equiv \mathtt{BP}\dot{q}^{(l-1)}$. It is clear that $\dot{q}^{(0)} \in \mathbf{\Gamma}$, so Proposition A.1 implies $\dot{q}^{(l)} \in \mathbf{\Gamma}$ for all $l \geq 1$, and furthermore that $(\dot{q}^{(l)})_{l\geq 1}$ forms an $\ell^1$ Cauchy sequence. By completeness of $\ell^1$ we conclude that there exists $\dot{q}^{(\infty)} = \dot{q}_\star \in \mathbf{\Gamma}$ such that $\mathtt{BP}\dot{q}_\star = \dot{q}_\star$ and $\|\dot{q}^{(l)} - \dot{q}_\star\| \to 0$ as $l \to \infty$. Applying Proposition A.1 again gives $\|\mathtt{BP}\dot{q} - \dot{q}_\star\| = O(k^2/2^k)\|\dot{q} - \dot{q}_\star\|$ for any $\dot{q} \in \mathbf{\Gamma}$, from which it follows that $\dot{q}_\star$ is the unique fixed point of $\mathtt{BP}$ in $\mathbf{\Gamma}$. □

***Proof of Proposition 5.5b*** For each $1 \leq T \leq \infty$, let $(\dot{q}_{\lambda,T})^{(l)}$ $(l \geq 0)$ be defined in the same way as $\dot{q}^{(l)}$ in the proof of Proposition A.1. It follows from the definition that $(\dot{q}_{\lambda,T})^{(l)} = (\dot{q}_{\lambda,\infty})^{(l)}$ for all $l \leq l_T$, where $l_T \equiv \ln T / \ln(dk)$. By the triangle inequality and Proposition 5.5a,

$$\|\dot{q}_{\lambda,T} - \dot{q}_{\lambda,\infty}\| \leq \|\dot{q}_{\lambda,T} - (\dot{q}_{\lambda,\infty})^{(l_T)}\| + \|(\dot{q}_{\lambda,\infty})^{(l_T)} - \dot{q}_{\lambda,\infty}\| \leq (C/2^k)^{l_T}$$

for some absolute constant $k$. The result follows assuming $k \geq k_0$. □

We now turn to the proof of Proposition A.1. We work with the non-normalized BP recursions $\dot{\mathtt{NB}} \equiv \dot{\mathtt{NB}}_{\lambda,T}$ and $\hat{\mathtt{NB}} \equiv \hat{\mathtt{NB}}_{\lambda,T}$, defined by substituting "$\cong$" with "$=$" in the definitions of $\dot{\mathtt{BP}}$ and $\hat{\mathtt{BP}}$ respectively. One can then recover $\dot{\mathtt{BP}}$, $\hat{\mathtt{BP}}$ from $\dot{\mathtt{NB}}$, $\hat{\mathtt{NB}}$ via

$$(\dot{\mathtt{BP}}\hat{p})(\dot{\sigma}) = \frac{(\dot{\mathtt{NB}}\hat{p})(\dot{\sigma})}{\sum_{\dot{\sigma}' \in \dot{\Omega}} (\dot{\mathtt{NB}}\hat{p})(\dot{\sigma}')}, \quad (\hat{\mathtt{BP}}\dot{p})(\hat{\sigma}) = \frac{(\hat{\mathtt{NB}}\dot{p})(\hat{\sigma})}{\sum_{\hat{\sigma}' \in \hat{\Omega}} (\hat{\mathtt{NB}}\dot{p})(\hat{\sigma}')}.$$

Let $\dot{p}$ be the reweighted measure defined by

$$\dot{p}(\dot{\sigma}) \equiv [\dot{p}(\dot{q})](\dot{\sigma}) \equiv \frac{\dot{q}(\dot{\sigma})}{1 - \dot{q}(\mathtt{r})}. \tag{71}$$

In the above we have assumed that the inputs to $\dot{\mathrm{BP}}$, $\hat{\mathrm{BP}}$, $\dot{\mathrm{NB}}$, $\hat{\mathrm{NB}}$ are probability measures; we now extend them in the obvious manner to nonnegative measures with strictly positive total mass.

Given two measures $r_1, r_2$ defined on any space $\mathcal{X}$, we denote $\Delta r(x) \equiv |r_1(x) - r_2(x)|$. We regard $\Delta r$ as a nonnegative measure on $\mathcal{X}$: for any subset $S \subseteq \mathcal{X}$,

$$\Delta r(S) = \sum_{x \in S} |r_1(x) - r_2(x)| \geq |r_1(S) - r_2(S)|,$$

where the inequality may be strict. For any nonnegative measure $\hat{r}$ on $\hat{\Omega}$, we abbreviate

$$\hat{m}^{\lambda} \hat{r}(\hat{\sigma}) \equiv \hat{m}(\hat{\sigma})^{\lambda} \hat{r}(\hat{\sigma}),$$
$$(1 - \hat{m})^{\lambda} \hat{r}(\hat{\sigma}) \equiv (1 - \hat{m}(\hat{\sigma}))^{\lambda} \hat{r}(\hat{\sigma}).$$

In what follows we will begin with two measures in $\mathbf{\Gamma}$, and show that they contract under one step of the BP recursion. Let $\hat{\mathrm{NB}}$ and $\dot{\mathrm{NB}}$ be the non-normalized single-copy BP recursions at parameters $\lambda, T$. Starting from $\dot{q}_i \in \mathbf{\Gamma}$ $(i = 1, 2)$, denote

$$\dot{p}_i \equiv \dot{p}(\dot{q}_i) \text{ (as defined by (71))},$$
$$\hat{p}_i \equiv \hat{\mathrm{NB}}(\dot{p}_i) \text{ and } \hat{p}_{i,\infty} \equiv \hat{\mathrm{NB}}_{\lambda,\infty}(\dot{p}_i),$$
$$\dot{p}_i^{\mathrm{u}} \equiv \dot{\mathrm{NB}}(\hat{p}_i) \text{ and } \tilde{q}_i \equiv \dot{\mathrm{BP}} \hat{p}_i = \mathrm{BP} \dot{q}_i.$$

With this notation in mind, the proof of Proposition A.1 is divided into four lemmas.

**Lemma A.2** (effect of reweighting) *Assuming $\dot{q}_1, \dot{q}_2 \in \mathbf{\Gamma}$, $\|\Delta \dot{p}\| = O(1)\|\dot{q}_1 - \dot{q}_2\|$, where $O(1)$ indicates a constant depending on the constant appearing in (68).*

**Lemma A.3** (clause BP) *Assuming $\dot{q}_1, \dot{q}_2 \in \mathbf{\Gamma}$,*

$$\begin{aligned}
\hat{m}^{\lambda} \hat{p}_i(s) &= 1 - 4/2^k + O(k/4^k), \\
\hat{m}^{\lambda} \hat{p}_i(f) &= \hat{m}^{\lambda} \hat{p}_i(s) + O(k/4^k), \\
\hat{m}^{\lambda} \hat{p}_i(b_1) &= 1 + O(k/2^k), \\
\hat{m}^{\lambda} \hat{p}_i(r_1) &= (2/2^k)[1 + O(k/2^k)].
\end{aligned} \tag{72}$$

*Further, writing $\Delta \hat{m}^{\lambda} \hat{p}(\cdot) \equiv \hat{m}^{\lambda}(\cdot)|\hat{p}_1(\cdot) - \hat{p}_2(\cdot)|$,*

$$\begin{aligned}
\Delta \hat{m}^{\lambda} \hat{p}(f) + \Delta \hat{m}^{\lambda} \hat{p}(r) &= O(k/2^k) \Delta \dot{p}(f), \\
\|\Delta \hat{m}^{\lambda} \hat{p}\| &= O(k^2/2^k) \|\Delta \dot{p}\|.
\end{aligned} \tag{73}$$

*(Recall that $\hat{p}(\hat{\sigma} \oplus 1) = \hat{p}(\hat{\sigma})$ and $\hat{m}(\hat{\sigma} \oplus 1) = 1 - \hat{m}(\hat{\sigma})$, so $(1-\hat{m})^{\lambda} \hat{p}(\hat{\sigma}) = \hat{m}^{\lambda} \hat{p}(\hat{\sigma} \oplus 1)$. As a result, the bounds for $\Delta \hat{m}^{\lambda} \hat{p}$ imply analogous bounds for $\Delta(1 - \hat{m})^{\lambda} \hat{p}$.)*

**Lemma A.4** (variable BP, non-normalized) *Assuming $\dot{q}_1, \dot{q}_2 \in \mathbf{\Gamma}$, we have*

$$\begin{bmatrix} \dot{p}_i^u(\mathtt{f}) \\ \dot{p}_i^u(\mathtt{r}) \end{bmatrix} = \begin{bmatrix} O(2^{-k}) \\ 1 + O(2^{-k}) \end{bmatrix} \dot{p}_i^u(\mathtt{b}), \quad \begin{bmatrix} \Delta\dot{p}^u(\mathtt{f}) \\ \Delta\dot{p}^u(\mathtt{b}) \\ \Delta\dot{p}^u(\mathtt{r}) \end{bmatrix} = \begin{bmatrix} O(k) \\ O(k2^k) \\ O(k2^k) \end{bmatrix} \|\Delta\hat{m}^\lambda \hat{p}\| \max_{i=1,2} \left\{ \dot{p}_i^u(\mathtt{b}) \right\}. \tag{74}$$

**Lemma A.5** (variable BP, normalized) *Assuming $\dot{q}_1, \dot{q}_2 \in \mathbf{\Gamma}$, we have $\tilde{q}_1, \tilde{q}_2 \in \mathbf{\Gamma}$ as well, with*

$$\|\tilde{q}_1 - \tilde{q}_2\| \lesssim k \|\Delta\hat{m}^\lambda \hat{p}\| .$$

***Proof of Proposition A.1*** Follows by combining the four preceding lemmas A.2–A.5.□

We now prove the four lemmas.

***Proof of Lemma A.2*** This follows from the elementary identity

$$\frac{a_1}{b_1} - \frac{a_2}{b_2} = \frac{1}{b_1}(a_1 - a_2) + \frac{b_2 - b_1}{b_1 b_2} a_2 . \tag{75}$$

together with (68). □

In the proof of the next two lemmas, the following elementary fact will be used repeatedly: suppose for $1 \le l \le m$ that we have nonnegative measures $a^l, b^l$ over a finite set $\mathcal{X}^l$. Then, denoting $\underline{\mathcal{X}} = \mathcal{X}^1 \times \cdots \times \mathcal{X}^m$, we have

$$\begin{aligned} \sum_{\underline{x} \in \underline{\mathcal{X}}} \left| \prod_{l=1}^m a^l(x^l) - \prod_{l=1}^m b^l(x^l) \right| &\le \sum_{l=1}^m \sum_{\underline{x} \in \underline{\mathcal{X}}} \left\{ \prod_{1 \le j < l} b^j(x^j) \right\} \left\{ \prod_{l < j \le m} a^j(x^j) \right\} \left| a^l(x^l) - b^l(x^l) \right| \\ &\le \sum_{l=1}^m \|a^l - b^l\| \prod_{j \ne l} \left( \|a^j\| + \|a^j - b^j\| \right). \end{aligned} \tag{76}$$

If all the $(\mathcal{X}^l, a^l, b^l)$ are the same $(\mathcal{X}, a, b)$, this reduces to the bound

$$\sum_{x_1, \ldots, x_m \in \mathcal{X}} \left| \prod_{i=1}^m a(x_i) - \prod_{i=1}^m b(x_i) \right| \le m \|a - b\| \left( \|a\| + \|a - b\| \right)^{m-1} . \tag{77}$$

In what follows we will abbreviate (for $\boldsymbol{x} \in \{\mathtt{0}, \mathtt{1}\}$)

$$\mathtt{a}_{\boldsymbol{x}} \equiv \left\{ \hat{\sigma} \in \hat{\Omega}_T : \underline{\hat{\sigma}} \in (\mathtt{g}_{\boldsymbol{x}})^{k-1} \text{ for all } \underline{\dot{\sigma}} \sim \hat{\sigma} \right\}. \tag{78}$$

***Proof of Lemma A.3*** From the definition, if $\dot{p} = \dot{p}(\dot{q})$ then

$$\dot{p}(\mathtt{b}) = \frac{\dot{q}(\mathtt{b})}{1 - \dot{q}(\mathtt{r})} = \frac{\dot{q}(\mathtt{b})}{\dot{q}(\mathtt{g})} = 1 - \dot{p}(\mathtt{f}) .$$

It follows that for any $\dot{q}_1, \dot{q}_2 \in \mathbf{\Gamma}$ we have $\Delta\dot{p}(\mathtt{b}) \le \Delta\dot{p}(\mathtt{f}) \le \dot{p}_1(\mathtt{f}) + \dot{p}_2(\mathtt{f}) = O(2^{-k})$. Another consequence of the definition of $\mathbf{\Gamma}$ is that $\|\Delta\dot{p}\| = O(1)$. We now control $\Delta\hat{m}^\lambda\hat{p}(\hat{\sigma})$, distinguishing a few cases:

1. We first consider $\hat{\sigma} \in \hat{\Omega}\backslash\{\mathtt{b}, \mathtt{s}\}$. For such $\hat{\sigma}$ we have

$$\Delta\hat{m}^\lambda\hat{p}(\hat{\sigma}) = \left|[\hat{m}(\hat{\sigma})\hat{z}(\hat{\sigma})]^\lambda \sum_{\underline{\dot{\sigma}}\sim\hat{\sigma}} \left( \prod_{j=2}^{k} \dot{p}_1(\dot{\sigma}_j) - \prod_{j=2}^{k} \dot{p}_2(\dot{\sigma}_j) \right)\right|,$$

   and it is easy to check that

$$\hat{m}(\hat{\sigma})\hat{z}(\hat{\sigma}) = 1 - \prod_{j=2}^{k} \dot{m}(\dot{\sigma}_j) \in [0, 1]\,.$$

   Moreover, any such $\hat{\sigma}$ must belong to $\mathtt{a}_0$ or $\mathtt{a}_1$. By summing over $\hat{\sigma} \in \mathtt{a}_0$ and applying (77) we have

$$\Delta\hat{m}^\lambda\hat{p}(\mathtt{a}_0) \le (k-1)\|(\dot{p}_1 - \dot{p}_2)\|_{\ell^1(\mathtt{g}_0)} \Big(\dot{p}_1(\mathtt{g}_0) + \Delta\dot{p}(\mathtt{f})\Big)^{k-2}\,.$$

   Recalling that $\dot{p}_1$ and $\dot{p}_2$ both lie in $\mathbf{\Gamma}$, in the above we have $\dot{p}_1(\mathtt{g}_0) + \Delta\dot{p}(\mathtt{f}) \le [1 + O(2^{-k})]/2$, as well as $\|(\dot{p}_1 - \dot{p}_2)\|_{\ell^1(\mathtt{b}_0,\mathtt{f})} = O(1)\Delta\dot{p}(\mathtt{f})$. Combining these gives

$$\Delta\hat{m}^\lambda\hat{p}(\mathtt{a}_0) = O(k/2^k)\Delta\dot{p}(\mathtt{f}),$$

   and the same bound holds for $\Delta\hat{m}^\lambda\hat{p}(\mathtt{a}_1)$.
2. Next consider $\hat{\sigma} = \mathtt{s}$, for which we have $\hat{m}(\hat{\sigma}) = 1/2$ and $\hat{z}(\hat{\sigma}) = 2$. Thus

$$\hat{m}^\lambda\hat{p}(\mathtt{s}) = 1 - (\dot{p}(\mathtt{g}_0))^{k-1} - (\dot{p}(\mathtt{g}_1))^{k-1} + \dot{p}(\mathtt{f})^{k-1}\,. \tag{79}$$

   Arguing as above gives $\Delta\hat{m}^\lambda\hat{p}(\mathtt{s}) = O(k/2^k)\Delta\dot{p}(\mathtt{f})$, proving the first half of (73).
3. Lastly consider $\hat{\sigma} \in \{\mathtt{b}_0, \mathtt{b}_1\}$. Recalling (69) we have $\Delta\hat{m}^\lambda\hat{p}(\mathtt{b}_0) = 0$, so let us take $\hat{\sigma} = \mathtt{b}_1$, and consider $\underline{\dot{\sigma}} \sim \mathtt{b}_1$. Note that if $\underline{\dot{\sigma}}$ has no entry in $\{\mathtt{r}\}$, then we also have $\underline{\dot{\sigma}} \sim \hat{\sigma}'$ for some $\hat{\sigma}' \in \{\mathtt{r}, \mathtt{f}\}$. Again making use of (69), this $\underline{\dot{\sigma}}$ gives the same contribution to $\hat{m}^\lambda\hat{p}_\infty(\hat{\sigma}')$ as to $\hat{m}^\lambda\hat{p}(\mathtt{b}_1)$. It follows that

$$\Delta\hat{m}^\lambda\hat{p}(\mathtt{b}_1) \le \Delta\hat{m}^\lambda\hat{p}_\infty(\mathtt{y}) + k\Big|\dot{p}_1(\mathtt{r}_0)\dot{p}_1(\mathtt{b}_1)^{k-2} - \dot{p}_2(\mathtt{r}_0)\dot{p}_2(\mathtt{b}_1)^{k-2}\Big|\,.$$

   The first term on the right-hand side captures the contribution from those $\underline{\dot{\sigma}}$ with no entry in $\{\mathtt{r}\}$, and by the preceding arguments it is $O(k/2^k)\Delta\dot{p}(\mathtt{f})$. It is easy to check that the second term is $O(k^2/2^k)\|\Delta\dot{p}\|$, which finishes the second part of (73).

Combining the above estimates proves (73). We next prove (72). For this purpose we introduce the notation $\mathtt{f}_{\geq 1}$ to refer to elements of $\dot{\Omega}$ or $\hat{\Omega}$ that contain at least one free variable. In particular, $\mathtt{f}_{\geq 1} \cap \hat{\Omega}$ is given by $\{\mathtt{f}\}\backslash\{\mathtt{s}\} \subseteq \mathtt{a}_0 \cup \mathtt{a}_1 \subseteq \hat{\Omega}$. Since $\dot{q}_i \in \mathbf{\Gamma}$, we must have from (68) that

$$\hat{m}^\lambda \hat{p}_i(\mathtt{f}_{\geq 1}) \leq 2 \sum_{l=1}^{k-1} \binom{k-1}{l} \dot{p}_i(\mathtt{f})^l \dot{p}_i(\mathtt{b}_0)^{k-1-l} \leq 2\dot{p}_i(\mathtt{b}_0)^{k-1} \sum_{l=1}^{k-1} \left(\frac{k\dot{p}_i(\mathtt{f})}{\dot{p}_i(\mathtt{b}_0)}\right)^l = O(k/4^k)\,. \tag{80}$$

On the other hand, we see from (79) that

$$\hat{m}^\lambda \hat{p}_i(\mathtt{s}) = 1 - 4/2^k + O(k/4^k)\,.$$

If $\underline{\dot{\sigma}} \sim \mathtt{b}_1$ has no entry in $\{\mathtt{r}\}$, then there must exist some $\hat{\sigma} \in \{\mathtt{f}\}$ such that $\underline{\dot{\sigma}} \sim \hat{\sigma}$ as well. Conversely, if $\hat{\sigma} \in \hat{\Omega}_T\backslash\{\mathtt{r}, \mathtt{b}\}$ and $\underline{\dot{\sigma}} \sim \hat{\sigma}$, then $\underline{\dot{\sigma}} \sim \mathtt{b}_1$, unless $\underline{\dot{\sigma}}$ has exactly one spin $\dot{\sigma}_i \in \{\mathtt{b}_0, \mathtt{f}\}$ with the remaining $k-2$ spins equal to $\mathtt{b}_1$.[1] It follows that

$$\begin{aligned}\hat{m}^\lambda \hat{p}_i(\mathtt{b}_1) = \hat{p}_i(\mathtt{b}_1) &= \hat{m}^\lambda \hat{p}_{i,\infty}(\mathtt{f}) + (k-1)\Big[\dot{p}_i(\mathtt{r}_0) - \dot{p}_i(\mathtt{g}_0)\Big]\dot{p}_i(\mathtt{b}_1)^{k-2} \\ &\leq \hat{m}^\lambda \hat{p}_{i,\infty}(\mathtt{f}) + (k-1)\dot{p}_i(\mathtt{r}_0)\dot{p}_i(\mathtt{b}_1)^{k-2} = 1 + O(k/2^k). \end{aligned} \tag{81}$$

For a lower bound it suffices to consider the contribution from clauses with all $k$ incident colors in $\{\mathtt{b}\}$:

$$\hat{m}^\lambda \hat{p}_i(\mathtt{b}_1) = \hat{p}_i(\mathtt{b}_1) \geq \dot{p}_i(\mathtt{b})^{k-1}[1 - O(k/2^k)] = 1 - O(k/2^k)\,. \tag{82}$$

Lastly, note by symmetry that

$$\hat{m}^\lambda \hat{p}_i(\mathtt{r}_1) = \hat{p}_i(\mathtt{r}_1) = \hat{p}_i(\mathtt{b}_0)^{k-1} = (2/2^k)\hat{p}_i(\mathtt{b})^{k-1}\,.$$

Combining these estimates proves (72). □

***Proof of Lemma A.4*** We control $\dot{p}^{\mathrm{u}}$ and $\Delta\dot{p}^{\mathrm{u}}$ in two cases.

1. First consider $\dot{\sigma} \in \dot{\Omega}\backslash\{\mathtt{r}, \mathtt{b}\}$. Up to permutation there is a unique $\underline{\hat{\sigma}} \in \{\mathtt{f}\}^{d-1}$ such that $\dot{\sigma} = \hat{T}(\underline{\hat{\sigma}})$. Let $\mathsf{comb}(\dot{\sigma})$ denote the number of distinct tuples $\underline{\hat{\sigma}}'$ that can be obtained by permuting the coordinates of $\underline{\hat{\sigma}}$. For this $\underline{\hat{\sigma}}$ we have

$$\prod_{j=2}^{d} \hat{m}(\hat{\sigma}_j)^\lambda \leq \dot{z}(\dot{\sigma})^\lambda \leq \prod_{j=2}^{d} \hat{m}(\hat{\sigma}_j)^\lambda + \prod_{j=2}^{d}(1 - \hat{m}(\hat{\sigma}_j))^\lambda, \tag{83}$$

where the rightmost inequality uses that $(a+b)^\lambda \leq a^\lambda + b^\lambda$ for $a, b \geq 0$ and $\lambda \in [0, 1]$. It follows that for $i = 1, 2$ we have

[1] The converse is not needed for the final bound, but we mention it for the sake of concreteness.

$$\mathsf{comb}(\dot\sigma)\prod_{j=2}^{d}\hat m\,\hat p_i(\hat\sigma_j)\le \dot p_i^{\mathrm u}(\dot\sigma)\le \mathsf{comb}(\dot\sigma)\left\{\prod_{j=2}^{d}\hat m^\lambda\,\hat p_i(\hat\sigma_j)+\prod_{j=2}^{d}(1-\hat m)^\lambda\,\hat p_i(\hat\sigma_j)\right\}.$$

It follows by symmetry that $\hat m^\lambda\,\hat p_i(\mathtt f)=(1-\hat m)^\lambda\,\hat p_i(\mathtt f)$, so

$$[\hat m^\lambda\,\hat p_i(\mathtt s)]^{d-1}\le \dot p_i^{\mathrm u}(\mathtt f)\le [\hat m^\lambda\,\hat p_i(\mathtt f)]^{d-1}+[(1-\hat m)^\lambda\,\hat p_i(\mathtt f)]^{d-1}=2[\hat m^\lambda\,\hat p_i(\mathtt f)]^{d-1}. \tag{84}$$

Making use of the symmetry together with (83) gives

$$\Delta\dot p^{\mathrm u}(\mathtt f)\le 2\sum_{\underline{\hat\sigma}\in(\hat\Omega_{\mathtt f})^{d-1}}\left|\prod_{j=2}^{d-1}\hat m^\lambda\,\hat p_1(\hat\sigma_j)-\prod_{j=2}^{d-1}\hat m^\lambda\,\hat p_2(\hat\sigma_j)\right|,$$

and applying (77) gives

$$\Delta\dot p^{\mathrm u}(\mathtt f)\lesssim d\|\Delta\hat m^\lambda\,\hat p\|\Big(\hat m^\lambda\,\hat p_1(\mathtt f)+\Delta\hat m^\lambda\,\hat p_1(\mathtt f)\Big)^{d-2}.$$

Combining (72) with the lower bound from (83) then gives

$$\Delta\dot p^{\mathrm u}(\mathtt f)\lesssim d\|\Delta\hat m^\lambda\,\hat p\|\max_{i=1,2}\left\{\dot p_i^{\mathrm u}(\mathtt f)\right\}.$$

2. ] Next consider $\dot\sigma\in\{\mathtt r,\mathtt b\}$: for $\boldsymbol x\in\{0,1\}$, note that $\dot p_i^{\mathrm u}(\mathtt r_{\boldsymbol x})=\hat p_i(\mathtt p_{\boldsymbol x})^{d-1}$, and

$$\frac{\dot p_i^{\mathrm u}(\mathtt r_{\boldsymbol x})-\dot p_i^{\mathrm u}(\mathtt b_{\boldsymbol x})}{\dot p_i^{\mathrm u}(\mathtt r_{\boldsymbol x})}=\frac{\hat p_i(\mathtt b_{\boldsymbol x})^{d-1}}{\hat p_i(\mathtt p_{\boldsymbol x})^{d-1}}=\left(1-\frac{\hat p_i(\mathtt r_{\boldsymbol x})}{\hat p_i(\mathtt p_{\boldsymbol x})}\right)^{d-1}=O(2^{-k}), \tag{85}$$

where the last estimate uses (72) and $d/k=2^{k-1}\ln 2+O(1)$. Applying (77) gives

$$\Delta\dot p^{\mathrm u}(\mathtt p_1)\lesssim d\|\hat m^\lambda\,\hat p\|\Big(\min_{i=1,2}\left\{\hat m^\lambda\,\hat p_i(\mathtt p_1)\right\}+\Delta\hat m^\lambda\,\hat p(\mathtt p_1)\Big)^{d-2}.$$

Suppose without loss that $\hat m^\lambda\,\hat p_1(\mathtt b_1)\le\hat m^\lambda\,\hat p_2(\mathtt b_1)$: then

$$\begin{aligned}\hat m^\lambda\,\hat p_1(\mathtt p_1)+\Delta\hat m^\lambda\,\hat p(\mathtt p_1)&=\hat m^\lambda\,\hat p_2(\mathtt b_1)+\hat m^\lambda\,\hat p_1(\mathtt r_1)+\Delta\hat m^\lambda\,\hat p(\mathtt r_1)\\&\le\hat m^\lambda\,\hat p_2(\mathtt p_1)+2\Delta\hat m^\lambda\,\hat p(\mathtt r_1),\end{aligned}$$

and substituting into the above gives

$$\Delta\dot p^{\mathrm u}(\mathtt p_1)\lesssim d\|\hat m^\lambda\,\hat p\|\Big(\max_{i=1,2}\left\{\hat m^\lambda\,\hat p_i(\mathtt p_1)\right\}+\Delta\hat m^\lambda\,\hat p(\mathtt r_1)\Big)^{d-2}.$$

From (73) and the definition (68) of $\mathbf{\Gamma}$ we have $\Delta\hat{m}^\lambda\hat{p}(\mathtt{r}_1) = O(k/2^k)\Delta\dot{p}(\mathtt{f}) = O(k/4^k)$. It follows from (85) that

$$\Delta\dot{p}^{\mathrm{u}}(\mathtt{p}_1) \lesssim d\|\Delta\hat{m}^\lambda\hat{p}\| \max_{i=1,2}\left\{\dot{p}_i^{\mathrm{u}}(\mathtt{b}_1)\right\}. \tag{86}$$

It remains to show $\dot{p}^{\mathrm{u}}(\mathtt{f})/\dot{p}^{\mathrm{u}}(\mathtt{b}) = O(2^{-k})$. From (81),

$$\hat{m}^\lambda\hat{p}_i(\mathtt{f}) - \hat{m}^\lambda\hat{p}_i(\mathtt{b}_1) \le \hat{m}^\lambda\hat{p}_{i,\infty}(\mathtt{f}) - \hat{m}^\lambda\hat{p}_i(\mathtt{b}_1) \le (k-1)\Big[\dot{p}_i(\mathtt{g}_0) - \dot{p}_i(\mathtt{r}_0)\Big]\dot{p}_i(\mathtt{b}_1)^{k-2},$$

and by definition of $\mathbf{\Gamma}$ the right-hand side is $O(k/4^k)\dot{p}_i(\mathtt{b})^{k-1}$. Now recall from (82) that $\hat{m}^\lambda\hat{p}_i(\mathtt{b}_1) \gtrsim \dot{p}_i(\mathtt{b})^{k-1}$. Combining these gives

$$\hat{m}^\lambda\hat{p}_i(\mathtt{f}) \le [1 + O(k/4^k)]\hat{m}^\lambda\hat{p}_i(\mathtt{b}_1). \tag{87}$$

Recalling (83), it follows that

$$\frac{\dot{p}_i^{\mathrm{u}}(\mathtt{f})}{\dot{p}_i^{\mathrm{u}}(\mathtt{b}_1)} \lesssim \left(\frac{\hat{m}^\lambda\hat{p}_i(\mathtt{f})}{\hat{m}^\lambda\hat{p}_i(\mathtt{p}_1)}\right)^{d-1} \lesssim \left(\frac{\hat{m}^\lambda\hat{p}_i(\mathtt{b}_1)}{\hat{m}^\lambda\hat{p}_i(\mathtt{p}_1)}\right)^{d-1} \lesssim 2^{-k},$$

where the last step uses (72). This concludes the proof. □

***Proof of Lemma A.5*** Denote $\tilde{q}_i \equiv \mathtt{BP}\dot{q}_i$ and $\Delta\tilde{q} \equiv |\tilde{q}_1 - \tilde{q}_2|$. We first check that $\tilde{q}_i$ lies in $\mathbf{\Gamma}$: the first condition of (68) follows from (74), and the second is automatically satisfied from the definition of $\dot{\mathtt{BP}}$. Next we bound $\Delta\tilde{q}$. With some abuse of notation, we shall write $\tilde{q}_i(\mathtt{X}) \equiv \tilde{q}_i(\mathtt{r}) - \tilde{q}_i(\mathtt{b})$ and

$$\Delta\tilde{q}(\mathtt{X}) \equiv |(\tilde{q}_1(\mathtt{r}) - \tilde{q}_1(\mathtt{b})) - (\tilde{q}_2(\mathtt{r}) - \tilde{q}_2(\mathtt{b}))|.$$

Let $\dot{p}_i^{\mathrm{u}}(\mathtt{X})$ and $\Delta\dot{p}^{\mathrm{u}}(\mathtt{X})$ be similarly defined. Arguing similarly as in the derivation of (86),

$$\Delta\dot{p}^{\mathrm{u}}(\mathtt{X}) = 2|\hat{p}_1(\mathtt{b}_1)^{d-1} - \hat{p}_2(\mathtt{b}_1)^{d-1}| \lesssim k\|\Delta\hat{m}^\lambda\hat{p}\| \max_{i=1,2}\left\{\dot{p}_i^{\mathrm{u}}(\mathtt{b})\right\} \tag{88}$$

Recalling $\|\tilde{q}_i\| = 1$, we have

$$\begin{aligned} 2\tilde{q}_i(\mathtt{r}) &= [1 - \tilde{q}_i(\mathtt{f})] + [\tilde{q}_i(\mathtt{r}) - \tilde{q}_i(\mathtt{b})] \text{ and} \\ 2\tilde{q}_i(\mathtt{b}) &= [1 - \tilde{q}_i(\mathtt{f})] - [\tilde{q}_i(\mathtt{r}) - \tilde{q}_i(\mathtt{b})], \text{ so} \\ \|\Delta\tilde{q}\| &\lesssim \Delta\tilde{q}(\mathtt{f}) + \Delta\tilde{q}(\mathtt{X}). \end{aligned}$$

If we take $a \in \{1, 2\}$ and $b = 2 - a$, and write $\dot{Z}_i \equiv \|\dot{p}_i^{\mathrm{u}}\|$, then

$$\Delta\tilde{q}(\mathtt{f}) + \Delta\tilde{q}(\mathtt{X}) \le \frac{\Delta\dot{p}^{\mathrm{u}}(\mathtt{f}) + \Delta\dot{p}^{\mathrm{u}}(\mathtt{X})}{\dot{Z}_a} + \frac{|\dot{Z}_a - \dot{Z}_b|}{\dot{Z}_a}\frac{[\dot{p}_b^{\mathrm{u}}(\mathtt{f}) + \dot{p}_b^{\mathrm{u}}(\mathtt{r}) - \dot{p}_b^{\mathrm{u}}(\mathtt{b})]}{\dot{Z}_b}.$$

If we take $a \in \arg\max_i \dot{p}_i^{\mathrm{u}}(\mathtt{b})$, then, by (74) and (88), the first term on the right-hand side is

$$\lesssim \frac{k\|\Delta\hat{m}^{\lambda}\hat{p}\|\dot{p}_a^{\mathrm{u}}(\mathtt{b})}{\dot{Z}_a} \lesssim k\|\Delta\hat{m}^{\lambda}\hat{p}\|,$$

where the rightmost inequality uses $\dot{Z}_i \geq \dot{p}_i^{\mathrm{u}}(\mathtt{b})$. As for the second term, (74) gives

$$\frac{|\dot{Z}_a - \dot{Z}_b|}{\dot{Z}_a} \lesssim d\|\Delta\hat{m}^{\lambda}\hat{p}\| \quad \text{and} \quad \frac{[\dot{p}_b^{\mathrm{u}}(\mathtt{f}) + \dot{p}_b^{\mathrm{u}}(\mathtt{r}) - \dot{p}_b^{\mathrm{u}}(\mathtt{b})]}{\dot{Z}_b} \lesssim 2^{-k}.$$

Combining these estimates yields the claimed bound. □

### A.2. Pair coloring recursions

In this section we analyze the BP recursions for the pair coloring model and prove the remaining assertions of Proposition 5.5. Recall that we assume $\dot{q} = \dot{q}^{\mathrm{av}}$ and $\hat{q} = \hat{q}^{\mathrm{av}}$, where these are now probability measures on $(\dot{\Omega}_T)^2$ and $(\hat{\Omega}_T)^2$ respectively. For any measure $p(x)$ defined on $x \equiv (x^1, x^2)$ in $(\dot{\Omega}_T)^2$ or $(\hat{\Omega}_T)^2$, define

$$(\mathfrak{f}p)(x) \equiv p(\mathfrak{f}x) \quad \text{where } \mathfrak{f}x \equiv x \oplus (0,1) \equiv (x^1, x^2 \oplus 1).$$

Recall from Sect. 5.3 the definition of $\mathbf{\Gamma}(c,\kappa)$. We will prove that

**Proposition A.6** *For any $c \in (0,1]$ and any $\dot{q}_1, \dot{q}_2 \in \mathbf{\Gamma}(c,1)$, we have $\mathtt{BP}\dot{q}_1, \mathtt{BP}\dot{q}_2 \in \mathbf{\Gamma}(1,1)$ and*

$$\|\mathtt{BP}\dot{q}_1 - \mathtt{BP}\dot{q}_2\| = O(k^4/2^k)\|\dot{q}_1 - \dot{q}_2\| + O(k^4/2^k)\sum_{i=1,2}\|\dot{q}_i - \mathfrak{f}\dot{q}_i\|. \tag{89}$$

Assuming this result, it is straightforward to deduce Proposition 5.5A:

***Proof of Proposition 5.5A*** Let $\dot{q}^{(0)}$ be the uniform probability measure on $\{\mathtt{b}_0, \mathtt{b}_1, \mathtt{r}_1, \mathtt{r}_0\}^2$, and define recursively $\dot{q}^{(l)} = \mathtt{BP}\dot{q}^{(l-1)}$ for $l \geq 1$. It is clear that $\dot{q}^{(0)} \in \mathbf{\Gamma}(1,1)$ and $\dot{q}^{(0)} = \mathfrak{f}\dot{q}^{(0)}$. Since $\dot{q}^{(l)} = \mathfrak{f}\dot{q}^{(l)}$ for all $l \geq 1$, it follows from (89) that $(\dot{q}^{(l)})_{l\geq 1}$ forms an $\ell^1$ Cauchy sequence. It follows by completeness of $\ell^1$ that $\dot{q}^{(l)}$ converges to a limit $\dot{q}^{(\infty)} = \dot{q}_\star \in \mathbf{\Gamma}(1,1)$, satisfying $\dot{q}_\star = \mathfrak{f}\dot{q}_\star = \mathtt{BP}\dot{q}_\star$. This implies that for any probability measure $\dot{q}$,

$$\|\dot{q} - \mathfrak{f}\dot{q}\| \leq \|\dot{q} - \dot{q}_\star\| + \|\dot{q}_\star - \mathfrak{f}\dot{q}\| = 2\|\dot{q} - \dot{q}_\star\|.$$

Applying (89) again gives

$$\|\mathtt{BP}\dot{q} - \dot{q}_\star\| = O(k^4/2^k)\|\dot{q} - \dot{q}_\star\| + O(k^4/2^k)\|\dot{q} - \mathfrak{f}\dot{q}\| = O(k^4/2^k)\|\dot{q} - \dot{q}_\star\|,$$

proving the claimed contraction estimate. Uniqueness of $\dot{q}_\star$ can be deduced from this contraction. □

We now turn to the proof of Proposition A.6; the proof of Proposition 5.5B is given after. Let $\dot{\mathtt{NB}}$, $\hat{\mathtt{NB}}$ now denote the non-normalized BP recursions for the pair model. Let $\mathtt{r}[\dot{\sigma}] \in \{0, 1, 2\}$ count the number of $\mathtt{r}$ spins in $\dot{\sigma}$, and let $\dot{p} \equiv \dot{p}(\dot{q})$ be the reweighted measure

$$\dot{p}(\dot{\sigma}) \equiv \frac{\dot{q}(\dot{\sigma})}{1 - \dot{q}(\mathtt{r}[\dot{\sigma}] > 0)}\,. \tag{90}$$

Recalling convention (69), we will denote

$$\hat{m}^{\lambda}\hat{r}(\hat{\sigma}^1, \hat{\sigma}^2) \equiv [\hat{m}(\hat{\sigma}^1)\hat{m}(\hat{\sigma}^2)]^{\lambda}\hat{r}(\hat{\sigma}^1, \hat{\sigma}^2)\,.$$

Let $\hat{\mathtt{NB}}$ and $\dot{\mathtt{NB}}$ be the non-normalized pair BP recursions at parameters $\lambda$, $T$. Starting from $\dot{q}_i \in \mathbf{\Gamma}(c, \kappa)$ $(i = 1, 2)$, we denote

$$\begin{aligned}
&\dot{p}_i \equiv \dot{p}(\dot{q}_i) \text{ (as defined by (refe:reweight.second))},\\
&\hat{p}_i \equiv \hat{\mathtt{NB}}(\dot{p}_i) \text{ and } \hat{p}_{i,\infty} \equiv \hat{\mathtt{NB}}_{\lambda,\infty}(\dot{p}_i),\\
&\dot{p}_i^{\mathrm{u}} \equiv \dot{\mathtt{NB}}(\hat{p}_i) \text{ and } \tilde{q}_i \equiv \dot{\mathtt{BP}}\hat{p}_i = \mathtt{BP}\dot{q}_i\,.
\end{aligned}$$

With this notation in mind, the proof of Proposition A.6 is divided into the following lemmas.

**Lemma A.7** (effect of reweighting) *Suppose $\dot{q}_1, \dot{q}_2 \in \mathbf{\Gamma}(c, \kappa)$ for $c \in (0, 1]$ and $\kappa \in [0, 1]$: then*

$$\begin{aligned}
\|\Delta\dot{p}\| &\equiv O(2^{2(1-\kappa)k})\|\Delta\dot{q}\|,\\
\|\dot{p}_i - \mathfrak{f}\dot{p}_i\| &\equiv O(2^{(1-\kappa)k})\|\dot{q}_i - \mathfrak{f}\dot{q}_i\|.
\end{aligned}$$

**Lemma A.8** (clause BP contraction) *Suppose $\dot{q}_1, \dot{q}_2 \in \mathbf{\Gamma}(c, \kappa)$ for $c \in (0, 1]$ and $\kappa \in [0, 1]$: then*

$$\begin{aligned}
\Delta\hat{m}^{\lambda}\hat{p}(\mathtt{yy}) &= O(k^3/2^k)\Delta\dot{p}(\mathtt{gg}) = O(k^3/2^{(1+c)k}),\\
\Delta\hat{m}^{\lambda}\hat{p}(\{\mathtt{br}, \mathtt{bf}_{\geq 1}\}) &= O(k^2/2^k)[\Delta\dot{p}(\mathtt{gg}) + 2^{-k}\Delta\dot{p}(\dot{\Omega}^2\backslash\{\mathtt{rr}\})] = O(k^3/2^{(1+c)k}),\\
\|\Delta\hat{m}^{\lambda}\hat{p}\| &= O(k^3/2^k)\|\Delta\dot{p}\| = O(k^3 2^{(1-2\kappa)k}),
\end{aligned} \tag{91}$$

*and the same estimates hold with $\mathfrak{f}\hat{p}$ in place of $\hat{p}$. For both $i = 1, 2$,*

$$\|\hat{m}^{\lambda}\hat{p}_i - \hat{m}^{\lambda}\mathfrak{f}\hat{p}_i\| = O(k^3/2^{(1+\kappa)k})\|\dot{p}_i - \mathfrak{f}\dot{p}_i\| = O(k^3/2^{2\kappa k})\|\dot{q}_i - \mathfrak{f}\dot{q}_i\|\,. \tag{92}$$

**Lemma A.9** (clause BP output values) *Suppose $\dot{q}_1, \dot{q}_2 \in \mathbf{\Gamma}(c, \kappa)$ for $c \in (0, 1]$ and $\kappa \in [0, 1]$. For $s, t \subseteq \hat{\Omega}$ let $st \equiv s \times t$. Then it holds for all $s, t \in \{\mathtt{r}_1, \mathtt{b}_1, \mathtt{f}, \mathtt{s}\}$ that*

$$\frac{\hat{m}^{\lambda}\hat{p}_i(s, t)}{(2/2^k)^{\mathtt{r}[s]+\mathtt{r}[t]}} = \begin{cases} 1 + O(k^2/2^k) & \text{if } \mathtt{r}[s] + \mathtt{r}[t] \leq 1,\\ 1 + O(k^2/2^{ck}) & \text{if } \mathtt{r}[s] + \mathtt{r}[t] = 2. \end{cases} \tag{93}$$

*Furthermore we have the bounds*

$$\hat{m}^{\lambda}\hat{p}_i(\mathfrak{f}_{\geq 1}t) + \hat{m}^{\lambda}\hat{p}_i(t\,\mathfrak{f}_{\geq 1}) \leq O(k/4^k) \text{ for all } t \in \{\mathtt{r}_1, \mathtt{b}_1, \mathfrak{f}, \mathfrak{s}\},$$
$$\hat{m}^{\lambda}\hat{p}_i(\{\mathfrak{f}\} \times \hat{\Omega}) - \hat{m}^{\lambda}\hat{p}_i(\{\mathtt{b}_1\} \times \hat{\Omega}) \leq O(k/4^k). \tag{94}$$

*The same estimates hold with $\mathfrak{f}\hat{p}_i$ in place of $\hat{p}_i$.*

**Lemma A.10** (variable BP) *Suppose $\dot{q}_1, \dot{q}_2 \in \mathbf{\Gamma}(c, \kappa)$ for $c \in (0, 1]$ and $\kappa \in [0, 1]$. Then $\mathtt{BP}\dot{q}_1, \mathtt{BP}\dot{q}_2 \in \mathbf{\Gamma}(c', 1)$ for $c' = \max\{0, 2\kappa - 1\}$, and*

$$\left\|\mathtt{BP}\dot{q}_1 - \mathtt{BP}\dot{q}_2\right\| \lesssim k\left\|\Delta\hat{m}^{\lambda}\hat{p} + \Delta\hat{m}^{\lambda}\mathfrak{f}\hat{p}\right\|\big) + k2^k \sum_{i=1,2}\left\|\hat{m}^{\lambda}\hat{p}_i - \hat{m}^{\lambda}\mathfrak{f}\hat{p}_i\right\|.$$

***Proof of Proposition A.6*** Follows by combining the preceding lemmas A.7–A.10. □

***Proof of Proposition 5.5B*** If $\dot{q} \in \mathbf{\Gamma}(c, 0)$ is a fixed point of BP, then it follows from Lemmas A.8–A.10 that we have $\dot{q} \in \mathbf{\Gamma}(c, 0) \cap \mathbf{\Gamma}(0, 1) = \mathbf{\Gamma}(c, 1)$. □

We now prove the three lemmas leading to Proposition A.6.

***Proof of Lemma A.7*** Applying (75) we have

$$|\dot{p}_1(\dot{\sigma}) - \dot{p}_2(\dot{\sigma})| \leq \frac{|\dot{q}_1(\dot{\sigma}) - \dot{q}_2(\dot{\sigma})|}{\dot{q}_1(\mathtt{gg})} + \frac{|\dot{q}_1(\mathtt{gg}) - \dot{q}_2(\mathtt{gg})|}{\dot{q}_1(\mathtt{gg})\dot{q}_2(\mathtt{gg})}\dot{q}_2(\dot{\sigma}),$$

and summing over $\dot{\sigma} \in \dot{\Omega}^2$ gives

$$\|\Delta\dot{p}\| \leq \frac{\|\dot{q}_1 - \dot{q}_2\|}{\dot{q}_1(\mathtt{gg})} + \frac{|\dot{q}_1(\mathtt{gg}) - \dot{q}_2(\mathtt{gg})|}{\dot{q}_1(\mathtt{gg})\dot{q}_2(\mathtt{gg})} \leq \frac{2\|\dot{q}_1 - \dot{q}_2\|}{\dot{q}_1(\mathtt{gg})\dot{q}_2(\mathtt{gg})}.$$

Since $\dot{q}_i \in \mathbf{\Gamma}$, we have, using $(1\mathbf{\Gamma})$ and $(2\mathbf{\Gamma})$,

$$\dot{p}_i(\dot{\Omega}^2\backslash\{\mathtt{rr}\}) = O(1), \quad \dot{p}_i(\mathtt{rr}) = O(2^{(1-\kappa)k}). \tag{95}$$

Consequently $\dot{q}_i(\mathtt{gg})^{-1} \leq O(1)2^{(1-\kappa)k}$, and the claimed bound on $\|\Delta\dot{p}\|$ follows. The bound on $\|\dot{p}_i - \mathfrak{f}\dot{p}_i\|$ follows by noting that if $\dot{q}_2 = \mathfrak{f}\dot{q}_1$, then $\dot{q}_1(\mathtt{gg}) = \dot{q}_2(\mathtt{gg})$. □

***Proof of Lemma A.8*** We will prove (91) for $\hat{p}_i$; the proof for $\mathfrak{f}\hat{p}_i$ is entirely similar. It follows from the symmetry $\dot{p}_i = (\dot{p}_i)^{\mathrm{av}}$ that for any $\boldsymbol{x}, \boldsymbol{y} \in \{0, 1\}$,

$$\left|\dot{p}_i(\mathtt{bb}) - 4\dot{p}_i(\mathtt{b}_{\boldsymbol{x}}\mathtt{b}_{\boldsymbol{y}})\right| = 2\left|\dot{p}_i(\mathtt{b}_{\boldsymbol{x}}\mathtt{b}_{\boldsymbol{y}\oplus 1}) - \dot{p}_i(\mathtt{b}_{\boldsymbol{x}}\mathtt{b}_{\boldsymbol{y}})\right| = 2\left|\dot{p}_i(\mathtt{b}_0\mathtt{b}_0) - \dot{p}_i(\mathtt{b}_0\mathtt{b}_1)\right|,$$

from which we obtain that

$$\Delta\dot{p}(\mathtt{bb}) \lesssim \max_{i=1,2}\left|\dot{p}_i(\mathtt{b}_0\mathtt{b}_0) - \dot{p}_i(\mathtt{b}_0\mathtt{b}_1)\right|.$$

Recall $\mathtt{g} = \{\mathtt{b}, \mathtt{f}\}$ and $\dot{p}_i(\mathtt{gg}) = 1$. Combining the above with the definition of $\mathbf{\Gamma}(c, \kappa)$ gives

$$\begin{aligned}\Delta\dot{p}(\mathtt{gg}) &\le \Delta\dot{p}(\mathtt{bb}) + \Delta\dot{p}(\mathtt{gf}) + \Delta\dot{p}(\mathtt{fg}) \\ &\le \sum_{i=1,2} \left\{ \left| \dot{p}_i(\mathtt{b}_0\mathtt{b}_0) - \dot{p}_i(\mathtt{b}_0\mathtt{b}_1) \right| + \dot{p}_i(\mathtt{gf}) + \dot{p}(\mathtt{fg}) \right\} = O(2^{-ck}). \quad (96)\end{aligned}$$

*Step I.* We first control $\Delta\hat{m}^\lambda\hat{p}(\hat{\sigma})$. By symmetry it suffices to analyze the BP recursion at a clause with all literals $\mathtt{L}_j = \mathtt{0}$. We distinguish the following cases of $\hat{\sigma} \in \hat{\Omega}^2$:

1. Recall $\mathtt{y} \equiv \mathtt{r} \cup \mathtt{f}$, and note $\{\mathtt{y}\}\backslash\{\mathtt{s}\} \subseteq \mathtt{a}_0 \cup \mathtt{a}_1$ (as defined by (78)). Thus

$$\Delta\hat{m}^\lambda\hat{p}(\{\mathtt{yy}\}\backslash\{\mathtt{ss}\}) \le \sum_{\boldsymbol{x}\in\{0,1\}} \left\{ \Delta\hat{m}^\lambda\hat{p}(\mathtt{a}_{\boldsymbol{x}} \times \{\mathtt{y}\}) + \Delta\hat{m}^\lambda\hat{p}(\{\mathtt{y}\} \times \mathtt{a}_{\boldsymbol{x}}) \right\}. \quad (97)$$

   For $\boldsymbol{x} \in \{0, 1\}$, consider $\hat{\sigma} \in \mathtt{a}_{\boldsymbol{x}} \times \{\mathtt{y}\}$: in order for $\underline{\dot{\sigma}} \in (\dot{\Omega}^2)^{k-1}$ to be compatible with $\hat{\sigma}$, it is necessary that $\dot{\sigma}_j \in A \equiv \{\mathtt{g}_{\boldsymbol{x}}\} \times \{\mathtt{g}\}$ for all $2 \le j \le k$. Combining with (77) gives

$$\Delta\hat{m}^\lambda\hat{p}(\mathtt{a}_{\boldsymbol{x}} \times \{\mathtt{y}\}) \le \sum_{\underline{\dot{\sigma}}\in A^{k-1}} \left| \prod_{j=2}^{k} \dot{p}_1(\dot{\sigma}_j) - \prod_{j=2}^{k} \dot{p}_2(\dot{\sigma}_j) \right| \le k\Delta\dot{p}(\mathtt{gg}) \Big( \dot{p}_1(A) + \Delta\dot{p}(\mathtt{gg}) \Big)^{k-2}.$$

   It follows from the definition of $\mathbf{\Gamma}(c, \kappa)$ that $\dot{p}_1(A) + \Delta\dot{p}(\mathtt{gg}) = \frac{1}{2} + O(2^{-ck})$, so we conclude

$$\Delta\hat{m}^\lambda\hat{p}(\{\mathtt{yy}\}\backslash\{\mathtt{ss}\}) = O(k/2^k)\Delta\dot{p}(\mathtt{gg})\,. \quad (98)$$

2. Now take $\hat{\sigma} = \mathtt{ss}$: for $\underline{\dot{\sigma}} \in (\dot{\Omega}^2)^{k-1}$ to be compatible with $\hat{\sigma}$, it is necessary that $\underline{\dot{\sigma}} \in \{\mathtt{yy}\}^{k-1}$. On the other hand, it is sufficient that $\underline{\dot{\sigma}} \in \{\mathtt{gg}\}^{k-1}$ does not belong to any of the sets $(A_0)^{k-1}$, $(A_1)^{k-1}$, $(B_0)^{k-1}$, $(B_1)^{k-1}$, where for $\boldsymbol{x} \in \{0, 1\}$ we define $A_{\boldsymbol{x}} \equiv \{\mathtt{b}_{\boldsymbol{x}}\mathtt{g}\} \cup \{\mathtt{fg}\}$ and $B_{\boldsymbol{x}} \equiv \{\mathtt{gb}_{\boldsymbol{x}}\} \cup \{\mathtt{gf}\}$. Therefore

$$\Delta\hat{m}^\lambda\hat{p}(\mathtt{ss}) \le \sum_{\boldsymbol{x}\in\{0,1\}} \sum_{\underline{\dot{\sigma}}\in(A_{\boldsymbol{x}})^{k-1}\cup(B_{\boldsymbol{x}})^{k-1}} \left| \prod_{j=2}^{k} \dot{p}_1(\dot{\sigma}_j) - \prod_{j=2}^{k} \dot{p}_2(\dot{\sigma}_j) \right| = O(k/2^k)\Delta\dot{p}(\mathtt{gg}),$$

   where the last estimate follows by the same argument that led to (98). This concludes the proof of the first line of (91).
3. Now consider $\hat{\sigma}$ with exactly one coordinate in $\{\mathtt{b}\}$, meaning the other must be in $\{\mathtt{y}\}$. Recalling convention (69), we assume without loss that $\hat{\sigma} \in \{\mathtt{b}_1\mathtt{y}\}$ and proceed to bound $\Delta\hat{m}^\lambda\hat{p}(\hat{\sigma})$. Let $\underline{\dot{\sigma}} \in (\dot{\Omega}^2)^{k-1}$ be compatible with $\hat{\sigma}$. There are two cases:

   a. If $\underline{\dot{\sigma}}$ has no entry in $\{\mathtt{r}\}$, it must also be compatible with some $\hat{\sigma}' \in \{\mathtt{yy}\}$, as long as we permit the possibility that $|(\hat{\sigma}')^1| > T$. Such $\underline{\dot{\sigma}}$ gives the same contribution to $\hat{m}^\lambda\hat{p}(\hat{\sigma})$ as to $\hat{m}^\lambda\hat{p}_\infty(\mathtt{yy})$. It follows from the preceding estimates

that the contribution to $\Delta\hat{m}^\lambda\hat{p}(\mathtt{b}_1\mathtt{y})$ from all such $\underline{\dot{\sigma}}$ is upper bounded by

$$\Delta\hat{m}^\lambda\hat{p}_\infty(\mathtt{yy}) = O(k/2^k)\Delta\dot{p}(\mathtt{gg}) \tag{99}$$

b. The only remaining possibility is that some permutation of $\underline{\dot{\sigma}}$ belongs to $A \times B^{k-2}$ for $A = \{\mathtt{r}_0\mathtt{g}\}$ and $B = \{\mathtt{b}_1\mathtt{g}\}$: the contribution to $\Delta\hat{m}^\lambda\hat{p}(\mathtt{b}_1\mathtt{y})$ from all such $\underline{\dot{\sigma}}$ is

$$\leq (k-1)\sum_{\underline{\dot{\sigma}}\in A\times B^{k-2}}\left|\prod_{j=2}^{k}\dot{p}_1(\dot{\sigma}_j) - \prod_{j=2}^{k}\dot{p}_2(\dot{\sigma}_j)\right| = O(k^2/2^k)\|\Delta\dot{p}\|, \tag{100}$$

where the last estimate follows using (76) and (95).

Combining the above estimates (and using the symmetry between $\mathtt{b}_1\mathtt{y}$ and $\mathtt{y}\mathtt{b}_1$) gives

$$\Delta\hat{m}^\lambda\hat{p}(\mathtt{b}_1\mathtt{y}) + \Delta\hat{m}^\lambda\hat{p}(\mathtt{y}\mathtt{b}_1) = O(k^2/2^k)\|\Delta\dot{p}\|\,. \tag{101}$$

If we further have $\hat{\sigma} \in \{\mathtt{b}_1\}\times\{\mathtt{r}, \mathtt{f}_{\geq 1}\}$, then, arguing as above, $\underline{\dot{\sigma}}$ either contributes to $\Delta\hat{m}^\lambda\hat{p}_\infty(\mathtt{y}\times\{\mathtt{r}, \mathtt{f}_{\geq 1}\})$, or else belongs to $A_{\boldsymbol{x}} \times (B_{\boldsymbol{x}})^{k-2}$ for $A_{\boldsymbol{x}} = \{\mathtt{r}_0\mathtt{g}_{\boldsymbol{x}}\}$, $B_{\boldsymbol{x}} = \{\mathtt{b}_1\mathtt{g}_{\boldsymbol{x}}\}$ and $\boldsymbol{x} \in \{0, 1\}$. The contribution from first case is bounded by (98). The contribution from the second case, using (76) and (95), is

$$\lesssim k\Delta\dot{p}(\dot{\Omega}^2\backslash\{\mathtt{rr}\})\Big(\max_{\boldsymbol{x}\in\{0,1\}}\dot{p}_1(B_{\boldsymbol{x}}) + \Delta\dot{p}(\mathtt{gg})\Big)^{k-2} = O(k^2/4^k)\Delta\dot{p}(\dot{\Omega}^2\backslash\{\mathtt{rr}\})\,.$$

The second claim of (91) follows by combining these estimates and recalling (96).

c. Lastly, consider $\hat{\sigma} \in \{\mathtt{bb}\}$. Without loss of generality, we take $\hat{\sigma} = \mathtt{b}_1\mathtt{b}_1$ and proceed to bound $\Delta\hat{m}^\lambda\hat{p}(\mathtt{b}_1\mathtt{b}_1)$. Let $\underline{\dot{\sigma}} \in (\dot{\Omega}^2)^{k-1}$ be compatible with $\hat{\sigma}$. We distinguish three cases:

a. For at least one $i \in \{1, 2\}$, $\dot{\sigma}^i$ contains no entry in $\{\mathtt{r}\}$. In this case $\underline{\dot{\sigma}}$ is also compatible with some $\hat{\sigma}' \in \{\mathtt{b}_1\mathtt{y}\}\cup\{\mathtt{y}\mathtt{b}_1\}$, as long as we permit the possibility that $|(\hat{\sigma}')^i| > T$. The contribution of all such $\underline{\dot{\sigma}}$ to $\Delta\hat{m}^\lambda\hat{p}(\mathtt{b}_1\mathtt{b}_1)$ is therefore upper bounded by

$$\Delta\hat{m}^\lambda\hat{p}_\infty(\mathtt{b}_1\mathtt{y}) + \Delta\hat{m}^\lambda\hat{p}_\infty(\mathtt{y}\mathtt{b}_1) = O(k^2/2^k)\|\Delta\dot{p}\|, \tag{102}$$

where the last step is by the same argument as for (101).

b. The next case is that $\underline{\dot{\sigma}}$ is a permutation of $(\mathtt{r}_0\mathtt{r}_0, (\mathtt{b}_1\mathtt{b}_1)^{k-2})$. The contribution to $\Delta\hat{m}^\lambda\hat{p}(\mathtt{b}_1\mathtt{b}_1)$ from this case is at most

$$(k-1)\left|\dot{p}_1(\mathtt{r}_0\mathtt{r}_0)\dot{p}_1(\mathtt{b}_1\mathtt{b}_1)^{k-2} - \dot{p}_2(\mathtt{r}_0\mathtt{r}_0)\dot{p}_2(\mathtt{b}_1\mathtt{b}_1)^{k-2}\right|.$$

Using (76) and the definition of $\mathbf{\Gamma}(c,\kappa)$, this is at most

$$O(k^2/4^k)\Big(\Delta\dot{p}(\mathtt{r}_0\mathtt{r}_0)+\dot{p}(\mathtt{r}_0\mathtt{r}_0)\cdot\Delta\dot{p}(\mathtt{b}_1\mathtt{b}_1)\Big)$$
$$=O(k^2/4^k)\|\dot{p}\|\|\Delta\dot{p}\|=O(k^2/2^{(1+\kappa)k})\|\Delta\dot{p}\|. \tag{103}$$

c. The last case is that $\underline{\dot{\sigma}}$ is a permutation of $(\mathtt{r}_0\mathtt{b}_1,\mathtt{b}_1\mathtt{r}_0,(\mathtt{b}_1\mathtt{b}_1)^{k-3})$. The contribution to $\Delta\hat{m}^\lambda\hat{p}(\mathtt{b}_1\mathtt{b}_1)$ from this case is at most

$$k^2\Big|\dot{p}_1(\mathtt{r}_0\mathtt{b}_1)\dot{p}_1(\mathtt{b}_1\mathtt{r}_0)\dot{p}_1(\mathtt{b}_1\mathtt{b}_1)^{k-3}-\dot{p}_2(\mathtt{r}_0\mathtt{b}_1)\dot{p}_2(\mathtt{b}_1\mathtt{r}_0)\dot{p}_2(\mathtt{b}_1\mathtt{b}_1)^{k-3}\Big|.$$

This is at most $O(k^2/4^k)\|\Delta\dot{p}\|$ by another application of (76) and the definition of $\mathbf{\Gamma}(c,\kappa)$.

The above estimates together give

$$\Delta\hat{m}^\lambda\hat{p}(\mathtt{b}_1\mathtt{b}_1)=O(k^2/2^k)\|\Delta\dot{p}\|, \tag{104}$$

where the main contribution comes from (102). Combining with the previous bound (101) yields the last part of (91).

*Step II.* Next we prove (92) by improving the preceding bounds in the special case that $\dot{p}_1=\dot{p}$ and $\dot{p}_2\equiv\mathfrak{f}\dot{p}$. Recall $\hat{p}_i\equiv\hat{\mathrm{NB}}(\dot{p}_i)$; it follows that $\hat{p}_2=\mathfrak{f}\hat{p}_1$. Thus, for any $\hat{\sigma}\in\hat{\Omega}^2$ with $\hat{\sigma}^2=\mathtt{s}$, we have $\hat{\sigma}=\mathfrak{f}\hat{\sigma}$, consequently $\hat{p}_2(\hat{\sigma})=\hat{p}_1(\mathfrak{f}\hat{\sigma})=\hat{p}_1(\hat{\sigma})$. For $\hat{\sigma}\in\hat{\Omega}^2$ with $\hat{\sigma}^1=\mathtt{s}$, we have $\hat{\sigma}=(\mathfrak{f}\hat{\sigma})\oplus\mathtt{1}$, so $\hat{p}_2(\hat{\sigma})=\hat{p}_1(\mathfrak{f}\hat{\sigma})=\hat{p}_1(\hat{\sigma})$, where the last step uses that $\hat{p}_1=(\hat{p}_1)^{\mathrm{av}}$. It follows that instead of (97) and (99) we have the improved bound

$$\Delta\hat{m}^\lambda\hat{p}_\infty(\mathtt{yy})=\Delta\hat{m}^\lambda\hat{p}_\infty(\{\mathtt{yy}\}\backslash(\{\mathtt{sy}\}\cup\{\mathtt{ys}\}))\le\sum_{\boldsymbol{x},\boldsymbol{y}\in\{0,1\}}\Delta\hat{m}^\lambda\hat{p}_\infty(\mathtt{a}_{\boldsymbol{x}}\times\mathtt{a}_{\boldsymbol{y}})$$
$$=O(k)\|\Delta\dot{p}\|\sum_{\boldsymbol{x},\boldsymbol{y}\in\{0,1\}}\Big(\dot{p}_1(\mathtt{g}_{\boldsymbol{x}},\mathtt{g}_{\boldsymbol{y}})+\Delta\dot{p}(\mathtt{gg})\Big)^{k-2}=O(k/4^k)\|\dot{p}-\mathfrak{f}\dot{p}\|.$$

Similarly, instead of (100) we would only have a contribution from $\underline{\dot{\sigma}}$ belonging to either $A_0\times(B_0)^{k-2}$ or $A_1\times(B_1)^{k-2}$, where $A_{\boldsymbol{x}}=\{\mathtt{r}_0\mathtt{g}_{\boldsymbol{x}}\}$ and $B_{\boldsymbol{x}}=\{\mathtt{b}_1\mathtt{g}_{\boldsymbol{x}}\}$. It follows that instead of (101) and (102) we have the improved bound

$$\Delta\hat{m}^\lambda\hat{p}_\infty(\mathtt{b}_1\mathtt{y})+\Delta\hat{m}^\lambda\hat{p}_\infty(\mathtt{y}\mathtt{b}_1)=O(k^4/4^k)\|\Delta\dot{p}\|\,.$$

Previously the main contribution in (104) came from (102), but now it comes instead from (103). This gives the improved bound $\Delta\hat{m}^\lambda\hat{p}(\mathtt{b}_1\mathtt{b}_1)=O(k^2/2^{(1+\kappa)k})$, which proves the first part of (92). The second part follows by applying Lemma A.7. □

***Proof of Lemma A.9*** We first prove (93). Assume $s, t \in \{\mathtt{b}_1, \mathtt{f}, \mathtt{s}\}$, and write $st \equiv s \times t \subseteq \hat{\Omega}^2$. Then for a lower bound we have

$$\hat{m}^\lambda \hat{p}_i(st) \geq [1 - O(k/2^k)]\dot{p}_i(\mathtt{bb})^{k-1} = 1 - O(k/2^k)\,.$$

for an upper bound we have

$$\begin{aligned}\hat{m}^\lambda \hat{p}_i(st) &\leq \dot{p}_i(\mathtt{gg})^{k-1} + k\dot{p}_i(\mathtt{r}_0\mathtt{g})\dot{p}_i(\mathtt{b}_1\mathtt{g})^{k-2} + k\dot{p}_i(\mathtt{gr}_0)\dot{p}_i(\mathtt{gb}_1)^{k-2} \\ &\quad + k\dot{p}_i(\mathtt{r}_0\mathtt{r}_0)\dot{p}_i(\mathtt{b}_1\mathtt{b}_1)^{k-2} + k^2\dot{p}_i(\mathtt{r}_0\mathtt{b}_1)\dot{p}_i(\mathtt{b}_1\mathtt{r}_0)\dot{p}_i(\mathtt{b}_1\mathtt{b}_1)^{k-3} \\ &= 1 + O(k^2/2^k).\end{aligned}$$

Writing $\mathtt{r}_1 t \equiv \mathtt{r}_1 \times t$ for $t \in \{\mathtt{b}_1, \mathtt{f}, \mathtt{s}\}$, a similar argument gives

$$\begin{aligned}\hat{m}^\lambda \hat{p}_i(\mathtt{r}_1 t) &\geq [1 - O(k/2^k)]\dot{p}_i(\mathtt{b}_0\mathtt{b})^{k-1} = [1 - O(k/2^k)] \cdot (2/2^k)\,, \\ \hat{m}^\lambda \hat{p}_i(\mathtt{r}_1 t) &\leq \dot{p}_i(\mathtt{b}_0\mathtt{g})^{k-1} + k\dot{p}_i(\mathtt{b}_0\mathtt{r}_0)\dot{p}_i(\mathtt{b}_0\mathtt{b}_1)^{k-2} = [1 - O(k/2^k)] \cdot (2/2^k)\,.\end{aligned} \tag{105}$$

Lastly, it is easily seen that

$$\hat{m}\hat{p}_i(\mathtt{r}_1\mathtt{r}_1) = \dot{p}_i(\mathtt{b}_0\mathtt{b}_0)^{k-1} = [1 - O(k/2^{ck})] \cdot (2/2^k)^2\,.$$

This concludes the proof of (93), and we turn next to the proof of (94). Arguing similarly as for (80) gives

$$\hat{m}^\lambda \hat{p}_i(\{\mathtt{ff}\}\backslash\{\mathtt{ss}\}) \leq \hat{m}^\lambda \hat{p}_i(\mathtt{f}_{\geq 1}\mathtt{f}) + \hat{m}^\lambda \hat{p}_i(\mathtt{ff}_{\geq 1}) = O(k/4^k)\,.$$

Next, suppose $\dot{\underline{\sigma}}$ is compatible with $\hat{\sigma} \in \mathtt{b}_1\mathtt{f}_{\geq 1}$: if $\dot{\underline{\sigma}}$ has no entry in $\{\mathtt{r}\}$, then it is also compatible with some $\hat{\sigma}' \in \mathtt{ff}_{\geq 1}$, provided we allow $|(\hat{\sigma}')^1| > T$. Therefore

$$\begin{aligned}&\hat{m}^\lambda \hat{p}_i(\mathtt{b}_1\mathtt{f}_{\geq 1}) - \hat{m}^\lambda \hat{p}_{i,\infty}(\mathtt{ff}_{\geq 1}) \\ &\quad \leq \sum_{\boldsymbol{y}\in\{0,1\}} \left[k\dot{p}_i(\mathtt{r}_0\mathtt{f})\dot{p}_i(\mathtt{b}_1\mathtt{g}_{\boldsymbol{y}})^{k-2} + k^2\dot{p}_i(\mathtt{r}_0\mathtt{b}_{\boldsymbol{y}})\dot{p}_i(\mathtt{b}_1\mathtt{f})\dot{p}_i(\mathtt{b}_1\mathtt{g}_{\boldsymbol{y}})^{k-3}\right],\end{aligned}$$

and by definition of $\boldsymbol{\Gamma}(c, \kappa)$ this is $O(k/4^k)$. Finally,

$$\hat{m}^\lambda \hat{p}_i(\mathtt{r}_1\mathtt{f}_{\geq 1}) \leq \sum_{\boldsymbol{y}\in\{0,1\}} k\dot{p}_i(\mathtt{b}_0\mathtt{f})\dot{p}_i(\mathtt{b}_0\mathtt{g}_{\boldsymbol{y}})^{k-2} = O(k/8^k),$$

which proves the first part of (94). For the second part, arguing as for (87), we have for any $\hat{\eta} \in \hat{\Omega}$ that

$$\hat{m}^\lambda \hat{p}_i(\mathtt{f}\hat{\eta}) - \hat{m}^\lambda \hat{p}_i(\mathtt{b}_1\hat{\eta}) \leq (k-1)\sum_{\dot{\underline{\sigma}}\sim\hat{\eta}} [\dot{p}_i(\mathtt{g}_0\dot{\sigma}_2) - \dot{p}_i(\mathtt{r}_0\dot{\sigma}_2)] \prod_{j=3}^{k} \dot{p}_i(\mathtt{b}_1\dot{\sigma}_j)\,.$$

Note that $\underline{\dot{\sigma}}$ has at most one entry in $\{\mathtt{r}\}$. If $\dot{\sigma}_2 = \mathtt{r}_0$, then $\dot{\sigma}_j = \mathtt{b}_1$ for all $j \geq 3$. Since $\dot{q}_i \in \mathbf{\Gamma}(c,\kappa)$ (which means also that $\dot{q}_i = (\dot{q}_i)^{\mathrm{av}}$), we have

$$\sum_{\underline{\dot{\sigma}} \sim \hat{\eta}} \mathbf{1}\{\dot{\sigma}_2 = \zeta\} \prod_{j=3}^{k} \dot{p}_i(\mathtt{b}_1 \dot{\sigma}_j) \leq \begin{cases} \dot{p}_i(\mathtt{b}_1\mathtt{b}_1)^{k-2} \leq O(4^{-k}) & \text{if } \zeta = \mathtt{r}_0, \\ \dot{p}_i(\mathtt{b}_1\mathtt{g})^{k-3} \leq O(2^{-k}) & \text{if } \zeta \in \dot{\Omega}\backslash\{\mathtt{r}_0\}. \end{cases}$$

On the other hand, $\dot{q}_i \in \mathbf{\Gamma}(c,\kappa)$ also implies

$$\dot{p}_i(\mathtt{g}_0\zeta) - \dot{p}_i(\mathtt{r}_0\zeta) \leq O(2^{-k})\dot{p}_i(\mathtt{b}_0\zeta) + \dot{p}_i(\mathtt{f}\zeta) \leq \begin{cases} O(1) & \text{if } \zeta = \mathtt{r}_0, \\ O(2^{-k}) & \text{if } \zeta \in \dot{\Omega}\backslash\{\mathtt{r}_0\}. \end{cases}$$

Combining these estimates and summing over $\hat{\eta} \in \hat{\Omega}$ proves the second part of (94).□

An immediate application of (93), which will be useful in the next proof, is that

$$\frac{\hat{m}^\lambda \hat{p}_i(\mathtt{r}_{\boldsymbol{x}}\hat{\eta})}{\hat{m}^\lambda \hat{p}_i(\mathtt{b}_{\boldsymbol{x}}\hat{\eta})} \geq [1 + O(k^2/2^k)] \cdot (2/2^k)\,. \tag{106}$$

for all $\hat{\eta} \in \{\mathtt{b}_0, \mathtt{b}_1, \mathtt{f}, \mathtt{s}\}$ and all $\boldsymbol{x} \in \{0,1\}$.

***Proof of Lemma A.10*** We divide the proof in two parts.
*Step I. Non-normalized messages.*

1. First consider $\dot{\sigma} \in \{\mathtt{ff}\}$. Recalling $(a+b)^\lambda \leq a^\lambda + b^\lambda$ for $a, b \geq 0$ and $\lambda \in [0,1]$,

$$\Delta\dot{p}^{\mathrm{u}}(\mathtt{ff}) \leq 2 \sum_{\hat{r} \in \{\hat{p}, \mathtt{f}\hat{p}\}} \sum_{\underline{\hat{\sigma}} \in \{\mathtt{ff}\}^{k-1}} \left| \prod_{j=2}^{d} \hat{m}^\lambda \hat{r}_1(\hat{\sigma}_j) - \prod_{j=2}^{d} \hat{m}^\lambda \hat{r}_2(\hat{\sigma}_j) \right|$$

where the $\hat{r} = \mathtt{f}\hat{p}$ term arises from the fact that

$$\hat{m}(\hat{\sigma}^1)^\lambda [1 - \hat{m}(\hat{\sigma}^2)]^\lambda \hat{p}(\hat{\sigma}) = \hat{m}(\hat{\sigma}^1)^\lambda \hat{m}(\hat{\sigma}^2 \oplus 1)^\lambda (\mathtt{f}\hat{p})(\mathtt{f}\hat{\sigma}) = \hat{m}^\lambda \mathtt{f}\hat{p}(\mathtt{f}\hat{\sigma})\,.$$

Applying (77) gives

$$\Delta\dot{p}^{\mathrm{u}}(\mathtt{ff}) = O(d) \sum_{\hat{r} \in \{\hat{p}, \mathtt{f}\hat{p}\}} \Delta\hat{m}^\lambda \hat{r}(\mathtt{ff}) \Big( \hat{m}^\lambda \hat{r}_1(\mathtt{ff}) + \Delta\hat{m}^\lambda \hat{r}(\mathtt{ff}) \Big)^{d-2}.$$

We have from (91) and (93) that $\hat{m}^\lambda \hat{p}_1(\mathtt{ff}) \asymp 1$ and $\Delta\hat{m}^\lambda \hat{p}(\mathtt{ff}) = O(k^3/2^{(1+c)k})$, so

$$\Delta\dot{p}^{\mathrm{u}}(\mathtt{ff}) = O(d) \| \Delta\hat{m}^\lambda \hat{p} + \Delta\hat{m}^\lambda \mathtt{f}\hat{p} \| \cdot \dot{p}_1^{\mathrm{u}}(\mathtt{ff})\,. \tag{107}$$

2. Next consider $\dot{\sigma} \in \{\mathtt{p}_1\mathtt{f}\}$. Let $\hat{r}_{\max}(\hat{\sigma}) \equiv \max_{i=1,2} \hat{r}_i(\hat{\sigma})$—in this notation,

$$\hat{r}_{\max}(\hat{\Omega}) = \sum_{\hat{\sigma} \in \hat{\Omega}} \max_{i=1,2} \hat{r}_i(\hat{\sigma}) \geq \max_{i=1,2} \sum_{\hat{\sigma} \in \hat{\Omega}} \hat{r}_i(\hat{\sigma}) = \max_{i=1,2} \hat{r}_i(\hat{\Omega})$$

where the inequality may be strict. Then

$$\Delta \dot{p}^{\mathrm{u}}(\mathtt{p}_1\mathtt{f}) = O(d) \sum_{\hat{r}\in\{\hat{p},\mathtt{f}\hat{p}\}} \Delta\hat{m}^{\lambda}\hat{r}(\mathtt{p}_1\mathtt{f})[\hat{m}^{\lambda}\hat{r}_{\max}(\mathtt{p}_1\mathtt{f})]^{d-2}.$$

Let $a \in \arg\max_i \hat{r}_i(\mathtt{b}_1\mathtt{s})$, so that

$$0 \le \hat{m}^{\lambda}\hat{r}_{\max}(\mathtt{p}_1\mathtt{f}) - \hat{m}^{\lambda}\hat{r}_a(\mathtt{p}_1\mathtt{f}) \le \Delta\hat{m}^{\lambda}\hat{r}(\mathtt{r}_1\mathtt{f}) + \Delta\hat{m}^{\lambda}\hat{r}(\mathtt{b}_1\mathtt{f}_{\ge 1}) = O(2^{-(1+c)k}),$$

where the last estimate is by (91) and (94). We also have from (93) that $\hat{m}^{\lambda}\hat{p}(\mathtt{p}_1\mathtt{f}) \ge \hat{m}^{\lambda}\hat{p}(\mathtt{b}_1\mathtt{f}) \asymp 1$, and it follows that

$$[\hat{m}^{\lambda}\hat{r}_{\max}(\mathtt{p}_1\mathtt{f})]^{d-2} \asymp [\hat{m}^{\lambda}\hat{r}_a(\mathtt{p}_1\mathtt{f})]^{d-1}. \tag{108}$$

Applying (93) and (94) again, we have (for $i = 1, 2$)

$$[\hat{m}^{\lambda}\hat{r}_i(\mathtt{p}_1\mathtt{f})]^{d-1} \asymp [\hat{m}^{\lambda}\hat{r}_i(\mathtt{p}_1\mathtt{s})]^{d-1}.$$

On the other hand, assuming $T \ge 1$, we have

$$\dot{p}_i^{\mathrm{u}}(\mathtt{r}_1\mathtt{f}) \ge [\hat{m}^{\lambda}\hat{r}_i(\mathtt{p}_1\mathtt{s})]^{d-1} - [\hat{m}^{\lambda}\hat{r}_i(\mathtt{b}_1\mathtt{s})]^{d-1} \asymp [\hat{m}^{\lambda}\hat{r}_i(\mathtt{p}_1\mathtt{s})]^{d-1}$$

where the last step follows by (106). Similarly,

$$\begin{aligned}\dot{p}_i^{\mathrm{u}}(\mathtt{r}_1\mathtt{f}) - \dot{p}_i^{\mathrm{u}}(\mathtt{b}_1\mathtt{f}) &= O(1) \sum_{\hat{r}\in\{\hat{p},\mathtt{f}\hat{p}\}} \hat{m}^{\lambda}\hat{r}_i(\mathtt{b}_1\mathtt{f})^{d-1} = O(2^{-k}) \sum_{\hat{r}\in\{\hat{p},\mathtt{f}\hat{p}\}} \hat{m}^{\lambda}\hat{r}_i(\mathtt{p}_1\mathtt{f})^{d-1} \\ &= O(2^{-k})\dot{p}_i^{\mathrm{u}}(\mathtt{r}_1\mathtt{f}) = O(2^{-k})\dot{p}_i^{\mathrm{u}}(\mathtt{b}_1\mathtt{f}),\end{aligned} \tag{109}$$

where the last step follows by rearranging the terms. Combining the above gives

$$\Delta \dot{p}^{\mathrm{u}}(\mathtt{p}_1\mathtt{f}) \le O(d)\|\Delta\hat{m}^{\lambda}\hat{p} + \Delta\hat{m}^{\lambda}\mathtt{f}\hat{p}\| \max_{i=1,2} \dot{p}_i^{\mathrm{u}}(\mathtt{b}_1\mathtt{f}). \tag{110}$$

Clearly, similar bounds hold if we replace $\mathtt{p}_1\mathtt{f}$ with any of $\mathtt{p}_0\mathtt{f}$, $\mathtt{f}\mathtt{p}_1$, or $\mathtt{f}\mathtt{p}_0$.

3. Lastly we bound $\Delta\dot{p}^{\mathrm{u}}(\mathtt{p}_{\boldsymbol{x}}\mathtt{p}_{\boldsymbol{x}})$ for $\boldsymbol{x}, \boldsymbol{y} \in \{0, 1\}$. As in the single-copy recursion, we denote

$$\begin{aligned}\dot{r}(\mathtt{X}_{\boldsymbol{x}}\dot{\sigma}) &\equiv \dot{r}(\mathtt{r}_{\boldsymbol{x}}\dot{\sigma}) - \dot{r}(\mathtt{b}_{\boldsymbol{x}}\dot{\sigma}),\\ \dot{r}(\dot{\sigma}\mathtt{X}_{\boldsymbol{x}}) &\equiv \dot{r}(\dot{\sigma}\mathtt{r}_{\boldsymbol{x}}) - \dot{r}(\dot{\sigma}\mathtt{b}_{\boldsymbol{x}}),\\ \dot{r}(\mathtt{X}_{\boldsymbol{x}}\mathtt{X}_{\boldsymbol{y}}) &\equiv \dot{r}(\mathtt{r}_{\boldsymbol{x}}\mathtt{r}_{\boldsymbol{y}}) - \dot{r}(\mathtt{r}_{\boldsymbol{x}}\mathtt{b}_{\boldsymbol{y}}) - \dot{r}(\mathtt{b}_{\boldsymbol{x}}\mathtt{r}_{\boldsymbol{y}}) + \dot{r}(\mathtt{b}_{\boldsymbol{x}}\mathtt{b}_{\boldsymbol{y}}).\end{aligned}$$

Applying (106) gives

$$\begin{aligned}\dot{p}_i^{\mathrm{u}}(\mathtt{X}_{\boldsymbol{x}}\mathtt{r}_{\boldsymbol{y}}) &= [\hat{p}_i(\mathtt{b}_{\boldsymbol{x}}\mathtt{p}_{\boldsymbol{y}})]^{d-1} = O(2^{-k})[\hat{p}_i(\mathtt{p}_{\boldsymbol{x}}\mathtt{p}_{\boldsymbol{y}})]^{d-1} = O(2^{-k})\dot{p}_i^{\mathrm{u}}(\mathtt{r}_{\boldsymbol{x}}\mathtt{r}_{\boldsymbol{y}}),\\ \dot{p}_i^{\mathrm{u}}(\mathtt{X}_{\boldsymbol{x}}\mathtt{X}_{\boldsymbol{y}}) &= [\hat{p}_i(\mathtt{b}_{\boldsymbol{x}}\mathtt{b}_{\boldsymbol{y}})]^{d-1} = O(2^{-k})\dot{p}_i^{\mathrm{u}}(\mathtt{r}_{\boldsymbol{x}}\mathtt{r}_{\boldsymbol{y}}).\end{aligned}$$

Combining the above estimates gives

$$\dot{p}_i^{\mathrm{u}}(\mathtt{r}_{\boldsymbol{x}}\mathtt{r}_{\boldsymbol{y}}) - \dot{p}_i^{\mathrm{u}}(\mathtt{b}_{\boldsymbol{x}}\mathtt{b}_{\boldsymbol{y}}) = \dot{p}_i^{\mathrm{u}}(\mathtt{X}_{\boldsymbol{x}}\mathtt{r}_{\boldsymbol{y}}) + \dot{p}_i^{\mathrm{u}}(\mathtt{r}_{\boldsymbol{x}}\mathtt{X}_{\boldsymbol{y}}) - \dot{p}_i^{\mathrm{u}}(\mathtt{X}_{\boldsymbol{x}}\mathtt{X}_{\boldsymbol{y}}) = O(2^{-k})\dot{p}_i^{\mathrm{u}}(\mathtt{r}_{\boldsymbol{x}}\mathtt{r}_{\boldsymbol{y}})\,.$$

Further, it follows from the BP equations that

$$\begin{aligned}&\max\{\dot{p}_i^{\mathrm{u}}(\mathtt{r}_{\boldsymbol{x}}\mathtt{X}_{\boldsymbol{y}}), \dot{p}_i^{\mathrm{u}}(\mathtt{b}_{\boldsymbol{x}}\mathtt{X}_{\boldsymbol{y}}), \dot{p}_i^{\mathrm{u}}(\mathtt{X}_{\boldsymbol{x}}\mathtt{r}_{\boldsymbol{y}}), \dot{p}_i^{\mathrm{u}}(\mathtt{X}_{\boldsymbol{x}}\mathtt{b}_{\boldsymbol{y}})\} \le \dot{p}_i^{\mathrm{u}}(\mathtt{r}_{\boldsymbol{x}}\mathtt{r}_{\boldsymbol{y}}) - \dot{p}_i^{\mathrm{u}}(\mathtt{b}_{\boldsymbol{x}}\mathtt{b}_{\boldsymbol{y}}),\\ &\text{so } \dot{p}_i^{\mathrm{u}}(st) = [1 + O(2^{-k})]\dot{p}_i^{\mathrm{u}}(\mathtt{b}_{\boldsymbol{x}}\mathtt{b}_{\boldsymbol{y}}) \text{ for all } s \in \{\mathtt{r}_{\boldsymbol{x}}, \mathtt{b}_{\boldsymbol{x}}\}, t \in \{\mathtt{r}_{\boldsymbol{y}}, \mathtt{b}_{\boldsymbol{y}}\}.\end{aligned} \tag{111}$$

Similarly, we can upper bound

$$\begin{aligned}\Delta\dot{p}^{\mathrm{u}}(\mathtt{p}_{\boldsymbol{x}}\mathtt{p}_{\boldsymbol{y}}) &\le 4[\Delta\dot{p}^{\mathrm{u}}(\mathtt{r}_{\boldsymbol{x}}\mathtt{r}_{\boldsymbol{y}}) + \Delta\dot{p}^{\mathrm{u}}(\mathtt{X}_{\boldsymbol{x}}\mathtt{r}_{\boldsymbol{y}}) + \Delta\dot{p}^{\mathrm{u}}(\mathtt{r}_{\boldsymbol{x}}\mathtt{X}_{\boldsymbol{y}}) + \Delta\dot{p}^{\mathrm{u}}(\mathtt{X}_{\boldsymbol{x}}\mathtt{X}_{\boldsymbol{y}})].\\ &\le O(d) \sum_{\hat{r}\in\{\hat{p},\mathfrak{f}\hat{p}\}} \sum_{\substack{s\in\{\mathtt{p}_{\boldsymbol{x}},\mathtt{b}_{\boldsymbol{x}}\}\\ t\in\{\mathtt{p}_{\boldsymbol{y}},\mathtt{b}_{\boldsymbol{y}}\}}} \|\Delta\hat{m}^{\lambda}\hat{r}\|[\hat{m}^{\lambda}\hat{r}_{\max}(st)]^{d-2}.\end{aligned} \tag{112}$$

For $\hat{r} \in \{\hat{p}, \mathfrak{f}\hat{p}\}$, let $a = \arg\max_{i=1,2} \hat{m}^{\lambda}\hat{r}_i(\mathtt{b}_1\mathtt{b}_1)$: then, for any $s \in \{\mathtt{p}_{\boldsymbol{x}}, \mathtt{b}_{\boldsymbol{x}}\}$, $t \in \{\mathtt{p}_{\boldsymbol{y}}, \mathtt{b}_{\boldsymbol{y}}\}$,

$$\begin{aligned}0 &\le \hat{m}^{\lambda}\hat{r}_{\max}(st) - \max_{i=1,2} \hat{m}^{\lambda}\hat{r}_i(st) \le \hat{m}^{\lambda}\hat{r}_{\max}(st) - \hat{m}^{\lambda}\hat{r}_a(st)\\ &\le O(1)\Delta\hat{m}^{\lambda}\hat{r}(\{\mathtt{pp}\}\backslash\{\mathtt{bb}\}) \le O(1/2^{(1+c)k}),\end{aligned}$$

where the last estimate is by (91). Combining with (72) and (111) gives

$$\sum_{\substack{s\in\{\mathtt{p}_{\boldsymbol{x}},\mathtt{b}_{\boldsymbol{x}}\}\\ t\in\{\mathtt{p}_{\boldsymbol{y}},\mathtt{b}_{\boldsymbol{y}}\}}} [\hat{m}^{\lambda}\hat{r}_{\max}(st)]^{d-2} = O(1)\Big[\max_{i=1,2} \hat{r}_i(\mathtt{p}_{\boldsymbol{x}}\mathtt{p}_{\boldsymbol{y}})\Big]^{d-1} = O(1) \max_{i=1,2} \dot{p}_i^{\mathrm{u}}(\mathtt{bb})\,.$$

Substituting into (112) gives

$$\Delta\dot{p}^{\mathrm{u}}(\mathtt{p}_{\boldsymbol{x}}\mathtt{p}_{\boldsymbol{y}}) \le O(d)\|\Delta\hat{m}^{\lambda}\hat{p} + \Delta\hat{m}^{\lambda}\mathfrak{f}\hat{p}\| \max_{i=1,2} \dot{p}_i^{\mathrm{u}}(\mathtt{bb})\,. \tag{113}$$

Further, for any $st \in \{\mathtt{r}_{\boldsymbol{x}}\mathtt{X}_{\boldsymbol{y}}, \mathtt{X}_{\boldsymbol{x}}\mathtt{r}_{\boldsymbol{y}}, \mathtt{X}_{\boldsymbol{x}}\mathtt{X}_{\boldsymbol{y}}\}$, we have

$$\Delta\dot{p}^{\mathrm{u}}(st) \le O(k)\|\Delta\hat{m}^{\lambda}\hat{p} + \Delta\hat{m}^{\lambda}\mathfrak{f}\hat{p}\| \max_{i=1,2} \dot{p}_i^{\mathrm{u}}(\mathtt{bb})\,. \tag{114}$$

Lastly, in the special case $\hat{p}_2 = \mathfrak{f}\hat{p}_1$, (113) reduces to

$$\begin{aligned}|\dot{p}_1^{\mathrm{u}}(\mathtt{b}_0\mathtt{b}_0) - \dot{p}_1^{\mathrm{u}}(\mathtt{b}_0\mathtt{b}_1)| &\le O(d)\|\hat{m}^{\lambda}\hat{p}_1 - \hat{m}^{\lambda}\mathfrak{f}\hat{p}_1\|\dot{p}_1^{\mathrm{u}}(\mathtt{bb})\\ &\le k^5 2^{(1-2\kappa)k}\|\dot{p}_i - \mathfrak{f}\dot{p}_i\|\,.\end{aligned} \tag{115}$$

where the last estimate is by (92).

*Step II. Normalized messages.* Recall $\tilde{q}_i \equiv \mathtt{BP}\dot{q}_i$. It remains to verify that $\tilde{q}_i \in \mathbf{\Gamma}(c', 1)$ with $c' = \max\{0, 2\kappa - 1\}$: recalling the definition of $\mathbf{\Gamma}$, this means

$$|p(\mathtt{b}_0\mathtt{b}_0) - p(\mathtt{b}_0\mathtt{b}_1)| \le (k^9/2^{c'k})p(\mathtt{bb}) \text{ and } p(\mathtt{ff}) + p(\{\mathtt{fr}, \mathtt{rf}\})/2^k + p(\mathtt{rr})/4^k = O(2^{-k})p(\mathtt{bb}); \tag{1$\mathbf{\Gamma}'$}$$

$$p(\mathtt{fr}) = O(2^{-k})p(\mathtt{bb}) \text{ and } p(\mathtt{rr}) = O(1)p(\mathtt{bb}); \tag{2$\mathbf{\Gamma}'$}$$

$$p(\mathtt{r}_{\boldsymbol{x}}\dot{\sigma}) \ge [1 - O(2^{-k})]p(\mathtt{b}_{\boldsymbol{x}}\dot{\sigma}) \text{ for all } \boldsymbol{x} \in \{0, 1\} \text{ and } \dot{\sigma} \in \dot{\Omega}. \tag{3$\mathbf{\Gamma}'$}$$

Condition (3$\mathbf{\Gamma}'$) is automatically satisfied due to the BP equations. The second part of (2$\mathbf{\Gamma}'$) follows from (111). The first part of (1$\mathbf{\Gamma}'$) holds trivially if $c' = 0$, and otherwise follows from (115). We claim that

$$\tilde{q}_i(\{\mathtt{rf}, \mathtt{fr}, \mathtt{ff}\}) = O(2^{-k})\tilde{q}_i(\mathtt{bb})\,. \tag{116}$$

This immediately gives the first part of (2$\mathbf{\Gamma}'$). Further, the BP equations give $\tilde{q}_i(\mathtt{bf}) \le \tilde{q}_i(\mathtt{rf})$ and $\tilde{q}_i(\mathtt{fb}) \le \tilde{q}_i(\mathtt{fr})$, so the second part of (1$\mathbf{\Gamma}'$) also follows. To see that (116) holds, note that the second part of (94) gives

$$\dot{p}_i^{\mathrm{u}}(\mathtt{ff}) \le O(1) \sum_{\hat{r}\in\{\hat{p},\mathtt{f}\hat{p}\}} [\hat{m}^{\lambda}\hat{r}_i(\mathtt{ff})]^{d-1} \le O(1) \sum_{\hat{r}\in\{\hat{p},\mathtt{f}\hat{p}\}} [\hat{m}^{\lambda}\hat{r}_i(\mathtt{b}_1\mathtt{b}_1)]^{d-1},$$
$$\dot{p}_i^{\mathrm{u}}(\mathtt{r}_1\mathtt{f}) \le O(1) \sum_{\hat{r}\in\{\hat{p},\mathtt{f}\hat{p}\}} [\hat{m}^{\lambda}\hat{r}_i(\mathtt{p}_1\mathtt{f})]^{d-1} \le O(1) \sum_{\hat{r}\in\{\hat{p},\mathtt{f}\hat{p}\}} [\hat{m}^{\lambda}\hat{r}_i(\mathtt{p}_1\mathtt{b}_1)]^{d-1}.$$

Combining with (106) gives $\dot{p}_i^{\mathrm{u}}(\{\mathtt{r}_1\mathtt{f}, \mathtt{ff}\}) = O(2^{-k})\dot{p}_i^{\mathrm{u}}(\mathtt{r}_1\mathtt{r}_1)$. Recalling (111) (and making use of symmetry) gives (116). Finally, we conclude the proof of the lemma by bounding the difference $\Delta\tilde{q} \equiv |\tilde{q}_1 - \tilde{q}_2|$. Recalling the definition of $\mathtt{X}_{\boldsymbol{x}}$, we have

$$\Delta\tilde{q}(\mathtt{pp}) \le O(1)\Delta\tilde{q}(\{\mathtt{bb}, \mathtt{rX}, \mathtt{Xr}, \mathtt{XX}\}),$$
$$\Delta\tilde{q}(\dot{\Omega}^2\backslash\{\mathtt{pp}\}) \le O(1)\Delta\tilde{q}(\{\mathtt{bf}, \mathtt{fb}, \mathtt{ff}, \mathtt{fX}, \mathtt{Xf}\}).$$

We next bound $\Delta\tilde{q}(\mathtt{bb})$, which is the sum of $\Delta\tilde{q}(\mathtt{b}_{\boldsymbol{x}}\mathtt{b}_{\boldsymbol{y}})$ over $\boldsymbol{x}, \boldsymbol{y} \in \{0, 1\}$. By symmetry let us take $\boldsymbol{x} = \boldsymbol{y} = 0$. Since $\tilde{q}_i = (\tilde{q}_i)^{\mathrm{av}}$, $\tilde{q}_i(\mathtt{b}_0\mathtt{b}_0) = \frac{1}{4}\tilde{q}_i(\mathtt{bb}) + \frac{1}{2}[\tilde{q}_i(\mathtt{b}_0\mathtt{b}_0) - \tilde{q}_i(\mathtt{b}_0\mathtt{b}_1)]$, so

$$\Delta\tilde{q}(\mathtt{b}_0\mathtt{b}_0) \le \tfrac{1}{4}|\tilde{q}_1(\mathtt{bb}) - \tilde{q}_2(\mathtt{bb})| + \tfrac{1}{2}\sum_{i=1,2} |\tilde{q}_i(\mathtt{b}_0\mathtt{b}_0) - \tilde{q}_i(\mathtt{b}_0\mathtt{b}_1)|\,.$$

Since the $\tilde{q}_i$ are normalized to be probability measures,

$$1 - \tilde{q}_i(\dot{\Omega}^2\backslash\{\mathtt{pp}\}) = \tilde{q}_i(\mathtt{pp}) = 2\tilde{q}_i(\mathtt{rX}) + 2\tilde{q}_i(\mathtt{Xr}) - 3\tilde{q}_i(\mathtt{XX}) + 4\tilde{q}_i(\mathtt{bb}),$$

from which it follows that

$$|\tilde{q}_1(\mathtt{bb}) - \tilde{q}_2(\mathtt{bb})| \lesssim |\tilde{q}_1(\dot{\Omega}^2 \backslash \{\mathtt{pp}\}) - \tilde{q}_2(\dot{\Omega}^2 \backslash \{\mathtt{pp}\})| + \Delta\tilde{q}(\{\mathtt{rX}, \mathtt{Xr}, \mathtt{XX}\}) .$$

Combining the above estimates gives

$$\|\Delta\tilde{q}\| \lesssim \Delta\tilde{q}(\mathtt{A}) + \sum_{i=1,2} |\tilde{q}_i(\mathtt{b}_0\mathtt{b}_0) - \tilde{q}_i(\mathtt{b}_0\mathtt{b}_1)|, \quad \mathtt{A} \equiv \{\mathtt{bf}, \mathtt{fb}, \mathtt{ff}, \mathtt{fX}, \mathtt{Xf}, \mathtt{rX}, \mathtt{Xr}, \mathtt{XX}\} .$$

Write $\dot{Z}_i \equiv \|\dot{p}_i^{\mathrm{u}}\|$. Taking $a \in \{1, 2\}$ and $b = 2 - a$, we find $\|\Delta\tilde{q}\| \le e_1 + e_2 e_3 + e_4$ with

$$\begin{aligned} e_1 &\equiv \frac{\Delta\dot{p}^{\mathrm{u}}(\mathtt{A})}{\dot{Z}_a}, \quad e_2 \equiv \frac{|\dot{Z}_1 - \dot{Z}_2|}{\dot{Z}_a} \\ &\le \frac{\|\Delta\dot{p}^{\mathrm{u}}\|}{\dot{Z}_a}, \quad e_3 \equiv \frac{\dot{p}_b^{\mathrm{u}}(\mathtt{A})}{\dot{Z}_b}, \\ e_4 &\equiv \sum_{i=1,2} \frac{|\dot{p}_i^{\mathrm{u}}(\mathtt{b}_0\mathtt{b}_0) - \dot{p}_i^{\mathrm{u}}(\mathtt{b}_0\mathtt{b}_1)|}{\dot{Z}_i} . \end{aligned}$$

It follows from (107), (110), (114) and (116), and taking $a = \arg\max_i \dot{p}_i^{\mathrm{u}}(\mathtt{bb})$, that

$$e_1 \lesssim \|\Delta\hat{m}^\lambda \hat{p} + \Delta\hat{m}^\lambda \mathfrak{f}\hat{p}\|(d/2^k) \max_{i=1,2} \dot{p}_i^{\mathrm{u}}(\mathtt{bb})/\dot{Z}_a \lesssim k\|\Delta\hat{m}^\lambda \hat{p} + \Delta\hat{m}^\lambda \mathfrak{f}\hat{p}\| .$$

Further, recalling (113) gives

$$e_2 \lesssim k2^k \|\Delta\hat{m}^\lambda \hat{p} + \Delta\hat{m}^\lambda \mathfrak{f}\hat{p}\| .$$

Combining (109), (111), and (116) gives $e_3 = O(2^{-k})$. Finally, (115) gives

$$e_4 \lesssim k2^k \|\hat{m}^\lambda \hat{p}_i - \hat{m}^\lambda \mathfrak{f}\hat{p}_i\| .$$

Combining the pieces together finishes the proof. □

## Appendix B: The 1RSB free energy

### B.1. Equivalence of recursions

In this section, we relate the coloring recursion (64) to the distributional recursion (10), and prove the following:

**Proposition B.1** *Let $\dot{q}_\lambda$ be the fixed point given by* Proposition 5.5a *for parameters $\lambda \in [0, 1]$ and $T = \infty$. Let $H_\lambda \equiv (\dot{H}_\lambda, \hat{H}_\lambda, \bar{H}_\lambda) \in \mathbf{\Delta}$ be the associated triple of measures defined by* Proposition 3.4. *We then have the identity $(\boldsymbol{s}(H_\lambda), \boldsymbol{\Sigma}(H_\lambda), \boldsymbol{F}(H_\lambda)) = (s_\lambda, \Sigma(s_\lambda), \mathfrak{F}(\lambda))$.*

In the course of the proof, we will obtain Proposition 1.2 as a corollary. Throughout the section we take $T = \infty$ unless explicitly indicated otherwise. We begin with some notations. Recall that $\mathscr{P}(\mathcal{X})$ is the space of probability measures on $\mathcal{X}$. Given $\dot{q} \in \mathscr{P}(\dot{\Omega})$, we define two associated measures $\dot{m}^{\lambda}\dot{q}$, $(1-\dot{m})^{\lambda}\dot{q}$ on $\dot{\Omega}$ by

$$(\dot{m}^{\lambda}\dot{q})(\dot{\sigma}) \equiv \dot{m}(\dot{\sigma})^{\lambda}\dot{q}(\dot{\sigma}), \quad ((1-\dot{m})^{\lambda}\dot{q})(\dot{\sigma}) \equiv (1-\dot{m}(\dot{\sigma}))^{\lambda}\dot{q}(\dot{\sigma}),$$

We let $\dot{\pi} \equiv \dot{\pi}(\dot{q})$ be the probability measure on $\dot{\mathscr{M}}\backslash\{\star\}$ given by

$$\dot{\pi}(\dot{\tau}) = \begin{cases} [1-\dot{q}(\mathtt{r})]^{-1}\dot{q}(\dot{\tau}) & \text{if } \dot{\tau} \in \dot{\mathscr{M}}\backslash\{0,1,\star\}, \\ [1-\dot{q}(\mathtt{r})]^{-1}\dot{q}(\mathtt{b}_{\boldsymbol{x}}) & \text{if } \dot{\tau} = \boldsymbol{x} \in \{0,1\}. \end{cases}$$

Recall from §A.1 the mappings $\dot{m} : \dot{\Omega} \to [0,1]$ and $\hat{m} : \hat{\Omega} \to [0,1]$. We then denote the pushforward measure $\dot{u} \equiv \dot{u}(\dot{q}) \equiv \dot{\pi} \circ \dot{m}^{-1}$, so that $\dot{u}$ belongs to the space $\mathscr{P}$ of discrete probability measures on $[0,1]$. Analogously, given $\hat{q} \in \mathscr{P}(\hat{\Omega})$, we define two associated measures $\hat{m}^{\lambda}\hat{q}$, $(1-\hat{m})^{\lambda}\hat{q}$ on $\hat{\Omega}$. We let $\hat{\pi} \equiv \hat{\pi}(\hat{q})$ be the probability measure on $\hat{\mathscr{M}}\backslash\{\star\}$ given by

$$\hat{\pi}(\hat{\tau}) \equiv \begin{cases} [1-\hat{q}(\mathtt{b})]^{-1}\hat{q}(\hat{\tau}) & \text{if } \hat{\tau} \in \hat{\mathscr{M}}\backslash\{0,1,\star\}, \\ [1-\hat{q}(\mathtt{b})]^{-1}\hat{q}(\mathtt{r}_{\boldsymbol{x}}) & \text{if } \hat{\tau} = \boldsymbol{x} \in \{0,1\}, \end{cases}$$

and we then denote $\hat{u} \equiv \hat{u}(\hat{q}) \equiv \hat{\pi} \circ \hat{m}^{-1}$, so that $\hat{u} \in \mathscr{P}$ also. The next two lemmas follow straightforwardly from the above definitions, and we omit their proofs:

**Lemma B.2** *Suppose $\dot{q} \in \mathscr{P}(\dot{\Omega})$ satisfies $\dot{q} = \dot{q}^{av}$ and*

$$\dot{m}^{\lambda}\dot{q}(\mathtt{f}) = \dot{q}(\mathtt{r}_1) - \dot{q}(\mathtt{b}_1) = \dot{q}(\mathtt{r}_0) - \dot{q}(\mathtt{b}_0) = (1-\dot{m})^{\lambda}\dot{q}(\mathtt{f}) \tag{117}$$

*Then $\hat{q} \equiv \hat{BP}\dot{q} \in \mathscr{P}(\hat{\Omega})$ must satisfy $\hat{q} = \hat{q}^{av}$ and*

$$\hat{m}^{\lambda}\hat{q}(\mathtt{f}) = \hat{q}(\mathtt{b}_1) = \hat{q}(\mathtt{b}_0) = (1-\hat{m})^{\lambda}\hat{q}(\mathtt{f}), \tag{118}$$

*Let $\hat{\mathtt{z}} \equiv (\hat{N}\hat{B}\dot{q})/(\hat{B}\hat{P}\dot{q})$ be the normalizing constant. Then $\dot{u} \equiv \dot{u}(\dot{q})$ and $\hat{u} \equiv \hat{u}(\hat{q})$ satisfy*

$$\hat{u} = \hat{\mathscr{R}}_{\lambda}(\dot{u}), \quad \hat{\mathscr{L}}_{\lambda}(\dot{u}) = \frac{\hat{\mathtt{z}}(1-\hat{q}(\mathtt{b}))}{(1-\dot{q}(\mathtt{r}))^{k-1}}. \tag{119}$$

**Lemma B.3** *Suppose $\hat{q} \in \mathscr{P}(\hat{\Omega})$ satisfies $\hat{q} = \hat{q}^{av}$ and (118). Then $\dot{q} \equiv \dot{BP}\hat{q} \in \mathscr{P}(\dot{\Omega})$ must satisfy $\dot{q} = \dot{q}^{av}$ and (117). Let $\dot{\mathtt{z}} \equiv (\dot{N}\dot{B}\hat{q})/(\dot{B}\dot{P}\hat{q})$ be the normalizing constant: then*

$$\dot{u} = \dot{\mathscr{R}}_{\lambda}(\hat{u}), \quad \dot{\mathscr{L}}_{\lambda}(\hat{u}) = \frac{\dot{\mathtt{z}}(1-\dot{q}(\mathtt{r}))}{(1-\hat{q}(\mathtt{b}))^{d-1}}. \tag{120}$$

***Proof of Proposition 1.2*** This is simply a rephrasing of the proof of Proposition 5.5a, using Lemma B.2 and Lemma B.3. □

We next prove Proposition B.1. In the remainder of this section, fix $\lambda \in [0,1]$ and $T = \infty$. Let $\dot{q} \equiv \dot{q}_\lambda$ be the fixed point of $\mathtt{BP} \equiv \mathtt{BP}_{\lambda,\infty}$ given by Proposition 5.5a. Let $\hat{q} \equiv \hat{q}_\lambda$ denote the image of $\dot{q}$ under the mapping $\hat{\mathtt{BP}} \equiv \hat{\mathtt{BP}}_{\lambda,\infty}$. Denote the associated normalizing constants

$$\hat{z} \equiv \hat{z}_\lambda \equiv (\hat{\mathtt{NB}}\dot{q})/(\hat{\mathtt{BP}}\dot{q}), \quad \dot{z} \equiv \dot{z}_\lambda \equiv (\dot{\mathtt{NB}}\hat{q})/(\dot{\mathtt{BP}}\hat{q})\,.$$

Let $H_\lambda \equiv (\dot{H}_\lambda, \hat{H}_\lambda, \bar{H}_\lambda)$ be the triple of associated measures defined as in Proposition 3.4, with normalizing constants $(\dot{\mathfrak{z}}_\lambda, \hat{\mathfrak{z}}_\lambda, \bar{\mathfrak{z}}_\lambda)$. Recall from (12) that $\mathfrak{F}(\lambda) = \ln \dot{\mathfrak{Z}}_\lambda + \alpha \ln \hat{\mathfrak{Z}}_\lambda - d \ln \bar{\mathfrak{Z}}_\lambda$. We now show that it coincides with $\boldsymbol{F}(H_\lambda)$:

**Lemma B.4** *Under the above notations,* $\boldsymbol{F}(H_\lambda) = \ln \dot{\mathfrak{z}}_\lambda + \alpha \ln \hat{\mathfrak{z}}_\lambda - d \ln \bar{\mathfrak{z}}_\lambda$, *and*

$$\bar{\mathfrak{Z}}_\lambda = \frac{\bar{\mathfrak{z}}_\lambda}{(1-\dot{q}_\lambda(r))(1-\hat{q}_\lambda(b))}, \quad \dot{\mathfrak{Z}}_\lambda = \frac{\dot{\mathfrak{z}}_\lambda}{(1-\hat{q}_\lambda(b))^d}, \quad \hat{\mathfrak{Z}}_\lambda = \frac{\hat{\mathfrak{z}}_\lambda}{(1-\dot{q}_\lambda(r))^k}. \tag{121}$$

*Consequently* $\mathfrak{F}(\lambda) = \boldsymbol{F}(H_\lambda)$.

***Proof*** It follows from the definition (43) (and recalling from Corollary 2.18 that $\hat{\Phi}(\underline{\sigma})^\lambda = \hat{F}(\underline{\sigma})^\lambda \hat{v}(\underline{\sigma})$) that

$$\boldsymbol{F}(H_\lambda) = \langle \ln(\dot{\Phi}^\lambda/\dot{H}), \dot{H}_\lambda\rangle + \alpha\langle \ln(\hat{\Phi}^\lambda/\hat{H}_\lambda), \hat{H}_\lambda\rangle + d\langle \ln(\bar{\Phi}^\lambda \bar{H}_\lambda), \bar{H}_\lambda\rangle\,.$$

Substituting in Definition 5.6 and rearranging gives

$$\begin{aligned}\boldsymbol{F}(H_\lambda) - \Big(\ln \dot{\mathfrak{z}}_\lambda + \alpha \ln \hat{\mathfrak{z}}_\lambda - d \ln \bar{\mathfrak{z}}_\lambda\Big) \\ = -\Big\langle \sum_{i=1}^{d} \ln \hat{q}_\lambda(\hat{\sigma}_i), \dot{H}_\lambda\Big\rangle - \alpha\Big\langle \sum_{i=1}^{k} \ln \dot{q}_\lambda(\dot{\sigma}_i), \hat{H}_\lambda\Big\rangle + d\langle \ln[\dot{q}_\lambda(\dot{\sigma})\hat{q}_\lambda(\hat{\sigma})], \bar{H}_\lambda\rangle.\end{aligned}$$

This equals zero since $H_\lambda \in \boldsymbol{\Delta}$. The proof of (121) is straightforward from the preceding definitions, and is omitted. □

***Proof of Proposition B.1*** By similar calculations as above, it is straightforward to verify that $s_\lambda = \boldsymbol{s}(H_\lambda)$. Since by definition $\mathfrak{F}(\lambda) = \lambda s_\lambda + \Sigma(s_\lambda)$ and $\boldsymbol{F}(H_\lambda) = \lambda \boldsymbol{s}(H_\lambda) + \boldsymbol{\Sigma}(H_\lambda)$, it follows that $\Sigma(s_\lambda) = \boldsymbol{\Sigma}(H_\lambda)$, concluding the proof. □

***Proof of Proposition 3.13*** Immediate consequence of Proposition B.1 together with Proposition 5.5b. □

### B.2. Large-*k* asymptotics

We now evaluate the large-$k$ asymptotics of the free energy, beginning with (12). Let $\dot{\mu}_\lambda$ be the probability measure on $[0,1]$ given by Proposition 1.2, and write $\hat{\mu}_\lambda \equiv \hat{\mathscr{R}}_\lambda(\dot{\mu}_\lambda)$. In what follows it will be useful to denote $\dot{\mu}_\lambda(\mathtt{f}) \equiv \dot{\mu}_\lambda((0,1))$, as well as

$$\psi_\lambda \equiv \int x^\lambda \mathbf{1}\{x \in (0,1)\}\dot{\mu}_\lambda(dx), \quad \rho_\lambda \equiv \int y^\lambda \mathbf{1}\{y \in (0,1)\backslash\{\tfrac{1}{2}\}\}\hat{\mu}_\lambda(dy)\,.$$

**Proposition B.5** *For $k \geq k_0$, $\alpha_{lbd} \leq \alpha = (2^{k-1} - c)\ln 2 \leq \alpha_{ubd}$, and $\lambda \in [0, 1]$,*

$$\ln \dot{\mathfrak{Z}}_\lambda = \ln 2 - (1 - 2^{\lambda-1})/2^k + d \ln \left(2^{-\lambda}\hat{\mu}_\lambda(\tfrac{1}{2}) + \hat{\mu}_\lambda(1) + \rho_\lambda\right) + \mathit{err}, \quad (122)$$

$$-d \ln \bar{\mathfrak{Z}} = -d \ln \left(2^{-\lambda}\hat{\mu}_\lambda(\tfrac{1}{2}) + \hat{\mu}(1) + \rho_\lambda\right) - (k \ln 2)[-\dot{\mu}_\lambda(\pounds) + 2\psi_\lambda] + \mathit{err}, \quad (123)$$

$$\alpha \ln \hat{\mathfrak{Z}} = \alpha \ln(1 - 2/2^k) + (k \ln 2)(-\dot{\mu}_\lambda(\pounds) + 2\psi_\lambda) + \mathit{err}, \quad (124)$$

*where* err *denotes any error bounded by $k^{O(1)}/4^k$. Altogether this yields*

$$\mathfrak{F}(\lambda) = f^{\mathrm{RS}}(\alpha) - (1 - 2^{\lambda-1})/2^k + \mathit{err} = [(2c - 1)\ln 2 - (1 - 2^{\lambda-1})]/2^k + \mathit{err}\,.$$

*On the other hand $\lambda s_\lambda = \lambda(\ln 2)2^{\lambda-1}/2^k + \mathit{err}$.*

***Proof of Proposition 1.4b*** Apply Proposition B.5: setting $\mathfrak{F}(\lambda) = \lambda s_\lambda$ gives

$$\alpha_\lambda = (2^{k-1} - c_\lambda)\ln 2 + \mathrm{err}, \quad c_\lambda = \frac{1}{2} + \frac{1 - 2^{\lambda-1}(1 - \lambda \ln 2)}{2 \ln 2}\,.$$

Substituting the special values $\lambda = 1$ and $\lambda = 0$ gives

$$c_{\mathrm{cond}} = c_1 = 1, \quad c_{\mathrm{sat}} = c_0 = \frac{1}{2} + \frac{1}{4 \ln 2},$$

verifying (1) and (16). □

***Proof of Proposition B.5*** Throughout the proof we abbreviate $\epsilon_k$ for a small error term which may change from one occurrence to the next, but is bounded throughout by $k^C/2^k$ for a sufficiently large absolute constant $C$. Note that

$$\hat{\mu}_\lambda(\tfrac{1}{2}) = 1 - 2 \cdot \frac{2^{1-\lambda}}{2^k} + \epsilon_k, \quad \hat{\mu}_\lambda(1) = \hat{\mu}_\lambda(0) = \frac{2^{1-\lambda}}{2^k} + \epsilon_k, \quad \hat{\mu}_\lambda((0,1)\backslash\{\tfrac{1}{2}\}) = \epsilon_k,$$

from which it follows that $\rho_\lambda = \epsilon_k$. Meanwhile, $\psi_\lambda \leq \dot{\mu}_\lambda(\mathtt{f})$, and we will show below that

$$\dot{\mu}_\lambda(\mathtt{f}) = \frac{2^{\lambda-1}}{2^k} + \epsilon_k\,. \quad (125)$$

□

*Estimate of* $\dot{\mathfrak{Z}}_\lambda$. Recall from the definition (11) that

$$\dot{\mathfrak{Z}}_\lambda = \int \left(\prod_{i=1}^d y_i + \prod_{i=1}^d (1 - y_i)\right)^\lambda \prod_{i=1}^d \hat{\mu}_\lambda(dy_i)\,.$$

Let $\dot{\mathfrak{Z}}_\lambda(\mathtt{f})$ denote the contribution to $\dot{\mathfrak{Z}}_\lambda$ from free variables, meaning $y_i \in (0,1)$ for all $i$. This can be decomposed further into the contribution $\dot{\mathfrak{Z}}_\lambda(\mathtt{f}_1)$ from isolated free variables (meaning $y_i = 1/2$ for all $i$) and the remainder $\dot{\mathfrak{Z}}_\lambda(\mathtt{f}_{\geq 2})$. We then calculate

$$\dot{\mathfrak{Z}}_\lambda(\mathtt{f}_1) = 2^\lambda \Big(2^{-\lambda}\hat{\mu}_\lambda(\tfrac{1}{2})\Big)^d .$$

This dominates the contribution from non-isolated free variables:

$$\begin{aligned}\dot{\mathfrak{Z}}_\lambda(\mathtt{f}_{\geq 2}) &= \sum_{j=1}^{d} \binom{d}{j} \Big( \int y^\lambda \mathbf{1}\{y \in (0,1)\backslash\{\tfrac{1}{2}\}\}\hat{\mu}_\lambda(dy)\Big)^j \Big(2^{-\lambda}\hat{\mu}_\lambda(\tfrac{1}{2})\Big)^{d-j} \\ &\leq O(1) d \hat{\mu}_\lambda((0,1)\backslash\{\tfrac{1}{2}\}) \Big(2^{-\lambda}\hat{\mu}_\lambda(\tfrac{1}{2})\Big)^d \leq \dot{\mathfrak{Z}}_\lambda(\mathtt{f}_1) k^{O(1)}/2^k .\end{aligned}$$

Next let $\dot{\mathfrak{Z}}_\lambda(\mathtt{1})$ denote the contribution from variables frozen to $\mathtt{1}$:

$$\begin{aligned}\dot{\mathfrak{Z}}_\lambda(\mathtt{1}) &= \Big( \int y^\lambda \hat{\mu}_\lambda(dy)\Big)^d - \Big( \int y^\lambda \mathbf{1}\{y \in (0,1)\}\hat{\mu}_\lambda(dy)\Big)^d \\ &= \Big(2^{-\lambda}\hat{\mu}_\lambda(\tfrac{1}{2}) + \hat{\mu}_\lambda(1) + \rho_\lambda\Big)^d - 2^{-\lambda}\dot{\mathfrak{Z}}_\lambda(\mathtt{f}_1) + \epsilon_k .\end{aligned}$$

The ratio of free to frozen variables is given by

$$\frac{\dot{\mathfrak{Z}}_\lambda(\mathtt{f})}{2\dot{\mathfrak{Z}}_\lambda(\mathtt{1})} = \frac{2^\lambda}{2}\left(\frac{\hat{\mu}_\lambda(\frac{1}{2})}{\hat{\mu}_\lambda(\frac{1}{2}) + 2^\lambda \hat{\mu}_\lambda(1)}\right)^d + \epsilon_k = \frac{2^{\lambda-1}}{2^k} + \epsilon_k .$$

Combining these yields (122). The proof of (125) is very similar.

*Estimate of* $\bar{\mathfrak{Z}}_\lambda$. Recall from the definition (11) that

$$\bar{\mathfrak{Z}}_\lambda = \int \Big(xy + (1-x)(1-y)\Big)^\lambda \dot{\mu}_\lambda(dx)\hat{\mu}_\lambda(dy) .$$

The contribution to $\bar{\mathfrak{Z}}$ from $x = 0$ or $x = 1$ is given by

$$\bar{\mathfrak{Z}}_\lambda(x=1) = \dot{\mu}_\lambda(1)\Big(2^{-\lambda}\hat{\mu}_\lambda(\tfrac{1}{2}) + \hat{\mu}_\lambda(1) + \rho_\lambda\Big) = \bar{\mathfrak{Z}}_\lambda(x=0) .$$

The contribution from $x \in (0,1)$ and $y = 1/2$ is given by

$$\bar{\mathfrak{Z}}_\lambda(x \in (0,1), y = 1/2) = \dot{\mu}_\lambda(\mathtt{f}) 2^{-\lambda}\hat{\mu}_\lambda(\tfrac{1}{2}) .$$

Lastly, the contribution from $x \in (0,1)$ and $y = 1$ is given by

$$\bar{\mathfrak{Z}}_\lambda(x \in (0,1), y = 1) = \hat{\mu}_\lambda(1)\psi_\lambda ,$$

and there is an equal contribution from the case $x \in (0,1)$ and $y = 0$. The contribution from the case that both $x, y \in (0,1)$ is $\leq k^{O(1)}/8^k$. Combining these estimates gives

$$\begin{aligned} d\ln \bar{\mathfrak{Z}}_\lambda &= d\ln\left(2^{-\lambda}\hat{\mu}_\lambda(\tfrac{1}{2}) + 2\dot{\mu}_\lambda(1)\hat{\mu}(1) + 2\dot{\mu}_\lambda(1)\rho_\lambda + 2\hat{\mu}(1)\psi_\lambda\right) + \epsilon_k \\ &= d\ln\left(2^{-\lambda}\hat{\mu}_\lambda(\tfrac{1}{2}) + \hat{\mu}(1) + \rho_\lambda\right) + d\ln\left(1 + \frac{\hat{\mu}(1)[-\dot{\mu}_\lambda(\mathtt{f}) + 2\psi_\lambda]}{2^{-\lambda}\hat{\mu}_\lambda(\frac{1}{2})}\right) + \epsilon_k. \end{aligned}$$

Recalling $\hat{\mu}_\lambda = \hat{\mathscr{R}}\dot{\mu}_\lambda$ gives

$$d\ln\left(1 + \frac{\hat{\mu}(1)[-\dot{\mu}_\lambda(\mathtt{f}) + 2\psi_\lambda]}{2^{-\lambda}\hat{\mu}_\lambda(\frac{1}{2})}\right) = d\dot{\mu}_\lambda(0)^{k-1}(-\dot{\mu}_\lambda(\mathtt{f}) + 2\psi_\lambda) + \epsilon_k,$$

and (123) follows.

*Estimate of* $\hat{\mathfrak{Z}}_\lambda$. Recall from the definition (11) that

$$\hat{\mathfrak{Z}}_\lambda = \int \left(1 - \prod_{i=1}^k x_i - \prod_{i=1}^k (1-x_i)\right) \prod_{i=1}^k \dot{\mu}_\lambda(x_i)\,.$$

Writing $\dot{\mu}_\lambda(0,\mathtt{f}) \equiv \dot{\mu}_\lambda([0,1))$, the contribution to $\hat{\mathfrak{Z}}$ from separating clauses is

$$1 - 2\dot{\mu}_\lambda(0,\mathtt{f})^k + \dot{\mu}_\lambda(\mathtt{f})^k = 1 - (2/2^k)(1 + k\dot{\mu}_\lambda(\mathtt{f})) + k^{O(1)}/8^k\,.$$

The contribution from clauses which are forcing to some variable that is not forced by any other clause is $2k\dot{\mu}_\lambda(0)^{k-1}\psi_\lambda$. The contribution from all other clause types is $\leq k^{O(1)}/8^k$, and (124) follows.

*Estimate of* $s_\lambda$. Recall from (13) the definition of $s_\lambda$. By similar considerations as above, it is straightforward to check that the total contribution from frozen variables, edges incident to frozen variables, and separating or forcing clauses is zero. The dominant term is the contribution of isolated free variables, and the estimate follows. □

### B.3. Properties of the complexity function

We conclude with a few basic properties of the complexity function $\Sigma(s)$, including a proof of Proposition 1.4a.

**Lemma B.6** *For fixed* $1 \leq T < \infty$, *the fixed point* $\dot{q}_{\lambda,T}$ *of* Proposition 5.5a *is continuously differentiable as a function of* $\lambda \in [0,1]$.

***Proof*** Fix $T < \infty$ and define $f_T[\dot{q},\lambda] \equiv \mathtt{BP}_{\lambda,T}[\dot{q}] - \dot{q}$ as the mapping from $\mathscr{P}(\dot{\Omega}_T) \times [0,1]$ to the set of signed measures on $\Omega_T$. Since function $\dot{z}(\dot{\sigma})$ ($\hat{z}(\hat{\sigma})$, respectively) can take only finitely many values on $\dot{\Omega}_T$ ($\hat{\Omega}_T$, respectively) and therefore must be uniformly bounded away from 0. It is straightforward to check that for any $\lambda \in [0,1]$,

$$f_T[\dot{q}_\star(\lambda,T),\lambda](\dot{\sigma}) = 0, \quad \forall \dot{\sigma} \in \Omega_T,$$

and is uniformly differentiable in a neighborhood of $\{(\dot{q}_\star(\lambda, T), \lambda) : \lambda \in [0, 1]\}$.

For any other $\dot{q}$ in the contraction region (68), Proposition A.1 guarantees that

$$\begin{aligned}\|f_T[\dot{q}, \lambda] - f_T[\dot{q}_\star(\lambda, T), \lambda]\| &\geq \|\dot{q} - \dot{q}_\star(\lambda)\| - \|\mathtt{BP}_{\lambda,T}[\dot{q}] - \mathtt{BP}_{\lambda,T}[\dot{q}_\star(\lambda, T)]\| \\ &\geq (1 - O(k^2 2^{-k}))\|\dot{q} - \dot{q}_\star(\lambda, T)\|.\end{aligned}$$

Therefore the Jacobian matrix

$$\Big(\frac{\partial f_T(\dot{\sigma}_i)}{\partial \dot{q}(\dot{\sigma}_j)}\Big)_{\dot{\Omega}\times\dot{\Omega}}$$

is invertible at each $(\dot{q}_\star(\lambda, T), \lambda)$. By implicit function theorem, $\dot{q}_\star(\lambda, T)$, as the solution of $f_T[\dot{q}, \lambda] = 0$, is uniformly differentiable in $\lambda$. □

Let us first fix $T < \infty$ and consider the clusters encoded by $T$-colorings. We have explicitly defined $\boldsymbol{\Sigma}(H)$ and $\boldsymbol{s}(H)$. Let $\mathcal{S}(s) \equiv \sup\{\boldsymbol{\Sigma}(H) : \boldsymbol{s}(H) = s\}$, with the convention that a supremum over an empty set is $-\infty$. Thus $\mathcal{S}(s)$ is a well-defined function which captures the spirit of the function $\Sigma(s)$ discussed in the introduction. (Note $\mathcal{S}$ implicitly depends on $T$ since the maximum is taken over empirical measures $H$ which are supported on $T$-colorings.) Recall that the physics approach ([31] and refs. therein) takes $\mathcal{S}(s)$ as a conceptual starting point. However, for purposes of explicit calculation the actual starting point is the Legendre dual

$$\mathcal{F}(\lambda) \equiv (-\mathcal{S})^\star(\lambda) = \sup_{s\in\mathbb{R}} \Big\{\lambda s + \mathcal{S}(s)\Big\} = \sup_H \boldsymbol{F}_\lambda(H),$$

where $\boldsymbol{F}_\lambda(H) \equiv \lambda \boldsymbol{s}(H) + \boldsymbol{\Sigma}(H)$. The replica symmetry breaking heuristic gives an explicit conjecture for $\mathcal{F}$. One then makes the assumption that $\mathcal{S}(s)$ is **concave** in $s$: this means it is the same as

$$\mathcal{R}(s) \equiv -\mathcal{F}^\star(s) = -(-\mathcal{S})^{\star\star}(s),$$

so if $\mathcal{S}$ is concave then it can be recovered from $\mathcal{F}$.

We do not have a proof that $\mathcal{S}(s)$ is concave for all $s$, but we will argue that this holds on the interval of $s$ corresponding to $\lambda \in [0, 1]$. Formally, for $\lambda \in [0, 1]$, we proved that $\boldsymbol{F}_\lambda(H)$ has a unique maximizer $H_\star \equiv H_\lambda$. This implies that there is a unique $s_\lambda$ which maximizes $\lambda s + \mathcal{S}(s)$, given by

$$s_\lambda = \boldsymbol{s}(H_\lambda)\,.$$

Recall that $H_\lambda$ and $s_\lambda$ both depend implicitly on $T$. We also have from Lemma B.6 that for any fixed $T < \infty$, $s_\lambda$ is continuous in $\lambda$, so it maps $\lambda \in [0, 1]$ onto some compact interval $\mathcal{I} \equiv [s_-, s_+]$. Define the modified function

$$\overline{\mathcal{S}}(s) \equiv \begin{cases} \mathcal{S}(s) & \text{if } s \in \mathcal{I}, \\ -\infty & \text{otherwise.} \end{cases}$$

**Lemma B.7** *For all $s \in \mathbb{R}$, $\overline{\mathcal{S}}(s) = -(-\overline{\mathcal{S}})^{\star\star}(s)$. Consequently the function $\overline{\mathcal{S}}$ is concave, and $s_\lambda$ is nondecreasing in $\lambda$.*

***Proof*** The function $-\mathcal{S}(s)$ has Legendre dual

$$\overline{\mathcal{F}}(\lambda) = \sup_{s\in\mathbb{R}} \left\{\lambda s + \overline{\mathcal{S}}(s)\right\} = \sup_{s\in\mathcal{I}} \left\{\lambda s + \mathcal{S}(s)\right\} \le \mathcal{F}(\lambda)\,.$$

For $\lambda \in [0, 1]$ it is clear that $\overline{\mathcal{F}}(\lambda) = \mathcal{F}(\lambda)$. It is straightforward to check that if $\lambda < 0$ then

$$\overline{\mathcal{F}}(\lambda) \le \max_{s\in\mathcal{I}} \lambda s + \max_{s\in\mathcal{I}} \mathcal{S}(s) = \lambda s_{\min} + \mathcal{S}(s_0),$$

so if $s < s_{\min}$ then

$$(-\overline{\mathcal{S}})^{\star\star}(s) = (\overline{\mathcal{F}})^{\star}(s) \ge \sup_{\lambda<0} \left\{\lambda s - \overline{\mathcal{F}}(\lambda)\right\} \ge \sup_{\lambda<0} \left\{\lambda(s - s_{\min}) - \mathcal{S}(s_0)\right\} = +\infty\,.$$

A symmetric argument shows that $(-\overline{\mathcal{S}})^{\star\star}(s) = +\infty$ also for $s > s_{\max}$. If $s \in \mathcal{I}$, we must have $s = s_{\lambda_\circ}$ for some $\lambda_\circ \in [0, 1]$, and so

$$(-\overline{\mathcal{S}})^{\star\star}(s) \ge \lambda_\circ s - \mathcal{F}(\lambda_\circ) = -\mathcal{S}(s)\,.$$

This proves $(-\overline{\mathcal{S}})^{\star\star}(s) \ge -\overline{\mathcal{S}}(s)$ for all $s \in \mathbb{R}$. On the other hand, it holds for any function $f$ that $f^{\star\star} \le f$, so we conclude $(-\overline{\mathcal{S}})^{\star\star}(s) = -\overline{\mathcal{S}}(s)$ for all $s \in \mathbb{R}$. This implies that $\overline{\mathcal{S}}$ is concave, concluding the proof. □

***Proof of Proposition 1.4a*** We can obtain $\Sigma(s)$ as the limit of $\overline{\mathcal{S}}(s)$ in the limit $T \to \infty$. It follows from Lemma B.7 together with Proposition 5.5b that it is strictly decreasing in $s$. □

## Appendix C: Constrained entropy maximization

In this section we review basic calculations for entropy maximization problems under affine constraints.

### C.1. Constraints and continuity

We will optimize a functional over nonnegative measures $\nu$ on a finite space $X$ (with $|X| = s$), subject to some affine constraints $M\nu = b$. We begin by discussing basic continuity properties. Denote

$$\mathbb{H}(b) \equiv \{\nu \ge 0\} \cap \{M\nu = b\} \subseteq \mathbb{R}^s\,.$$

Let $\Delta \equiv \{\nu \geq 0\} \cap \{\langle \mathbf{1}, \nu \rangle = 1\}$, and let $\boldsymbol{B}$ denote the space of $b \in \mathbb{R}^r$ for which

$$\varnothing \neq \mathbb{H}(b) \subseteq \Delta\,.$$

Then $\boldsymbol{B}$ is contained in the image of $\Delta$ under $M$, so $\boldsymbol{B}$ is a compact subset of $\mathbb{R}^r$.

**Proposition C.1** *If $\boldsymbol{F}$ is any continuous function on $\Delta$ and*

$$F(b) = \max\{\boldsymbol{F}(\nu) : \nu \in \mathbb{H}(b)\}, \tag{126}$$

*then $F$ is (uniformly) continuous over $b \in \boldsymbol{B}$.*

Proposition C.1 is a straightforward consequence of the following two lemmas.

**Lemma C.2** *For $b \in \boldsymbol{B}$ and any vector $u$ in the unit sphere $\mathbb{S}^{r-1}$, let*

$$d(b, u) \equiv \inf\{t \geq 0 : b + tu \notin \boldsymbol{B}\}\,.$$

*There exists $\delta = \delta(b) > 0$ such that*

$$d(b, u) \in \{0\} \cup [\delta, \infty) \quad \textit{for all } b \in \boldsymbol{B}\,.$$

***Proof*** $\boldsymbol{B}$ is a polytope, so it can be expressed as the intersection of finitely many closed half-spaces $H_1, \ldots, H_k$, where $H_i = \{x \in \mathbb{R}^r : \langle a_i, x \rangle \leq c_i\}$. Consequently there is at least one index $1 \leq i \leq k$ such that

$$d(b, u) = \inf\{t \geq 0 : b + tu \notin H_i\}\,.$$

It follows that $\langle a_i, u \rangle > 0$ and

$$d(b, u) = \frac{c_i - \langle a_i, b \rangle}{\langle a_i, u \rangle} \geq \frac{c_i - \langle a_i, b \rangle}{|a_i|} = d(b, \partial H_i)$$

where $d(b, \partial H_i)$ is the distance between $b$ and the boundary of $H_i$. In particular, $d(b, u) > 0$ if and only if $\langle a_i, b \rangle < c_i$, which in turn holds if and only if $d(b, \partial H_i) > 0$. It follows that for all $u \in \mathbb{S}^{r-1}$ we have $d(b, u) \in \{0\} \cup [\delta, \infty)$ with

$$\delta = \delta(b) = \min\{d(b, \partial H_i) : d(b, \partial H_i) > 0\};$$

$\delta$ is a minimum over finitely many positive numbers so it is also positive. □

**Lemma C.3** *The set-valued function $\mathbb{H}$ is continuous on $\boldsymbol{B}$ with respect to the Hausdorff metric $d_{\mathcal{H}}$, that is to say, if $b_n \in \boldsymbol{B}$ with $\lim_{n\to\infty} b_n = b$ then*

$$\lim_{n\to\infty} d_{\mathcal{H}}(\mathbb{H}(b_n), \mathbb{H}(b)) = 0\,.$$

***Proof*** Recall that the Hausdorff distance between two subsets $X$ and $Y$ of a metric space is

$$d_{\mathcal{H}}(X, Y) = \inf\{\epsilon \geq 0 : X \subseteq Y^{\epsilon} \text{ and } Y \subseteq X^{\epsilon}\},$$

where $X^{\epsilon}$, $Y^{\epsilon}$ are the $\epsilon$-thickenings of $X$ and $Y$. Any sequence $\nu_n \in \mathbb{H}(b_n)$ converges along subsequences to limits $\nu \in \mathbb{H}(b)$, so for all $\epsilon > 0$ there exists $n_0(\epsilon)$ large enough that

$$\mathbb{H}(b_n) \subseteq (\mathbb{H}(b))^{\epsilon}, \quad n \geq n_0(\epsilon)\,.$$

In the other direction, we now argue that if $\nu \in \mathbb{H}(b)$ and $b' = b + tu$ for $u \in \mathbb{S}^{r-1}$ and $t$ a small positive number, then we can find $\nu' \in \mathbb{H}(b')$ which is close to $\nu$. For $u \in \mathbb{S}^{r-1}$ let $d(b, u)$ be as in Lemma C.2, and take $\nu(b, u)$ to be any fixed element of $\mathbb{H}(b + d(b, u)u)$ (which by definition is nonempty). Since we consider $b' = b + tu$ for $t > 0$, we can assume that $d(b, u)$ is positive, hence $\geq \delta(b)$ by Lemma C.2. We can express $b' = b + tu$ as the convex combination

$$b' = (1 - \epsilon)b + \epsilon[b + d(b, u)u], \quad \epsilon = \frac{t}{d(b, u)} = \frac{|b' - b|}{d(b, u)} \leq \frac{|b' - b|}{\delta}\,.$$

Then $\nu' = (1 - \epsilon)\nu + \epsilon\nu(b, u) \in \mathbb{H}(b')$, so

$$|\nu' - \nu| = \epsilon|\nu(b, u) - \nu| \leq \frac{(\operatorname{diam} \Delta)|b - b'|}{\delta}$$

This implies $H(b) \subseteq (H(b_n))^{\epsilon}$ for large enough $n$, and the result follows. □

***Proof of Proposition C.1*** Take $\nu \in \mathbb{H}(b)$ so that $F(b) = \boldsymbol{F}(\nu)$. If $b' = b + tu \in \boldsymbol{B}$ for some $u \in \mathbb{S}^{r-1}$, then Lemma C.3 implies that we can find $\nu' \in \mathbb{H}(b')$ with $|\nu' - \nu| = o_t(1)$, where $o_t(1)$ indicates a function tending to zero in the limit $t \downarrow 0$, uniformly over $u \in \mathbb{S}^{r-1}$. It follows that $\boldsymbol{F}(\nu) = \boldsymbol{F}(\nu') + o_t(1)$, since $\boldsymbol{F}$ is uniformly continuous on $\Delta$ by the Heine–Cantor theorem. Therefore

$$F(b) = \boldsymbol{F}(\nu) = \boldsymbol{F}(\nu') + o_t(1) \leq F(b') + o_t(1)\,.$$

By the same argument $F(b') \leq F(b) + o_t(1)$, concluding the proof. □

When solving (126) for a **fixed** value of $b \in \boldsymbol{B}$, it will be convenient to make the following reduction:

***Remark C.4*** Suppose $M$ is an $r \times s$ matrix where $s = |X|$. We can assume without loss that $M$ has full rank $r$, since otherwise we can eliminate redundant constraints. We consider only $b \in \boldsymbol{B}$, meaning $\varnothing \neq \mathbb{H}(b) \subseteq \Delta$. The affine space $\{M\nu = b\}$ has dimension $s - r$; we assume this is positive since otherwise $\mathbb{H}(b)$ would be a single point. Then, if $\mathbb{H}(b)$ does not contain an interior point of $\{\nu \geq 0\}$, it must be that

$$X_{\circ} \equiv \{x \in X : \exists \nu \in \{\nu \geq 0\} \cap \{M\nu = b\} \text{ so that } \nu(x) > 0\}$$

is a nonempty subset of $X$. In this case, it is equivalent to solve the optimization problem over measures $\nu_\circ$ on the reduced alphabet $X_\circ$, subject to constraints $M'\nu_\circ = b$ where $M'$ is the submatrix of $M$ formed by the columns indexed by $X_\circ$. Then, by construction, the space

$$\mathbb{H}_\circ(b) = \{\nu_\circ \geq 0\} \cap \{M'\nu_\circ = b\}$$

contains an interior point of $\{\nu_\circ \geq 0\}$. The matrix $M'$ is $r \times s_\circ$ where $s_\circ = |X_\circ|$; and if $M'$ is not of rank $r$ then we can again remove redundant constraints, replacing $M'$ with an $r_\circ \times s_\circ$ submatrix $M_\circ$ which has full rank $r_\circ$. We emphasize that the final matrix $M_\circ$ depends on $b$. In conclusion, when solving (126) for a fixed $b \in \boldsymbol{B}$, we may assume with no essential loss of generality that the original matrix $M$ is $r \times s$ with full rank $r$, and that $\mathbb{H}(b) = \{\nu \geq 0\} \cap \{M\nu = b\}$ contains an interior point of $\{\nu \geq 0\}$. It follows that this space has dimension $s - r > 0$, and its boundary is contained in the boundary of $\{\nu \geq 0\}$.

### C.2. Entropy maximization

We now restrict (126) to the case of functionals $\boldsymbol{F}$ which are **concave** on the domain $\{\nu \geq 0\}$. It is straightforward to verify from definitions that the optimal value $F(b)$ is (weakly) concave in $b$. Recall that the convex conjugate of a function $f$ on domain $C$ is the function $f^\star$ defined by

$$f^\star(x^\star) = \sup\{\langle x^\star, x\rangle - f(x) : x \in C\}\,.$$

Denote $G(\gamma) \equiv (-\boldsymbol{F})^\star(M^t\gamma)$, and consider the Lagrangian functional

$$\mathcal{L}(\gamma; b) = \sup\{\boldsymbol{F}(\nu) + \langle\gamma, M\nu - b\rangle : \nu \geq 0\} = -\langle\gamma, b\rangle + G(\gamma)\,.$$

It holds for any $\gamma \in \mathbb{R}^r$ that $\mathcal{L}(\gamma; b) \geq F(b)$, so

$$F(b) \leq \inf\{\mathcal{L}(\gamma; b) : \gamma \in \mathbb{R}^r\} = -G^\star(b)\,. \tag{127}$$

Now assume $\psi$ is a positive function on $X$, and consider (126) for the special case

$$\boldsymbol{F}(\nu) = \mathcal{H}(\nu) + \langle\nu, \ln\psi\rangle = \sum_{x\in X} \nu(x) \ln\frac{\psi(x)}{\nu(x)}\,. \tag{128}$$

We remark that the supremum in $(-\mathcal{H})^\star(\nu^\star) = \sup\{\langle\nu^\star, \nu\rangle + \mathcal{H}(\nu) : \nu \geq 0\}$ is uniquely attained by the measure $\nu^{\text{op}}(x) = \exp\{-1 + \nu^\star(x)\}$, yielding

$$(-\mathcal{H})^\star(\nu^\star) = \langle\nu^{\text{op}}(\nu^\star), 1\rangle = \sum_x \exp\{-1 + \nu^\star(x)\}\,.$$

This gives the explicit expression

$$G(\gamma) = (-\boldsymbol{F})^\star(M^t\gamma) = (-\mathcal{H})^\star(\ln\psi + M^t\gamma) = \sum_x \psi(x)\exp\{-1 + (M^t\gamma)(x)\}\,. \tag{129}$$

**Lemma C.5** *Assume $\psi$ is a strictly positive function on a set $X$ of size $s$ and that $M$ is $r \times s$ with rank $r$. Then the function $G(\gamma)$ of (129) is strictly convex in $\gamma$.*

***Proof*** Let $\nu \equiv \nu(\gamma)$ denote the measure on $X$ defined by

$$\nu(x) = \psi(x)\exp\{-1 + (M^t\gamma)(x)\},$$

and write $\langle f(x)\rangle_\nu \equiv \langle f, \nu\rangle$. The Hessian matrix $H \equiv \mathrm{Hess}\, G(\gamma)$ has entries

$$H_{i,j} = \frac{\partial^2 \mathcal{L}(\gamma; b)}{\partial\gamma_i\partial\gamma_j} = \sum_{x\in X} \nu(x) M_{i,x}M_{j,x} = \langle M_{i,x}M_{j,x}\rangle_\nu\,.$$

Let $M_x$ denote the vector-valued function $(M_{i,x})_{i\le r}$, so

$$\alpha^t H\alpha = \langle(\alpha^t M_x)^2\rangle_\nu\,.$$

This is zero if and only if $\nu(\{x \in X : \alpha^t M_x = 0\}) = 1$. Since $\nu$ is a positive measure, this can only happen if $\alpha^t M_x = 0$ for all $x \in X$, but this contradicts the assumption that $M$ has rank $r$. This proves that $H$ is positive-definite, so $G$ is strictly convex in $\gamma$. □

**Proposition C.6** *Let $b \in \boldsymbol{B}$ such that $\mathbb{H}(b) = \{\nu \ge 0\}\cap\{M\nu = b\}$ contains an interior point of $\{\nu \ge 0\}$, and consider the optimization problem (126) for $\boldsymbol{F}$ as in (128). For this problem, the inequality (127) becomes an equality,*

$$F(b) = \inf\{\mathcal{L}(\gamma; b) : \gamma \in \mathbb{R}^r\} = -G^\star(b)\,.$$

*Further, $\mathcal{L}(\gamma; b)$ is strictly convex in $\gamma$, and its infimum is achieved by a unique $\gamma = \gamma(b)$. The optimum value of (126) is uniquely attained by the measure $\nu = \nu^{op}(b)$ defined by*

$$\nu(x) = \psi(x)\exp\{-1 + (M^t\gamma)(x)\}\,. \tag{130}$$

*For any $\mu \in \mathbb{H}(b)$, $\boldsymbol{F}(\nu) - \boldsymbol{F}(\mu) = \mathcal{D}_{\mathrm{KL}}(\mu|\nu) \gtrsim \|\nu - \mu\|^2$. Finally, in a neighborhood of $b$ in $\boldsymbol{B}$, $\gamma'(b)$ is defined and $F(b)$ is strictly concave in $b$.*

***Proof*** Under the assumptions, the boundary of the set $\mathbb{H}(b)$ is contained in the boundary of $\{\nu \ge 0\}$. The entropy $\mathcal{H}$ has unbounded gradient at this boundary, so for $\boldsymbol{F}$ as in (128), the optimization problem (126) must be solved by a strictly positive measure $\nu > 0$. Since $\nu > 0$, we can differentiate in the direction of any vector $\delta$ with $M\delta = 0$ to find

$$0 = \frac{d}{dt}\Big[\mathcal{H}(\nu + t\delta) + \langle\ln\psi, \nu + t\delta\rangle\Big]\Big|_{t=0} = \langle\delta, -1 - \ln\nu + \ln\psi\rangle\,.$$

Recalling Remark C.4, we assume without loss that $M$ is $r \times s$ with rank $r$, since otherwise we can eliminate redundant constraints. Then, since $M\delta = 0$, for any $\gamma \in \mathbb{R}^r$ we have

$$0 = \langle \delta, \epsilon \rangle \quad \text{where } \epsilon = -1 - \ln \nu + \ln \psi + M^t \gamma \,.$$

We can then solve for $\gamma$ so that $M\epsilon = 0$:[2]

$$\gamma = (MM^t)^{-1} M(\ln \nu - \ln \psi + 1) \,.$$

Setting $\delta = \epsilon$ in the above gives $0 = \|\epsilon\|^2$, therefore we must have $\epsilon = 0$. This proves the existence of $\gamma = \gamma(b) \in \mathbb{R}^r$ such that (126) is optimized by $\nu = \nu^{\text{op}}(b)$, as given by (130). The optimal value of (126) is then

$$\begin{aligned} F(b) &= \langle 1, \nu^{\text{op}}(b) \rangle - \langle M^t \gamma(b), \nu^{\text{op}}(b) \rangle \\ &= \sum_x \psi(x) \exp\{-1 + (M^t \gamma)(x)\} - \langle \gamma, b \rangle \Big|_{\gamma = \gamma(b)} = \mathcal{L}(\gamma(b), b). \end{aligned}$$

In view of (127), this proves that in fact

$$-G^\star(b) = \inf\{\mathcal{L}(\gamma, b) : \gamma \in \mathbb{R}^r\} = \min\{\mathcal{L}(\gamma, b) : \gamma \in \mathbb{R}^r\} = \mathcal{L}(\gamma(b), b) = F(b)$$

as claimed. Now recall from Lemma C.5 that $G(\gamma)$ is strictly convex, which implies that $\mathcal{L}(\gamma; b)$ is strictly convex in $\gamma$. Thus $\gamma = \gamma(b)$ is the unique stationary point of $\mathcal{L}(\gamma; b)$.

These conclusions are valid under the assumption that $\mathbb{H}(b)$ contains an interior point of $\{\nu \geq 0\}$, which is valid in a neighborhood of $b$ in $\boldsymbol{B}$. Throughout this neighborhood, $\gamma(b)$ is defined by the stationarity condition $b = G'(\gamma)$. Differentiating again with respect to $\gamma$ gives

$$b'(\gamma) = \text{Hess}\, G(\gamma), \quad \gamma'(b) = [\text{Hess}\, G(\gamma(b))]^{-1} \,. \tag{131}$$

We also find (in this neighborhood) that

$$F'(b) = -\gamma(b), \quad F''(b) = -\gamma'(b) = -[\text{Hess}\, G(\gamma(b))]^{-1},$$

so $F$ is strictly concave. It remains to prove that $\boldsymbol{F}(\nu) - \boldsymbol{F}(\mu) = \mathcal{D}_{\text{KL}}(\mu|\nu)$. (The estimate $\mathcal{D}_{\text{KL}}(\mu|\nu) \gtrsim \|\mu - \nu\|^2$ is well known and straightforward to verify.) For any measure $\mu$,

$$-\mathcal{D}_{\text{KL}}(\mu|\nu) = \mathcal{H}(\mu) + \langle \mu, \ln(\psi \exp\{-1 + M^t \gamma\}) \rangle \,.$$

[2] The matrix $MM^t$ is invertible: if $MM^t x = 0$ then $M^t x \in \ker M = (\text{im}\, M^t)^\perp$. On the other hand clearly $M^t x \in \text{im}\, M^t$, so $M^t x \in (\text{im}\, M^t) \cap (\text{im}\, M^t)^\perp = \{0\}$. Therefore $x \in \ker M^t$, but $M^t$ is injective by assumption.

Applying this with $\mu = \nu$ gives

$$0 = -\mathcal{D}_{\mathrm{KL}}(\nu|\nu) = \mathcal{H}(\nu) + \langle \nu, \ln(\psi \exp\{-1 + M^t\gamma\})\rangle\,.$$

Subtracting these two equations gives

$$-\mathcal{D}_{\mathrm{KL}}(\mu|\nu) = \mathcal{H}(\mu) - \mathcal{H}(\nu) + \langle \mu - \nu, \ln\psi\rangle + \langle \mu - \nu, \ln(\exp\{-1 + M^t\gamma\})\rangle\,.$$

If $M\nu = M\nu = b$ then the last term vanishes, giving $-\mathcal{D}_{\mathrm{KL}}(\mu|\nu) = \boldsymbol{F}(\mu) - \boldsymbol{F}(\nu)$. □

***Remark C.7*** Our main application of Proposition C.6 is for the depth-one tree $\mathcal{D}$ of Fig. 6. In the notation of the current section, $X$ is the space of valid $T$-colorings $\underline{\sigma}$ of $\mathcal{D}$, and $\psi : X \to (0, \infty)$ is defined by

$$\psi(\underline{\sigma}) = \boldsymbol{w}_{\mathcal{D}}(\underline{\sigma})^{\lambda} = \left\{\dot{\Phi}(\underline{\sigma}_{\delta v}) \prod_{a\in\partial v} [\bar{\Phi}(\sigma_{av})\hat{\Phi}(\underline{\sigma}_{\delta a})]\right\}^{\lambda}.$$

We then wish to solve the optimization problem (126) for $\boldsymbol{F}(\nu)$ as in (128), under the constraint that $\nu$ has marginals $\dot{h}^{\mathrm{tr}}(\dot{\sigma})$ on the boundary edges $\delta\mathcal{D}$. This can be expressed as $M\nu = \dot{h}$ where $M$ has rows indexed by the spins $\dot{\sigma} \in \dot{\Omega}$, columns indexed by valid $T$-colorings $\underline{\eta} \equiv \underline{\eta}_{\mathcal{D}}$ of $\mathcal{D}$: the $(\dot{\sigma}, \underline{\eta})$ entry of $M$ is given by

$$M(\dot{\sigma}, \underline{\eta}) = |\delta\mathcal{D}|^{-1} \sum_{e\in\delta\mathcal{D}} \mathbf{1}\{\dot{\eta}_e = \dot{\sigma}\}\,.$$

Recall Remark C.4, let $\dot{\Omega}_+ = \{\dot{\sigma} \in \dot{\Omega} : \dot{h}^{\mathrm{tr}}(\dot{\sigma}) > 0\}$, and $X_\circ = \{\underline{\eta} \in X : M(\dot{\sigma}, \underline{\eta}) = 0\ \forall \dot{\sigma} \notin \dot{\Omega}\}$. Let $M_+$ be the $\dot{\Omega}_+ \times X_\circ$ submatrix of $M$, and set $\dot{q}(\dot{\sigma}) = 0$ for all $\dot{\sigma} \notin \dot{\Omega}_+$. Next, in the matrix $M_+$, if the $\dot{\eta}$ row is a linear combination of other rows, then set $\dot{q}(\dot{\eta}) = 1$ and remove this row. Repeat until we arrive at an $\dot{\Omega}_\circ \times X_\circ$ matrix $M_\circ$ of full rank $r_\circ = |\dot{\Omega}_\circ|$. The original problem reduces to an optimization over $\{\nu_\circ \geq 0\} \cap \{M_\circ\nu_\circ = b_\circ\}$ where $b_\circ$ denotes the entries of $b$ indexed by $\dot{\Omega}_\circ$. It follows from Proposition C.6 that the unique maximizer of (126) is the measure $\nu = \nu^{\mathrm{op}}(b)$ given by

$$\nu(\underline{\sigma}) = \frac{1}{Z}\boldsymbol{w}_{\mathcal{D}}(\underline{\sigma})^{\lambda} = \frac{1}{Z}\left\{\dot{\Phi}(\underline{\sigma}_{\delta v}) \prod_{a\in\partial v} [\bar{\Phi}(\sigma_{av})\hat{\Phi}(\underline{\sigma}_{\delta a})]\right\}^{\lambda} \prod_{e\in\delta\mathcal{D}} \dot{q}(\sigma_e)\,.$$

Note however that if $M_+$ is not of full rank then $\dot{q}$ need not be unique.

## Appendix D: Pairs of intermediate or large overlap

In this section we prove Proposition 3.7, which states that the first moment of $\boldsymbol{Z} = \boldsymbol{Z}_{\lambda,T}$ is dominated by separable colorings provided $0 \leq \lambda \leq 1$.

### D.1. Intermediate overlap

We first show that configurations with "intermediate" overlap are negligible. This can be done with quite crude estimates, working with NAE- SAT solutions rather than colorings.

**Lemma D.1** *Consider random regular* NAE- SAT *at clause density* $\alpha \geq 2^{k-1}\ln 2 - O(1)$. *On* $\mathscr{G} = (V, F, E, \underline{L})$, *let* $Z^2[\rho]$ *count the number of pairs* $\underline{x}, \underline{\acute{x}} \in \{0,1\}^V$ *of valid* NAE- SAT *solutions which agree on* $\rho$ *fraction of variables. Then*

$$\mathbb{E}Z^2[\rho] \leq (\mathbb{E}Z)\exp\Big\{n\Big[H(\rho) - (\ln 2)\pi(\rho) + O(1/2^k)\Big]\Big\},$$

*for* $\pi(\rho) \equiv 1 - \rho^k - (1-\rho)^k$.

***Proof*** For $\underline{u} \in \{0,1\}^V$, let $I^{\mathrm{NAE}}(\underline{u};\mathscr{G})$ be the indicator that $\underline{u}$ is a valid NAE- SAT solution on $\mathscr{G}$. Fix any pair of vectors $\underline{x}, \underline{\acute{x}} \in \{0,1\}^V$ which agree on $\rho$ fraction of variables:

$$\mathbb{E}Z^2[\rho] = 2^n\binom{n}{n\rho}\mathbb{E}[I^{\mathrm{NAE}}(\underline{x};\mathscr{G})I^{\mathrm{NAE}}(\underline{\acute{x}};\mathscr{G})] = (\mathbb{E}Z)\binom{n}{n\rho}\mathbb{E}[I^{\mathrm{NAE}}(\underline{\acute{x}};\mathscr{G}) \mid I^{\mathrm{NAE}}(\underline{x};\mathscr{G}) = 1].$$

Given $\underline{x}, \underline{\acute{x}}$, let $M \equiv M(\underline{x}, \underline{\acute{x}})$ count the number of clauses $a \in F$ where

$$|\{e \in \delta a : x_{v(e)} = \acute{x}_{v(e)}\}| \notin \{0, k\}.$$

In each of these clauses, there are $2^k - 2$ literal assignments $\underline{\mathtt{L}}_{\delta a}$ which are valid for $\underline{x}$. Out of these, exactly $2^k - 4$ are valid also for $\underline{\acute{x}}$. If we define i.i.d. binomial random variables $D_a \sim \mathrm{Bin}(k,\rho)$, indexed by $a \in F$, then

$$\mathbb{P}(M = m\gamma) = \mathbb{P}\Big(\sum_{a\in F}\mathbf{1}\{D_a \notin \{0,k\}\}\,\Big|\,\sum_{a\in F}D_a = mk\rho\Big).$$

The $(D_a)_{a\in F}$ sum to $mk\rho$ with probability which is polynomial in $n$, so

$$\mathbb{P}(M = m\gamma) \leq n^{O(1)}\mathbb{P}(\mathrm{Bin}(m,\pi) = m\gamma)$$

with $\pi = \pi(\rho)$ as in the statement of the lemma. Therefore

$$\mathbb{E}[I^{\mathrm{NAE}}(\underline{\acute{x}};\mathscr{G}) \mid I^{\mathrm{NAE}}(\underline{x};\mathscr{G}) = 1] \leq n^{O(1)}\mathbb{E}\left[\left(\frac{2^k-4}{2^k-2}\right)^X\right]$$

for $X \sim \mathrm{Bin}(m,\rho)$. It is easily seen that the above is $\leq \exp\{-m\pi/2^{k-1}\}$, and the claimed bound follows, using the lower bound on $\alpha = m/n$. □

**Corollary D.2** *Let $\psi(\rho) = H(\rho) - (\ln 2)\pi(\rho)$. Then $\psi(\rho) \le -2k/2^k$ for all $\rho$ in*

$$[\exp\{-k/(\ln k)\}, \tfrac{1}{2}(1 - k/2^{k/2})] \cup [\tfrac{1}{2}(1 + k/2^{k/2}), 1 - \exp\{-k/(\ln k)\}]\,.$$

*Assuming $\alpha = m/n \ge 2^{k-1}\ln 2 - O(1)$, $\mathbb{E}Z^2[\rho] \le \exp\{-nk/2^k\}$ for all such $\rho$.*

***Proof*** Note that $H(\frac{1+\epsilon}{2}) \le \ln 2 - \epsilon^2/2$. If $(k \ln k)/2^k \le \epsilon \le 1/k$, then

$$\psi(\tfrac{1+\epsilon}{2}) \le -\epsilon^2/2 + O(k\epsilon/2^k) \le -\epsilon^2/3\,.$$

Both $H(\frac{1+\epsilon}{2})$ and $\pi(\frac{1+\epsilon}{2})$ are symmetric about $\epsilon = 0$, and decreasing on the interval $0 \le \epsilon \le 1$. It follows that for any $0 \le a \le b \le 1$,

$$\max_{a \le \epsilon \le b} \psi(\tfrac{1+\epsilon}{2}) \le H(\tfrac{1+a}{2}) - (\ln 2)\pi(\tfrac{1+b}{2})\,.$$

With this in mind, if $1/k \le \epsilon \le 1 - 5(\ln k)/k$,

$$\psi(\tfrac{1+\epsilon}{2}) \le -(2k^2)^{-1} + O(k^{-5/2}) \le -(4k^2)^{-1}\,.$$

If $1 - 5(\ln k)/k \le \epsilon \le 1 - (\ln k)^3/k^2$,

$$\psi(\tfrac{1+\epsilon}{2}) \le O(1)(\ln k)^2/k - \Omega(1)(\ln k)^3/k \le -\Omega(1)(\ln k)^3/k\,.$$

Finally, if $1 - (\ln k)^3/k^2 \le \epsilon \le 1 - \exp\{-2k/(\ln k)\}$, then

$$\psi(\tfrac{1+\epsilon}{2}) \le O(1)\epsilon k/(\ln k) - \Omega(1)\epsilon k \le -\Omega(1)\epsilon k\,.$$

Combining these estimates proves the claimed bound on $\psi(\rho)$. The assertion for $\mathbb{E}[Z^2(\rho)]$ then follows by substituting into Lemma D.1, and noting that $\mathbb{E}Z \le \exp\{O(n/2^k)\}$. □

### D.2. Large overlap

In what follows, we restrict consideration to a small neighborhood $\mathbf{N}$ of $H_\star$. We abbreviate $\underline{\sigma} \in H$ if $H(\mathcal{G}, \underline{\sigma}) = H$, and $\underline{\sigma} \in \mathbf{N}$ if $H(\mathcal{G}, \underline{\sigma}) \in \mathbf{N}$. Recall that we write $\underline{\sigma}' \succcurlyeq \underline{\sigma}$ if the number of free variables in $\underline{x}(\underline{\sigma}')$ upper bounds the number in $\underline{x}(\underline{\sigma})$. We also write $H' \succcurlyeq H$ if $\underline{\sigma}' \succcurlyeq \underline{\sigma}$ for any (all) $\underline{\sigma} \in H$ and $\underline{\sigma}' \in H'$. Let $\mathbf{Z}_{\mathrm{ns}}(H, H')$ count the colorings $\underline{\sigma} \in H$ such that

$$\left|\left\{\underline{\sigma}' \in H' : \mathsf{sep}(\underline{\mathfrak{o}}, \underline{\mathfrak{o}}') \le \exp\{-\mathrm{k}/(\ln \mathrm{k})\}\right\}\right| \ge !(\mathrm{n}),$$

for $\omega(n) = \exp\{(\ln n)^4\}$. (Although we will not write it explicitly, it should be understood that $\mathbf{Z}_{\mathrm{ns}}(H, H')$ depends on $\mathcal{G}$, since both $\underline{\sigma}, \underline{\sigma}'$ are required to be valid colorings of $\mathcal{G}$.) Let $\mathbf{Z}_{\mathrm{ns}}(\mathbf{N})$ denote the sum of $\mathbf{Z}_{\mathrm{ns}}(H; H')$ over all pairs $H, H' \in \mathbf{N}$ with $H' \succcurlyeq H$. Let $\mathbf{Z}(\mathbf{N})$ denote the sum of $\boldsymbol{Z}(H)$ over all $H \in \mathbf{N}$.

**Proposition D.3** *There exists a small enough positive constant $\epsilon_{\max}(k)$ such that, if* $\mathbf{N}$ *is the $\epsilon$-neighborhood of $H_\star$ for any $\epsilon \le \epsilon_{\max}$, then*

$$\mathbb{E}\boldsymbol{Z}_{ns}(\mathbf{N}) \le \mathbb{E}\boldsymbol{Z}(\mathbf{N}) \exp\{-(\ln n)^2\}\,.$$

***Proof*** By definition,

$$\boldsymbol{Z}_{\rm ns}(\mathbf{N}) = \sum_{H\in\mathbf{N}} \boldsymbol{Z}_{\succcurlyeq}(H), \quad \boldsymbol{Z}_{\succcurlyeq}(H) \equiv \sum_{H'\in\mathbf{N}} \mathbf{1}\{H' \succcurlyeq H\}\mathbf{Z}_{\rm ns}(H, H')\,.$$

It suffices to show that for every $H \in \mathbf{N}$, $\mathbb{E}\boldsymbol{Z}_{\succcurlyeq}(H) \le \mathbb{E}\boldsymbol{Z}(H)\exp\{-2(\ln n)^2\}$. Note that the total number of empirical measures $H'$ is at most $n^c$ for some constant $c(k, T)$. Let $\boldsymbol{E}$ denote the set of pairs $(\mathscr{G}, \underline{\sigma})$ for which

$$\left|\left\{\underline{\sigma}' \in \mathbf{N} : \underline{\sigma}' \succcurlyeq \underline{\sigma} \text{ and } \mathsf{sep}(\underline{\mathfrak{o}}, \underline{\mathfrak{o}}') \le \exp\{-\mathrm{k}/(\ln \mathrm{k})\}\right\}\right| \ge !(\mathrm{n})\,.$$

(Again, it is understood that both $\underline{\sigma}, \underline{\sigma}'$ must be valid colorings of $\mathscr{G}$.) Then

$$\boldsymbol{Z}_{\succcurlyeq}(H) \le n^c \sum_{\underline{\sigma}\in H} \mathbf{1}\{(\mathscr{G}, \underline{\sigma}) \in \boldsymbol{E}\}\,.$$

Consequently, in order to show the required bound on $\mathbb{E}\boldsymbol{Z}_{\succcurlyeq}(H)$, it suffices to show

$$\mathbb{P}^H(\boldsymbol{E}) \le n^{-c}\exp\{-2(\ln n)^2\}, \tag{132}$$

where $\mathbb{P}^H$ is a "planted" measure on pairs $(\mathscr{G}, \underline{\sigma})$: to sample from $\mathbb{P}^H$, we start with a set $V$ of $n$ isolated variables each with $d$ incident half-edges, and a set $F$ of $m$ isolated clauses each with $k$ incident half-edges. Assign colorings of the half-edges,

$$\underline{\sigma}_\delta \equiv (\underline{\sigma}_{\delta V}, \underline{\sigma}_{\delta F}) \quad \text{where } \underline{\sigma}_{\delta V} \equiv (\underline{\sigma}_{\delta v})_{v\in V},\ \underline{\sigma}_{\delta F} \equiv (\underline{\sigma}_{\delta a})_{a\in F},$$

which are uniformly random subject to the empirical measure $H$. Then $\underline{\sigma}_\delta$ is the "planted" coloring: conditioned on it, we sample uniformly at random a graph $\mathscr{G}$ such that $\underline{\sigma}_\delta$ becomes a valid coloring $\underline{\sigma}$ on $\mathscr{G}$. The resulting pair $(\mathscr{G}, \underline{\sigma})$ is a sample from $\mathbb{P}^H$.

Suppose $(\mathscr{G}, \underline{\sigma}) \in \boldsymbol{E}$. The total number of configurations $\underline{\sigma}'$ with $\mathsf{sep}(\underline{\mathfrak{o}}, \underline{\mathfrak{o}}') \le$ is at most $(cn)^{n\delta}$, which is $\ll \omega(n)$ if $\delta \le n^{-1}(\ln n)^2$. This implies that there must exist $\underline{\sigma}' \in \mathbf{N}$ such that $\underline{\sigma}' \succcurlyeq \underline{\sigma}$ and

$$n^{-1}(\ln n)^2 \le \mathsf{sep}(\underline{\mathfrak{o}}, \underline{\mathfrak{o}}') \le \exp\{-\mathrm{k}/(\ln \mathrm{k})\}\,.$$

It follows that

$$S \equiv \{v \in V : x_v(\underline{\sigma}) \in \{0, 1\} \text{ and } x_v(\underline{\sigma}') \ne x_v(\underline{\sigma})\}$$

has size $|S| \equiv ns$ for $s \in [(2n)^{-1}(\ln n)^2, \exp\{-k/(\ln k)\}]$. The set $S$ is **internally forced** in $\underline{\sigma}$: for every $v \in S$, any clause forcing to $v$ must have another edge connecting to $S$. Formally, let $\mathtt{R}_U$ (resp. $\mathtt{B}_U$) count the number of $\{\mathtt{r}\}$-colored (resp. $\{\mathtt{b}\}$-colored) edges incident to a subset of vertices $U$. Let $I_S$ be the indicator that all variables in $S$ are forced. For any fixed $S \subseteq V$,

$$\mathbb{P}^H(S \text{ internally forced}) \le \mathbb{E}_{\mathbb{P}^H}\left[ I_S k^{\mathtt{R}_S} \frac{(\mathtt{B}_S)_{\mathtt{R}_S}}{(\mathtt{B}_F)_{\mathtt{R}_S}} \right] \le \mathbb{E}_{\mathbb{P}^H}[I_S (4ks)^{\mathtt{R}_S}] .$$

In the first inequality, the factor $k^{\mathtt{R}_S}$ accounts for the choice, for each $S$-incident $\{\mathtt{r}\}$-colored edge $e$, of another edge $e'$ sharing the same clause. The factor $(\mathtt{B}_S)_{\mathtt{R}_S}/(\mathtt{B}_F)_{\mathtt{R}_S}$ then accounts for the chance that the chosen edge $e'$ (which must have color in $\{\mathtt{b}\}$) will also be $S$-incident. The second inequality follows by noting that we certainly have $\mathtt{B}_S \le nsd$, and for $H$ near $H_\star$ we also clearly have $\mathtt{B}_F \ge nd/4$.

To bound the above, we can work with a slightly different measure $\mathbb{Q}^H$: instead of sampling $\underline{\sigma}_\delta$ subject to $H$, we can simply sample variable-incident colorings $\underline{\sigma}_{\delta v}$ i.i.d. from $\dot{H}$, and clause-incident colorings $\underline{\sigma}_{\delta a}$ i.i.d. from $\hat{H}$. On the event $\mathsf{MARG}$ that the resulting $\underline{\sigma}_\delta$ has empirical measure $H$, we sample the graph $\mathscr{G}$ according to $\mathbb{P}^H(\mathscr{G}|\underline{\sigma}_\delta)$, and otherwise we set $\mathscr{G} = \varnothing$. Then, since $\mathbb{Q}^H(\mathsf{MARG}) \ge \mathrm{n}^{-\mathrm{c}}$ (adjusting $c$ as needed), we have

$$\mathbb{P}^H((\mathscr{G}, \underline{\sigma})) = \mathbb{Q}^H((\mathscr{G}, \underline{\sigma}) \,|\, \mathsf{MARG}) \le \mathrm{n}^{\mathrm{c}}\, \mathbb{Q}^{\mathrm{H}}((\mathscr{G}, \underline{\mathfrak{o}}); \mathsf{MARG}) .$$

Let us abbreviate $\dot{H}(\ell)$ for the probability under $\dot{H}$ that $\underline{\sigma}$ has $\ell$ entries in $\{\mathtt{r}\}$: then

$$\mathbb{E}_{\mathbb{P}^H}[I_S(4ks)^{\mathtt{R}_S}] \le n^c\, \mathbb{E}_{\mathbb{Q}^H}[I_S(4ks)^{\mathtt{R}_S}; \mathsf{MARG}] \le \mathrm{n}^{\mathrm{c}} \left( \sum_{\mathfrak{`} \ge 1} \mathrm{H}(\mathfrak{`})(4\mathrm{ks})^{\mathfrak{`}} \right)^{\mathrm{ns}} . \quad (133)$$

For $H$ sufficiently close to $H_\star$, we will have

$$\dot{H}(\ell) \le 2\dot{H}_\star(\ell) \le 2\binom{d}{\ell} \frac{\hat{q}_\star(\mathtt{r}_1)^\ell \hat{q}_\star(\mathtt{b}_1)^{d-\ell}}{[\hat{q}_\star(\mathtt{r}_1) + \hat{q}_\star(\mathtt{b}_1)]^d - \hat{q}_\star(\mathtt{b}_1)^d} .$$

It follows that the right-hand side of (133) is (for some absolute constant $\delta$)

$$\le n^c\, 2^{ns} \left( \frac{[\hat{q}_\star(\mathtt{r}_1) \cdot 4ks + \hat{q}_\star(\mathtt{b}_1)]^d - \hat{q}_\star(\mathtt{b}_1)^d}{[\hat{q}_\star(\mathtt{r}_1) + \hat{q}_\star(\mathtt{b}_1)]^d - \hat{q}_\star(\mathtt{b}_1)^d} \right)^{ns} \le n^c s^{ns} 2^{-\delta kns},$$

where the last inequality uses that $s \le \exp\{-k/(\ln k)\}$. Summing over $S$ gives

$$\mathbb{P}^H(\boldsymbol{E}) \le \max_{s \ge (2n)^{-1}(\ln n)^2} n^c 2^{-\delta kns/2} \le \exp\{-\Omega(1) k (\ln n)^2\} .$$

This implies (132); and the claimed result follows as previously explained. □

***Proof of Proposition 3.7*** Follows by combining Corollary D.2 and Proposition D.3. □

## Appendix E: Free energy upper bound

For a family of spin systems that includes NAE- SAT, an interpolative calculation gives an upper bound for the free energy on Erdős-Rényi graphs ([26,43], cf. [30]). These bounds build on earlier work [29] concerning the subadditivity of the free energy in the Sherrington–Kirkpatrick model, which was later generalized to a broad class of models [4,5,12,27]. (Although these results are closely related, we remark that interpolation gives quantitative bounds whereas subadditivity does not.) To prove the upper bound in Theorem 1, we establish the analogue of [26,43] for random **regular** graphs. Although the main concern of this paper is the NAE- SAT model, we give the bound for a more general class of models, which may be of independent interest.

### E.1. Basic interpolation bound

Recall $\mathcal{G} = (V, F, E)$ denotes a $(d, k)$-regular bipartite graph (without edge literals). We consider measures defined on vectors $\underline{x} \in \mathcal{X}^V$ where $\mathcal{X}$ is some fixed alphabet of finite size. Fix also a finite index set $S$. Suppose we have (random) vectors $b \in \mathbb{R}^S$ and $f \in \mathcal{F}(\mathcal{X})^S$, where $\mathcal{F}(\mathcal{X})$ denotes the space of functions $\mathcal{X} \to \mathbb{R}_{\geq 0}$. Independently of $b$, let $f_1, \ldots, f_k$ be i.i.d. copies of $f$, and define the random function

$$\theta(\underline{x}) \equiv \sum_{s \in S} b_s \prod_{j=1}^{k} f_{s,j}(x_j). \tag{134}$$

Let $h$ be another (random) element of $\mathcal{F}(\mathcal{X})$. Assume there is a constant $\epsilon > 0$ so that

$$\epsilon \leq \{h, 1 - \theta\} \leq \frac{1}{\epsilon} \quad \text{almost surely.} \tag{135}$$

Note we do not require the $b_s$ to be nonnegative; however, we assume that

$$b^p(\underline{s}) \equiv \mathbb{E}\Big[\prod_{\ell=1}^{p} b_{s_\ell}\Big] \geq 0 \quad \text{for any } p \geq 1, \underline{s} \equiv (s_1, \ldots, s_p) \in S^p. \tag{136}$$

Let $\mathscr{G}$ denote the graph $\mathcal{G}$ labelled by a vector $((h_v)_{v \in V}, (\theta_a)_{a \in F})$ of independent functions, where the $h_v$ are i.i.d. copies of $h$ and the $\theta_a$ are i.i.d. copies of $\theta$. For $a \in F$ we abbreviate $\underline{x}_{\delta a} \equiv (x_{v(e)})_{e \in \delta a} \in \mathcal{X}^k$, and we consider the (random) Gibbs measure

$$\mu_{\mathscr{G}}(\underline{x}) \equiv \frac{1}{Z(\mathscr{G})} \prod_{v \in V} h_v(x_v) \prod_{a \in F} [1 - \theta_a(\underline{x}_{\delta a})] \tag{137}$$

where $Z(\mathscr{G})$ is the normalizing constant. Now let $\mathscr{G}$ be the random $(d, k)$-regular graph on $n$ variables, together with the random function labels. We write $\mathbb{E}_n$ for expectation

over the law of $\mathscr{G}$, and define the (logarithmic) free energy of the model to be

$$F_n \equiv \frac{1}{n}\mathbb{E}_n \ln Z(\mathscr{G})\,.$$

***Example E.1*** *(positive temperature* NAE- SAT*)* Let $\mathcal{X} = \{0, 1\}$, and let $\underline{\mathtt{L}} \equiv (\mathtt{L}_i)_{i\le k}$ be a sequence of i.i.d. Bernoulli$(1/2)$ random variables. The positive-temperature NAE- SAT model corresponds to taking $h \equiv 1$ and

$$\theta(\underline{x}) \equiv (1 - e^{-\beta})\Big(\prod_{i=1}^{k} \frac{\mathtt{L}_1 \oplus x_i}{2} + \prod_{i=1}^{k} \frac{1 \oplus \mathtt{L}_i \oplus x_i}{2}\Big)$$

where $\beta \in (0, \infty)$ is the inverse temperature. In this model, each violated clause incurs a multiplicative penalty $e^{-\beta}$.

***Example E.2*** *(positive-temperature coloring)* Let $\mathcal{X} = [q]$. The positive-temperature coloring model (i.e., anti-ferromagnetic Potts model) on a $k$-uniform hypergraph corresponds to $h \equiv 1$ and

$$\theta(\underline{x}) \equiv (1 - e^{-\beta}) \sum_{s=1}^{q} \mathbf{1}\{x_1 = \cdots = x_k = s\}$$

where $\beta \in (0, \infty)$ is the inverse temperature. In this model, each monochromatic (hyper)edge incurs a multiplicative penalty $e^{-\beta}$.

The following theorem is a random regular graph analog of [43, Thm. 3]. (We stated our result for a slightly more general class of models than considered in [43]; however the main result of [43] extends to these models with only minor modifications.)

**Theorem E.3** *Consider a (random) Gibbs measure* (137) *satisfying assumptions* (134)–(136)*, and consider the (nonasymptotic) free energy* $F_n \equiv n^{-1}\mathbb{E}_n \ln Z(\mathscr{G})$*. Let*

$$\begin{aligned}
\mathcal{M}_0 &\equiv \textit{space of probability measures over } \mathcal{X},\\
\mathcal{M}_1 &\equiv \textit{space of probability measures over } \mathcal{M}_0,\\
\mathcal{M}_2 &\equiv \textit{space of probability measures over } \mathcal{M}_1.
\end{aligned}$$

*For* $\zeta \in \mathcal{M}_2$*, let* $\underline{\eta} \equiv (\eta_{a,j})_{a\ge 0, j\ge 0}$ *be an array of i.i.d. samples from* $\zeta$*. For each index* $(a, j)$ *let* $\rho_{a,j}$ *be a conditionally independent sample from* $\eta_{a,j}$*, and denote*

$\underline{\rho} \equiv (\rho_{a,j})_{a\geq 0, j\geq 0}$. *Let* $(h\rho)_{a,j}(x) \equiv h_{a,j}(x)\rho_{a,j}(x)$, *define random variables*

$$\boldsymbol{u}_a(x) \equiv \sum_{\underline{x}\in\mathcal{X}^k} \mathbf{1}\{x_1 = x\}[1-\theta_a(\underline{x})]\prod_{j=2}^{k}(h\rho)_{a,j}(x_j),$$

$$\boldsymbol{u}_a \equiv \sum_{\underline{x}\in\mathcal{X}^k} [1-\theta_a(\underline{x})]\prod_{j=1}^{k}(h\rho)_{a,j}(x_j).$$

*For any* $\lambda \in (0,1)$ *and any* $\zeta \in \mathcal{M}_2$,

$$F_n \leq \lambda^{-1}\mathbb{E}\ln\mathbb{E}'\Big[\Big(\sum_{x\in\mathcal{X}} h(x)\prod_{a=1}^{d}\boldsymbol{u}_a(x)\Big)^{\lambda}\Big] - (k-1)\alpha\lambda^{-1}\mathbb{E}\ln\mathbb{E}'[(\boldsymbol{u}_0)^{\lambda}] + O_\epsilon(n^{-1/3})$$

*where* $\mathbb{E}'$ *denotes the expectation over* $\underline{\rho}$ *conditioned on all else, and* $\mathbb{E}$ *denotes the overall expectation.*

***Remark E.4*** In the statistical physics framework, elements $\rho \in \mathcal{M}_0$ correspond to belief propagation messages for the underlying model, which has state space $\mathcal{X}$. Elements $\eta \in \mathcal{M}_1$ correspond to belief propagation messages for the 1RSB model (termed "auxiliary model" in [33, Ch. 19]), which has state space $\mathcal{M}_0$. The informal picture is that the $\eta$ associated to variable $x$ is determined by the geometry of the local neighborhood of $x$ — that is to say, the randomness of $\zeta$ reflects the randomness in the geometry of the $R$-neighborhood of a uniformly randomly variable in the graph. In random regular graphs this randomness is degenerate—the $R$-neighborhood of (almost) every vertex is simply a regular tree. It is therefore expected that the best upper bound in Theorem E.3 can be achieved with $\zeta$ a point mass.

### E.2. Replica symmetric bound

Along the lines of [43], we first prove a weaker "replica symmetric" version of Theorem E.3. Afterwards we will apply it to obtain the full result.

**Theorem E.5** *In the setting of* Theorem E.3, *define*

$$\Phi_V \equiv \mathbb{E}\ln\Big(\sum_{x\in\mathcal{X}} h(x)\prod_{a=1}^{d}\boldsymbol{u}_a(x)\Big), \quad \Phi_F \equiv (k-1)\alpha\mathbb{E}\ln(\boldsymbol{u}_0)\,.$$

*Then* $F_n \leq \Phi_V - \Phi_F - O_\epsilon(n^{-1/3})$.

Inspired by the proof of [12], we prove Theorem E.5 by a combinatorial interpolation between two graphs, $\mathcal{G}_{-1}$ and $\mathcal{G}_{nd+1}$. The initial graph $\mathcal{G}_{-1}$ will have free energy $\Phi_V$, and the final graph $\mathcal{G}_{nd+1}$ will have free energy $F_n + \Phi_F$. We will show that, up to $O_\epsilon(n^{1/3})$ error, the free energy of $\mathcal{G}_{-1}$ will be larger than that of $\mathcal{G}_{nd+1}$, from which the bound of Theorem E.5 follows.

To begin, we take $\mathscr{G}_{-1}$ to be a factor graph consisting of $n$ disjoint trees (Fig. 7a). Each tree is rooted at a variable $v$ which joins to $d$ clauses. Each of these clauses then joins to $k-1$ more variables, which form the leaves of the tree. We write $V$ for the root variables, $A$ for the clauses, and $U$ for the leaf variables. Note $|V| = n$, $|A| = nd$, and $|U| = nd(k-1)$.

Independently of all else, take a vector of i.i.d. samples $(\eta_u, \rho_u)_{u\in U}$ where $\eta_u$ is a sample from $\zeta$, and $\rho_u$ is a sample from $\eta_u$.[3] As before, the variables and clauses in $\mathscr{G}_{-1}$ are labelled independently with functions $h_v$ and $\theta_a$. We now additionally assign to each $u \in U$ the label $(\eta_u, \rho_u)$. Let $(h\rho)_u(x) \equiv h_u(x)\rho_u(x)$. We consider the factor model on $\mathscr{G}_{-1}$ defined by

$$\mu_{\mathscr{G}_{-1}}(\underline{x}) = \frac{1}{Z(\mathscr{G}_{-1})} \prod_{v\in V} h_v(x_v) \prod_{a\in A} [1 - \theta_a(\underline{x}_{\delta a})] \prod_{u\in U} (h\rho)_u(x_u)\,.$$

We now define the interpolating sequence of graphs $\mathscr{G}_{-1}, \mathscr{G}_0, \ldots, \mathscr{G}_{nd+1}$. Fix $m' \equiv 2n^{2/3}$. The construction proceeds by adding and removing clauses. Whenever we remove a clause $a$, the edges $\delta a$ are left behind as $k$ unmatched edges in the remaining graph. Whenever we add a new clause $b$, we label it with a fresh sample $\theta_b$ of $\theta$. The graph $\mathscr{G}_r$ has clauses $F_r$ which can be partitioned into $A_{U,r}$ (clauses involving $U$ only), $A_{V,r}$ (clauses involving $V$ only), and $A_r$ (clauses involving both $U$ and $V$). We will define below a certain sequence of events $\mathsf{COUP}_\mathrm{r}$. Let $\mathsf{COUP}_{\leq \mathrm{r}}$ be the event that $\mathsf{COUP}_\mathrm{s}$ occurs for all $0 \leq s \leq r$. The event $\mathsf{COUP}_{\leq -1}$ occurs vacuously, so $\mathbb{P}(\mathsf{COUP}_{\leq -1}) = 1$. With this notation in mind, the construction goes as follows:

1. Starting from $\mathscr{G}_{-1}$, choose a uniformly random subset of $m'$ clauses from $F_{-1} = A_{-1} = A$, and remove them to form the new graph $\mathscr{G}_0$.
2. For $0 \leq r \leq nd - m' - 1$, we start from $\mathscr{G}_r$ and form $\mathscr{G}_{r+1}$ as follows.
   a. If $\mathsf{COUP}_{\leq \mathrm{r}-1}$ succeeds, choose a uniformly random clause $a$ from $A_r$, and remove it to form the new graph $\mathscr{G}_{r,\circ}$. Let $\delta' U_{r,\circ}$ and $\delta' V_{r,\circ}$ denote the unmatched half-edges incident to $U$ and $V$ respectively in $\mathscr{G}_{r,\circ}$, and define the event
   $$\mathsf{COUP}_\mathrm{r} \equiv \{\min\{{}'\mathrm{U}_{\mathrm{r},\circ}, {}'\mathrm{V}_{\mathrm{r},\circ}\} \geq \mathrm{k}\}\,.$$
   If instead $\mathsf{COUP}_{\leq \mathrm{r}-1}$ fails, then $\mathsf{COUP}_{\leq \mathrm{r}}$ fails by definition.
   b. If $\mathsf{COUP}_{\leq \mathrm{r}}$ fails, let $\mathscr{G}_{r+1} = \mathscr{G}_r$. If $\mathsf{COUP}_{\leq \mathrm{r}}$ succeeds, then with probability $1/k$ take $k$ half-edges from $\delta' V_{r,\circ}$ and join them into a new clause $c$. With the remaining probability $(k-1)/k$ take $k$ half-edges from $\delta' U_{r,\circ}$ and join them into a new clause $c$.
   c. For $nd - m' \leq r \leq nd - 1$ let $\mathscr{G}_{r+1} = \mathscr{G}_r$. Starting from $\mathscr{G}_{nd}$, remove all the clauses in $A_{nd}$. Then connect (uniformly at random) all remaining unmatched $V$-incident edges into clauses. Likewise, connect all remaining unmatched $U$-incident edges into clauses. Denote the resulting graph $\mathscr{G}_{nd+1}$.

[3] For the proof of Theorem E.5 it is equivalent to sample $\rho$ from $\eta^{\mathrm{av}} \equiv \int \eta\, d\zeta$.

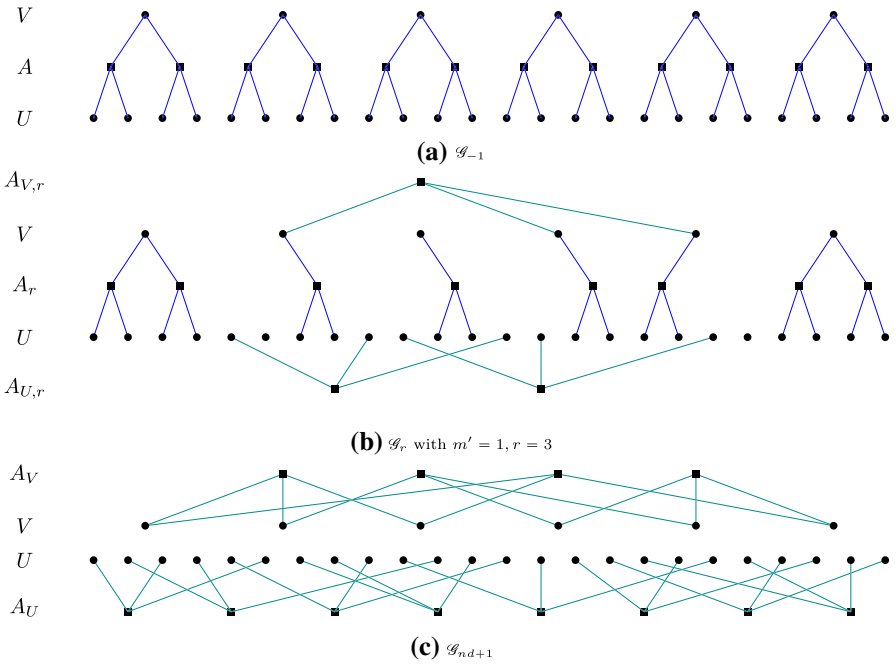



**Fig. 7** Interpolation with $d = 2, k = 3, n = 6$

By construction, $\mathscr{G}_{nd+1}$ consists of two disjoint subgraphs, which are the induced subgraphs $\mathscr{G}_U, \mathscr{G}_V$ of $U, V$ respectively. Note that $\mathscr{G}_V$ is distributed as the random graph $\mathscr{G}$ of interest, while $\mathscr{G}_U$ consists of a collection of $nd(k-1)/k = n\alpha(k-1)$ disjoint trees.

**Lemma E.6** *Under the construction above,*

$$\mathbb{E} \ln Z(\mathscr{G}_0) \geq \mathbb{E} \ln Z(\mathscr{G}_{nd}) - O_\epsilon(n^{1/3}), \tag{138}$$

*where the expectation* $\mathbb{E}$ *is over the sequence of random graphs* $(\mathscr{G}_r)_{-1 \leq r \leq nd+1}$.

***Proof*** Let $\mathscr{F}_{r,\circ}$ be the $\sigma$-field generated by $\mathscr{G}_{r,\circ}$, and write $\mathbb{E}_{r,\circ}$ for expectation conditioned on $\mathscr{F}_{r,\circ}$. One can rewrite (138) as

$$\mathbb{E} \ln \frac{Z(\mathscr{G}_0)}{Z(\mathscr{G}_{nd})} = \sum_{r=0}^{nd-1} \mathbb{E}\Delta_r, \quad \Delta_r \equiv \mathbb{E}_{r,\circ} \ln \frac{Z(\mathscr{G}_r)}{Z(\mathscr{G}_{r,\circ})} - \mathbb{E}_{r,\circ} \ln \frac{Z(\mathscr{G}_{r+1})}{Z(\mathscr{G}_{r,\circ})}.$$

In particular, $\Delta_r = 0$ if the coupling fails. Therefore it suffices to show that $\Delta_r$ is positive conditioned on $\mathsf{COUP}_{\leq r}$.[4] First we compare $\mathscr{G}_r$ and $\mathscr{G}_{r,\circ}$. Conditioned on $\mathscr{F}_{r,\circ}$, we know $\mathscr{G}_{r,\circ}$. From $\mathscr{G}_{r,\circ}$ we can obtain $\mathscr{G}_r$ by adding a single clause $a \equiv a_r$, together

[4] The event $\mathsf{COUP}_{\leq r}$ is measurable with respect to $\mathscr{F}_{r,\circ}$, since $\delta' V_{r,\circ}, \delta' U_{r,\circ}$ would remain less than $k$ if the coupling fails at an earlier iteration.

with a random label $\theta_a$ which is a fresh copy of $\theta$. To choose the unmatched edges $\delta a = (e_1, \dots, e_k)$ which are combined into the clause $a$, we take $e_1$ uniformly at random from $\delta' V_{r,\circ}$, then take $\{e_2, \dots, e_k\}$ a uniformly random subset of $\delta' U_{r,\circ}$. Let $\mu_{r,\circ}$ be the Gibbs measure on $\mathscr{G}_{r,\circ}$ (ignoring unmatched half-edges). Let $\underline{\underline{x}} \equiv (\underline{x}, \underline{x}^1, \underline{x}^2, \dots)$ be an infinite sequence of i.i.d. samples from $\mu_{r,\circ}$, and write $\langle\cdot\rangle_{r,\circ}$ for the expectation with respect to their joint law. Then

$$\mathbb{E}_{r,\circ} \ln \frac{Z(\mathscr{G}_r)}{Z(\mathscr{G}_{r,\circ})} = \mathbb{E}_{r,\circ} \ln(1 - \langle \theta(\underline{x}_{\delta a})\rangle_{r,\circ}) = \sum_{p\ge 1} \frac{1}{p} \mathscr{A}_p, \quad \mathscr{A}_p \equiv \mathbb{E}_{r,\circ}\Big[\Big\langle \prod_{\ell=1}^{p} \theta(\underline{x}^{\ell}_{\delta a})\Big\rangle_{r,\circ}\Big].$$

We have $\mathbb{E}_{r,\circ} = \mathbb{E}_a \mathbb{E}_\theta$ where $\mathbb{E}_a$ is expectation over the choice of $\delta a$, and $\mathbb{E}_\theta$ is expectation over the choice of $\theta$. Under $\mathbb{E}_a$, the edges $(e_2, \dots, e_k)$ are weakly dependent, since they are required to be distinct elements of $\delta' U_{r,\circ}$. We can consider instead sampling $e_2, \dots, e_k$ uniformly **with replacement** from $\delta' U_{r,\circ}$, so that $e_1, \dots, e_k$ are independent conditional on $\mathscr{F}_{r,\circ}$; let $\mathbb{E}_{a,\mathrm{ind}}$ denote expectation with respect to this choice of $\delta a$. Under $\mathbb{E}_{a,\mathrm{ind}}$ the chance of a collision $e_i = e_j$ $(i \le j)$ is $O(k^2/|\delta' U_{r,\circ}|)$. Recalling $1 - \theta \ge \epsilon$ almost surely, we have

$$\mathscr{A}_{p,\mathrm{ind}} \equiv \mathbb{E}_{a,\mathrm{ind}} \mathbb{E}_\theta \Big[\Big\langle \prod_{\ell=1}^{p} \theta(\underline{x}^{\ell}_{\delta a})\Big\rangle_{r,\circ}\Big] = \mathscr{A}_p + O(1)(1-\epsilon)^p \min\Big\{\frac{k^2}{|\delta' U_{r,\circ}|}, 1\Big\}.$$

Recall from (134) the product form of $\theta$, and let $\mathbb{E}_f$ denote expectation over the law of $f \equiv (f_s)_{s\in S}$. Then, with $b^p(\underline{s})$ as defined in (136), we have

$$\begin{aligned}\mathscr{A}_{p,\mathrm{ind}} &= \sum_{\underline{s}\in S^p} b^p(\underline{s}) \Big\langle \mathbb{E}_{a,\mathrm{ind}} \Big\{ \prod_{j=1}^{k} \mathbb{E}_f \Big[ \prod_{\ell=1}^{p} f_{s_\ell}(x^{\ell}_{e_j}) \Big] \Big\} \Big\rangle_{r,\circ} \\ &= \sum_{\underline{s}\in S^p} b^p(\underline{s}) \langle I_{V,\underline{s}}(\underline{\underline{x}}) I_{U,\underline{s}}(\underline{\underline{x}})^{k-1} \rangle_{r,\circ},\end{aligned}$$

where, for $W = U$ or $W = V$, we define

$$I_{W,\underline{s}}(\underline{\underline{x}}) \equiv \frac{1}{|\delta' W_{r,\circ}|} \sum_{e\in\delta' W_{r,\circ}} \mathbb{E}_f \Big[ \prod_{\ell=1}^{p} f_{s_\ell}(x^{\ell}_{e}) \Big].$$

Summing over $p \ge 1$ gives that, on the event $\mathsf{COUP}_{\le \mathrm{r}}$,

$$\begin{aligned}\mathbb{E}_{r,\circ} \ln \frac{Z(\mathscr{G}_r)}{Z(\mathscr{G}_{r,\circ})} &= \sum_{p\ge 1} \frac{1}{p} \sum_{\underline{s}\in S^p} b^p(\underline{s}) \mathbb{E}_{r,\circ} \langle I_{V,\underline{s}}(\underline{\underline{x}}) I_{U,\underline{s}}(\underline{\underline{x}})^{k-1} \rangle_{r,\circ} + \mathsf{err}_{r,1}, \\ \text{where } |\mathsf{err}_{r,1}| &\le O_\epsilon(1) \min\Big\{\frac{k^2}{|\delta' U_{r,\circ}|}, 1\Big\}.\end{aligned}$$

A similar comparison between $\mathscr{G}_{r+1}$ and $\mathscr{G}_{r,\circ}$ gives

$$\mathbb{E}_{r,\circ} \ln \frac{Z(\mathscr{G}_r)}{Z(\mathscr{G}_{r,\circ})} = \sum_{p\geq 1} \frac{1}{p} \mathbb{E}_{r,\circ}\Big[ \sum_{\underline{s}\in S^p} b^p(\underline{s}) \Big\langle \frac{k-1}{k} I_{U,\underline{s}}(\underline{\underline{x}})^k + \frac{1}{k} I_{V,\underline{s}}(\underline{\underline{x}})^k \Big\rangle_{r,\circ}\Big] + \mathsf{err}_{r,2},$$

$$|\mathsf{err}_{r,2}| \leq O_\epsilon(1) \min\left\{ \frac{k^2}{\min\{|\delta' U_{r,\circ}|, |\delta' V_{r,\circ}|\}}, 1 \right\}.$$

We now argue that the sum of the error terms $\mathsf{err}_{r,1}$, $\mathsf{err}_{r,2}$, over $0 \leq r \leq nd-1$, is small in expectation. First note that for a constant $C = C(k,\epsilon)$,

$$\sum_{r=0}^{nd-1} \mathbb{E}[\mathsf{err}_{r,1} + \mathsf{err}_{r,2}] \leq Cn\Big[ n^{-2/3} + \mathbb{P}\Big( \min\{|\delta' V_{r,\circ}|, |\delta' V_{r,\circ}|\} \leq n^{2/3} \text{ for some } r \leq nd \Big)\Big].$$

The process $(|\delta' V_{r,\circ}|)_{r\geq 0}$ is an unbiased random walk started from $m'+1 = 2n^{2/3}+1$. In each step it goes up by 1 with chance $(k-1)/k$, and down by $k-1$ with chance $1/k$; it is absorbed if it hits $k$ before time $nd-m'$. Similarly, $(|\delta' U|_{r,\circ})_{r\geq 0}$ is an unbiased random walk started from $(m'+1)(k-1)$ with an absorbing barrier at $k$. By the Azuma–Hoeffding bound, there is a constant $c = c(k)$ such that

$$\mathbb{P}(|\delta' V_{r,\circ}| \leq |\delta' V_{0,\circ}| - n^{2/3}) + \mathbb{P}(|\delta' U_{r,\circ}| \leq |\delta' U_{0,\circ}| - n^{2/3}) \leq \exp\{-cn^{1/3}\}$$

Taking a union bound over $r$ shows that with very high probability, neither of the walks $|\delta' V_{r,\circ}|$, $|\delta' U_{r,\circ}|$ is absorbed before time $nd-m'$, and (adjusting the constant $C$ as needed)

$$\sum_{r=0}^{nd-1} \mathbb{E}[\mathsf{err}_{r,1} + \mathsf{err}_{r,2}] \leq Cn^{1/3}.$$

Altogether this gives

$$\mathbb{E} \ln \frac{Z(\mathscr{G}_0)}{Z(\mathscr{G}_{nd})} - O_\epsilon(n^{1/3})$$
$$= \sum_{r=0}^{nd-1} \sum_{p\geq 1} \frac{1}{p} \sum_{\underline{s}} b^p(\underline{s}) \mathbb{E}_{r,\circ} \Big\langle I_{V,\underline{s}}(\underline{\underline{x}}) I_{U,\underline{s}}(\underline{\underline{x}})^{k-1} - \frac{k-1}{k} I_{U,\underline{s}}(\underline{\underline{x}})^{k-1} - \frac{1}{k} I_{V,\underline{s}}(\underline{\underline{x}})^{k-1} \Big\rangle_{r,\circ}.$$

Using the fact that $x^k - kxy^{k-1} + (k-1)y^k \geq 0$ for all $x, y \in \mathbb{R}$ and even $k \geq 2$, or $x, y \geq 0$ and odd $k \geq 3$ finishes the proof. □

**Corollary E.7** *In the setting of* Lemma E.6*,*

$$\mathbb{E} \ln Z(\mathscr{G}_{-1}) \geq \mathbb{E} \ln Z(\mathscr{G}_{nd+1}) - O_\epsilon(n^{2/3}),$$

*where the expectation* $\mathbb{E}$ *is over the sequence of random graphs* $(\mathscr{G}_r)_{-1\leq r\leq nd+1}$.

***Proof*** Adding or removing a clause can change the partition function by at most a multiplicative constant (depending on $\epsilon$). On the event that the coupling succeeds for all $r$,

$$\left|\ln \frac{Z(\mathscr{G}_0)}{Z(\mathscr{G}_{-1})}\right| + \left|\ln \frac{Z(\mathscr{G}_{nd+1})}{Z(\mathscr{G}_{nd})}\right| = O_\epsilon(m') = O_\epsilon(n^{2/3})\,.$$

On the event that the coupling fails, the difference is crudely $O_\epsilon(n)$. We saw in the proof of Lemma E.6 that the coupling fails with probability exponentially small in $n$, so altogether we conclude

$$\mathbb{E}\left|\ln \frac{Z(\mathscr{G}_0)}{Z(\mathscr{G}_{-1})}\right| + \mathbb{E}\left|\ln \frac{Z(\mathscr{G}_{nd+1})}{Z(\mathscr{G}_{nd})}\right| = O_\epsilon(n^{2/3})\,.$$

Combining with the result of Lemma E.6 proves the claim. □

***Proof of Theorem E.5*** In the interpolation, the initial graph $\mathscr{G}_{-1}$ consists of $n$ disjoint trees $T_v$, each rooted at a variable $v \in V$. Thus

$$n^{-1}\mathbb{E}\ln Z(\mathscr{G}_{-1}) = \mathbb{E}\ln Z(T_v) = \mathbb{E}\ln\left(\sum_{x\in\mathcal{X}} h_v(x)\prod_{a=1}^{d}\boldsymbol{u}_a(x)\right).$$

The final graph $\mathscr{G}_{nd+1}$ is comprised of two disjoint subgraphs—one subgraph $\mathscr{G}_V$ has the same law as the graph $\mathscr{G}$ of interest, while the other subgraph $\mathscr{G}_U = (U, F_U, E_U)$ consists of $n\alpha(k-1)$ disjoint trees $S_c$, each rooted at a clause $c \in A_U$. Thus

$$n^{-1}\mathbb{E}\ln Z(\mathscr{G}_{nd+1}) = \alpha(k-1)\mathbb{E}\ln Z(S_c) + n^{-1}\mathbb{E}\ln Z(\mathscr{G}) = \alpha(k-1)\mathbb{E}\ln \boldsymbol{u}_0 + F_n\,.$$

The theorem follows by substituting these into the bound of Corollary E.7. □

### E.3. Proof of 1RSB bound

For the proof of Theorem E.3, we take $\mathscr{G}_{-1}$ as before and modify it as follows. Where previously each $u \in U$ had spin value $x_u \in \mathcal{X}$, it now has the augmented spin $(x_u, \gamma_u)$ where $\gamma$ goes over the positive integers. Let $\underline{\gamma} \equiv (\gamma_u)_u$. Next, instead of labeling $u$ with $(h_u, \eta_u, \rho_u)$ as before, we now label it with $(h_u, \eta_u, (\rho_u^\gamma)_{\gamma\geq 1})$ where $(\rho_u^\gamma)_{\gamma\geq 1}$ is an infinite sequence of i.i.d. samples from $\eta_u$. Lastly, we join all variables in $U$ to a new clause $a_*$ (Fig. 8), which is labelled with the function

$$\varphi_{a_*}(\underline{\gamma}) = \sum_{\gamma\geq 1} z_\gamma \prod_{u\in U}\mathbf{1}\{\gamma_u = \gamma\}$$

for some sequence of (random) weights $(z_\gamma)_{\gamma\geq 1}$. Let $\mathscr{H}_{-1}$ denote the resulting graph.

Given $\mathscr{H}_{-1}$, let $\mu_{\mathscr{H}_{-1}}$ be the associated Gibbs measure on configurations $(\underline{\gamma}, \underline{x})$. Due to the definition of $\varphi_{a_*}$, the support of $\mu_{\mathscr{H}_{-1}}$ contains only those configurations

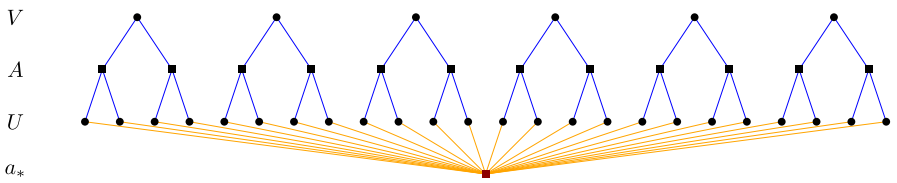



**Fig. 8** $\mathscr{H}_{-1}$

where all the $\gamma_u$ share a common value $\gamma$, in which case we denote $(\underline{\gamma}, \underline{x}) \equiv (\gamma, \underline{x})$. Explicitly,

$$\mu_{\mathscr{H}_{-1}}(\gamma, \underline{x}) = \frac{1}{Z(\mathscr{H}_{-1})} z_\gamma \prod_{v \in V} h_v(x_v) \prod_{a \in A} [1 - \theta_a(\underline{x}_{\delta a})] \prod_{u \in U} (\rho^\gamma h)_u(x_u) \,.$$

We can then define an interpolating sequence $\mathscr{H}_{-1}, \ldots, \mathscr{H}_{nd+1}$ precisely as in the proof of Theorem E.5, leaving $a_*$ untouched. Let $\mathscr{G}_r$ denote the graph $\mathscr{H}_r$ without the clause $a_*$, and let $Z_\gamma(\mathscr{G}_r)$ denote the partition function on $\mathscr{G}_r$ restricted to configurations where $\gamma_u = \gamma$ for all $u$. Then, for each $0 \le r \le nd + 1$,

$$Z(\mathscr{H}_r) = \sum_\gamma z_\gamma Z_\gamma(\mathscr{G}_r) \,.$$

The proofs of Lemma E.6 and Corollary E.7 carry over to this setting with essentially no changes, giving

**Corollary E.8** *Under the assumptions above,*

$$\mathbb{E} \ln Z(\mathscr{H}_{-1}) \ge \mathbb{E} \ln Z(\mathscr{H}_{nd+1}) - O_\epsilon(n^{2/3}),$$

*where the expectation* $\mathbb{E}$ *is over the sequence of random graphs* $(\mathscr{H}_r)_{-1 \le r \le nd+1}$.

The result of Corollary E.8 applies for any $(z_\gamma)_{\gamma \ge 1}$. Now take $(z_\gamma)_{\gamma \ge 1}$ to be a Poisson–Dirichlet process with parameter $\lambda \in (0, 1)$.[5] The process has the following invariance property (see e.g. [41, Ch. 2]):

**Proposition E.9** *Let* $(z_\gamma)_{\gamma \ge 1}$ *be a Poisson–Dirichlet process with parameter* $\lambda \in (0, 1)$. *Independently, let* $(\xi_\gamma)_{\gamma \ge 1}$ *be a sequence of i.i.d. positive random variables with finite second moment. Then the two sequences* $(z_\gamma \xi_\gamma)_{\gamma \ge 1}$ *and* $(z_\gamma (\mathbb{E}\xi_1^\lambda)^{1/\lambda})_{\gamma \ge 1}$ *have the same distribution, and consequently*

$$\mathbb{E} \ln \sum_{\gamma \ge 1} z_\gamma \xi_\gamma = \frac{1}{\lambda} \ln \mathbb{E}\xi^\lambda \,.$$

[5] That is to say, let $(w_\gamma)_{\gamma \ge 1}$ be a Poisson point process on $\mathbb{R}_{>0}$ with intensity measure $w^{-(1+\lambda)}\, dw$. Let $W$ denote their sum, which is finite almost surely. Assume the points of $w_\gamma$ are arranged in decreasing order, and write $z_\gamma \equiv w_\gamma / W$. Then $(z_\gamma)_{\gamma \ge 1}$ is distributed as a Poisson–Dirichlet process with parameter $\lambda$.

***Proof of Theorem E.3*** Consider $\underline{Z}(\gamma) \equiv (Z_\gamma(\mathscr{G}_r))_{-1\le r\le nd+1}$. If we condition on everything else except for the $\rho$'s, then $(\underline{Z}(\gamma))_{\gamma\ge 1}$ is an i.i.d. sequence indexed by $\gamma$. Let $\mathbb{E}_{z,\rho}$ denote expectation over the $z$'s and $\rho$'s, conditioned on all else: then applying Proposition E.9 gives

$$n^{-1}\mathbb{E}\ln Z(\mathscr{H}_{-1}) = (n\lambda)^{-1}\mathbb{E}\ln\mathbb{E}_{z,\rho}[Z(\mathscr{G}_{-1})^\lambda] = \lambda^{-1}\mathbb{E}\ln\mathbb{E}_{z,\rho}\Big[\Big(\sum_{x\in\mathfrak{X}} h(x)\prod_{a=1}^{d}\boldsymbol{u}_a(x)\Big)^\lambda\Big],$$
$$n^{-1}\mathbb{E}\ln Z(\mathscr{H}_{nd+1}) = F_n + \lambda^{-1}\mathbb{E}\ln\mathbb{E}_{z,\rho}[(\boldsymbol{u}_0)^\lambda].$$

Combining with Corollary E.8 proves the result. □

### E.4. Extension to higher levels of RSB

We finally explain that Theorem E.3 can be extended relatively easily to cover the scenario of $r$-step replica symmetry breaking. Before stating the result, we define some notations (mainly following notation of [41, §2.3]). Let $\mathbb{N}$ be the set of positive integers and $\mathbb{N}^r$ be its $r$-fold product; in particular, $\mathbb{N}^0 \equiv \{\varnothing\}$. We consider arrays indexed by the set

$$\mathcal{A} \equiv \bigcup_{p=0}^{r}\mathbb{N}^p\,.$$

We view $\mathcal{A}$ as a depth-$r$ infinitary tree rooted at $\varnothing$. For $0 \le p \le r-1$, each vertex $\gamma = (\gamma_1,\dots,\gamma_p) \in \mathbb{N}^p$ has children $\gamma n \equiv (\gamma_1,\dots,\gamma_s,n) \in \mathbb{N}^{s+1}$. The leaves of the tree are in the last level $\mathbb{N}^r$. For $\gamma\in\mathbb{N}^p$ write $|\gamma| \equiv p$, and let $\mathsf{p}(\gamma)$ be the path between the root and $\gamma$ (not inclusive):

$$\mathsf{p}(\gamma) \equiv \Big\{\gamma_1, (\gamma_1,\gamma_2),\dots,(\gamma_1,\dots,\gamma_{p-1})\Big\}\,.$$

Fix a sequence of parameters $\underline{m} = (m_1,\dots,m_r)$ satisfying

$$0 < m_0 < \cdots < m_{r-1} < 1. \tag{139}$$

For each $\gamma\in\mathcal{A}\backslash\mathbb{N}^r$, let $\Pi_\gamma$ be (independently of all else) a Poisson–Dirichlet point process with parameter $m_{|\gamma|}$. Let $(u_{\gamma n})_{n\in\mathbb{N}}$ be the points of $\Pi_\gamma$ arranged in decreasing order. As $\gamma$ goes over all of $\mathcal{A}\setminus\mathbb{N}^r$, we obtain an array $(u_\beta)_{\beta\in\mathcal{A}\setminus\mathbb{N}^0}$. Let

$$w_\gamma \equiv \prod_{\beta\in\mathsf{p}(\gamma)} u_\beta\,.$$

The **Ruelle probability cascade of parameter** $\underline{m}$ (hereafter RPC($\underline{m}$)) is defined as the $\mathbb{N}^r$-indexed array

$$\nu_\gamma \equiv \frac{w_\gamma}{\sum_{\beta\in\mathbb{N}^r} w_\beta}.$$

For the validity of the definition, see for instance [41, Lem. 2.4]. As in the 1RSB setting, we plan to apply Theorem E.5 to the modified graph $\mathscr{H}_{-1}$, where we "glue" multiple weighted copies of $\mathscr{G}_{-1}$'s together via the extra clause $a_\star$. The only difference is that now the copies of $\mathscr{G}_{-1}$ are indexed by $\gamma \in \mathbb{N}^r$ instead of $\mathbb{N}$. More precisely, the extra spin at each vertex $u \in U$ will take a value $\gamma \in \mathbb{N}^r$; the label at each vertex $u \in U$ will be $(h_u, \eta_u, (\rho_u^\gamma)_{\gamma\in\mathcal{A}})$; and the function at $a_\star$ will be

$$\phi_{a_\star}(\underline{\gamma}) = \sum_{\gamma\in\mathbb{N}^r} z_\gamma \prod_{u\in U} \mathbf{1}\{\gamma_u = \gamma\}, \tag{140}$$

where $(z_\gamma)_{\gamma\in\mathbb{N}^r}$ is a $\mathbb{N}^r$-indexed random array representing the weight of copy $\gamma \in \mathbb{N}^r$. In the proof, we will choose $(z_\gamma)_{\gamma\in\mathbb{N}^r}$ according to the RPC($\underline{m}$) law.

We now specify the labels $(\rho^\gamma)_{\gamma\in\mathcal{A}}$ that will be used in the proof. Recall that $\mathcal{M}_0$ is the space of probability measures on the alphabet $\mathcal{X}$. We recursively define $\mathcal{M}_r$, for $1 \le r \le p$, to be the space of probability measures on $\mathcal{M}_{r-1}$. Now fix an element $\rho^\varnothing = \zeta \in \mathcal{M}_r$. For each $0 \le p \le r-1$ and $\gamma \in \mathbb{N}^p$, suppose inductively that we have constructed $\rho^\gamma \in \mathcal{M}_{r-p}$. We then take $\rho^{\gamma n} \in \mathcal{M}_{r-p-1}$ for $n \in \mathbb{N}$ as i.i.d. samples $\rho^\gamma \in \mathcal{M}_{r-p}$ in $\mathcal{M}_{r-p-1}$. The process terminates with the construction of $\rho^\gamma \in \mathcal{X}$ for each $\gamma \in \mathbb{N}^r$. Define the $\sigma$-field

$$\mathcal{F}_p \equiv \sigma\left(\left((\rho^\gamma)_{\gamma\in\mathbb{N}^s}\right)_{s\le p}\right),$$

and write $\mathbb{E}_p$ for expectation conditional on $\mathcal{F}_p$. For any deterministic function $V(u, \rho)$, any random variable $U$ independent of $(\rho^\gamma)_{\gamma\in\mathbb{N}^r}$, and any sequence of parameters $\underline{m}$ satisfying (139), consider the random array $(V^\gamma)_{\gamma\in\mathbb{N}^r} \equiv (V(U, \rho^\gamma))_{\gamma\in\mathbb{N}^r}$. Let $\mathsf{T}_r(V) = V(U, \rho^{\underline{1}})$, and for $0 \le p \le r-1$ let

$$\mathsf{T}_p(V) = \left\{\mathbb{E}_p\left(\mathsf{T}_{p+1}(V)\right)^{m_p}\right\}^{1/m_p}$$

The resulting operator $\mathsf{T}_0$ depends implicitly on the distribution of $U$, measure $\rho^\varnothing \in \mathcal{M}_r$ and parameter $\underline{m}$. The following lemma is a well-known property of the RPC.

**Lemma E.10** ([43, Prop. 2]) *Let $(z_\gamma)_{\gamma\in\mathbb{N}^r}$ be the RPC with parameter $\underline{m}$. Under the notations above,*

$$\mathbb{E}\ln \mathsf{T}_0(V) = \mathbb{E}\ln \sum_{\gamma\in\mathbb{N}^r} z_\gamma V^\gamma.$$

The next result generalizes Theorem E.3.

**Theorem E.11** *Consider a (random) Gibbs measure (137) satisfying assumptions (134)–(136). Write* $(h\rho)_{a,j}(x) \equiv h_{a,j}(x)\rho_{a,j}(x)$. *For each* $a \in F$ *we define*

$$\boldsymbol{u}_a(x) \equiv \sum_{\underline{x} \in \mathcal{X}^k} \mathbf{1}\{x_1 = x\}[1 - \theta_a(\underline{x})] \prod_{j=2}^{k} (h\rho)_{a,j}(x_j),$$

$$\boldsymbol{u}_a \equiv \sum_{\underline{x} \in \mathcal{X}^k} [1 - \theta_a(\underline{x})] \prod_{j=1}^{k} (h\rho)_{a,j}(x_j)\,.$$

*Note that* $\boldsymbol{u}_a(x)$ *and* $\boldsymbol{u}_a$ *are deterministic functions of the variables* $\big(\theta_a, (h_{a,j})_{j\in[k]}, (\rho_{a,j})_{j\in[k]}\big)$. *Let*

$$\boldsymbol{v} \equiv \sum_{x \in \mathcal{X}} h(x) \prod_{a=1}^{d} \boldsymbol{u}_a(x).$$

*For any* $\zeta \in \mathcal{M}_r$ *and sequence* $\underline{m}$ *satisfying (139), let* $(\rho^\gamma)_{\gamma\in\mathbb{N}^r}$ *be constructed as above, and let* $(\rho^\gamma_{a,j})_{\gamma\in\mathbb{N}^r}$ *be i.i.d. copies indexed by* $(a, j)$. *Define* $\mathsf{T}_0$ *similarly as above, using the* $\sigma$*-fields*

$$\mathcal{F}_p = \sigma\Big((\rho^\gamma_{a,j})_{\gamma\in\mathbb{N}^s} : a \in F,\, j \in [k],\, s \le p\Big)\,.$$

*Then the nonasymptotic free energy* $F_n \equiv n^{-1}\mathbb{E}_n \ln Z(\mathscr{G})$ *satisfies the bound*

$$F_n \le \mathbb{E} \ln \mathsf{T}_0(\boldsymbol{v}) - (k-1)\alpha \mathbb{E} \ln \mathsf{T}_0(\boldsymbol{u}_0) + O_\epsilon\Big(\frac{1}{n^{1/3}}\Big)$$

*where* $\mathbb{E}$ *denotes the expectation over* $(\theta_a)_{a\in F}$ *and* $(h_{a,j})_{a\in F, j\in[k]}$.

***Proof*** As outlined above, we consider the modified graph $\mathscr{H}_{-1}$ where each vertex $u \in U$ is independently labeled with $(h_u, \eta_u, (\rho^\gamma_u)_{\gamma\in\mathcal{A}})$ and the extra clause $a_\star$ is labeled with the function defined in (140). In this setting, each $u \in U$ has spin value $(\gamma, x) \in \mathbb{N}^r \times \mathcal{X}$. Since we are interested only in configurations $(\underline{\gamma}, \underline{x})$ such $\gamma_u \equiv \gamma$ for all $u \in U$, we write $(\gamma, \underline{x})$ instead of $(\underline{\gamma}, \underline{x})$ and define the Gibbs measure as

$$\mu_{\mathscr{H}_{-1}}(\gamma, \underline{x}) = \frac{1}{Z(\mathscr{H}_{-1})} z_\gamma \prod_{v\in V} h_v(x_v) \prod_{a\in A} [1 - \theta_a(\underline{x}_{\delta a})] \prod_{u\in U} (\rho^\gamma h)_u(x_u)\,.$$

Sample the weights $(z_\gamma)_{\gamma\in\mathbb{N}^r}$ according the law RPC($\underline{m}$). The result then follows by the proof of Theorem E.3, with Lemma E.10 replacing the role of Proposition E.9. □

## References


1. Achlioptas, D., Coja-Oghlan, A.: Algorithmic barriers from phase transitions. In: Proceedings of 49th FOCS, pp. 793–802. IEEE (2008)
2. Achlioptas, D., Coja-Oghlan, A., Ricci-Tersenghi, F.: On the solution-space geometry of random constraint satisfaction problems. Random Struct. Algorithm **38**(3), 251–268 (2011)
3. Achlioptas, D., Moore, C.: Random $k$-SAT: two moments suffice to cross a sharp threshold. SIAM J. Comput. **36**(3), 740–762 (2006)
4. Abbe, E., Montanari, A.: On the concentration of the number of solutions of random satisfiability formulas. arXiv:1006.3786, (2010)
5. Abbe, E., Montanari, A.: On the concentration of the number of solutions of random satisfiability formulas. Random Struct. Algorithm **45**(3), 362–382 (2014)
6. Achlioptas, D., Ricci-Tersenghi, F.: On the solution-space geometry of random constraint satisfaction problems. In: Proceedings of 38th STOC, pp. 130–139. ACM, New York (2006)
7. Aizenman, M., Sims, R., Starr, S.L.: Extended variational principle for the Sherrington-Kirkpatrick spin-glass model. Phys. Rev. B **68**(21), 214403 (2003)
8. Bapst, V., Coja-Oghlan, A.: The condensation phase transition in the regular $k$-SAT model. In: Approximation, Randomization, and Combinatorial Optimization. Algorithms and techniques, volume 60 of LIPIcs. Leibniz Int. Proc. Inform., pages Art. No. 22, 18. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern (2016)
9. Bapst, V., Coja-Oghlan, A.: Harnessing the Bethe free energy. Random Struct. Algorithm **49**(4), 694–741 (2016)
10. Bapst, V., Coja-Oghlan, A., Hetterich, S., Raßmann, F., Vilenchik, D.: The condensation phase transition in random graph coloring. Commun. Math. Phys. **341**(2), 543–606 (2016)
11. Bapst, V., Coja-Oghlan, A., Raßmann, F.: A positive temperature phase transition in random hypergraph 2-coloring. Ann. Appl. Probab. **26**(3), 1362–1406 (2016)
12. Bayati, M., Gamarnik, D., Tetali, P.: Combinatorial approach to the interpolation method and scaling limits in sparse random graphs. Ann. Probab. **41**(6), 4080–4115 (2013)
13. Braunstein, A., Mézard, M., Zecchina, R.: Survey propagation: an algorithm for satisfiability. Random Struct. Algorithm **27**(2), 201–226 (2005)
14. Coja-Oghlan, A., Krzą kała, F., Perkins, W., Zdeborová, L.: Information-theoretic thresholds from the cavity method. In: Proceedings of 49th STOC, pp. 146–157. ACM, New York (2017)
15. Coja-Oghlan, A., Krzakala, F., Perkins, W., Zdeborová, L.: Information-theoretic thresholds from the cavity method. Adv. Math. **333**, 694–795 (2018)
16. Coja-Oghlan, A., Perkins, W.: Belief propagation on replica symmetric random factor graph models. Ann. Inst. Henri Poincaré D **5**(2), 211–249 (2018)
17. Coja-Oghlan, A., Panagiotou, K.: Catching the $k$-NAE-SAT threshold. In: Proceedings of 44th STOC, pp. 899–907. ACM, New York (2012)
18. Coja-Oghlan, A., Panagiotou, K.: The asymptotic $k$-SAT threshold. Adv. Math. **288**, 985–1068 (2016)
19. Coja-Oghlan, A., Perkins, W., Skubch, K.: Limits of discrete distributions and Gibbs measures on random graphs. Eur. J. Combin. **66**, 37–59 (2017)
20. Coja-Oghlan, A., Zdeborová, L.: The condensation transition in random hypergraph 2-coloring. In: Proceedings of 23rd SODA, pp. 241–250. ACM, New York (2012)
21. Decelle, A., Krzą kała, F., Moore, C., Zdeborová, L.: Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. Phys. Rev. E **84**(6), 66106 (2011)
22. Ding, J., Sly, A., Sun, N.: Satisfiability threshold for random regular NAE-SAT. In: Proceedings of 46th STOC. ACM, New York (2014)
23. Ding, J., Sly, A., Sun, N.: Proof of the satisfiability conjecture for large $k$. In: Proceedings of 47th STOC, pp. 59–68. ACM, New York (2015)
24. Ding, J., Sly, A., Sun, N.: Maximum independent sets on random regular graphs. Acta Math. **217**(2), 263–340 (2016)
25. Ding, J., Sly, A., Sun, N.: Satisfiability threshold for random regular NAE-SAT. Commun. Math. Phys. **341**(2), 435–489 (2016)
26. Franz, S., Leone, M.: Replica bounds for optimization problems and diluted spin systems. J. Stat. Phys. **111**(3–4), 535–564 (2003)
27. Gamarnik, D.: Right-convergence of sparse random graphs. Probab. Theory Relat. Fields **160**(1–2), 253–278 (2014)

28. Gerschenfeld, A., Montanari, A.: Reconstruction for models on random graphs. In: Proceedings of 48th FOCS, pp. 194–204. IEEE (2007)
29. Guerra, F., Toninelli, F.L.: Infinite volume limit and spontaneous replica symmetry breaking in mean field spin glass models. Ann. Henri Poincaré **4**(suppl. 1), S441–S444 (2003)
30. Guerra, F.: Broken replica symmetry bounds in the mean field spin glass model. Commun. Math. Phys. **233**(1), 1–12 (2003)
31. Krząkała, F., Montanari, A., Ricci-Tersenghi, F., Semerjian, G., Zdeborová, L.: Gibbs states and the set of solutions of random constraint satisfaction problems. Proc. Natl. Acad. Sci. USA **104**(25), 10318–10323 (2007)
32. Massoulié, L.: Community detection thresholds and the weak Ramanujan property. In: Proceedings of 46th STOC, pp. 694–703. ACM, New York (2014)
33. Mézard, M., Montanari, A.: Information, Physics, and Computation. Oxford University Press, Oxford, Oxford Graduate Texts (2009)
34. Maneva, E., Mossel, E., Wainwright, M.J.: A new look at survey propagation and its generalizations. J. ACM **54**(4), 41 (2007)
35. Mézard, M., Mora, T., Zecchina, R.: Clustering of solutions in the random satisfiability problem. Phys. Rev. Lett. **94**(19), 197205 (2005)
36. Mertens, S., Mézard, M., Zecchina, R.: Threshold values of random $k$-SAT from the cavity method. Random Struct. Algorithm **28**(3), 340–373 (2006)
37. Mossel, E., Neeman, J., Sly, A.: Reconstruction and estimation in the planted partition model. Probab. Theory Relat. Fields **162**(3–4), 431–461 (2015)
38. Montanari, A., Ricci-Tersenghi, F., Semerjian, G.: Clusters of solutions and replica symmetry breaking in random $k$-satisfiability. J. Stat. Mech. **2008**(04), P04004 (2008)
39. Montanari, A., Restrepo, R., Tetali, P.: Reconstruction and clustering in random constraint satisfaction problems. SIAM J. Discrete Math. **25**(2), 771–808 (2011)
40. Nam, D., Sly, A., Sohn, Y.: One-step replica symmetry breaking of random regular nae-sat. arXiv:2011.14270, (2020)
41. Panchenko, D.: The Sherrington0-Kirkpatrick Model. Springer Monographs in Mathematics, Springer, New York (2013)
42. Parisi, G.: On local equilibrium equations for clustering states. arXiv:cs/0212047, (2002)
43. Panchenko, D., Talagrand, M.: Bounds for diluted mean-fields spin glass models. Probab. Theory Relat. Fields **130**(3), 319–336 (2004)
44. Zdeborova, L., Krząkała, F.: Phase transitions in the coloring of random graphs. Phys. Rev. E **76**(3), 31131 (2007)