THE NUMBER OF SOLUTIONS FOR RANDOM REGULAR NAE-SAT

ALLAN SLY*, NIKE SUN†, AND YUMENG ZHANG
University of California, Berkeley

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.

1. INTRODUCTION

In a general random constraint satisfaction problem (CSP), there are n variables taking values in a finite alphabet $\mathcal{X}$, subject to a random collection 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 major obstacles to mathematical analysis. On this front, major advances were achieved by statistical physicists, who developed powerful analytic heuristics to shed light on the behavior of random CSPs ([KMR+07] and references therein). Their insights and methods are fundamental to the current understanding of random CSPs.

One prominent application of the physics heuristic is in giving explicit predictions for the locations of satisfiability thresholds in a large class of random CSPs ([MMZ06] and others). Some of these thresholds are established rigorously in recent works [DSS14, DSS13, DSS15]. However, this 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 Section 2.) See [AM06] 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

Date: 28 April 2016. Research supported in part by *NSF DMS-1208338, DMS-1352013, Sloan Fellowship, and †NSF MSPRF.

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, $\alpha = d/k$. The *partition function* of the model, denoted $Z \equiv Z_n$, is simply the number of valid NAE-SAT assignments for an instance on n variables. It is conjectured that for each $k \geqslant 3$, 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 (as $n \to \infty$, with k fixed). This has been proved [DSS14] for all k exceeding an absolute constant k_0, together with an explicit formula for α_{sat} which matches the physics prediction. The exact formula is rather intricate so we omit it here, and note only its approximate value

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

where ϵ_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*

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

One of the intriguing predictions from the physics analysis [ZK07, MRS08] is that there is a critical value α_{cond} strictly below α_{sat}, such that $\mathsf{f}(\alpha)$ and $\mathsf{f}^{\text{RS}}(\alpha)$ agree up to $\alpha = \alpha_{\text{cond}}$ and diverge thereafter. Since f^{RS} is analytic, f must be non-analytic at α_{cond}. This is the *condensation* or *Kauzmann transition*, and will be further described below. For $\alpha \in (\alpha_{\text{cond}}, \alpha_{\text{sat}})$ it is conjectured that $\mathsf{f}(\alpha)$ takes a value $\mathsf{f}^{\text{1RSB}}(\alpha)$ strictly below $\mathsf{f}^{\text{RS}}(\alpha)$. The function $\mathsf{f}^{\text{1RSB}}(\alpha)$ is explicit, although not extremely simple: it is derived via the heuristic of *one-step replica symmetry breaking* (1RSB), and is presented below in Definition 1.3. Our main result is to prove this prediction for large k.

Theorem 1. *In random regular k-NAE-SAT with $k \geqslant k_0$, for all $\alpha < \alpha_{\text{sat}}(k)$ 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 [DSS14]). The result of Theorem 1 is already proved [DSS14] for $\alpha \leqslant \alpha_{\text{lbd}} \equiv (2^{k-1} - 2)\ln 2$, so we restrict our attention to $\alpha \in (\alpha_{\text{lbd}}, \alpha_{\text{sat}})$, which is a strict superset of the condensation regime $(\alpha_{\text{cond}}, \alpha_{\text{sat}})$. Of course, for $\alpha > \alpha_{\text{sat}}$, we already know $\mathsf{f}(\alpha) = 0$. The case $\alpha = \alpha_{\text{sat}}$ can arise only if $d_{\text{sat}}(k) \equiv k\alpha_{\text{sat}}(k)$ is integer-valued for some k. We have no reason to believe that this ever occurs; if however it does miraculously occur then the probability for $Z > 0$ is bounded away from both zero and one. In this case, our methods would show that $Z^{1/n}$ does not concentrate around a single value but rather on two values, zero and $\lim_{\alpha \uparrow \alpha_{\text{sat}}} \mathsf{f}^{\text{1RSB}}(\alpha)$.

The condensation transition has been actively studied in recent work. The existence of a condensation phenomenon was first established for random NAE-SAT [CP12], and has since been found in random regular NAE-SAT and independent set [DSS14, DSS13]. It has been demonstrated to occur even at positive temperature in the problem of hypergraph bicoloring (which is very similar to NAE-SAT) [BCR14]. However, determining the precise location of α_{cond} is challenging, and was first achieved for the random graph coloring model [BCH+16]

by an impressive and technically challenging analysis. Subsequent work pinpoints α_{cond} for random regular k-SAT (which again is very similar to NAE-SAT) [BC15a]. 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 statistical physics heuristic, the random regular NAE-SAT model has a single level of replica symmetry breaking (1RSB). We refer the reader to [MM09, Ch. 19] for an expository account. We now summarize some of the key phenomena that are predicted from the 1RSB framework [ZK07, KMR+07, MRS08]. While part of the following discussion remains conjectural, much of it is rigorously established by the present paper. For this discussion we focus on the leading exponential terms and ignore $\exp\{o(n)\}$ corrections.

Recall that we consider NAE-SAT model with k fixed and parameter $\alpha = d/k$. Abbreviate $\mathtt{0} \equiv$ TRUE, $\mathtt{1} \equiv$ FALSE. For small α, almost all of the solutions lie in a single well-connected subset of $\{\mathtt{0},\mathtt{1}\}^n$. This holds until a *clustering transition* α_{d}, above which the solution space becomes broken up into exponentially many well-separated components, or *clusters*. For k large, α_{d} is very small relative to α_{sat}. For α above α_{d}, the number of clusters of size $\exp\{ns\}$ has mean value $\exp\{n\Sigma(s;\alpha)\}$, and further is concentrated about this mean; Σ is the "cluster complexity function." It is common to abbreviate $\Sigma(s) \equiv \Sigma(s;\alpha)$. Summing this prediction over cluster sizes s gives that the total number Z of NAE-SAT solutions has mean

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

where $s_1 = \arg\max[s+\Sigma(s)]$, and we write $\doteq$ to indicate equality up to $\exp\{o(n)\}$ factors. It is predicted that Σ is continuous and strictly concave in s, and also that $s+\Sigma(s)$ has a unique maximizer s_1 with $\Sigma'(s_1) = -1$. Note that we have the dependence $s_1 = s_1(\alpha)$, and $\Sigma(s_1) = \Sigma(s_1(\alpha);\alpha)$.

Under the 1RSB framework, physicists propose an explicit (conjectural) formula for Σ. For NAE-SAT and related models, this explicit calculation reveals another critical value $\alpha_{\text{cond}} \in (\alpha_{\text{d}}, \alpha_{\text{sat}})$, characterized as

$$\alpha_{\text{cond}} = \inf\{\alpha \geqslant \alpha_{\text{d}} : \Sigma(s_1(\alpha);\alpha) < 0\}.$$

For $\alpha > \alpha_{\text{cond}}$, $\mathbb{E}Z$ is dominated by clusters of size $\exp\{ns_1\}$, whose mean number $\exp\{n\Sigma(s_1)\}$ is exponentially small, meaning they are highly unlikely to appear in a typical realization. Instead, a typical realization is dominated by clusters of size $s_{\max}$ where

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

Since $\Sigma(s_{\max}) = 0$, it follows that with high probability

$$Z \doteq \exp\{n[s_{\max}+\Sigma(s_{\max})]\} = \exp\{ns_{\max}\}.$$

According to this picture, we will have (with high probability) $Z \doteq \mathbb{E}Z$ for $\alpha \leqslant \alpha_{\text{cond}}$, and $Z \ll \mathbb{E}Z$ for $\alpha > \alpha_{\text{cond}}$. Thus, for $\alpha > \alpha_{\text{cond}}$, the first moment $\mathbb{E}Z$ fails to capture the typical behavior of Z. This difficulty persists up to and beyond the satisfiability threshold

$$\alpha_{\text{sat}} = \inf\{\alpha \geqslant \alpha_{\text{cond}} : \max_s \Sigma(s;\alpha) < 0\}$$

— indeed, it is well known that there is a non-trivial interval $(\alpha_{\text{sat}}, \alpha_1)$ in which $\mathbb{E}Z \gg 1$ even though $Z = 0$ with high probability.

1.3. **The tilted cluster partition function.** Once the function $\Sigma(s;\alpha)$ is determined, it becomes straightforward to derive α_{cond}, α_{sat}, and $\mathsf{f}(\alpha)$. However, prior works have not taken the approach of actually computing Σ. Indeed, α_{sat} was determined [DSS14] by an analysis involving only $\max_s \Sigma(s;\alpha)$, which contains less information than the full curve Σ. In related models, the determination of α_{cond} [BCH+16, BC15a] also avoids Σ, going instead through the so-called "planted model." In order to obtain Σ, consider the λ-tilted partition function

$$\boldsymbol{Z}_\lambda \equiv \sum_{\boldsymbol{\gamma}} |\boldsymbol{\gamma}|^\lambda \tag{3}$$

where the sum is taken over all clusters $\boldsymbol{\gamma}$. According to the physics heuristic as described above, $\mathbb{E}\boldsymbol{Z}_\lambda \doteq \exp\{n\mathfrak{F}(\lambda)\}$ where $\mathfrak{F}$ is the Legendre dual of $-\Sigma$:

$$\mathfrak{F}(\lambda) \equiv (-\Sigma)^\star(\lambda) \equiv \max_s[\lambda s + \Sigma(s)].$$

The physics approach to computing Σ is to first compute $\mathfrak{F}$, and then set $\Sigma = -\mathfrak{F}^\star$. Note that by differentiating $\mathfrak{F}(\lambda) = n^{-1}\ln \mathbb{E}\boldsymbol{Z}_\lambda$ we find that $\mathfrak{F}$ is convex in λ, so the resulting Σ will indeed be concave.

The computation of $\mathfrak{F}(\lambda)$ may seem at first glance quite intractable. Indeed, the reason for NAE-SAT solutions to occur in clusters is that a typical solution has a positive density of variables which are *free*, meaning their value can be changed without violating any clause. Each cluster (connected component of NAE-SAT solutions) may be a complicated subset of $\{\mathtt{0},\mathtt{1}\}^n$ — changing the value at one free variable may affect whether its neighbors are free, so a cluster need not be a simple subcube of $\{\mathtt{0},\mathtt{1}\}^n$. We then wish to sum over the cluster sizes raised to non-integer powers.

However, in the regime of interest $\alpha \geqslant \alpha_{\text{lbd}}$ (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 can be encoded by a configuration $\underline{x} \in \{\mathtt{0},\mathtt{1},\mathtt{f}\}^n$ (representing its circumscribed subcube, so $x_v = \mathtt{f}$ indicates a free variable) with no essential loss of information. We call $\underline{x}$ the *frozen configuration* representing the cluster. It turns out that the frozen configurations can be regarded as the solutions of a certain CSP lifted from the original NAE-SAT problem — so the physics heuristics can be applied again to the new CSP. Variations on this idea appear in several places in the physics literature; in the specific context of random CSPs we refer to [Par02, BMZ05, MMW07].

Analyzing the *number* of frozen configurations — corresponding to (3) with $\lambda = 0$ — leads to the sharp satisfiability threshold for this model [DSS14]. To analyze (3) for general λ requires a deeper investigation of the arrangement of free and frozen variables in the frozen configurations $\underline{x}$. In fact, the majority of free variables are simply isolated vertices. A smaller fraction occur in linked pairs, and 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 $z(\boldsymbol{T})$ count the number of NAE-SAT assignments on $\boldsymbol{T}$ which are consistent with the frozen boundary. Then the total size of the cluster represented by $\underline{x}$ is simply the product of $z(\boldsymbol{T})$ over all the free components $\boldsymbol{T}$ of $\underline{x}$.

The random NAE-SAT graph has few short cycles, so almost all of the free components are *trees*, and so their weights $z(\boldsymbol{T})$ can be evaluated recursively by the method of *belief propagation* (BP). To implement this, we must replace variable spins by "messages," which are indexed by the directed edges of the graph and so are more natural for tree recursions.

The message $\mathtt{m}_{v\to a}$ from variable v to clause a represents the state of v "in absence of a." It is also necessary to introduce a richer alphabet of symbols for these messages, replacing $\{\mathtt{0},\mathtt{1},\mathtt{f}\}$ with probability measures on $\{\mathtt{0},\mathtt{1}\}$ (where any non-degenerate measure will project to $\mathtt{f}$). Thus the message $\mathtt{m}_{v\to a}$ represents the distribution at v (within the cluster) in absence of clause a. The messages are related to one another via local consistency equations, which are precisely the BP equations. The configuration $\underline{\mathtt{m}}$ encodes the same cluster as $\underline{x}$, with the key advantage that *the cluster size can be readily deduced from* $\underline{\mathtt{m}}$*, as a certain product of local functions.* For the cluster size raised to power λ, simply raise each local function to power λ. Thus the configurations $\underline{\mathtt{m}}$ with λ-tilted weights form a *spin system* (Markov random field), whose partition function is the quantity of interest (3). The new spin system is sometimes termed the "auxiliary model" [MM09, Ch. 19].

1.4. **One-step replica symmetry breaking.** Above, we asserted informally that each BP solution $\underline{\mathtt{m}}$ encodes a cluster of NAE-SAT solutions. An important caveat is that this is only rigorous if the free variables in $\underline{\mathtt{m}}$ occur in trees, separated by frozen regions where we must have messages $\mathtt{m}_{v\to a}$ that are degenerate (supported on either on $\mathtt{0}$ or on $\mathtt{1}$). Otherwise, one always has the trivial "replica symmetric" BP solution where every $\mathtt{m}_{v\to a}$ is $\mathrm{unif}(\{\mathtt{0},\mathtt{1}\})$, and this is not a "meaningful" solution for large α. One way to understand this is via the physics calculation of $\mathtt{f}^{\mathrm{RS}}(\alpha)$, which we now describe by way of motivating the more complicated expression for $\mathtt{f}^{\mathrm{1RSB}}(\alpha)$.

Given a random regular NAE-SAT instance $\mathscr{G}$ on n variables, choose k uniformly random variables $v_1,\dots,v_k$, and assume for simplicity that no two of these share a clause. Then (1) remove the k variables along with their kd incident clauses, producing an instance $\mathscr{G}''$, and (2) add $d(k-1)$ new clauses to $\mathscr{G}''$, producing $\mathscr{G}'$. Then $\mathscr{G}'$ is distributed as a random regular NAE-SAT instance on $n-k$ variables. If the free energy exists, then

$$\mathtt{f}(\alpha)^n \doteq Z \doteq [Z(\mathscr{G})/Z(\mathscr{G}')]^{n/k}. \tag{4}$$

Suppose u is a variable in $\mathscr{G}'$ of degree $d-1$, meaning it was a neighbor of a clause a which was deleted from $\mathscr{G}$. The interpretation of $\underline{\mathtt{m}}$ is that in $\mathscr{G}''$, the spin at u has law $\mathtt{m}_{u\to a}$, and the different $u's$ are independent. If every $\mathtt{m}_{u\to a}$ is $\mathrm{unif}(\{\mathtt{0},\mathtt{1}\})$, then

$$\left(\frac{Z(\mathscr{G})}{Z(\mathscr{G}'')}\right)^{1/k} = 2(1-2/2^k)^d, \quad \left(\frac{Z(\mathscr{G}')}{Z(\mathscr{G}'')}\right)^{1/k} = (1-2/2^k)^{\alpha(k-1)}, \tag{5}$$

Taking the ratio of these and substituting into (4) gives the prediction $\mathtt{f}(\alpha)\doteq\mathtt{f}^{\mathrm{RS}}(\alpha)$, which we know to be false for large α. Thus the replica symmetric $\underline{\mathtt{m}}$ gives the incorrect prediction. The reason for this failure is that in reality the u's are *not* independent in $\mathscr{G}''$, but rather are significantly correlated even though they are typically far apart in $\mathscr{G}''$. This phenomenon of long-range dependence may be taken as a definition of replica symmetry breaking, and it is expected to occur precisely for $\alpha>\alpha_{\mathrm{cond}}$.

The idea of 1RSB is that, in passing from the original NAE-SAT model to the (seemingly far more complicated) "auxiliary model" of weighted BP solutions, we in fact return to replica symmetry, provided

$$\Sigma(s_\lambda)>0 \quad \text{for} \quad s_\lambda \equiv \arg\max_s\{\lambda s+\Sigma(s)\}. \tag{6}$$

That is, for such λ, the auxiliary model is predicted to have correlation decay, in contrast with the long-range correlations of the original model. The implication is that in this context, the above heuristic ((4) and (5)) *is* expected to yield the correct answer. The replica symmetric

BP solution for the auxiliary model will be a certain measure $\dot{q}_\lambda$ over messages $\mathtt{m}$. Taking $\dot{q}_{v\to a} \equiv \dot{q}_\lambda$ is the precise analogue, in the auxiliary model, of taking $\mathtt{m}_{v\to a} \equiv \mathrm{unif}(\{\mathtt{0},\mathtt{1}\})$ on every $v \to a$ in the original model. Under the assumption that the auxiliary model has strong correlation decay, (4) and (5) give an expression for $\mathfrak{F}(\lambda)$ in terms of $\dot{q}_\lambda$.

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

$$\begin{aligned}\hat{\mathscr{R}}_\lambda\mu(B) &\equiv \hat{\mathscr{Z}}(\mu)^{-1}\int\Big(2-\prod_{i=1}^{k-1}x_i-\prod_{i=1}^{k-1}(1-x_i)\Big)^\lambda \mathbf{1}\Big\{\frac{1-\prod_{i=1}^{k-1}x_i}{2-\prod_{i=1}^{k-1}x_i-\prod_{i=1}^{k-1}(1-x_i)}\in B\Big\}\prod_{i=1}^{k-1}\mu(dx_i),\\ \dot{\mathscr{R}}_\lambda\mu(B) &\equiv \dot{\mathscr{Z}}(\mu)^{-1}\int\Big(\prod_{i=1}^{d-1}y_i+\prod_{i=1}^{d-1}(1-y_i)\Big)^\lambda \mathbf{1}\Big\{\frac{\prod_{i=1}^{d-1}y_i}{\prod_{i=1}^{d-1}y_i+\prod_{i=1}^{d-1}(1-y_i)}\in B\Big\}\prod_{i=1}^{d-1}\mu(dy_i),\end{aligned} \tag{7}$$

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 in the regime

$$(2^{k-1}-2)\ln 2 \equiv \alpha_{\mathrm{lbd}} \leqslant \alpha \leqslant \alpha_{\mathrm{ubd}} \equiv 2^{k-1}\ln 2,$$

which we recall is a superset of $(\alpha_{\mathrm{cond}}, \alpha_{\mathrm{sat}})$.

Proposition 1.2. *For any $\lambda \in [0,1]$, let $\dot{\mu}_{\lambda,l} \in \mathscr{P}$ be the sequence of probability measures defined by $\dot{\mu}_{\lambda,0} \equiv \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$ and $\dot{\mu}_{\lambda,l+1} = \mathscr{R}_\lambda\dot{\mu}_{\lambda,l}$ for all $l \geqslant 1$. Let*

$$S_l \equiv (\mathrm{supp}\,\dot{\mu}_{\lambda,l})\backslash(\mathrm{supp}(\dot{\mu}_{\lambda,0}+\ldots+\dot{\mu}_{\lambda,l-1})),$$

so S_l 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 $k \geqslant k_0$ and $\alpha_{\mathrm{lbd}} \leqslant \alpha \leqslant \alpha_{\mathrm{ubd}}$, in the limit $l \to \infty$, $\dot{\mu}_{\lambda,l}$ converges in the ℓ^1 sequence space to a limit $\dot{\mu}_\lambda \in \mathscr{P}$ satisfying $\dot{\mu}_\lambda = \mathscr{R}_\lambda\dot{\mu}_\lambda$ and

$$\dot{\mu}_\lambda((0,1)) \leqslant 7/2^k, \quad \dot{\mu}_\lambda(dx) = \dot{\mu}_\lambda(d(1-x)).$$

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 $\hat{\mu}_\lambda \equiv \mathscr{R}_\lambda\dot{\mu}_\lambda$, let $\dot{w}_\lambda, \hat{w}_\lambda, \bar{w}_\lambda \in \mathscr{P}$ be defined by

$$\begin{aligned}\dot{w}_\lambda(B) &= (\dot{\mathfrak{Z}}_\lambda)^{-1}\int\Big(\prod_{i=1}^{d}y_i+\prod_{i=1}^{d}(1-y_i)\Big)^\lambda \mathbf{1}\Big\{\prod_{i=1}^{d}y_i+\prod_{i=1}^{d}(1-y_i)\in B\Big\}\prod_{i=1}^{d}\hat{\mu}_\lambda(dy_i),\\ \hat{w}_\lambda(B) &= (\hat{\mathfrak{Z}}_\lambda)^{-1}\int\Big(1-\prod_{i=1}^{k}x_i-\prod_{i=1}^{k}(1-x_i)\Big)^\lambda \mathbf{1}\Big\{1-\prod_{i=1}^{k}x_i-\prod_{i=1}^{k}(1-x_i)\in B\Big\}\prod_{i=1}^{k}\dot{\mu}_\lambda(dx_i),\\ \bar{w}_\lambda(B) &= (\bar{\mathfrak{Z}}_\lambda)^{-1}\iint\Big(xy+(1-x)(1-y)\Big)^\lambda \mathbf{1}\Big\{xy+(1-x)(1-y)\in B\Big\}\dot{\mu}_\lambda(dx)\hat{\mu}_\lambda(dy),\end{aligned} \tag{8}$$

with $\dot{\mathfrak{Z}}_\lambda, \hat{\mathfrak{Z}}_\lambda, \bar{\mathfrak{Z}}_\lambda$ the normalizing constants. The analogue of (5) for this model is

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

and substituting into (4) gives the 1RSB prediction $\boldsymbol{Z}_\lambda \doteq \exp\{\mathfrak{F}(\lambda)\}$ where

$$\mathfrak{F}(\lambda) \equiv \mathfrak{F}(\lambda;\alpha) \equiv \ln \dot{\mathfrak{Z}}_\lambda + \alpha \ln \hat{\mathfrak{Z}}_\lambda - k\alpha \ln \bar{\mathfrak{Z}}_\lambda. \tag{9}$$

Further, the maximizer of (6) is predicted to be given by

$$s_\lambda \equiv s_\lambda(\alpha) \equiv \int \ln(x)\dot{w}_\lambda(dx) + \alpha \int \ln(x)\hat{w}_\lambda(dx) - k\alpha \int \ln(x)\bar{w}_\lambda(dx). \tag{10}$$

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

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

This yields the predicted thresholds

$$\begin{aligned} \alpha_{\mathrm{cond}} &\equiv \sup\{\alpha : \Sigma(s_1;\alpha) > 0\}, \\ \alpha_{\mathrm{sat}} &\equiv \sup\{\alpha : \Sigma(s_0;\alpha) > 0\}, \end{aligned}$$

and 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 $\mathsf{f}^{\mathrm{1RSB}}(\alpha)$ is defined as

$$\mathsf{f}^{\mathrm{1RSB}}(\alpha) = \begin{cases} \mathsf{f}^{\mathrm{RS}}(\alpha) = 2(1-2/2^k)^\alpha & \alpha \leqslant \alpha_{\mathrm{cond}}, \\ \exp[\sup\{s : \Sigma(s) \geqslant 0\}] & \alpha_{\mathrm{cond}} \leqslant \alpha < \alpha_{\mathrm{sat}}, \\ 0 & \alpha > \alpha_{\mathrm{sat}}. \end{cases} \tag{11}$$

(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. *Consider $\alpha \in A \equiv [\alpha_{\mathrm{lbd}}, \alpha_{\mathrm{ubd}}] \cap (k^{-1}\mathbb{Z})$. For $k \geqslant k_0$ and $\alpha \in A$, the function $\Sigma(s) \equiv \Sigma(s;\alpha)$ is well-defined, continuous, and strictly decreasing in s, so that $\mathsf{f}^{\mathrm{RS}}(\alpha)$ is well-defined.*

Proposition 1.5. *For $k \geqslant k_0$ and $\lambda \in [0,1]$, $\Sigma(s_\lambda;\alpha) \equiv \mathfrak{F}(\lambda) - \lambda s_\lambda$ is strictly decreasing as a function of $\alpha \in A$. There is a unique $\alpha_\lambda \in A$ such that $\Sigma(s_\lambda;\alpha)$ is non-negative for all $\alpha \leqslant \alpha_\lambda$, and is negative for all $\alpha > \alpha_\lambda$. In particular*

$$\alpha_{\mathrm{cond}} = \alpha_1 = (2^{k-1}-1)\ln 2 + \mathsf{err}, \quad \alpha_{\mathrm{sat}} = \alpha_0 = \left(2^{k-1} - \frac{1}{2} - \frac{1}{4\ln 2}\right)\ln 2 + \mathsf{err}.$$

We remark that the asymptotic expansion of α_{sat} matches the previously mentioned result (1) from [DSS14]. The asymptotic expansion of α_{cond} matches an earlier result of [COZ12], which was obtained for a slightly different but closely related model.

1.6. **Proof approach.** Since $\mathsf{f} = \mathsf{f}(\alpha)$ is *a priori* not well-defined, the statement $\mathsf{f} \leqslant \mathsf{g}$ means formally that for all $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(Z^{1/n} \geqslant \mathsf{g} + \epsilon) = 0.$$

With this notation in mind, we will prove separately the upper bound $\mathsf{f}(\alpha) \leqslant \mathsf{f}^{\mathrm{1RSB}}(\alpha)$ and the matching lower bound $\mathsf{f}(\alpha) \geqslant \mathsf{f}^{\mathrm{1RSB}}(\alpha)$. 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 in Section 8 by an interpolation argument. This builds on similar bounds for spin glasses on Erdős–Rényi graphs [FL03, PT04], together with ideas from [BGT13] for interpolation in random regular models. Write $Z_n(\beta)$ for 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 μ ranging over probability measures on $[0,1]$. We then choose μ according to Proposition 1.2, and take $\beta \to \infty$ to obtain the desired bound $\mathsf{f}(\alpha) \leqslant \mathsf{f}^{\mathrm{1RSB}}(\alpha)$.

Most of the paper is devoted to establishing the matching lower bound. The proof is inspired by the physics picture described above, and at a high level proceeds as follows. Take any λ for which the (predicted) value of $\Sigma(s_\lambda)$ is non-negative, and let $\boldsymbol{Y}_\lambda$ be the number of clusters of size $\doteq \exp\{ns_\lambda\}$. The informal statement of what we show is that

$$\boldsymbol{Y}_\lambda \doteq \exp\{n[\lambda s_\lambda + \Sigma(s_\lambda)]\}. \tag{12}$$

Adjusting λ as indicated by (11) then proves the desired bound $\mathtt{f}(\alpha) \geqslant \mathtt{f}^{\text{1RSB}}(\alpha)$.

Proving a formalized version of (12) 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 free trees with fewer than T variables, which limits the distance that information can propagate between free variables. We prove a version of (12) for every fixed T, and show that this yields the sharp lower bound in the limit $T \to \infty$. The proof of (12) 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. [DSS14, 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, 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.

1.7. **Open problems.** Beyond the free energy, it remains a challenge to establish the full picture predicted by statistical physicists for $\alpha \leqslant \alpha_{\text{sat}}$. We refer the reader to several recent works targeted at a broad class of models in the regime $\alpha \leqslant \alpha_{\text{cond}}$ [BC15b, CPS15, CP16b]. In the condensation regime $(\alpha_{\text{cond}}, \alpha_{\text{sat}})$, 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, and so would be quite interesting to prove for models in the condensation regime.

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 [DKMZ11]. A great deal is now known rigorously for the case of two blocks [Mas14, MNS15], 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.

Acknowledgements. We are grateful to Amir Dembo, Jian Ding, Andrea Montanari, and Lenka Zdeborová for helpful conversations. We also wish to acknowledge the hospitality of the Simons Institute at Berkeley where part of this work was completed.

2. Combinatorial model

A *not-all-equal*-SAT (NAE-SAT) problem instance is naturally encoded by a bipartite graph $\mathscr{G}$, as follows. The vertex set of $\mathscr{G}$ is divided into a set $V = \{v_1, \dots, v_n\}$ of variables and a set $F = \{a_1, \dots, a_m\}$ of clauses. All vertices are labelled, and the edge set E joins variables to clauses. For each $e \in E$ we let $v(e)$ denote the incident variable, and $a(e)$ the incident clause. The edge e comes with a literal $\mathtt{L}_e \in \{\mathtt{0}, \mathtt{1}\}$, indicating that $v(e)$ participates affirmatively ($\mathtt{L}_e = \mathtt{0}$) or negatively ($\mathtt{L}_e = \mathtt{1}$) in $a(e)$. We permit $\mathscr{G}$ to have multi-edges; in particular it is possible that a is joined to v by two edges $e', e'' \in E$, whose literals may or may not agree. We assume the graph is (d, k)-regular: each variable has d incident edges, and each clause has k incident edges, so $|E| = nd = mk$. Formally, we regard the edge set E as a permutation $\mathfrak{m}$ of $[nd]$, as follows. The i-th variable v_i has d incident half-edges, labelled

$$\dot{e}_{d(i-1)+1}, \dots, \dot{e}_{di}.$$

The i-th clause a_i has k incident half-edges, labelled

$$\hat{e}_{k(i-1)+1}, \dots, \hat{e}_{ki}.$$

An edge then consists of a pair of half-edges $(\dot{e}, \hat{e})$, and we take $E = \{(\dot{e}_i, \dot{e}_{\mathfrak{m}(i)}) : i \in [nd]\}$. For $v \in V$ we write δv for the ordered d-tuple of edges incident to v:

$$\delta v_i = ((\dot{e}_{d(i-1)+1}, \hat{e}_{\mathfrak{m}(d(i-1)+1)}), \dots, (\dot{e}_{di}, \hat{e}_{\mathfrak{m}(di)})).$$

For $a \in F$ we write δa for the ordered k-tuple of edges incident to a:

$$\delta a_i = ((\dot{e}_{\mathfrak{m}^{-1}(k(i-1)+1)}, \hat{e}_{k(i-1)+1}), \dots, (\dot{e}_{\mathfrak{m}^{-1}(ki)}, \hat{e}_{ki})).$$

Throughout this paper we denote $\mathscr{G} = (V, F, E)$ where it is understood that E corresponds to a permutation $\mathfrak{m}$ of $[nd]$, and includes the literals $\underline{\mathtt{L}}$. We also write

$$\mathscr{G} \equiv (\mathcal{G}, \underline{\mathtt{L}}) \tag{13}$$

where $\mathcal{G}$ denotes the graph forgetting the edge labels $\underline{\mathtt{L}}$. We define all edges to have length $\frac{1}{2}$, so two variables $v \neq v'$ lie at unit distance if and only if they appear in the same clause.

Definition 2.1. An NAE-SAT solution for $\mathscr{G} = (V, F, E)$ is any $\underline{\boldsymbol{x}} \in \{\mathtt{0}, \mathtt{1}\}^V$ such that

$$\text{for all } a \in F,\ (\mathtt{L}_e \oplus \boldsymbol{x}_{v(e)})_{e \in \delta a} \text{ is neither identically } \mathtt{0} \text{ nor identically } \mathtt{1}.$$

Let $\mathsf{SOL}(\mathscr{G}) \subseteq \{\mathtt{0}, \mathtt{1}\}^V$ denote the set of all NAE-SAT solutions of $\mathscr{G}$, and define a graph on $\mathsf{SOL}(\mathscr{G})$ by connecting any pair of solutions at Hamming distance one. The connected components of this graph are the *clusters* of NAE-SAT solutions.

2.1. Frozen and warning configurations.

We begin by reviewing two standard encodings (see [Par02, BMZ05, MMW07, MM09, DSS14]) of NAE-SAT solution clusters, via frozen configurations and warning configurations.

Definition 2.2. On $\mathscr{G} = (V, F, E)$, we say that $\underline{x} \in \{\mathtt{0}, \mathtt{1}, \mathtt{f}\}^V$ is a valid *frozen configuration* if (with the convention $\mathtt{1} \oplus \mathtt{f} = \mathtt{0} \oplus \mathtt{f} = \mathtt{f}$)

1. For all $a \in F$, $(\mathtt{L}_e \oplus x_{v(e)})_{e \in \delta a}$ is neither identically $\mathtt{0}$ nor identically $\mathtt{1}$; and

2. For all $v \in V$, $x_v \in \{\mathtt{0}, \mathtt{1}\}$ if and only if there exists some $e \in \delta v$ such that
$$(\mathtt{L}_{e'} \oplus x_{v(e')})_{e' \in \delta a(e) \setminus e} \text{ is identically equal to } \mathtt{L}_e \oplus x_v \oplus \mathtt{1}. \tag{14}$$
If no such $e \in \delta v$ exists then $x_v = \mathtt{f}$.

It is well known that on any given problem instance $\mathscr{G} = (V, F, E)$, 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 [Par02], as follows. Start by setting $\underline{x} = \underline{\boldsymbol{x}}$. Then, whenever $x_v \in \{\mathtt{0}, \mathtt{1}\}$ but there exists no $e \in \delta v$ such that (14) 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 (Definition 2.1).

We say that an NAE-SAT solution $\underline{\boldsymbol{x}}$ *extends* a frozen configuration $\underline{x}$ if $\boldsymbol{x}_v = x_v$ whenever $x_v \in \{\mathtt{0}, \mathtt{1}\}$. Let $\mathsf{size}(\underline{x})$ count the number of such extensions. The purpose of this section is to define (under a certain restriction) an alternative combinatorial representation $\underline{\sigma}$ of $\underline{x}$ — which we call a *coloring* — from which $\mathsf{size}(\underline{x})$ can be easily calculated. We will explain the correspondence between $\underline{x}$ and $\underline{\sigma}$ in a few stages:
$$\begin{aligned} &\text{frozen configurations } \underline{x} \\ &\leftrightarrow \text{warning configurations } \underline{y} \\ &\leftrightarrow \text{message configurations } \underline{\tau} \\ &\leftrightarrow \text{colorings } \underline{\sigma}. \end{aligned} \tag{15}$$

The first step $\underline{x} \leftrightarrow \underline{y}$ is quite standard: $\underline{y}$ takes values in M^E where $M = \{\mathtt{0}, \mathtt{1}, \mathtt{f}\}^2$. Each $e \in E$ has a pair of warnings $y_e \equiv (\dot{y}_e, \hat{y}_e)$ where $\dot{y}_e$ represents the variable-to-clause warning along e, and $\hat{y}_e$ represents the clause-to-variable warning along e. The warnings must satisfy some local equations, as follows:

Definition 2.3. On $\mathscr{G} = (V, F, E)$, $\underline{y} \in M^E$ is a valid *warning configuration* if for all $e \in E$,
$$\begin{aligned} \dot{y}_e &= \dot{\Upsilon}((\hat{y}_{e'})_{e' \in \delta v(e) \setminus e}) \text{ and} \\ \hat{y}_e &= \mathtt{L}_e \oplus \hat{\Upsilon}((\mathtt{L}_{e'} \oplus \dot{y}_{e'})_{e' \in \delta a(e) \setminus e}) \end{aligned}$$
where $\dot{\Upsilon} : \{\mathtt{0}, \mathtt{1}, \mathtt{f}\}^{d-1} \to \{\mathtt{0}, \mathtt{1}, \mathtt{f}, \varnothing\}$ and $\hat{\Upsilon} : \{\mathtt{0}, \mathtt{1}, \mathtt{f}\}^{k-1} \to \{\mathtt{0}, \mathtt{1}, \mathtt{f}\}$ are defined by
$$\dot{\Upsilon}(\underline{\hat{y}}) = \begin{cases} \mathtt{0} & \mathtt{0} \in \{\hat{y}_i\} \subseteq \{\mathtt{0}, \mathtt{f}\}; \\ \mathtt{1} & \mathtt{1} \in \{\hat{y}_i\} \subseteq \{\mathtt{1}, \mathtt{f}\}; \\ \mathtt{f} & \{\hat{y}_i\} = \mathtt{f}; \\ \varnothing & \text{otherwise.} \end{cases} \qquad \hat{\Upsilon}(\underline{\dot{y}}) = \begin{cases} \mathtt{0} & \{\dot{y}_i\} = \{\mathtt{1}\}; \\ \mathtt{1} & \{\dot{y}_i\} = \{\mathtt{0}\}; \\ \mathtt{f} & \text{otherwise.} \end{cases}$$
(For $\underline{y}$ to be valid, we require that no edge e has $\dot{y}_e = \varnothing$.)

It is well known that there is a bijection
$$\left\{ \begin{matrix} \text{frozen configurations} \\ \underline{x} \in \{\mathtt{0}, \mathtt{1}, \mathtt{f}\}^V \end{matrix} \right\} \leftrightarrow \left\{ \begin{matrix} \text{warning configurations} \\ \underline{y} \in M^E \end{matrix} \right\}.$$
The mapping from $\underline{x}$ to $\underline{y}$ is as follows: for any v and any $e \in \delta v$ such that (14) holds, set $\hat{y}_e = x_v \in \{\mathtt{0}, \mathtt{1}\}$. In all other cases set $\hat{y}_e = \mathtt{f}$. If any entry of $(\hat{y}_{e'})_{e' \in \delta v(e) \setminus e}$ is not $\mathtt{f}$, then it must equal $x_{v(e)}$, and in this case set $\dot{y}_e = x_{v(e)}$. Otherwise, set $\dot{y}_e = \mathtt{f}$.

2.2. **Message configurations.** We shall now restrict consideration to frozen configurations without "free cycles" (defined below), and decompose $\mathtt{f}$ into a more refined set of "messages."

Definition 2.4. Let $\underline{x} \in \{\mathtt{0},\mathtt{1},\mathtt{f}\}^V$ be a valid frozen configuration on $\mathscr{G} = (V,F,E)$. 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

$$\mathtt{L}_{e'} \oplus x_{v(e')} = \mathtt{L}_{e''} \oplus x_{v(e'')} \oplus 1 \neq \mathtt{f}.$$

In particular, a forcing clause is also separating. A cycle is a sequence of edges

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

where, taking indices modulo 2ℓ, 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 is *free* if all its variables are free and all its clauses are non-separating.

Definition 2.5. Let $\underline{x}$ be a frozen configuration on $\mathscr{G} = (V,F,E)$. Let H be the subgraph of $\mathscr{G}$ induced by the free variables and non-separating clauses of $\underline{x}$. If $\underline{x}$ has no free cycles, then H is a disjoint union of tree components $\boldsymbol{t}$, which we term the *free trees* of $\underline{x}$. For each $\boldsymbol{t}$, let $\boldsymbol{T}$ be the subgraph of $\mathscr{G}$ induced by the depth-one neighborhood of $\boldsymbol{t}$, which may contain cycles. The subgraphs $\boldsymbol{T}$ 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.

In the message configuration $\underline{\tau} \in \mathscr{M}^E$, each edge $e \in E$ has a pair of messages $\tau_e \equiv (\dot{\tau}_e, \hat{\tau}_e)$, where each message is a rooted tree. To motivate the formal definition, consider the situation that e belongs to a free piece $\boldsymbol{T}$ which is a tree. We define one-sided versions $\dot{\boldsymbol{T}}_e$ and $\hat{\boldsymbol{T}}_e$: 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. 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. We regard $\dot{\boldsymbol{T}}_e$ and $\hat{\boldsymbol{T}}_e$ as being rooted at $a(e)$ and $v(e)$ respectively. Informally, $\dot{\tau}_e$ encodes the isomorphism class of $\dot{\boldsymbol{T}}_e$ while $\hat{\tau}_e$ encodes the isomorphism class of $\hat{\boldsymbol{T}}_e$. However the situation is more subtle if the edge has warning $\mathtt{f}$ in one direction but $\mathtt{0}/\mathtt{1}$ in the reverse direction; minor complications also arise relating to the edge literals and the presence of cycles. We now make a formal definition which takes these issues into account.

It will be convenient to let $\mathtt{E}$ indicate a directed edge, pointing from tail vertex $t(\mathtt{E})$ to head vertex $h(\mathtt{E})$. If e is the undirected version of $\mathtt{E}$, then we let

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

We will make a definition such that either $\tau_{\mathtt{E}}$ is a bipartite factor tree, or $\tau_{\mathtt{E}} = \star$. 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 vertex of the tree is required to have degree one, and should be thought of as corresponding to $h(\mathtt{E})$.

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 one of its endpoints — the root is the incident clause in the case of $\dot{\tau}$, the incident variable in the case of $\hat{\tau}$. We use $\square$ to stand for the tree consisting of a single unlabelled edge, rooted at the incident variable. Given a collection of rooted trees $t_1, \dots, t_\ell$ whose roots $o_1, \dots, o_\ell$ are all of the same type (either all variable or all clauses), we define $t = \mathsf{join}(t_1, \dots, t_\ell)$ 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 as the o_i, and o' is given the opposite type, so the resulting tree t is a bipartite factor graph rooted at a vertex of degree one. Let $\dot{\mathscr{M}}$ and $\hat{\mathscr{M}}$ denote the possible values of $\dot{\tau}_e$ and $\hat{\tau}_e$ respectively. Write

$$\dot{\Omega}_{\mathtt{f}} \equiv \dot{\mathscr{M}} \backslash \{\mathtt{0}, \mathtt{1}, \star\}, \quad \hat{\Omega}_{\mathtt{f}} \equiv \hat{\mathscr{M}} \backslash \{\mathtt{0}, \mathtt{1}, \star\}.$$

In particular, $\square \in \hat{\Omega}_{\mathtt{f}}$. We will see below what other elements belong to $\dot{\Omega}_{\mathtt{f}}$ and $\hat{\Omega}_{\mathtt{f}}$.

Definition 2.6. On $\mathscr{G} = (V, F, E)$, $\underline{\tau} \in \mathscr{M}^E$ is a valid *message configuration* if for all $e \in E$,

$$\dot{\tau}_e = \dot{\mathtt{T}}((\hat{\tau}_{e'})_{e' \in \delta v(e) \backslash e}) \text{ and}$$
$$\hat{\tau}_e = \mathtt{L}_e \oplus \hat{\mathtt{T}}((\mathtt{L}_{e'} \oplus \dot{\tau}_{e'})_{e' \in \delta a(e) \backslash e})$$

where $\dot{\mathtt{T}} : \hat{\mathscr{M}}^{d-1} \to \dot{\mathscr{M}}$ and $\hat{\mathtt{T}} : \dot{\mathscr{M}}^{k-1} \to \hat{\mathscr{M}}$ are defined by

$$\dot{\mathtt{T}}(\underline{\hat{\tau}}) = \begin{cases} \mathtt{0} & \mathtt{0} \in \{\hat{\tau}_i\} \subseteq \dot{\mathscr{M}} \backslash \{\mathtt{1}\}; \\ \mathtt{1} & \mathtt{1} \in \{\hat{\tau}_i\} \subseteq \dot{\mathscr{M}} \backslash \{\mathtt{0}\}; \\ \mathsf{join}\{\hat{\tau}_i\} & \{\hat{\tau}_i\} \subseteq \hat{\Omega}_{\mathtt{f}}; \\ \star & \star \in \{\hat{\tau}_i\} \subseteq \{\star\} \cup \hat{\Omega}_{\mathtt{f}}; \\ \varnothing & \text{otherwise;} \end{cases} \qquad \hat{\mathtt{T}}(\underline{\dot{\tau}}) = \begin{cases} \mathtt{0} & \{\dot{\tau}_i\} = \{\mathtt{1}\}; \\ \mathtt{1} & \{\dot{\tau}_i\} = \{\mathtt{0}\}; \\ \square & \{\mathtt{0}, \mathtt{1}\} \subseteq \{\dot{\tau}_i\}; \\ \mathsf{join}\{\dot{\tau}_i\} & \{\mathtt{0}\} \neq \{\dot{\tau}_i\} \subseteq \{\mathtt{0}\} \cup \dot{\Omega}_{\mathtt{f}} \\ & \text{or } \{\mathtt{1}\} \neq \{\dot{\tau}_i\} \subseteq \{\mathtt{1}\} \cup \dot{\Omega}_{\mathtt{f}}; \\ \star & \text{otherwise.} \end{cases}$$

For $\underline{\tau}$ to be valid, we require for all $e \in E$ that $\dot{\tau}_e \neq \varnothing$, and further if one of $\dot{\tau}_e, \hat{\tau}_e$ equals $\star$ then the other must be in $\{\mathtt{0}, \mathtt{1}\}$.

Given a frozen configuration $\underline{x}$ we define the message configuration $\underline{\tau}$ in a recursive manner. If $y_{\mathtt{E}} \in \{\mathtt{0}, \mathtt{1}\}$ then set $\tau_{\mathtt{E}} = y_{\mathtt{E}}$. If

$$\{\mathtt{0}, \mathtt{1}\} \subseteq \{\mathtt{L}_{e'} \oplus \dot{y}_{e'}\}_{e' \in \delta a(e) \backslash e}$$

then set $\hat{\tau}_e = \square$. Let $\mathtt{F}$ denote the reversal of $\mathtt{E}$, and let $\delta_{\mathtt{E}}$ denote the set of directed edges $\mathtt{E}'$ pointing towards $t(\mathtt{E})$ (including $\mathtt{F}$). Then, whenever $\tau_{\mathtt{E}}$ is undefined but $\tau_{\mathtt{E}'}$ is defined for all $\mathtt{E}' \in \delta_{\mathtt{E}} \backslash \mathtt{F}$, set

$$\tau_{\mathtt{E}} \equiv \begin{cases} \dot{\mathtt{T}}((\hat{\tau}_{e'})_{e' \in \delta v(e) \backslash e}) & \text{if } t(\mathtt{E}) \text{ is a variable;} \\ \mathtt{L}_e \oplus \hat{\mathtt{T}}((\mathtt{L}_{e'} \oplus \dot{\tau}_{e'})_{e' \in \delta a(e) \backslash e}) & \text{if } t(\mathtt{E}) \text{ is a clause;} \end{cases}$$

Repeat until no further updates are possible. At the end of this procedure, if any $\tau_{\mathtt{E}}$ remains undefined then set it to $\star$.

Lemma 2.7. *Let $\underline{x} \in \{\mathtt{0}, \mathtt{1}, \mathtt{f}\}^V$ be a valid frozen configuration on $\mathscr{G} = (V, F, E)$ which has no free cycles. Then $\underline{x}$ maps under the above procedure to a valid message configuration $\underline{\tau}$.*

Proof. Suppose $\tau_{\mathtt{E}} = \star$, and let $\mathtt{F}$ denote the reversal of $\mathtt{E}$. From the above construction, it must be that $y_{\mathtt{E}} = \mathtt{f}$ and $\tau_{\mathtt{E}'} = \star$ for some $\mathtt{E}' \in \delta_{\mathtt{E}} \backslash \mathtt{F}$. Consequently $\mathtt{E}$ must belong to a cycle of directed edges

$$\mathtt{E}_1 \mathtt{E}_2 \ldots \mathtt{E}_{2k} \mathtt{E}_1$$

with all the $\tau_{\mathtt{E}_i}$ equal to $\star$. Whenever $\mathtt{E}$ points from a separating clause a to free variable v, we must have $\tau_{\mathtt{E}} = \square$. 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 $v = t(\mathtt{E}_1)$. Let $\mathtt{F}_i$ denote the reversal of $\mathtt{E}_i$: since $x_v \neq \mathtt{f}$ but $y_{\mathtt{E}_1} = \mathtt{f}$, it must be that $y_{\mathtt{F}_1} = x_v$. This means that the clause $a = h(\mathtt{E}_1) = t(\mathtt{F}_1)$ is forcing to v, so in particular $y_{\mathtt{F}_2} \in \{\mathtt{0}, \mathtt{1}\}$. Continuing in this way we see that $y_{\mathtt{F}_i} \in \{\mathtt{0}, \mathtt{1}\}$ for all i, and it follows that $\underline{\tau}$ is a valid message configuration. $\square$

Lemma 2.8. *There is a bijection*

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

Proof. Given $\underline{x}$, let $\underline{y}$ and $\underline{\tau}$ be the corresponding warning and message configurations. The mapping from $\underline{y}$ to $\underline{\tau}$ is clearly injective. Since $\underline{x} \leftrightarrow \underline{y}$, the mapping from $\underline{x}$ to $\underline{\tau}$ is also injective. To see that it is surjective, let $\underline{\tau}$ be any message configuration. Projecting $\{\star\} \cup \dot{\Omega}_{\mathtt{f}} \mapsto \mathtt{f}$ and $\{\star\} \cup \hat{\Omega}_{\mathtt{f}} \mapsto \mathtt{f}$ 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. Suppose for the sake of contradiction that there exists a cycle of directed edges

$$\mathtt{E}_1\mathtt{E}_2 \ldots \mathtt{E}_{2k}\mathtt{E}_1$$

where all the variables are free and all the clauses are non-separating. Writing $\mathtt{F}_i$ for the reversal of $\mathtt{E}_i$, we see that all the messages $\tau_{\mathtt{E}_i}, \tau_{\mathtt{F}_i}$ must lie in $\{\star\} \cup \dot{\Omega}_{\mathtt{f}} \cup \hat{\Omega}_{\mathtt{f}}$. In fact, none of the messages can be $\star$, since in that case we require the message in the reverse direction to be in $\{\mathtt{0},\mathtt{1}\}$. Therefore all the messages are in $\dot{\Omega}_{\mathtt{f}} \cup \hat{\Omega}_{\mathtt{f}}$. By definition of $\dot{\mathtt{T}}$ and $\hat{\mathtt{T}}$, $\tau_{\mathtt{E}_i}$ must be a proper subtree of $\tau_{\mathtt{E}_{i+1}}$ for all i, with indices modulo $2k$. Going around the cycle we find that $\tau_{\mathtt{E}_1}$ is a proper subtree of $\tau_{\mathtt{E}_{2k+1}} = \tau_{\mathtt{E}_1}$, which gives the required contradiction. □

2.3. **Bethe formula.** The messages $\dot{\tau}_e, \hat{\tau}_e$ can be used to define probability measures $\dot{\boldsymbol{m}}_e, \hat{\boldsymbol{m}}_e$ on $\{\mathtt{0},\mathtt{1}\}$ where

$$\begin{aligned}\dot{\boldsymbol{m}}_e &\equiv \dot{\boldsymbol{m}}(\dot{\tau}_e) \text{ represents the law of } v(e) \text{ in absence of } a(e);\\ \hat{\boldsymbol{m}}_e &\equiv \hat{\boldsymbol{m}}(\hat{\tau}_e) \text{ represents the law of } v(e) \text{ in absence of } \delta v(e)\backslash e.\end{aligned}$$

If $\dot{\tau}_e \neq \star$, then there will be a normalizing constant $\dot{z}_e$ such that

$$\dot{\boldsymbol{m}}_e(x) = \frac{1}{\dot{z}_e} \prod_{e' \in \delta v(e)\backslash e} \hat{\boldsymbol{m}}_{e'}(x) \quad \text{for } x \in \{\mathtt{0},\mathtt{1}\}.$$

Similarly, let $I^{\mathrm{NAE}}(\underline{x})$ be the indicator that the entries of $\underline{x}$ are not all equal: if $\hat{\tau}_e \neq \star$ then there will be a normalizing constant $\hat{z}_e$ such that

$$\hat{\boldsymbol{m}}_e(x) = \frac{1}{\hat{z}_e} \sum_{\underline{x}_{\delta a(e)\backslash e}} I^{\mathrm{NAE}}(x \oplus \mathtt{L}_e, (\underline{x} \oplus \mathtt{L})_{\delta a(e)\backslash e}) \prod_{e' \in \delta a(e)\backslash e} \dot{\boldsymbol{m}}(x_{e'}) \quad \text{for } x \in \{\mathtt{0},\mathtt{1}\}.$$

In what follows we usually represent a probability measure on $\{\mathtt{0},\mathtt{1}\}$ by the probability assigned to $\mathtt{1}$, writing $\dot{m} \equiv \dot{\boldsymbol{m}}(\mathtt{1})$ and $\hat{m} \equiv \hat{\boldsymbol{m}}(\mathtt{1})$. Explicitly, $\dot{m}(\dot{\tau})$ and $\hat{m}(\hat{\tau})$ can be defined recursively, starting from the base cases

$$\dot{m}(\mathtt{1}) = \hat{m}(\mathtt{1}) = 1, \quad \dot{m}(\mathtt{0}) = \hat{m}(\mathtt{0}) = 0.$$

If $\dot{\tau} \in \dot{\Omega}_{\mathtt{f}}$ equals $\dot{\mathtt{T}}(\hat{\tau}_1, \ldots, \hat{\tau}_{d-1})$ where none of the $\hat{\tau}_i$ are $\star$, then set

$$\dot{m}(\dot{\tau}) = \frac{1}{\dot{z}(\dot{\tau})} \prod_{i=1}^{d-1} \hat{m}(\hat{\tau}_i), \quad \dot{z}(\dot{\tau}) = \prod_{i=1}^{d-1} \hat{m}(\hat{\tau}_i) + \prod_{i=1}^{d-1} (1 - \hat{m}(\hat{\tau}_i)), \tag{16}$$

where we note that $(\hat{\tau}_1, \ldots, \hat{\tau}_{d-1})$ can be recovered from $\dot{\tau}$ modulo permutation of the indices, so $\dot{z}(\dot{\tau})$ is well-defined. Similarly, if $\hat{\tau} \in \hat{\Omega}_{\mathtt{f}}$ equals $\hat{\mathtt{T}}(\dot{\tau}_1, \ldots, \dot{\tau}_{k-1})$ where none of the $\dot{\tau}_i$ are $\star$, then set

$$\hat{m}(\hat{\tau}) = \frac{1}{\hat{z}(\hat{\tau})} \left(1 - \prod_{i=1}^{k-1} \dot{m}(\dot{\tau}_i)\right), \quad \hat{z}(\hat{\tau}) = 2 - \prod_{i=1}^{k-1} \dot{m}(\dot{\tau}_i) - \prod_{i=1}^{k-1} (1 - \dot{m}(\dot{\tau}_i)). \tag{17}$$

Finally, we will see below that for our purposes we can take $\dot{m}(\star), \hat{m}(\star)$ to be any fixed values in $(0,1)$. We arbitrarily set $\dot{m}(\star) = \frac{1}{2} = \hat{m}(\star)$.

Lemma 2.9. *Suppose on $\mathscr{G} = (V, F, E)$ that $\underline{\tau}$ is a valid message configuration, and let $\underline{x}$ be the corresponding frozen configuration (which has no free cycles). Suppose $\boldsymbol{T}$ is a free piece of $\underline{x}$, and let $\boldsymbol{t}$ be the free tree inside $\boldsymbol{T}$. Let $\mathsf{size}(\underline{x}; \boldsymbol{T})$ count the number of valid* NAE-SAT *assignments which extend $\underline{x}$ on $\boldsymbol{T}$. Then*

$$\mathsf{size}(\underline{x}; \boldsymbol{T}) = \prod_{v \in \boldsymbol{t} \cap V} \dot{\varphi}(\hat{\underline{\tau}}_{\boldsymbol{t} \cap \delta v}) \prod_{a \in \boldsymbol{t} \cap F} \hat{\varphi}^{\mathrm{lit}}((\dot{\underline{\tau}} \oplus \underline{\mathsf{L}})_{\delta a}) \prod_{e \in \boldsymbol{t} \cap E} \bar{\varphi}(\tau_e) \tag{18}$$

where $\bar{\varphi}(\dot{\tau}, \hat{\tau}) \equiv [\dot{m}(\dot{\tau})\hat{m}(\hat{\tau}) + (1 - \dot{m}(\dot{\tau}))(1 - \hat{m}(\hat{\tau}))]^{-1}$,

$$\hat{\varphi}^{\mathrm{lit}}(\dot{\tau}_1, \dots, \dot{\tau}_k) = 1 - \prod_{i=1}^{k} \dot{m}(\dot{\tau}_i) - \prod_{i=1}^{k} (1 - \dot{m}(\dot{\tau}_i)),$$

and for any $\ell \geqslant 0$ we define

$$\dot{\varphi}(\hat{\tau}_1, \dots, \hat{\tau}_\ell) = \prod_{i=1}^{\ell} \hat{m}(\hat{\tau}_i) + \prod_{i=1}^{\ell} (1 - \hat{m}(\hat{\tau}_i)).$$

We take the convention that the empty product equals one, so if $\ell = 0$ then $\dot{\varphi} = 2$. The number of valid NAE-SAT *assignments extending $\underline{x}$ is given by*

$$\mathsf{size}(\underline{x}) = \prod_{\boldsymbol{T} \in \underline{x}} \mathsf{size}(\underline{x}; \boldsymbol{T}) \tag{19}$$

where the product is taken over all free pieces (Definition 2.5) $\boldsymbol{T}$ of $\underline{x}$.

Proof. The first claim (18) is a well-known calculation; see e.g. [MM09, Ch. 14]. The product formula (19) then follows from the fact that different free trees are disjoint. □

Corollary 2.10. *Suppose on $\mathscr{G} = (V, F, E)$ that $\underline{\tau}$ is a valid message configuration, and let $\underline{x}$ be the corresponding frozen configuration. Then*

$$\mathsf{size}(\underline{x}) = \prod_{v \in V} \dot{\varphi}(\hat{\underline{\tau}}_{\delta v}) \prod_{a \in F} \hat{\varphi}^{\mathrm{lit}}((\dot{\underline{\tau}} \oplus \underline{\mathsf{L}})_{\delta a}) \prod_{e \in \boldsymbol{t} \cap E} \bar{\varphi}(\tau_e);$$

and this identity holds for any choices of $\dot{m}(\star), \hat{m}(\star) \in (0,1)$.

Proof. Let V' denote the set of free variables, and F' the set of non-separating clauses. For each $v \in V'$ let $\boldsymbol{t}(v)$ denote the (unique) free tree containing v. Rearranging the product formula (19) gives

$$\mathsf{size}(\underline{x}) = \prod_{v \in V'} \left\{ \dot{\varphi}(\hat{\underline{\tau}}_{\boldsymbol{t}(v) \cap \delta v}) \prod_{e \in \boldsymbol{t}(v) \cap \delta v} \bar{\varphi}(\tau_e) \right\} \prod_{a \in F'} \hat{\varphi}^{\mathrm{lit}}((\dot{\underline{\tau}} \oplus \underline{\mathsf{L}})_{\delta a}).$$

If e joins a free variable v to a separating clause a, then $\hat{m}(\hat{\tau}_e) = \frac{1}{2} = \bar{\varphi}(\tau_e)^{-1}$, so

$$\dot{\varphi}(\hat{\underline{\tau}}_{\boldsymbol{t}(v) \cap \delta v}) = \dot{\varphi}(\hat{\underline{\tau}}_{\delta v}) 2^{|\delta v \setminus \boldsymbol{t}|} = \dot{\varphi}(\hat{\underline{\tau}}_{\delta v}) \prod_{e \in \delta v \setminus \boldsymbol{t}(v)} \bar{\varphi}(\tau_e).$$

Substituting into the above proves that

$$\mathsf{size}(\underline{x}) = \prod_{v \in V'} \left\{ \dot{\varphi}(\hat{\underline{\tau}}_{\delta v}) \prod_{e \in \delta v} \bar{\varphi}(\tau_e) \right\} \prod_{a \in F'} \hat{\varphi}^{\mathrm{lit}}((\dot{\underline{\tau}} \oplus \underline{\mathsf{L}})_{\delta a}). \tag{20}$$

For $v \notin V'$ (meaning $x_v \in \{\mathtt{0},\mathtt{1}\}$), partition δv into

$$\delta v(\mathtt{r}) = \{e \in \delta v : \hat{y}_e = x_v\}, \quad \delta v(\mathtt{b}) = \{e \in \delta v : \hat{y}_e = \mathtt{f}\}.$$

Say without loss that $x_v = \mathtt{1}$: since $\hat{m}(\hat{\tau}_e) = 1$ for all $e \in \delta v(\mathtt{r})$, we have

$$\dot{\varphi}(\underline{\hat{\tau}}_{\delta v}) = \prod_{e\in\delta v} \hat{m}(\hat{\tau}_e) + \prod_{e\in\delta v}(1-\hat{m}(\hat{\tau}_e)) = \prod_{e\in\delta v(\mathtt{b})} \hat{m}(\hat{\tau}_e) = \prod_{e\in\delta v(\mathtt{b})} \bar{\varphi}(\tau_e)^{-1}. \tag{21}$$

Some of the messages $\hat{\tau}_e$ incoming to v may equal $\star$, but the above identity holds for any choice of $\hat{m}(\star) \in (0,1)$. Likewise, if a is a separating clause which is non-forcing, then some of the messages $\dot{\tau}_e$ incoming to a may equal $\star$, but

$$\hat{\varphi}^{\mathrm{lit}}((\underline{\dot{\tau}} \oplus \underline{\mathtt{L}})_{\delta a}) = 1 \tag{22}$$

for any choice of $\dot{m}(\star) \in (0,1)$. Finally, if a is forcing in the direction of edge e, then

$$\hat{\varphi}^{\mathrm{lit}}((\underline{\dot{\tau}} \oplus \underline{\mathtt{L}})_{\delta a}) = \bar{\varphi}(\tau_e)^{-1} = \begin{cases} \dot{m}(\dot{\tau}_e) & \text{if } x_{v(e)} = \mathtt{1}; \\ 1-\dot{m}(\dot{\tau}_e) & \text{if } x_{v(e)} = \mathtt{0}; \end{cases} \tag{23}$$

including in the case that $\dot{\tau}_e = \star$. It follows from (21), (22), and (23) that

$$\prod_{v\in V\setminus V'} \left\{ \dot{\varphi}(\underline{\hat{\tau}}_v) \prod_{e\in\delta v} \bar{\varphi}(\tau_e) \right\} \prod_{a\in F\setminus F'} \hat{\varphi}^{\mathrm{lit}}((\underline{\dot{\tau}} \oplus \underline{\mathtt{L}})_{\delta a}) = 1,$$

and multiplying with (20) proves the claim. □

2.4. **Colorings.** We now define the last step of (15). Recall $\underline{\tau} \in \mathscr{M}^E$, and let $\Omega_{\mathtt{f}} \subseteq \mathscr{M}$ denote the subset of values $\tau = (\dot{\tau}, \hat{\tau}) \in \mathscr{M}$ for which $\dot{\tau} \in \dot{\Omega}_{\mathtt{f}}$ and $\hat{\tau} \in \hat{\Omega}_{\mathtt{f}}$. Then the colorings will be configurations $\underline{\sigma} \in \Omega^E$ where

$$\Omega \equiv \{\mathtt{r}_0, \mathtt{r}_1, \mathtt{b}_0, \mathtt{b}_1\} \cup \Omega_{\mathtt{f}}.$$

We define a mapping $\mathtt{s} : \mathscr{M} \to \Omega$ by

$$\mathtt{s}(\tau) = \begin{cases} \mathtt{r}_0 & \hat{\tau} = \mathtt{0}; \\ \mathtt{r}_1 & \hat{\tau} = \mathtt{1}; \\ \mathtt{b}_0 & \hat{\tau} \neq \mathtt{0} \text{ and } \dot{\tau} = \mathtt{0}; \\ \mathtt{b}_1 & \hat{\tau} \neq \mathtt{1} \text{ and } \dot{\tau} = \mathtt{1}; \\ \tau & \text{otherwise.} \end{cases} \tag{24}$$

Note that if $\tau = (\dot{\tau}, \hat{\tau})$ with $\dot{\tau} = \star$, then $\hat{\tau}$ must equal some $x \in \{\mathtt{0},\mathtt{1}\}$, and so we set $\sigma(\tau) = \mathtt{r}_x$. Likewise if $\hat{\tau} = \star$ then $\dot{\tau}$ must equal some $x \in \{\mathtt{0},\mathtt{1}\}$ and so we set $\sigma(\tau) = \mathtt{b}_x$. If $\sigma = \tau \in \Omega_{\mathtt{f}}$ we write $(\dot{\sigma}, \hat{\sigma}) \equiv (\dot{\tau}, \hat{\tau})$; otherwise we write $(\dot{\sigma}, \hat{\sigma}) \equiv (\sigma, \sigma)$. We write $\dot{\Omega}, \hat{\Omega}$ for the possible values of $\dot{\sigma}, \hat{\sigma}$. The map $\mathtt{s}$ is not one-to-one, and we shall denote

$$\begin{aligned}
\dot{\tau}^{\mathrm{pos}}(\dot{\sigma}) &= \{\dot{\tau} \in \dot{\mathscr{M}} : (\dot{\tau}, \hat{\tau}) \in \mathtt{s}^{-1}(\dot{\sigma}, \hat{\sigma}) \text{ for some } \hat{\tau} \in \hat{\mathscr{M}}, \hat{\sigma} \in \hat{\Omega}\}, \\
\hat{\tau}^{\mathrm{pos}}(\hat{\sigma}) &= \{\hat{\tau} \in \hat{\mathscr{M}} : (\dot{\tau}, \hat{\tau}) \in \mathtt{s}^{-1}(\dot{\sigma}, \hat{\sigma}) \text{ for some } \dot{\tau} \in \dot{\mathscr{M}}, \dot{\sigma} \in \dot{\Omega}\}, \\
\dot{\sigma}^{\mathrm{pos}}(\dot{\tau}) &= \{\dot{\sigma} \in \dot{\Omega} : (\dot{\sigma}, \hat{\sigma}) = \mathtt{s}(\dot{\tau}, \hat{\tau}) \text{ for some } \hat{\tau} \in \hat{\mathscr{M}}, \hat{\sigma} \in \hat{\Omega}\}, \\
\hat{\sigma}^{\mathrm{pos}}(\hat{\tau}) &= \{\hat{\sigma} \in \hat{\Omega} : (\dot{\sigma}, \hat{\sigma}) = \mathtt{s}(\dot{\tau}, \hat{\tau}) \text{ for some } \dot{\tau} \in \dot{\mathscr{M}}, \dot{\sigma} \in \dot{\Omega}\}.
\end{aligned}$$

The following definition is derived from Definition 2.6.

Definition 2.11. On $\mathscr{G} = (V, F, E)$, $\underline{\sigma} \in \Omega^E$ is a valid *coloring* if for all $e \in E$,

$$\dot{\sigma}_e \in \dot{\mathtt{S}}((\hat{\sigma}_{e'})_{e' \in \delta v(e) \setminus e}) \text{ and}$$
$$\hat{\sigma}_e \in \mathtt{L}_e \oplus \hat{\mathtt{S}}((\mathtt{L}_{e'} \oplus \dot{\sigma}_{e'})_{e' \in \delta a(e) \setminus e})$$

where $\dot{\mathtt{S}} : \hat{\Omega}^{d-1} \to 2^{\dot{\Omega}}$ and $\hat{\mathtt{S}} : \dot{\Omega}^{k-1} \to 2^{\hat{\Omega}}$ are defined by

$$\dot{\mathtt{S}}(\underline{\hat{\sigma}}) = \dot{\sigma}^{\mathrm{pos}} \circ \dot{\mathtt{T}} \circ \hat{\tau}^{\mathrm{pos}}(\underline{\hat{\sigma}}) = \{\dot{\sigma} : \dot{\sigma} \in \dot{\sigma}^{\mathrm{pos}}(\dot{\mathtt{T}}(\underline{\hat{\tau}})) \text{ for any } \underline{\hat{\tau}} \text{ with } \hat{\tau}_i \in \hat{\tau}^{\mathrm{pos}}(\hat{\sigma}_i)\ \forall i\},$$
$$\hat{\mathtt{S}}(\underline{\dot{\sigma}}) = \hat{\sigma}^{\mathrm{pos}} \circ \hat{\mathtt{T}} \circ \dot{\tau}^{\mathrm{pos}}(\underline{\dot{\sigma}}) = \{\hat{\sigma} : \hat{\sigma} \in \hat{\sigma}^{\mathrm{pos}}(\hat{\mathtt{T}}(\underline{\dot{\tau}})) \text{ for any } \underline{\dot{\tau}} \text{ with } \dot{\tau}_i \in \dot{\tau}^{\mathrm{pos}}(\dot{\sigma}_i)\ \forall i\}.$$

An equivalent characterization is that $\underline{\sigma}$ is a valid coloring if and only if

$$\prod_{v \in V} \dot{I}(\underline{\sigma}_{\delta v}) \prod_{a \in F} \hat{I}^{\mathrm{lit}}((\underline{\sigma} \oplus \underline{\mathtt{L}})_{\delta a}) = 1 \tag{25}$$

where $\dot{I} : \Omega^d \to \{0, 1\}$ and $\hat{I}^{\mathrm{lit}} : \Omega^k \to \{0, 1\}$ are given by

$$\dot{I}(\underline{\sigma}) \equiv \prod_{i=1}^{d-1} \mathbf{1}\{\dot{\sigma}_i \in \dot{\mathtt{S}}((\hat{\sigma}_j)_{j \neq i})\}, \quad \hat{I}^{\mathrm{lit}}(\underline{\sigma}) \equiv \prod_{i=1}^{k-1} \mathbf{1}\{\hat{\sigma}_i \in \hat{\mathtt{S}}((\dot{\sigma}_j)_{j \neq i})\}.$$

This builds on a related encoding introduced by [CP16a]. More explicitly, we have

$$\dot{I}(\underline{\sigma}) = \begin{cases} 1 & \mathtt{r}_0 \in \{\sigma_i\} \subseteq \{\mathtt{r}_0, \mathtt{b}_0\}, \\ 1 & \mathtt{r}_1 \in \{\sigma_i\} \subseteq \{\mathtt{r}_1, \mathtt{b}_1\}, \\ 1 & \{\sigma_i\} \subseteq \Omega_{\mathtt{f}} \text{ and} \\ & \qquad \dot{\sigma}_i = \dot{\mathtt{T}}((\hat{\sigma}_j)_{j \neq i})\ \forall i, \\ 0 & \text{otherwise;} \end{cases} \qquad \hat{I}^{\mathrm{lit}}(\underline{\sigma}) = \begin{cases} 1 & \exists i : \sigma_i = \mathtt{r}_0 \text{ and } \{\sigma_j\}_{j \neq i} = \{\mathtt{b}_1\}, \\ 1 & \exists i : \sigma_i = \mathtt{r}_1 \text{ and } \{\sigma_j\}_{j \neq i} = \{\mathtt{b}_0\}, \\ 1 & \{\sigma_i\} \cap \{\mathtt{r}_0, \mathtt{r}_1\} = \varnothing, \\ & \qquad \nexists i : \{\sigma_j\}_{j \neq i} = \{\mathtt{b}_0\} \text{ or } \{\mathtt{b}_1\}, \text{ and} \\ & \qquad \hat{\sigma}_i \in \{\mathtt{b}_0, \mathtt{b}_1, \hat{\mathtt{T}}((\dot{\tau}^{\mathrm{pos}}(\dot{\sigma}_j))_{j \neq i})\}\ \forall i, \\ 0 & \text{otherwise.} \end{cases}$$

In the definition of $\hat{I}^{\mathrm{lit}}$, we note that if $\{\sigma_i\} \cap \{\mathtt{r}_0, \mathtt{r}_1\} = \varnothing$, then $\dot{\tau}^{\mathrm{pos}}(\dot{\sigma}_i)$ is a singleton for each i. If $\{\sigma_j\}_{j \neq i}$ is neither $\{\mathtt{b}_0\}$ nor $\{\mathtt{b}_1\}$, then we have $\hat{\mathtt{T}}((\dot{\tau}^{\mathrm{pos}}(\dot{\sigma}_j))_{j \neq i}) \in \hat{\Omega}_{\mathtt{f}}$.

One purpose of this encoding is to take advantage of some of the cancellations seen in the proof of Corollary 2.10. It follows easily from the definition that we have a bijection

$$\left\{ \begin{matrix} \text{message configurations} \\ \underline{\tau} \in \mathscr{M}^E \end{matrix} \right\} \leftrightarrow \left\{ \begin{matrix} \text{colorings} \\ \underline{\sigma} \in \Omega^E \end{matrix} \right\},$$

The following is a straightforward consequence of Lemma 2.9:

Lemma 2.12. *Suppose $\underline{\sigma}$ be a valid coloring on $\mathscr{G} = (V, F, E)$. Let $\underline{\tau}$ be the corresponding message configuration, and $\underline{x}$ the corresponding frozen configuration. Then $\mathsf{size}(\underline{x}) \equiv \mathsf{size}(\underline{\sigma})$ is given by the formula*

$$\mathsf{size}(\underline{\sigma}) = \boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G}}(\underline{\sigma}) \equiv \prod_{v \in V} \dot{\Phi}(\underline{\sigma}_{\delta v}) \prod_{a \in F} \hat{\Phi}^{\mathrm{lit}}((\underline{\sigma} \oplus \underline{\mathtt{L}})_{\delta a}) \prod_{e \in E} \bar{\Phi}(\sigma_e)$$

where Φ agrees with φ for v, a, e belonging to free trees, and is one otherwise. More precisely, $\dot{\Phi} : \Omega^d \to \mathbb{R}_{\geqslant 0}$ is given by

$$\dot{\Phi}(\underline{\sigma}) = \begin{cases} 0 & \dot{I}(\underline{\sigma}) = 0; \\ 1 & \dot{I}(\underline{\sigma}) = 1 \text{ and } \{\sigma_i\} \text{ contains } \mathtt{r}_0 \text{ or } \mathtt{r}_1; \\ \dot{\varphi}(\underline{\hat{\tau}}) & \text{otherwise, meaning } v \in V'; \end{cases}$$

note in the last case that each σ_i can be mapped to a unique $\hat{\tau}_i$, so the value of $\dot{\varphi}(\underline{\hat{\tau}})$ is well-defined. Similarly, $\hat{\Phi}^{\mathrm{lit}} : \Omega^k \to \mathbb{R}_{\geqslant 0}$ is given by

$$\hat{\Phi}^{\mathrm{lit}}(\underline{\sigma}) = \begin{cases} 0 & \hat{I}^{\mathrm{lit}}(\underline{\sigma}) = 0; \\ 1 & \hat{I}^{\mathrm{lit}}(\underline{\sigma}) = 1 \text{ and } \{\sigma_i\} \text{ contains } \mathtt{r}_0 \text{ or } \mathtt{r}_1; \\ 1 & \hat{I}^{\mathrm{lit}}(\underline{\sigma}) = 1 \text{ and } \{\mathtt{b}_0, \mathtt{b}_1\} \subseteq \{\sigma_i\}; \\ \hat{\varphi}^{\mathrm{lit}}(\underline{\dot{\tau}}) & \text{otherwise, meaning } a \in F'; \end{cases}$$

note again in the last case that each σ_i can be mapped to a unique $\dot{\tau}_i$, so the value of $\hat{\varphi}^{\mathrm{lit}}(\underline{\dot{\tau}})$ is well-defined. Finally, $\bar{\Phi} : \Omega \to \mathbb{R}_{\geqslant 0}$ is given by

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

Proof. This is essentially a rewriting of (20). □

According to the above definitions, if $\underline{\sigma}$ is not a valid coloring, then $\boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G}}(\underline{\sigma}) \equiv 0$. For $\sigma \in \Omega$ let $|\sigma|$ count the number of free variables encoded by σ. Thus $|\sigma| = 0$ if and only if $\sigma \notin \Omega_{\mathtt{f}}$. On $\mathscr{G} = (V, F, E)$ we say that $\underline{\sigma}$ is a valid *T-coloring* if $|\sigma_e| \leqslant T$ for all $e \in E$. We write $\boldsymbol{I}_{\mathscr{G},T}(\underline{\sigma})$ for the indicator that $\underline{\sigma}$ is a valid T-coloring of $\mathscr{G}$, and let

$$\boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G},T}(\underline{\sigma}) \equiv \boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G}}(\underline{\sigma}) \boldsymbol{I}_{\mathscr{G},T}(\underline{\sigma}).$$

Recall from Lemma 2.12 the product formula for $\boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G}}(\underline{\sigma})$, and note that an analogous formula for $\boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G},T}(\underline{\sigma})$ is obtained by simply replacing $\bar{\Phi}$ with the modified factor $\bar{\Phi}_T$, where

$$\bar{\Phi}_T(\sigma) \equiv \bar{\Phi}(\sigma) \mathbf{1}\{|\sigma| \leqslant T\}.$$

We then define $\boldsymbol{Z}_{\lambda,T}$ to be the partition function of λ-tilted T-colorings,

$$\boldsymbol{Z}_{\lambda,T} = \sum_{\underline{\sigma} \in \Omega^E} \boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G},T}(\underline{\sigma})^{\lambda}. \tag{26}$$

Thus $\boldsymbol{Z}_{\lambda,T}$ is a function of the NAE-SAT problem instance $\mathscr{G} = (V, F, E)$. Clearly $\boldsymbol{Z}_{\lambda,T}$ is nondecreasing in T, and we write $\boldsymbol{Z}_{\lambda,\infty} \equiv \boldsymbol{Z}_{\lambda}$ for the sum over all valid colorings with no size truncation. The following gives the formal version of (3) which we will work with in the proof of the free energy lower bound.

Proposition 2.13. *On $\mathscr{G} = (V, F, E)$ let $\mathscr{C}(\mathscr{G})$ denote the collection of* NAE-SAT *clusters, so each $\boldsymbol{\gamma} \in \mathscr{C}(\mathscr{G})$ is a subset of $\{0, 1\}^V$. Then, for all $0 \leqslant T \leqslant \infty$,*

$$\boldsymbol{Z}_{\lambda,T} \leqslant \sum_{\boldsymbol{\gamma} \in \mathscr{C}(\mathscr{G})} |\boldsymbol{\gamma}|^{\lambda}.$$

Proof. This is a direct consequence of Lemma 2.12. □

3. Proof outline

Having formally set up our combinatorial model encoding the clusters of NAE-SAT solutions (Proposition 2.13), we now proceed to outline the proof of Theorem 1. The basic approach will be to show concentration for $\boldsymbol{Z}_{\lambda,T}$ via the second moment method.

3.1. **Averaging over edge literals.** In the setting of NAE-SAT, we can take advantage of the following simplification:

Remark 3.1. For any function $g : \{0,1\}^t \to \mathbb{R}$, let $\mathbb{E}^{\mathrm{lit}} g$ denote the average value of $g(\underline{\mathtt{L}})$ over all $\underline{\mathtt{L}} \in \{0,1\}^t$. Recalling from (13) the notation $\mathscr{G} = (\mathcal{G}, \underline{\mathtt{L}})$, if $\underline{\sigma}$ is any coloring of the edges of $\mathcal{G}$, then the average of $\boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G},T}(\underline{\sigma})$ over all $\underline{\mathtt{L}}$ is given by

$$\mathbb{E}^{\mathrm{lit}}[\boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G},T}(\underline{\sigma})^\lambda] = \boldsymbol{w}_{\mathcal{G},T}(\underline{\sigma})^\lambda \equiv \left\{ \prod_{v\in V} \dot{\Phi}(\underline{\sigma}_{\delta v}) \prod_{a\in F} \hat{\Phi}(\underline{\sigma}_{\delta a}) \prod_{e\in E} \bar{\Phi}_T(\sigma_e) \right\}^\lambda, \tag{27}$$

with $\hat{\Phi}(\underline{\sigma}) \equiv (\mathbb{E}^{\mathrm{lit}}[\hat{\Phi}(\underline{\sigma} \oplus \underline{\mathtt{L}})^\lambda])^{1/\lambda}$. A similar simplification holds in the second moment, where we consider pairs $\underline{\sigma} \equiv (\underline{\sigma}^1, \underline{\sigma}^2)$ with weights $\boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G},T}(\underline{\sigma}) \equiv \boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G},T}(\underline{\sigma}^1)\boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G},T}(\underline{\sigma}^2)$:

$$\mathbb{E}^{\mathrm{lit}}[\boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G},T}(\underline{\sigma})^\lambda] = \boldsymbol{w}_{\mathcal{G},T}(\underline{\sigma})^\lambda \equiv \left\{ \prod_{v\in V} \dot{\Phi}_2(\underline{\sigma}_{\delta v}) \prod_{a\in F} \hat{\Phi}_2(\underline{\sigma}_{\delta a}) \prod_{e\in E} \bar{\Phi}_{T,2}(\sigma_e) \right\}^\lambda \tag{28}$$

where $\dot{\Phi}_2(\underline{\sigma}) \equiv \dot{\Phi}(\underline{\sigma}^1)\dot{\Phi}(\underline{\sigma}^2)$, $\bar{\Phi}_{T,2}(\sigma) \equiv \bar{\Phi}_T(\sigma^1)\bar{\Phi}_T(\sigma^2)$, and

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

Let us emphasize that $\hat{\Phi}$ and $\hat{\Phi}_2$ depend on λ, although we suppress it from the notation.

Clearly the weight $\boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G},T}(\underline{\sigma})$ depends on $\underline{\mathtt{L}}$, since $\underline{\sigma}$ need not even be a valid coloring for all choices of $\underline{\mathtt{L}}$. However, the following lemma shows that, as long as $\underline{\sigma}$ remains valid, the size of its encoded cluster remains the same:

Lemma 3.2. *Given $\mathcal{G}$, let $\boldsymbol{w}^{\max}_{\mathcal{G},T}(\underline{\sigma})$ denote the maximum of $\boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G},T}(\underline{\sigma})$ over all $\mathscr{G} = (\mathcal{G}, \underline{\mathtt{L}})$. For any $\mathscr{G} = (\mathcal{G}, \underline{\mathtt{L}})$, $\boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G},T}(\underline{\sigma})$ is either zero or equal to $\boldsymbol{w}^{\max}_{\mathcal{G},T}(\underline{\sigma})$.*

Proof. We claim that for all $\underline{\sigma}, \underline{\mathtt{L}}$ we have the factorization

$$\begin{aligned} \hat{\Phi}^{\mathrm{lit}}(\underline{\sigma} \oplus \underline{\mathtt{L}}) &= \hat{I}^{\mathrm{lit}}(\underline{\sigma} \oplus \underline{\mathtt{L}}) \hat{\Phi}^{\max}(\underline{\sigma}), \text{ where} \\ \hat{\Phi}^{\max}(\underline{\sigma}) &\equiv \max\{\hat{\Phi}^{\mathrm{lit}}(\underline{\sigma} \oplus \underline{\mathtt{L}}) : \underline{\mathtt{L}} \in \{0,1\}^k\}. \end{aligned}$$

To see this, note that for $\zeta \in \Omega^{d-1}$ and $\xi \in \Omega^{k-1}$, if $\dot{I}(\sigma, \zeta) = 1$ and $\hat{I}^{\mathrm{lit}}(\sigma, \xi) = 1$, then

$$\begin{aligned} \dot{\Phi}(\sigma, \zeta)\bar{\Phi}_T(\sigma) = \dot{z}(\dot{\sigma}) &\equiv \begin{cases} \dot{z}(\dot{\tau}) & \text{if } \dot{\sigma} = \dot{\tau}, \\ 1 & \text{otherwise;} \end{cases} \\ \hat{\Phi}^{\mathrm{lit}}(\sigma, \xi)\bar{\Phi}_T(\sigma) = \hat{z}(\hat{\sigma}) &\equiv \begin{cases} \hat{z}(\hat{\tau}) & \text{if } \hat{\sigma} = \hat{\tau}, \\ 1 & \text{otherwise.} \end{cases} \end{aligned} \tag{29}$$

In particular, since $\hat{z}(\hat{\sigma}) = \hat{z}(\hat{\sigma} \oplus \mathtt{1})$, we see that the claim holds with $\hat{\Phi}^{\max}(\underline{\sigma}) = \hat{z}(\hat{\sigma}_i)/\bar{\Phi}_T(\sigma_i)$ for any $1 \leqslant i \leqslant k$. The lemma then follows: either $\boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G},T}(\underline{\sigma})$ is zero, or it equals

$$\prod_v \dot{\Phi}(\underline{\sigma}_{\delta v}) \prod_a \hat{\Phi}^{\max}(\underline{\sigma}_{\delta a}) \prod_e \bar{\Phi}_T(\sigma_e) = \boldsymbol{w}^{\max}_{\mathcal{G},T}(\underline{\sigma}),$$

as claimed. □

Lemma 3.2 says that, in averaging over the literals, we do not lose any essential information on the cluster size. For $\underline{\sigma} \in \Omega^k$, let

$$\hat{v}(\underline{\sigma}) \equiv \mathbb{E}^{\mathrm{lit}}[\hat{I}^{\mathrm{lit}}(\underline{\sigma} \oplus \underline{\mathtt{L}})] \tag{30}$$

denote the fraction of $\underline{\mathtt{L}} \in \{0,1\}^k$ which are compatible with $\underline{\sigma}$. Then Lemma 3.2 gives

$$\boldsymbol{w}_{\mathcal{G},T}(\underline{\sigma})^\lambda = \boldsymbol{w}^{\max}_{\mathcal{G},T}(\underline{\sigma})^\lambda \boldsymbol{p}_{\mathcal{G}}(\underline{\sigma}), \quad \boldsymbol{p}_{\mathcal{G}}(\underline{\sigma}) \equiv \prod_{a\in F} \hat{v}(\underline{\sigma}). \tag{31}$$

We will see below that, thanks to this simplification, we can extract the desired information from the averaged weights $\boldsymbol{w}_{\mathcal{G},T}(\underline{\sigma})$, without referring to the edge literals $\underline{\mathtt{L}}$.

Definition 3.3. On a bipartite factor graph $\mathcal{G}$ (without edge literals), the *factor model* with specification $g \equiv (\dot{g}, \hat{g}, \bar{g})$ is the probability measure $\nu_{\mathcal{G}}$ on configurations $\underline{\xi} \in \mathscr{X}^E$ defined by

$$\nu_{\mathcal{G}}(\underline{\xi}) = \frac{1}{Z} \prod_{v \in V} \dot{g}(\xi_{\delta v}) \prod_{a \in F} \hat{g}(\xi_{\delta a}) \prod_{e \in E} \bar{g}(\xi_e), \tag{32}$$

with Z the normalizing constant.

The measure (27) on T-colorings is a factor model with specification $(\dot{\Phi}, \hat{\Phi}, \bar{\Phi}_T)^\lambda$. The measure (28) on pairs of T-colorings is a factor model with specification $(\dot{\Phi}_2, \hat{\Phi}_2, \bar{\Phi}_{T,2})^\lambda$. To distinguish between the two cases, we sometimes refer to (27) as the "first-moment" or "single-copy" model, and refer to (28) as the "second-moment" or "pair" model. In much of what follows, we treat these two in a unified manner under the general framework (32).

3.2. **Empirical measures and moments.** We will decompose colorings $\underline{\sigma}$ according to their empirical measure H, defined as follows:

Definition 3.4. Given a coloring $\underline{\sigma}$ on $\mathscr{G} = (\mathcal{G}, \underline{\mathtt{L}})$, let

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

Note that the validity of $\underline{\sigma}$ on $\mathscr{G}$ clearly depends on $\underline{\mathtt{L}}$, but we can regard H as a function of $(\mathcal{G}, \underline{\sigma})$ only. We therefore write

$$H \equiv H(\mathscr{G}, \underline{\sigma}) \equiv H(\mathcal{G}, \underline{\sigma}) \equiv (\dot{H}, \hat{H}, \bar{H}),$$

and we term this the *empirical measure* of $\underline{\sigma}$ on $\mathcal{G}$.

If $\mathbf{H}$ is any subset of empirical measures H, we write $\underline{\sigma} \in \mathbf{H}$ to indicate that $H(\mathcal{G}, \underline{\sigma}) \in \mathbf{H}$, and let $\boldsymbol{Z}_{\lambda,T}(\mathbf{H})$ denote the contribution to $\boldsymbol{Z}_{\lambda,T}$ from (valid) colorings $\underline{\sigma} \in \mathbf{H}$. If $\mathbf{H}$ is a singleton $\{H\}$, then we write $\underline{\sigma} \in H$ to indicate $H(\mathcal{G}, \underline{\sigma}) = H$, and let $\boldsymbol{Z}_{\lambda,T}(H)$ denote the contribution from all colorings $\underline{\sigma} \in H$. Much of the paper concerns the calculation of first and second moments for $\boldsymbol{Z}_{\lambda,T}(H)$.

First note that for any pair $(\mathcal{G}, \underline{\sigma})$ with $H(\mathcal{G}, \underline{\sigma}) = H$, the weight $\boldsymbol{w}_{\mathcal{G},T}(\underline{\sigma})$ is the same and depends only on T, $\mathcal{G}$, and H. In fact, the weight equals $\boldsymbol{w}_{\mathcal{G}}(\underline{\sigma}) \equiv \boldsymbol{w}_{\mathcal{G},\infty}(\underline{\sigma})$ if the support of $\bar{H}$ is contained in Ω_T, and equals zero otherwise. From now on we assume $\bar{H}$ is supported within Ω_T, so $\boldsymbol{w}_{\mathcal{G},T}(\underline{\sigma}) = \boldsymbol{w}_{\mathcal{G}}(\underline{\sigma})$ depends only on $(\mathcal{G}, H)$, and can be denoted $\boldsymbol{w}_{\mathcal{G}}(H)$. Further, we see in (31) that, as long as $\operatorname{supp} \bar{H} \subseteq \Omega_T$, the weights $\boldsymbol{w}^{\max}_{\mathcal{G},T}(\underline{\sigma})$ and $\boldsymbol{p}_{\mathcal{G}}(\underline{\sigma})$ also depend only on $(\mathcal{G}, H)$, so we can rewrite (31) as

$$\boldsymbol{w}_{\mathcal{G}}(H)^\lambda = \boldsymbol{w}^{\max}_{\mathcal{G}}(H)^\lambda \boldsymbol{p}(H). \tag{33}$$

In what follows, for ease of notation we will often suppress the dependence on λ and T, and write simply $\boldsymbol{Z} \equiv \boldsymbol{Z}_{\lambda,T}$.

In fact we have a quite explicit expression for $\mathbb{E}\boldsymbol{Z}(H)$, as follows. We will use the usual multi-index notations, in particular, if π is a probability measure on a space X, we write

$$\binom{n}{n\pi} \equiv n! \Big/ \prod_{x \in X} (n\pi(x))!.$$

It follows straightforwardly from the definition of the random regular NAE-SAT graph that

$$\mathbb{E}\boldsymbol{Z}(H) = \left\{ \binom{n}{n\dot{H}} \binom{m}{m\hat{H}} \Big/ \binom{nd}{nd\bar{H}} \right\} \boldsymbol{w}_{\mathcal{G}}(H)^{\lambda}. \tag{34}$$

We write $\mathcal{H}(\pi) = -\langle \pi, \ln \pi \rangle$ for the Shannon entropy of π. Applying Stirling's formula gives the following:

Lemma 3.5. *For any fixed $H \equiv (\dot{H}, \hat{H}, \bar{H})$, we have in the limit of large n that*

$$\mathbb{E}\boldsymbol{Z}(H) \asymp n^{-\wp(H)/2} \exp\{n\boldsymbol{F}(H)\}$$

where for an empirical measure $H = (\dot{H}, \hat{H}, \bar{H})$ we define

$$\begin{aligned}
\boldsymbol{v}(H) &\equiv (d/k)\langle \ln \hat{v}, \hat{H} \rangle = n^{-1} \ln \boldsymbol{p}(H), \\
\boldsymbol{s}(H) &\equiv \langle \ln \dot{\Phi}, \dot{H} \rangle + (d/k)\langle \ln \hat{\Phi}^{\max}, \hat{H} \rangle + d\langle \ln \bar{\Phi}, \bar{H} \rangle = n^{-1} \ln \boldsymbol{w}^{\max}(H), \\
\boldsymbol{\Sigma}(H) &\equiv \mathcal{H}(\dot{H}) + (d/k)\mathcal{H}(\hat{H}) - d\mathcal{H}(\bar{H}) + \boldsymbol{v}(H), \\
\boldsymbol{F}(H) &\equiv \boldsymbol{\Sigma}(H) + \boldsymbol{s}(H)\lambda, \\
\wp(H) &\equiv |\operatorname{supp} \dot{H}| + |\operatorname{supp} \hat{H}| - |\operatorname{supp} \bar{H}| - 1.
\end{aligned} \tag{35}$$

3.3. **Outline of first moment.** The function $\boldsymbol{F}(H)$ is difficult to optimize directly, and we combine a few techniques in order to analyze it. In view of the result of [DSS14] (see Remark 1.1), we restrict consideration to the regime

$$(2^{k-1} - 2)\ln 2 \equiv \alpha_{\mathrm{lbd}} \leqslant d \leqslant \alpha_{\mathrm{ubd}} \equiv 2^{k-1} \ln 2. \tag{36}$$

In this regime, we use *a priori* estimates to show that the optimal H must lie in a certain restricted set $\mathbf{N}_{\circ}$. We then show that in the restricted set, a certain block optimization procedure converges to a unique, and explicit, optimizer $H_{\star}$. The convergence of the block optimization is based on a certain contraction estimate for the belief propagation recursion, which we describe below.

First, to describe the set $\mathbf{N}_{\circ}$, let us abbreviate $\bar{H}(\mathtt{r})$ and $\bar{H}(\mathtt{f})$ for the mass assigned by $\bar{H}$ to the sets $\{\mathtt{r}\} \equiv \{\mathtt{r}_0, \mathtt{r}_1\}$ and $\{\mathtt{f}\} \equiv \Omega_{\mathtt{f}}$; and let $\mathbf{N}_{\circ}$ denote the set of H such that

$$\max\{\bar{H}(\mathtt{f}), \bar{H}(\mathtt{r})\} \leqslant 7/2^k. \tag{37}$$

The following *a priori* estimate shows that in the regime (36), the measures $H \notin \mathbf{N}_{\circ}$ give a negligible contribution to the first moment.

Lemma 3.6. *Let $\boldsymbol{Z}((\mathbf{N}_{\circ})^c)$ be the contribution to $\boldsymbol{Z} = \boldsymbol{Z}_{\lambda,T}$ from empirical measures $H \notin \mathbf{N}_{\circ}$. For $k \geqslant k_0$, α satisfying (36), and $0 \leqslant \lambda \leqslant 1$, $\mathbb{E}\boldsymbol{Z}((\mathbf{N}_{\circ})^c)$ is exponentially small in n.*

Proof. In view of Proposition 2.13, for $0 \leqslant \lambda \leqslant 1$ we have

$$\boldsymbol{Z}((\mathbf{N}_{\circ})^c) \leqslant Z^{\mathtt{free}} + Z^{\mathtt{red}}$$

where $Z^{\mathtt{free}}$ (resp. $Z^{\mathtt{red}}$) counts NAE-SAT solutions $\underline{\boldsymbol{x}} \in \{0,1\}^V$ which map — via coarsening and the bijection (15) — to warning configurations $\underline{y}$ with density of $\mathtt{free}$ (resp. $\mathtt{red}$) edges $\geqslant 7/2^k$. For α satisfying (36), $\mathbb{E}Z^{\mathtt{f}}$ is exponentially small in n by [DSS14, Propn. 2.2]. As for $Z^{\mathtt{red}}$, let us say that an edge $e \in E$ is *blocked* under $\underline{\boldsymbol{x}} \in \{0,1\}^V$ if

$$\mathtt{L}_e \oplus x_{v(e)} = \mathtt{1} \oplus \mathtt{L}_{e'} \oplus x_{v(e')} \quad \text{for all } e' \in \delta a(e) \backslash e.$$

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

$$\mathbb{E}Z^{\mathtt{red}} \leqslant (\mathbb{E}Z)\mathbb{P}\Big(\mathrm{Bin}(m,\theta) \geqslant 7nd/2^k\Big),$$

which is exponentially small in n by a standard Chernoff bound, in combination with the trivial bound $\mathbb{E}Z \leqslant 2^n$. □

Lemma 3.6 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 non-negative, since otherwise $\mathbb{E}\boldsymbol{Z}$ is exponentially small in n and there is nothing to prove. From this we conclude that any maximizer H of $\boldsymbol{F}$ must lie in $\mathbf{N}_\circ$. By a block optimization procedure in $\mathbf{N}_\circ$, we prove

Proposition 3.7 (*proved in Section 6*)**.** *Assuming the global maximum of $\boldsymbol{F}$ is non-negative, the unique maximizer of $\boldsymbol{F}$ is a point $H_\star$ in the interior of $\mathbf{N}_\circ$. Further, there is a positive constant $\epsilon = \epsilon(k,\lambda,T)$ so that for $\|H - H_\star\| \leqslant \epsilon$, $\boldsymbol{F}(H) \leqslant \boldsymbol{F}(H_\star) - \epsilon\|H - H_\star\|^2$. Explicitly,*

$$\dot{H}_\star(\dot{\zeta}) = \frac{\dot{\Phi}(\dot{\zeta})^\lambda}{\dot{Z}_\star}\prod_{i=1}^d \hat{q}_\star(\dot{\zeta}_i), \quad \hat{H}_\star(\hat{\xi}) = \frac{\hat{\Phi}(\hat{\xi})^\lambda}{\hat{Z}_\star}\prod_{i=1}^d \dot{q}_\star(\hat{\xi}_i), \quad \bar{H}_\star(\bar{\sigma}) = \frac{\bar{\Phi}(\bar{\sigma})^{-\lambda}}{\bar{Z}_\star}\dot{q}_\star(\dot{\sigma})\hat{q}_\star(\hat{\sigma}), \tag{38}$$

where $\dot{q}_\star$ is the fixed point of $\mathtt{BP}_{\lambda,T}$ given by Proposition 4.2, $\hat{q}_\star = \hat{\mathtt{BP}}_{\lambda,T}(\dot{q}_\star)$, and $\dot{Z}_\star$, $\hat{Z}_\star$, $\bar{Z}_\star$ are the normalizing constants such that $\dot{H}_\star, \hat{H}_\star, \bar{H}_\star$ are probability measures.

A straightforward consequence of the above is that we can compute the first moment of $\boldsymbol{Z}$ up to constant factors. More formally, define the neighborhood

$$\mathbf{N} = \{H : \|H - H_\star\| \leqslant n^{-1/3}\} \subseteq \mathbf{N}_\circ.$$

We say $\underline{\sigma} \in \mathbf{N}$ if $H(\mathcal{G}, \underline{\sigma}) \in \mathbf{N}$, and let $\boldsymbol{Z}(\mathbf{N})$ be the contribution to $\boldsymbol{Z}$ from colorings $\underline{\sigma} \in \mathbf{N}$. In the following, let $\dot{s} \equiv \dot{s}(T)$ count the number of d-tuples $\underline{\sigma} \in (\Omega_T)^d$ for which $\dot{\Phi}(\underline{\sigma}) > 0$. Let $\hat{s} \equiv \hat{s}(T)$ count the number of k-tuples $\underline{\sigma} \in (\Omega_T)^k$ for which $\hat{\Phi}(\underline{\sigma}) > 0$. Let $\bar{s} \equiv |\Omega_T|$, and denote $\wp \equiv \dot{s} + \hat{s} - \bar{s} - 1$.

Corollary 3.8. *In the setting of Proposition 3.7,*

$$\mathbb{E}\boldsymbol{Z}(\mathbf{N}) \asymp \mathbb{E}\boldsymbol{Z} \asymp \exp\{\boldsymbol{F}(H_\star)\}.$$

Proof. In an pair empirical measure $H = (\dot{H}, \hat{H}, \bar{H})$, the edge marginal $\bar{H}$ can be determined from either the variable or the clause measure:

$$nd\bar{H}(\bar{\sigma}) = \sum_{\dot{\zeta}} n\dot{H}(\dot{\zeta})\dot{M}(\bar{\sigma},\dot{\zeta}) = \sum_{\hat{\xi}} m\hat{H}(\hat{\xi})\hat{M}(\bar{\sigma},\hat{\xi}) \tag{39}$$

where $\dot{M} \in \mathbb{R}^{\bar{s}\times\dot{s}}$ and $\hat{M} \in \mathbb{R}^{\bar{s}\times\hat{s}}$ are defined by

$$\dot{M}(\bar{\sigma},\dot{\zeta}) = \sum_{i=1}^d \mathbf{1}\{\dot{\zeta}_i = \bar{\sigma}\}, \quad \hat{M}(\bar{\sigma},\hat{\xi}) = \sum_{i=1}^k \mathbf{1}\{\hat{\xi}_i = \bar{\sigma}\}.$$

The $(\dot{s}+\hat{s})$-dimensional vector $(\dot{H}, \hat{H})$ gives rise to a valid empirical measure on the graph $\mathcal{G}$ if and only if

(i) $\langle \mathbf{1}, \dot{H}\rangle = 1$;
(ii) $(n\dot{H}, m\hat{H})$ lies in the kernel of the $\bar{s} \times (\dot{s}+\hat{s})$ matrix $M \equiv \begin{pmatrix}\dot{M} & -\hat{M}\end{pmatrix}$;
(iii) $(n\dot{H}, m\hat{H})$ is integer-valued;
(iv) $\dot{H}, \hat{H} \geqslant 0$.

One can verify that the matrix M is of full rank, from which it follows that the space of $(\dot{H}, \hat{H})$ satisfying (i) and (ii) has dimension $\wp$. In Lemma 5.6 we will show that M satisfies a stronger condition, which implies that the space of $(\dot{H}, \hat{H})$ satisfying (i), (ii), and (iii) is an affine translation of $(n^{-1}\mathbb{Z})^{\wp}$, where the coefficients of the transformation are bounded. It then follows by combining Lemma 3.5 and Proposition 3.7 that

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

The contribution to $\mathbb{E}\boldsymbol{Z}$ from $H \notin \mathbf{N}$ is negligible, so the above estimate holds as well with $\mathbb{E}\boldsymbol{Z}(\mathbf{N})$ in place of $\mathbb{E}\boldsymbol{Z}$. □

3.4. **Second moment of correlated pairs.** We will show in Section 10 that for fixed $\lambda \in [0,1]$, the pair $(\boldsymbol{s}(H_\star), \boldsymbol{\Sigma}(H_\star))$ converges as $T \to \infty$ to a limit $(s_\lambda, \Sigma(s_\lambda))$, which matches the physics 1RSB prediction. We then consider the second moment only for colorings in $\mathbf{N}$, beginning with the following definition (following [CP16a]) which is intended to address the contribution from pairs of colorings with large correlation.

Definition 3.9. Given a coloring $\underline{\sigma}$ of $\mathcal{G}$, write $\underline{x}(\underline{\sigma}) \equiv (x_v(\underline{\sigma}))_{v\in V}$ for the corresponding frozen configuration. For two colorings $\underline{\sigma}, \underline{\sigma}'$ of $\mathcal{G}$, let

$$\delta(\underline{\sigma}, \underline{\sigma}') \equiv |\{v \in V : x_v(\underline{\sigma}) \neq x_v(\underline{\sigma}')\}|/|V|.$$

Let $I_{\text{sep}} \equiv [(1 - k^4/2^{k/2})/2, (1 + k^4/2^{k/2})/2]$. 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 say that a coloring $\underline{\sigma} \in \mathbf{N}$ is *separable* if

$$|\{\underline{\sigma}' \in \mathbf{N} : \underline{\sigma}' \succcurlyeq \underline{\sigma} \text{ and } \delta(\underline{\sigma}, \underline{\sigma}') \notin I_{\text{sep}}\}| \leqslant \exp\{(\ln n)^4\},$$

where it is understood that both $\underline{\sigma}, \underline{\sigma}'$ must be valid colorings.

Proposition 3.10 (*proved in Section 7*)**.** *If $\boldsymbol{S}(\mathbf{N})$ is the contribution to $\boldsymbol{Z}(\mathbf{N})$ from separable colorings, then $\mathbb{E}\boldsymbol{S}(\mathbf{N}) = (1 - o(1))\mathbb{E}\boldsymbol{Z}(\mathbf{N})$.*

In the second moment, we continue to write $H \equiv (\dot{H}, \hat{H}, \bar{H})$ for the empirical measure, with the understanding that it now refers to pair colorings $\underline{\sigma} = (\underline{\sigma}^1, \underline{\sigma}^2)$. Thus $\dot{H}$ is in this context a measure on $(\dot{\Omega}^d)^2$, and so on. If we wish to emphasize that we are in the second moment setting, we will refer to H as the *pair* empirical measure. The single-copy marginals of H are defined as $H^j = (\dot{H}^j, \hat{H}^j, \bar{H}^j)$ for $j = 1, 2$ where

$$\dot{H}^j(\zeta) = \sum_{\underline{\sigma}^1, \underline{\sigma}^2} \dot{H}(\underline{\sigma}^1, \underline{\sigma}^2)\mathbf{1}\{\underline{\sigma}^j = \zeta\},$$

and similarly for $\hat{H}^j, \bar{H}^j$. To calculate the second moment of $\boldsymbol{Z}(\mathbf{N})$, we must understand all pair empirical measures H in the set

$$\mathbf{N}_2 \equiv \{H : H^1, H^2 \in \mathbf{N}\}.$$

The purpose of Definition 3.9 is to allow us to make a further restriction: we compute the second moment of $\boldsymbol{S}(\mathbf{N})$ rather than of $\boldsymbol{Z}(\mathbf{N})$. Any $\underline{\sigma} = (\underline{\sigma}^1, \underline{\sigma}^2)$ with pair empirical measure H will have the same value $\delta(\underline{\sigma}^1, \underline{\sigma}^2) = \delta$, so we can define $\delta(H) = \delta$. Let

$$\mathbf{N}_{\text{sep}} \equiv \{H \in \mathbf{N}_2 : \delta(H) \in I_{\text{sep}}\}, \quad \mathbf{N}_{\text{ns}} \equiv \mathbf{N}_2 \backslash \mathbf{N}_{\text{sep}}.$$

Lemma 3.11. *If $\boldsymbol{S}^2(\mathbf{N}_{\text{ns}})$ is the contribution to $\boldsymbol{S}(\mathbf{N})^2$ from pair empirical measures $H \in \mathbf{N}_{\text{ns}}$, then $\mathbb{E}[\boldsymbol{S}^2(\mathbf{N}_{\text{ns}})] \dot{\leqslant} \exp\{n\boldsymbol{s}(H_\star)\lambda + o(n)\}\,\mathbb{E}\boldsymbol{Z}$.*

Proof. Denote $\underline{\sigma} \in \boldsymbol{S}(\mathbf{N})$ if $\underline{\sigma}$ contributes to $\boldsymbol{S}(\mathbf{N})$, meaning that $\underline{\sigma}$ is separable and has empirical measure in $\mathbf{N}$. Then, by symmetry,

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

where the last step is by the definition of separability. The result follows easily by noting that $\boldsymbol{S}(\mathbf{N}) \leqslant \boldsymbol{Z}$. □

Corollary 3.12. *For any $\lambda \in [0,1]$ with $\Sigma(s_\lambda) > 0$, there exists $T(\lambda)$ large enough such that for all $T \geqslant T(\lambda)$, the ratio*

$$\frac{\mathbb{E}[\boldsymbol{S}^2(\mathbf{N}_{\mathrm{ns}})]}{(\mathbb{E}\boldsymbol{Z}(\mathbf{N}))^2}$$

decays exponentially with n.

Proof. By Lemma 3.11 and and Proposition 3.7,

$$\frac{\mathbb{E}[\boldsymbol{S}^2(\mathbf{N}_{\mathrm{ns}})]}{(\mathbb{E}\boldsymbol{Z}(\mathbf{N}))^2} \leqslant \frac{\exp\{n\boldsymbol{s}(H_\star)\lambda + o(n)\}}{\mathbb{E}\boldsymbol{Z}(\mathbf{N})} = \exp\{-n\boldsymbol{\Sigma}(H_\star) + o(n)\}.$$

Since for fixed λ the pair $(\boldsymbol{s}(H_\star), \boldsymbol{\Sigma}(H_\star))$ converges in the limit $T \to \infty$ to $(s_\lambda, \Sigma(s_\lambda))$, for $T \geqslant T(\lambda)$ the above ratio decays exponentially with n, concluding the proof. □

3.5. **Second moment of uncorrelated pairs.** The derivation of Lemma 3.5 applies equally well to the second moment, giving the expansion

$$\mathbb{E}[\boldsymbol{Z}^2(H)] \asymp n^{-\wp(H)} \exp\{n\boldsymbol{F}_2(H)\}$$

where H is the empirical measure for pair colorings, and $\wp(H), \boldsymbol{F}_2(H)$ are defined explicitly as follows. Recalling (30), for $\underline{\sigma} \in \Omega^{2k}$ let

$$\hat{v}_2(\underline{\sigma}) \equiv \mathbb{E}^{\mathrm{lit}}[\hat{I}^{\mathrm{lit}}(\underline{\sigma}^1 \oplus \underline{\mathtt{L}})\hat{I}^{\mathrm{lit}}(\underline{\sigma}^2 \oplus \underline{\mathtt{L}})].$$

For a pair empirical measure H with single-copy marginals H^1, H^2 we have (cf. (35))

$$\begin{aligned}\boldsymbol{v}_2(H) &\equiv (d/k)\langle \ln \hat{v}_2, \hat{H}\rangle, \\ \boldsymbol{s}_2(H) &\equiv \boldsymbol{s}(H^1) + \boldsymbol{s}(H^2), \\ \boldsymbol{\Sigma}_2(H) &\equiv \mathcal{H}(\dot{H}) + (d/k)\mathcal{H}(\hat{H}) - d\mathcal{H}(\bar{H}) + \boldsymbol{v}_2(H), \\ \boldsymbol{F}_2(H) &\equiv \boldsymbol{\Sigma}_2(H) + \boldsymbol{s}_2(H)\lambda, \\ \wp(H) &\equiv |\operatorname{supp}\dot{H}| + |\operatorname{supp}\hat{H}| - |\operatorname{supp}\bar{H}| - 1.\end{aligned} \tag{40}$$

We will show that the maximizer for $\boldsymbol{F}_2$ can be described in terms of the maximizer $H_\star$ of $\boldsymbol{F}$. To this end, we will say that a measure $\hat{K}$ on pairs $(\underline{\xi}, \underline{\mathtt{L}})$ is *factorized* if

(i) the marginal $\hat{K}$ on $\underline{\mathtt{L}}$ is uniform over $\{\mathtt{0},\mathtt{1}\}^k$, and
(ii) for each $\underline{\xi}$ the conditional measure $\hat{K}(\underline{\mathtt{L}}|\underline{\xi})$ is uniform on $\{\underline{\mathtt{L}} : \hat{I}^{\mathrm{lit}}(\underline{\xi} \oplus \underline{\mathtt{L}}) = 1\}$.

From (38) and Lemma 3.2, $\hat{H}_\star$ is the marginal on $\underline{\xi}$ of the probability measure $\hat{K}_\star$ on pairs $(\underline{\xi}, \underline{\mathtt{L}}) \in \Omega^k \times \{\mathtt{0},\mathtt{1}\}^k$ defined by

$$\hat{K}_\star(\underline{\xi}, \underline{\mathtt{L}}) \cong \hat{I}^{\mathrm{lit}}(\underline{\xi} \oplus \underline{\mathtt{L}})\hat{g}(\underline{\xi}), \quad \text{where } \hat{g}(\underline{\xi}) = \hat{\Phi}^{\max}(\underline{\xi}) \prod_{i=1}^k \dot{q}_\star(\xi_i).$$

We will characterize $\dot{q}_\star$ in detail below, but for now we note that it has the symmetry $\dot{q}_\star(\dot{\sigma}) = \dot{q}_\star(\dot{\sigma} \oplus \mathtt{1})$, which implies $\hat{g}(\underline{\xi}) = \hat{g}(\underline{\xi} \oplus \underline{\mathtt{L}})$ for any $\underline{\mathtt{L}} \in \{\mathtt{0},\mathtt{1}\}^k$. It follows from this that the measure $\hat{K}_\star$ is indeed factorized.

Lemma 3.13. *Assume we have empirical measures $H^j = (\dot{H}^j, \hat{H}^j, \bar{H}^j)$ $(j = 1, 2)$, such that $\hat{H}^j$ is the marginal on $\underline{\xi}$ of an $\underline{\mathtt{L}}$-factorized measure $\hat{K}^j$. Suppose $H = (\dot{H}, \hat{H}, \bar{H})$ where $\dot{H}, \bar{H}$ are the product measures $\dot{H}^1 \otimes \dot{H}^2$ and $\bar{H}^1 \otimes \bar{H}^2$, and*

$$\hat{H}(\underline{\xi}) = \mathbb{E}^{\mathrm{lit}}[\hat{K}^1(\underline{\xi}^1|\underline{\mathtt{L}})\hat{K}^2(\underline{\xi}^2|\underline{\mathtt{L}})].$$

Then $\boldsymbol{F}_2(H) = \boldsymbol{F}(H^1) + \boldsymbol{F}(H^2)$.

Proof. From the definitions we have

$$(k/d)[\boldsymbol{F}_2(H) - \boldsymbol{F}(H^1) - \boldsymbol{F}(H^2)] = \mathcal{H}(\hat{H}) + \langle \ln \hat{v}_2, \hat{H}\rangle - \sum_{j=1,2}[\mathcal{H}(\hat{H}^j) + \langle \ln \hat{v}, \hat{H}^j\rangle].$$

From the assumption, $\hat{H}$ is the marginal on $\underline{\xi}$ of the measure $\hat{K}(\underline{\xi}, \underline{\mathtt{L}}) = 2^{-k}\hat{K}^1(\underline{\xi}^1|\underline{\mathtt{L}})\hat{K}^2(\underline{\xi}^2|\underline{\mathtt{L}})$. Note that the marginal of $\hat{K}$ on $\underline{\mathtt{L}}$ is uniform, and $\hat{K}(\underline{\mathtt{L}}|\underline{\xi}^1, \underline{\xi}^2)$ is uniform on $\underline{\mathtt{L}}$ compatible with both $\underline{\xi}^1, \underline{\xi}^2$. Therefore, letting $(\underline{\xi}^1, \underline{\xi}^2, \underline{\mathtt{L}})$ denote a random sample from $\hat{K}$,

$$\mathcal{H}(\hat{H}) + \langle \ln \hat{v}_2, \hat{H}\rangle = \mathcal{H}(\underline{\xi}^1, \underline{\xi}^2|\underline{\mathtt{L}}) + \mathcal{H}(\underline{\mathtt{L}}) - \mathcal{H}(\underline{\mathtt{L}}|\underline{\xi}^1, \underline{\xi}^2) + \langle \ln \hat{v}_2, \hat{H}\rangle = \mathcal{H}(\underline{\xi}^1, \underline{\xi}^2|\underline{\mathtt{L}}).$$

Applying conditional independence gives

$$\mathcal{H}(\underline{\xi}^1, \underline{\xi}^2|\underline{\mathtt{L}}) = \sum_{j=1,2}[\mathcal{H}(\underline{\xi}^j) + \mathcal{H}(\underline{\mathtt{L}}|\underline{\xi}^j) - \mathcal{H}(\underline{\mathtt{L}})] = \sum_{j=1,2}[\mathcal{H}(\hat{H}^j) + \langle \ln \hat{v}, \hat{H}^j\rangle],$$

which proves $\boldsymbol{F}_2(H) = \boldsymbol{F}(H^1) + \boldsymbol{F}(H^2)$. □

Proposition 3.14 (*proved in Section 6*)**.** *The unique maximizer of $\boldsymbol{F}_2$ in $\mathbf{N}_{\mathrm{sep}}$ is the pair empirical measure $H_\otimes = (\dot{H}_\otimes, \hat{H}_\otimes, \bar{H}_\otimes)$ given by $\dot{H}_\otimes = \dot{H}_\star \otimes \dot{H}_\star$, $\bar{H}_\otimes = \bar{H}_\star \otimes \bar{H}_\star$, and*

$$\hat{H}_\otimes = \mathbb{E}^{\mathrm{lit}}[\hat{K}_\star(\cdot|\underline{\mathtt{L}}) \otimes \hat{K}_\star(\cdot|\underline{\mathtt{L}})].$$

Further, there is a positive constant $\epsilon = \epsilon(k, \lambda, T)$ so that for $\|H - H_\otimes\| \leqslant \epsilon$,

$$\boldsymbol{F}_2(H) \leqslant \boldsymbol{F}_2(H_\otimes) - \epsilon\|H - H_\otimes\|^2.$$

Corollary 3.15. *There exists a constant $C = C(k, \lambda, T)$ such that*

$$\mathbb{E}[\boldsymbol{Z}^2(\mathbf{N}_{\mathrm{sep}})] \leqslant C(\mathbb{E}\boldsymbol{Z}(\mathbf{N}))^2$$

Proof. Recall from Corollary 3.8 the definition of $(\dot{s}, \hat{s}, \bar{s})$ for the single-copy model, and define $(\dot{s}_2, \hat{s}_2, \bar{s}_2)$ analogously for the pair model. Let $\wp_2 \equiv \dot{s}_2 + \hat{s}_2 - \bar{s}_2 - 1$. For any fixed $H^1, H^2 \in \mathbf{N}$, the set of pair empirical measures H with single-copy marginals (H^1, H^2) spans a space of dimension $\wp_2 - 2\wp$. Thus, writing $\mathbb{E}[\boldsymbol{Z}^2(H^1, H^2)]$ for the second-moment contribution from such measures, it follows from Proposition 3.14 and Lemma 5.6 that

$$\mathbb{E}[\boldsymbol{Z}^2(H^1, H^2)] \asymp n^{-\wp}\exp\{n\boldsymbol{F}_2(H_\otimes)\}.$$

Summing over $(H^1, H^2) \in \mathbf{N}_{\mathrm{sep}}$ then gives

$$\mathbb{E}[\boldsymbol{Z}^2(\mathbf{N}_{\mathrm{sep}})] \asymp n^{-\wp}\exp\{n\boldsymbol{F}_2(H_\otimes)\},$$

which in turn is $\asymp (\mathbb{E}\boldsymbol{Z}(\mathbf{N}))^2$ by Proposition 3.7 and Lemma 3.13. □

3.6. **Conclusion of main result.** We now explain that the main theorem follows from the preceding assertions.

Corollary 3.16. *For any $\lambda \in [0,1]$ with $\Sigma(s_\lambda) > 0$, there exists $T(\lambda)$ large enough such that for all $T \geqslant T(\lambda)$, and for n sufficiently large,*

$$\mathbb{E}[\boldsymbol{S}(\mathbf{N})^2] \leqslant C(\mathbb{E}\boldsymbol{S}(\mathbf{N}))^2$$

for a constant $C = C(k, \lambda, T)$.

Proof. Since $\boldsymbol{S} \leqslant \boldsymbol{Z}$, we can bound

$$\mathbb{E}[\boldsymbol{S}(\mathbf{N})^2] \leqslant \mathbb{E}[\boldsymbol{Z}^2(\mathbf{N}_{\mathrm{sep}})] + \mathbb{E}[\boldsymbol{Z}^2(\mathbf{N}_{\mathrm{ns}})].$$

By Corollaries 3.12 and 3.15, the above is bounded by a constant times $(\mathbb{E}\boldsymbol{Z}(\mathbf{N}))^2$, which in turn is bounded by a constant times $(\mathbb{E}\boldsymbol{S}(\mathbf{N}))^2$ by Proposition 3.10. □

Corollary 3.16 implies $\mathbb{P}(\boldsymbol{S}(\mathbf{N}) \geqslant \delta\mathbb{E}\boldsymbol{S}(\mathbf{N})) \geqslant \delta$ for some positive constant δ. By adapting methods of [DSS14] we can strengthen this to

Proposition 3.17. *In the setting of Corollary 3.16, $\boldsymbol{S}(\mathbf{N})$ concentrates around its mean in the sense that $\lim_{\epsilon \downarrow 0} \liminf \mathbb{P}(\epsilon \leqslant \boldsymbol{S}(\mathbf{N})/\mathbb{E}\boldsymbol{S}(\mathbf{N}) \leqslant \epsilon^{-1}) = 1$.*[1]

Proof. This is a straightforward consequence of the method described in [DSS14, §6]. □

Corollary 3.18. *For $k \geqslant k_0$, $\mathsf{f}(\alpha) \geqslant \mathsf{f}^{\mathrm{1RSB}}(\alpha)$ for all $\alpha_{\mathrm{lbd}} \leqslant \alpha < \alpha_{\mathrm{sat}}$.*

Proof. Follows by combining Corollary 3.16 and Proposition 3.17. □

The proofs of the above propositions occupies Sections 4 through 7, with the contraction estimates deferred to Section 9. In Section 8 we will show

Proposition 3.19. *For $k \geqslant k_0$, it holds for all $\alpha < \alpha_{\mathrm{sat}}$ that $\mathsf{f}(\alpha) \leqslant \mathsf{f}^{\mathrm{1RSB}}(\alpha)$.*

Proof of Theorem 1. Follows by combining Corollary 3.18 and Proposition 3.19. □

4. Tree recursions

4.1. **Belief propagation.** We now describe the *belief propagation* (BP) recursions for this model. In the standard formulation (see e.g. [MM09, Ch. 14]), this is a pair of relations for two probability measures $\dot{\boldsymbol{q}}, \hat{\boldsymbol{q}}$ over Ω:

$$\dot{\boldsymbol{q}}(\sigma) = [\dot{\boldsymbol{B}}(\hat{\boldsymbol{q}})](\sigma) = \frac{1}{\dot{\boldsymbol{z}}}\bar{\Phi}_T(\sigma)^\lambda \sum_{\sigma_2,\dots,\sigma_d} \dot{\Phi}(\sigma, \sigma_2, \dots, \sigma_d)^\lambda \prod_{i=2}^{d} \hat{\boldsymbol{q}}(\sigma_i)$$

$$\hat{\boldsymbol{q}}(\sigma) = [\hat{\boldsymbol{B}}(\dot{\boldsymbol{q}})](\sigma) = \frac{1}{\hat{\boldsymbol{z}}}\bar{\Phi}_T(\sigma)^\lambda \sum_{\sigma_2,\dots,\sigma_k} \hat{\Phi}(\sigma, \sigma_2, \dots, \sigma_k)^\lambda \prod_{i=2}^{k} \dot{\boldsymbol{q}}(\sigma_i)$$

where $\dot{\boldsymbol{z}}, \hat{\boldsymbol{z}}$ are the normalizing constants ensuring that the outputs are probability measures. The first equation above is the *variable recursion*, and the second is the *clause recursion*. A standard simplification (see e.g. [MM09, Ch. 19]) is to assume a one-sided dependence:

$$\dot{\boldsymbol{q}}(\sigma) \cong \dot{q}(\dot{\sigma}) \text{ and } \hat{\boldsymbol{q}}(\sigma) \cong \hat{q}(\hat{\sigma}). \tag{41}$$

[1]The upper bound follows trivially from Markov's inequality, so the task is to show the lower bound.

where $\dot{q}, \hat{q}$ are probability measures on $\dot{\Omega}, \hat{\Omega}$, and $\cong$ denotes equivalence up to normalization. To see that this restriction makes sense, we note the following lemma which confirms that the restriction is preserved under the BP mapping:

Lemma 4.1. *The restriction* (41) *is preserved under the* BP *mapping, that is, if* $\dot{\boldsymbol{q}}$ *depends only on* $\dot{\sigma}$ *then* $\hat{\boldsymbol{B}}(\dot{\boldsymbol{q}})$ *depends only on* $\hat{\sigma}$*; and if* $\hat{\boldsymbol{q}}$ *depends only on* $\hat{\sigma}$ *then* $\dot{\boldsymbol{B}}(\hat{\boldsymbol{q}})$ *depends only on* $\dot{\sigma}$*.*

Proof. Suppose $\hat{\boldsymbol{q}}$ depends only on $\hat{\sigma}$, so $\hat{\boldsymbol{q}}(\sigma) \cong \hat{q}(\hat{\sigma})$, and consider the variable BP mapping $\dot{\boldsymbol{B}}$. If $\sigma \notin \Omega_{\mathtt{f}}$ then $\hat{\sigma}$ is uniquely determined by $\dot{\sigma}$, so there is nothing to prove. Therefore we need only consider the case that $\sigma \in \Omega_{\mathtt{f}}$. In order for $\dot{I}(\sigma, \sigma_2, \dots, \sigma_d) = 1$, we must have

$$\dot{\sigma} = \dot{\mathtt{T}}(\hat{\sigma}_2, \dots, \hat{\sigma}_d); \tag{42}$$

note that this condition does not depend on $\hat{\sigma}$. Further, given $(\sigma, \hat{\sigma}_2, \dots, \hat{\sigma}_d)$ satisfying (42), there is a unique choice of $(\dot{\sigma}_2, \dots, \dot{\sigma}_d)$ for which $\dot{I}(\sigma, \sigma_2, \dots, \sigma_d) = 1$; it is determined by the relation $\dot{\sigma}_i = \dot{\mathtt{T}}((\hat{\sigma}_j)_{j \neq i})$. In this case, applying (29) gives

$$\dot{\Phi}(\sigma, \sigma_2, \dots, \sigma_d)\bar{\Phi}(\sigma) = \dot{z}(\dot{\sigma}),$$

which also does not depend on $\hat{\sigma}$. It follows that

$$[\dot{\boldsymbol{B}}(\hat{\boldsymbol{q}})](\sigma) \cong \dot{z}(\dot{\sigma})^{\lambda} \sum_{\hat{\sigma}_2, \dots, \hat{\sigma}_d} \mathbf{1}\{\dot{\sigma} = \dot{\mathtt{T}}((\hat{\sigma}_i)_{i \geqslant 2})\} \prod_{i=2}^{d} \hat{q}(\hat{\sigma}_i).$$

The right-hand side does not depend on $\hat{\sigma}$, which proves the claim concerning $\dot{\boldsymbol{B}}$.

Similarly, suppose $\dot{\boldsymbol{q}}$ depends only on $\dot{\sigma}$, so $\dot{\boldsymbol{q}}(\sigma) \cong \dot{q}(\dot{\sigma})$, and consider the clause mapping $\hat{\boldsymbol{B}}$. Again, if $\sigma \notin \Omega_{\mathtt{f}}$ then there is nothing to prove, so suppose $\sigma \in \Omega_{\mathtt{f}}$. Then, in order for $\hat{I}^{\mathrm{lit}}((\sigma, \sigma_2, \dots, \sigma_k) \oplus \underline{\mathtt{L}}) = 1$, we must have

$$\hat{\sigma} = \mathtt{L}_1 \oplus \hat{\mathtt{T}}((\dot{\sigma}_i \oplus \mathtt{L}_i)_{i \geqslant 2}); \tag{43}$$

note that this condition does not depend on $\dot{\sigma}$. Further, given $(\sigma, \dot{\sigma}_2, \dots, \dot{\sigma}_k, \underline{\mathtt{L}})$ satisfying (43), there is a unique choice of $(\hat{\sigma}_2, \dots, \hat{\sigma}_k)$ for which $\hat{I}^{\mathrm{lit}}((\sigma, \sigma_2, \dots, \sigma_k) \oplus \underline{\mathtt{L}}) = 1$; it is determined by the mapping $\hat{\mathtt{T}}$. In this case, applying (29) gives

$$\hat{\Phi}((\sigma, \sigma_2, \dots, \sigma_k) \oplus \underline{\mathtt{L}})\bar{\Phi}(\sigma) = \hat{z}(\hat{\sigma}),$$

which also does not depend on $\dot{\sigma}$. It follows that

$$[\hat{\boldsymbol{B}}(\dot{\boldsymbol{q}})](\sigma) \cong \hat{z}(\hat{\sigma})^{\lambda} \sum_{\underline{\mathtt{L}}} \sum_{\dot{\sigma}_2, \dots, \dot{\sigma}_d} \mathbf{1}\{\hat{\sigma} = \mathtt{L}_1 \oplus \hat{\mathtt{T}}((\dot{\sigma}_i \oplus \mathtt{L}_i)_{i \geqslant 2})\} \prod_{i=2}^{k} \dot{q}(\dot{\sigma}_i).$$

The right-hand side does not depend on $\dot{\sigma}$, which proves the claim concerning $\hat{\boldsymbol{B}}$. □

Lemma 4.1 verifies that the one-sided dependence is preserved under the BP recursion, and from now on we always assume (41). In this setting, $\dot{\boldsymbol{B}}$ and $\hat{\boldsymbol{B}}$ reduce to mappings

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

(Generally we will fix λ, T and suppress them from the notation.) We also denote

$$\mathtt{BP} \equiv \dot{\mathtt{BP}} \circ \hat{\mathtt{BP}} \equiv \mathtt{BP}_{\lambda,T}. \tag{44}$$

Note that $\dot{q}$ is a measure on spins $\dot{\sigma} \in \dot{\Omega}$. In the introduction we discussed probability measures over messages $\dot{\mathtt{m}}$; this can be recovered by taking $\dot{q}(\{\dot{\sigma} : \dot{m}(\dot{\sigma}) = \dot{\mathtt{m}}\})$. As in Proposition 1.2 we consider $\dot{\mathscr{P}}(\dot{\Omega})$ and $\dot{\mathscr{P}}(\hat{\Omega})$ as ℓ^1 sequence spaces.

In the context of NAE-SAT, a useful observation is that the BP recursion has an averaging property, as follows. Since in the clause recursion we average over the clause literals $\underline{\mathtt{L}}$, we can make the change of variables $\tau_i = \mathtt{L}_1 \oplus \sigma_i \oplus \mathtt{L}_i$ for $i \geqslant 2$, which yields

$$\begin{aligned}\hat{\boldsymbol{B}}(\dot{\boldsymbol{q}}) &\cong \bar{\Phi}_T(\sigma)^\lambda \sum_{\tau_2,\dots,\tau_k} \tfrac{1}{2} \sum_{\mathtt{L}_1} \hat{\Phi}^{\mathrm{lit}}((\sigma,\tau_2,\dots,\tau_k) \oplus \mathtt{L}_1)^\lambda \prod_{i=2}^k \left\{ \tfrac{1}{2} \sum_{\mathtt{L}_i} \dot{\boldsymbol{q}}(\tau_i \oplus \mathtt{L}_i \oplus \mathtt{L}_1) \right\} \\ &= \bar{\Phi}_T(\sigma)^\lambda \sum_{\tau_2,\dots,\tau_k} \hat{\Phi}^{\mathrm{lit}}(\sigma,\tau_2,\dots,\tau_k)^\lambda \prod_{i=2}^k \dot{\boldsymbol{q}}^{\mathrm{avg}}(\tau_i) = \hat{\boldsymbol{B}}(\dot{\boldsymbol{q}}^{\mathrm{avg}})\end{aligned}$$

where $\dot{\boldsymbol{q}}^{\mathrm{avg}}(\sigma) \equiv \frac{1}{2}[\dot{\boldsymbol{q}}(\sigma) + \dot{\boldsymbol{q}}(\sigma \oplus \mathtt{1})]$. Therefore, under assumption (41),

$$\dot{\mathtt{BP}}\hat{\mathtt{BP}}\dot{q} = \dot{\mathtt{BP}}\hat{\mathtt{BP}}\dot{q}^{\mathrm{avg}}, \quad \text{and consequently} \quad \mathtt{BP}\dot{q} = \mathtt{BP}\dot{q}^{\mathrm{avg}}.$$

We are primarily interested in fixed points of the mapping $\mathtt{BP}$, in which case we can restrict attention to measures satisfying $\dot{q} = \dot{q}^{\mathrm{avg}}$.

The BP recursions for the pair model are completely analogous to those of the single-copy model. They can be simplified to a pair of mappings

$$\begin{aligned}\dot{\mathtt{BP}}_2 &: \dot{\mathscr{P}}(\hat{\Omega}^2) \to \dot{\mathscr{P}}(\dot{\Omega}^2), \\ \hat{\mathtt{BP}}_2 &: \dot{\mathscr{P}}(\dot{\Omega}^2) \to \dot{\mathscr{P}}(\hat{\Omega}^2);\end{aligned}$$

and once again $\mathtt{BP}_2 \equiv \dot{\mathtt{BP}}_2 \circ \hat{\mathtt{BP}}_2$ satisfies the averaging property $\mathtt{BP}_2(\dot{q}) = \mathtt{BP}_2(\dot{q}^{\mathrm{avg}})$ where

$$\dot{q}^{\mathrm{avg}}(\dot{\sigma}^1,\dot{\sigma}^2) = \tfrac{1}{2}\dot{q}(\dot{\sigma}^1,\dot{\sigma}^2) + \tfrac{1}{2}\dot{q}(\dot{\sigma}^1 \oplus \mathtt{1}, \dot{\sigma}^2 \oplus \mathtt{1}).$$

In what follows we will drop the subscript and write simply $\dot{\mathtt{BP}}, \hat{\mathtt{BP}}, \mathtt{BP}$; it will be clear from context whether we are in the single-copy or pair setting.

4.2. **Contraction estimate.** A key step in the proof is to (explicitly) define a subset $\boldsymbol{\Gamma} \subseteq \dot{\mathscr{P}}(\dot{\Omega})$ on which we have a contraction estimate of the form $\|\mathtt{BP}\dot{q} - \dot{q}_\star\|_1 \leqslant c\|\dot{q} - \dot{q}_\star\|_1$ for a constant $c < 1$, in both first- and second-moment settings. We remark that it suffices to prove such an estimate for measures $\dot{q} = \dot{q}^{\mathrm{avg}}$, since for general $\dot{q}$ it implies

$$\|\mathtt{BP}\dot{q} - \dot{q}_\star\|_1 = \|\mathtt{BP}\dot{q}^{\mathrm{avg}} - \dot{q}_\star\|_1 \leqslant c\|\dot{q}^{\mathrm{avg}} - \dot{q}_\star\|_1 \leqslant c\|\dot{q} - \dot{q}_\star\|_1.$$

Thus it will be sufficient to define $\boldsymbol{\Gamma}$ as a subset of measures satisfying $\dot{q} = \dot{q}^{\mathrm{avg}}$. Let us abbreviate $\{\mathtt{r}\} \equiv \{\mathtt{r}_0, \mathtt{r}_1\}$ and $\{\mathtt{b}\} \equiv \{\mathtt{b}_0, \mathtt{b}_1\}$. We also abbreviate $\{\mathtt{f}\} \equiv \dot{\Omega}_{\mathtt{f}}$ in the context of $\dot{q}$, and $\{\mathtt{f}\} \equiv \hat{\Omega}_{\mathtt{f}}$ in the context of $\hat{q}$. For the first moment analysis, we define $\boldsymbol{\Gamma}$ to be the set of measures $\dot{q}$ supported on $\dot{\Omega}_T$, satisfying $\dot{q} = \dot{q}^{\mathrm{avg}}$, such that

$$\begin{aligned}&\dot{q}(\mathtt{r}) + 2^k \dot{q}(\mathtt{f}) = O(1)\dot{q}(\mathtt{b}), \\ &\dot{q}(\mathtt{b})[1 - O(2^{-k})] \leqslant \dot{q}(\mathtt{r}).\end{aligned} \tag{45}$$

For the second moment analysis, we define $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}(c,\kappa)$ to be the set of $\dot{q}$ supported on $(\dot{\Omega}_T)^2$, satisfying $\dot{q} = \dot{q}^{\mathrm{avg}}$, such that

$$\begin{aligned}&\text{(A)} \quad \textstyle\sum_{\dot{\sigma} \notin \{\mathtt{bb}\}} (2^{-k})^{\mathtt{r}[\dot{\sigma}]} p(\dot{\sigma}) = O(2^{-k}) p(\mathtt{bb}), \quad |p(\mathtt{b_0 b_0}) - p(\mathtt{b_0 b_1})| \leqslant (k^9/2^{ck}) p(\mathtt{bb}), \\ &\text{(B)} \quad p(\{\mathtt{rf}, \mathtt{fr}\}) = O(2^{-\kappa k}) p(\mathtt{bb}), \quad p(\mathtt{rr}) = O(2^{(1-\kappa)k}) p(\mathtt{bb}), \\ &\text{(C)} \quad p(\mathtt{r}_x \dot{\sigma}) \geqslant [1 - O(2^{-k})] p(\mathtt{b}_x \dot{\sigma}) \text{ and} \\ &\qquad p(\dot{\sigma} \mathtt{r}_x) \geqslant [1 - O(2^{-k})] p(\dot{\sigma} \mathtt{b}_x) \text{ for all } x \in \{\mathtt{0},\mathtt{1}\} \text{ and } \dot{\sigma} \in \dot{\Omega}.\end{aligned} \tag{46}$$

Proposition 4.2. *In the first moment, let* $\mathtt{BP} \equiv \mathtt{BP}_{\lambda,T}$ *for* $\lambda \in [0,1]$ *and* $1 \leqslant T \leqslant \infty$. *There is a unique* $\dot{q}_\star \equiv \dot{q}_{\lambda,T} \in \mathbf{\Gamma}$ *satisfying* $\dot{q}_\star = \mathtt{BP}\dot{q}_\star$. *If* $\dot{q}$ *is any element of* $\mathbf{\Gamma}$, *then* $\mathtt{BP}\dot{q} \in \mathbf{\Gamma}$ *also, with* $\|\mathtt{BP}\dot{q} - \dot{q}_\star\|_1 = O(k^2/2^k)\|\dot{q} - \dot{q}_\star\|_1$.

Proposition 4.3 (second moment contraction)**.** *In the second moment, let* $\mathtt{BP} \equiv \mathtt{BP}_{\lambda,T}$ *for* $\lambda \in [0,1]$ *and* $1 \leqslant T \leqslant \infty$. *There is a unique* $\dot{q}_\star \equiv \dot{q}_{\lambda,T} \in \mathbf{\Gamma}(1,1)$ *satisfying* $\dot{q}_\star = \mathtt{BP}\dot{q}_\star$. *Further, for* $c \in (0,1]$ *and* $k \geqslant k_0(c)$, *there is no other fixed point of* BP *in* $\mathbf{\Gamma}(c,1)$*: if* $\dot{q} \in \mathbf{\Gamma}(c,1)$ *then* $\mathtt{BP}\dot{q} \in \mathbf{\Gamma}(1,1)$, *with* $\|\mathtt{BP}\dot{q} - \dot{q}_\star\|_1 = O(k^4/2^k)\|\dot{q} - \dot{q}_\star\|_1$.

We will also make use of the following lemma which says that if $\dot{q}$ is a BP fixed point, then showing (46) with $\kappa = 0$ implies the stronger bound with $\kappa = 1$:

Lemma 4.4. *In the second moment, if for some* $c \in (0,1]$ *we have* $\dot{q} \in \mathbf{\Gamma}(c,0)$ *and* $\dot{q} = \mathtt{BP}(\dot{q})$, *then in fact* $\dot{q} \in \mathbf{\Gamma}(c,1)$.

The proofs of Proposition 4.2 and 4.3 and of Lemma 4.4 are deferred to Section 9. In the next sections we apply them to compute the first and second moments of $\boldsymbol{Z}_{\lambda,T}(H)$.

5. Reduction to tree optimization

In this section we prove a key reduction for the proofs of Propositions 3.7 and 3.14, concerning the optimization of $\boldsymbol{F}$ and its second-moment analogue $\boldsymbol{F}_2$. As we have already commented, direct analysis of these functions is in general quite challenging. Instead, we first rely on other means to restrict the set of empirical measures — the set $\mathbf{N}_\circ$ in the first moment, and the set $\mathbf{N}_{\mathrm{sep}}$ in the second moment. With this restriction, we can successfully optimize $\boldsymbol{F}$ and $\boldsymbol{F}_2$ through a related, but simpler, optimization problem on trees. In this section we explain this reduction.

Definition 5.1. The tree analogues of $\mathbf{\Sigma}, \mathbf{\Sigma}_2$ (from (35) and (40)) are defined as

$$\begin{aligned}\mathbf{\Theta}(H) &\equiv \mathcal{H}(\dot{H}) + d\mathcal{H}(\hat{H}) - d\mathcal{H}(\bar{H}) + \boldsymbol{v}(H)\\ \mathbf{\Theta}_2(H) &\equiv \mathcal{H}(\dot{H}) + d\mathcal{H}(\hat{H}) - d\mathcal{H}(\bar{H}) + \boldsymbol{v}_2(H)\end{aligned}$$

(where H denotes a single-copy empirical measure in the first line, and a pair empirical measure in the second). The tree analogues of $\boldsymbol{F}, \boldsymbol{F}_2$ are defined as

$$\begin{aligned}\mathbf{\Lambda}(H) &\equiv \mathbf{\Theta}(H) + \lambda \boldsymbol{s}(H),\\ \mathbf{\Lambda}_2(H) &\equiv \mathbf{\Theta}_2(H) + \lambda \boldsymbol{s}_2(H).\end{aligned}$$

Given H, let $\dot{h}^{\mathrm{tree}}(H)$ be the measure on $\dot{\sigma}$ defined by

$$[\dot{h}^{\mathrm{tree}}(H)](\dot{\sigma}) \equiv (k-1)^{-1} \sum_{\underline{\xi} \in \Omega^k} \sum_{j=2}^{k} \hat{H}(\underline{\sigma}) \mathbf{1}\{\xi_j = \dot{\sigma}\}.$$

We then let

$$\begin{aligned}\mathbf{\Lambda}^{\mathrm{opt}}(\dot{h}) &\equiv \sup\{\mathbf{\Lambda}(H) : \dot{h}^{\mathrm{tree}}(H) = \dot{h}\}, \qquad \mathbf{\Xi}(H) \equiv \mathbf{\Lambda}^{\mathrm{opt}}(\dot{h}^{\mathrm{tree}}(H)) - \mathbf{\Lambda}(H);\\ \mathbf{\Lambda}_2^{\mathrm{opt}}(\dot{h}) &\equiv \sup\{\mathbf{\Lambda}_2(H) : \dot{h}^{\mathrm{tree}}(H) = \dot{h}\}, \quad \mathbf{\Xi}_2(H) \equiv \mathbf{\Lambda}_2^{\mathrm{opt}}(\dot{h}^{\mathrm{tree}}(H)) - \mathbf{\Lambda}_2(H).\end{aligned}$$

Note that $\mathbf{\Xi}, \mathbf{\Xi}_2$ are non-negative functions.

Definition 5.2. For $\underline{\sigma} \in \Omega^k$ and $j \in [k]$ define the rotation

$$\underline{\sigma}^{(j)} \equiv (\sigma_j, \dots, \sigma_k, \sigma_1, \dots, \sigma_{j-1}).$$

We let $\hat{H}^{\mathrm{sym}}(\underline{\sigma})$ denote the average of $\hat{H}(\underline{\sigma}^{(j)})$ over $j \in [k]$, and write $H^{\mathrm{sym}} \equiv (\dot{H}, \hat{H}^{\mathrm{sym}}, \bar{H})$.

Theorem 5.3. *For ϵ small enough, and with H^{sym} as in Definition 5.2,*

$$\boldsymbol{F}(H) \leqslant \max\{\boldsymbol{F}(H') : \|H' - H\|_1 \leqslant \epsilon(dk)^{2T}\} - \epsilon \cdot \boldsymbol{\Xi}(H^{\mathrm{sym}}),$$
$$\boldsymbol{F}_2(H) \leqslant \max\{\boldsymbol{F}_2(H') : \|H' - H\|_1 \leqslant \epsilon(dk)^{2T}\} - \epsilon \cdot \boldsymbol{\Xi}_2(H^{\mathrm{sym}}).$$

For the sake of exposition, we will give the proof of Theorem 5.3 for $\boldsymbol{F}$ only; the assertion for $\boldsymbol{F}_2$ follows from the same argument with essentially no modifications. The interpretation of $\boldsymbol{\Lambda}$ will emerge during the proof, which occupies the remainder of this section. Informally, while $\boldsymbol{F}$ refers to a graph optimization problem which need not be concave, $\boldsymbol{\Lambda}$ refers to an entropy maximization problem on colorings of a finite tree, which becomes a tractable problem. Once we have proved Theorem 5.3 it remains to analyze the functions $\boldsymbol{\Lambda}, \boldsymbol{\Lambda}_2$, which will be done in Section 6.

5.1. **Tree updates.** We prove Theorem 5.3 by analyzing one step of a certain Markov chain. To define the chain we require a certain update function for colorings on trees, which we now describe.

Definition 5.4. A *directed tree* is a bipartite tree $\boldsymbol{n}$ rooted at an edge $e_\circ$ which has a single incident vertex $x_\circ$. All edges e of $\boldsymbol{n}$ are labelled with literals $\mathtt{L}_e \in \{\mathtt{0}, \mathtt{1}\}$. We let $\mathcal{L}(\boldsymbol{n})$ denote the boundary edges of $\boldsymbol{n}$ other than $e_\circ$. We call $\boldsymbol{n}$ a *variable-to-clause tree* if $x_\circ$ is a variable; otherwise we call it a *clause-to-variable tree*. We say that $\underline{\sigma} \in \Omega^{E(\boldsymbol{n})}$ is a valid T-coloring of the tree $\boldsymbol{n}$ if the weight

$$\boldsymbol{w}^{\mathrm{lit}}_{\boldsymbol{n},T}(\underline{\sigma}) \equiv \prod_{v \in V(\boldsymbol{n})} \dot{\Phi}(\underline{\sigma}_{\delta v}) \prod_{a \in F(\boldsymbol{n})} \hat{\Phi}^{\mathrm{lit}}((\underline{\sigma} \oplus \underline{\mathtt{L}})_{\delta a}) \prod_{e \in E(\boldsymbol{n})} \bar{\Phi}_T(\sigma_e)$$

is positive.

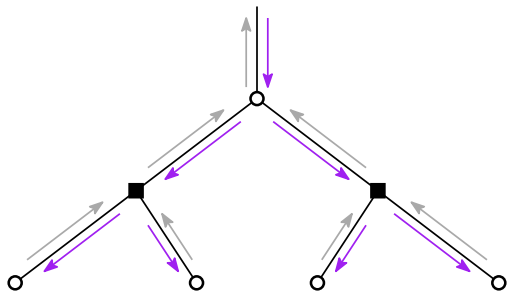


FIGURE 1. A variable-to-clause tree $\boldsymbol{n}$ (Definition 5.4).

We always visualize the tree $\boldsymbol{n}$ as in Figure 1, with the root edge 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. Now suppose $\underline{\sigma}$ is a valid T-coloring of a directed tree $\boldsymbol{n}$ with root spin $\sigma_{e_\circ} = \sigma$, and consider updating to a new root spin $\zeta \in \Omega_T$. If σ and ζ agree in the upward direction of $e_\circ$, then there is a unique valid coloring

$$\underline{\zeta} = \mathsf{update}(\underline{\sigma}, \zeta; \boldsymbol{n}) \in \Omega^{E(\boldsymbol{n})}$$

which has root spin ζ, and agrees with $\underline{\sigma}$ in all the upward colors. Indeed, the only possibility for $\sigma \neq \zeta$ is that both $\sigma, \zeta \in \Omega_{\mathtt{f}}$. It is then clear that $\mathsf{update}(\underline{\sigma}, \zeta; \boldsymbol{n})$ is uniquely defined by recursively applying the mappings $\dot{\mathtt{T}}$ and $\hat{\mathtt{T}}$, starting from the root and continuing downwards.

Since we assumed that $\underline{\sigma}$ was a valid T-coloring and $\zeta \in \Omega_T$, it is easy to verify that the resulting $\underline{\zeta}$ is also a valid T-coloring, so the update procedure respects the restriction to Ω_T.

From now on we assume all edge colors belong to Ω_T, and for the most part we drop T from the notation.

Lemma 5.5. *If $\underline{\sigma}$ is a valid coloring of the directed tree $\boldsymbol{n}$, and $\underline{\zeta} = \mathsf{update}(\underline{\sigma}, \zeta; \boldsymbol{n})$ agrees with $\underline{\sigma}$ on the boundary edges $\mathcal{L}(\boldsymbol{n})$, then*

$$\boldsymbol{w}_{\boldsymbol{n}}^{\mathrm{lit}}(\underline{\sigma}) = \boldsymbol{w}_{\boldsymbol{n}}^{\mathrm{lit}}(\underline{\zeta}).$$

Proof. For each vertex $x \in \boldsymbol{n}$, let $e(x)$ denote the parent edge of x (the unique edge of $\boldsymbol{n}$ which lies above x). We then have

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

For a variable v in $\boldsymbol{n}$ with $e(v) = e$, it follows from (29) that

$$\dot{\Phi}(\underline{\sigma}_{\delta v}) \bar{\Phi}(\sigma_e) = \dot{z}(\dot{\sigma}_e) = \dot{z}(\dot{\zeta}_e) = \dot{\Phi}(\underline{\zeta}_{\delta v}) \bar{\Phi}(\zeta_e).$$

Likewise, at a clause a in $\boldsymbol{n}$ with $e(a) = e$, it follows from (29) that

$$\hat{\Phi}^{\mathrm{lit}}((\underline{\sigma} \oplus \underline{\mathtt{L}})_{\delta a}) \bar{\Phi}(\underline{\sigma}_e) = \hat{z}(\hat{\sigma}_e) = \hat{z}(\hat{\zeta}_e) = \hat{\Phi}^{\mathrm{lit}}((\underline{\zeta} \oplus \underline{\mathtt{L}})_{\delta a}) \bar{\Phi}(\zeta_e).$$

Recalling that $\underline{\sigma}$ and $\underline{\zeta}$ agree on $\mathcal{L}(\boldsymbol{n})$, we have $\boldsymbol{w}_{\boldsymbol{n}}^{\mathrm{lit}}(\underline{\sigma}) = \boldsymbol{w}_{\boldsymbol{n}}^{\mathrm{lit}}(\underline{\zeta})$ as claimed. □

Lemma 5.6. *Let $\dot{M}, \hat{M}$ be as defined in Corollary 3.8, and let $\dot{M}_2, \hat{M}_2$ be their analogues in the pair model. For any $\sigma, \sigma' \in \Omega$ there exists an integer-valued vector $(\dot{H}, \hat{H})$ so that*

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

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

Proof. We define a graph on Ω by putting an edge between σ and σ' if there exist valid colorings $\underline{\sigma}, \underline{\sigma}'$ on some directed tree $\boldsymbol{n}$ which take values σ, σ' on the root edge $e_\circ$, but agree on the boundary edges $\mathcal{L}(\boldsymbol{n})$. If σ, σ' are connected in this way, then taking

$$\dot{H}(\underline{\zeta}) = \sum_{v \in V(\boldsymbol{n})} \mathbf{1}\{\underline{\sigma}_{\delta v} = \underline{\zeta}\} - \sum_{v \in V(\boldsymbol{n})} \mathbf{1}\{(\underline{\sigma}')_{\delta v} = \underline{\zeta}\},$$

$$\hat{H}(\underline{\xi}) = \sum_{a \in F(\boldsymbol{n})} \mathbf{1}\{\underline{\sigma}_{\delta a} = \underline{\xi}\} - \sum_{a \in F(\boldsymbol{n})} \mathbf{1}\{(\underline{\sigma}')_{\delta a} = \underline{\xi}\}.$$

gives $\dot{M}\dot{H} - \hat{M}\hat{H} = \mathbf{1}_\sigma - \mathbf{1}_{\sigma'}$ as required. It therefore suffices to show that the graph we have defined on Ω is connected (hence complete).

If $\dot{\sigma} = \dot{\sigma}'$, it is clear that σ and σ' can be connected via colorings $\underline{\sigma}, \underline{\sigma}'$ of some variable-to-clause tree $\boldsymbol{n}$, with $\underline{\sigma}' = \mathsf{update}(\underline{\sigma}, \zeta; \boldsymbol{n})$. Similarly, if $\hat{\sigma} = \hat{\sigma}'$, then σ and σ' can be connected using a clause-to-variable tree. This implies that $\Omega_{\mathtt{f}}$ is connected.

Next, it is also easy to see that if $\sigma = \mathtt{r}_x$ and $\sigma' = \mathtt{b}_x$, then they can be connected via a depth-one variable-to-clause tree. Similarly, if $\sigma = \mathtt{b}_x$ and $\sigma' = (\dot{\tau}, \square)$ for any $\dot{\tau} \in \dot{\Omega}_{\mathtt{f}}$, then they can be connected via a depth-one clause-to-variable tree. It follows that Ω is indeed connected, which proves the assertion concerning $(\dot{M}, \hat{M})$. The proof for $(\dot{M}_2, \hat{M}_2)$ is very similar. □

5.2. **Markov chain.** We now define a Markov chain on tuples $(\mathscr{G}, \underline{\sigma}, Y)$ where $\mathscr{G} = (V, F, E)$ is a (d, k)-regular NAE-SAT instance, $\underline{\sigma}$ is a valid T-coloring on $\mathscr{G}$, and $Y \subseteq V$ is a subset of variables such that

$$\begin{aligned} &\text{(i)} \quad \text{for all } v \in Y, \text{ the neighborhood } B_{2T}(v) \text{ is a tree, and} \\ &\text{(ii)} \quad \text{each pair of variables } v \neq v' \text{ in } Y \text{ lies at graph distance at least } 4T. \end{aligned} \tag{47}$$

(Recall that each variable-clause edge is defined to have length $\frac{1}{2}$.) For $v \in Y$ let $\mathscr{N}(v)$ denote the depth-one neighborhood of v, excluding the variables at unit distance from v. Let $\mathscr{N} \equiv \mathscr{N}(Y)$ denote the (disjoint) union of the graphs $\mathscr{N}(v)$, $v \in Y$:

$$\mathscr{N} = (\mathcal{N}, \underline{\mathtt{L}}_{\mathcal{N}})$$

where $\mathcal{N}$ denotes the graph without the edge literals, and $\underline{\mathtt{L}}_{\mathcal{N}}$ denotes the vector of $|Y|dk$ edge literals. Let $\underline{\sigma}_{\mathcal{N}}$ be the restriction of $\underline{\sigma}$ to the edges incident to vertices of $\mathcal{N}$, and define

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

Let V_∂ denote the vertices of $\mathcal{G} \backslash \mathcal{N}$ (including the variables at unit distance from Y), and let F_∂ denote the clauses of $\mathcal{G} \backslash \mathcal{N}$. Let E_∂ denote the set of all edges incident to $V_\partial \cup F_\partial$, and let $\underline{\sigma}_\partial$ denote the restriction of $\underline{\sigma}$ to E_∂. Define

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

Then the overall weight $\boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G}}(\underline{\sigma})$ of $\underline{\sigma}$ factorizes as

$$\boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G}}(\underline{\sigma}) = \boldsymbol{w}^{\mathrm{lit}}_{\partial}(\underline{\sigma}_{\partial}) \boldsymbol{w}^{\mathrm{lit}}_{\mathcal{N}}(\underline{\sigma}_{\mathcal{N}} | \underline{\mathtt{L}}_{\mathcal{N}}). \tag{48}$$

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

1. Sample a new pair $(\underline{\mathtt{L}}'_{\mathcal{N}}, \underline{\zeta}_{\mathcal{N}})$ from the probability measure

$$p((\underline{\mathtt{L}}'_{\mathcal{N}}, \underline{\zeta}_{\mathcal{N}}) \,|\, (\underline{\mathtt{L}}_{\mathcal{N}}, \underline{\sigma}_{\mathcal{N}})) = \frac{1}{z} \mathbf{1}\{\dot{h}^{\mathrm{tree}}(\underline{\sigma}_{\delta\mathcal{N}}) = \dot{h}^{\mathrm{tree}}(\underline{\zeta}_{\delta\mathcal{N}})\} \boldsymbol{w}^{\mathrm{lit}}_{\mathcal{N}}(\underline{\zeta}_{\mathcal{N}} | \underline{\mathtt{L}}'_{\mathcal{N}})^{\lambda}$$

   where z denotes the normalizing constant, which depends on $|\mathcal{N}|$ and $\dot{h}^{\mathrm{tree}}(\underline{\sigma}_{\delta\mathcal{N}})$.
2. If $e = (\dot{e}, \hat{e})$ then denote

$$\dot{\sigma}_e \equiv \dot{\sigma}(\dot{e}) \equiv \dot{\sigma}(\hat{e}).$$

   Each edge $e \in E$ pairs some $\dot{e}_i$ with some $\hat{e}_{\mathfrak{m}(i)}$, for some permutation $\mathfrak{m} : [nd] \to [nd]$. Let B denote the subset of indices $i \in [nd]$ such that $(\dot{e}_i, \hat{e}_{\mathfrak{m}(i)}) \in \delta\mathcal{N}$. Now consider the set $\mathcal{M} = \mathcal{M}(\mathscr{G}, Y, \underline{\sigma}, \underline{\zeta})$ of permutations $\mathfrak{m}' : [nd] \to [nd]$ such that

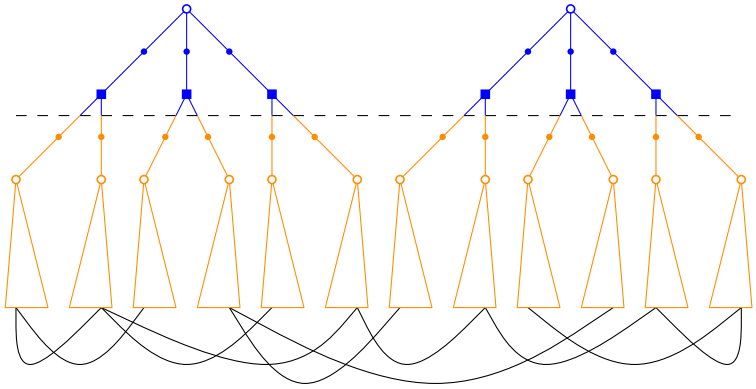
$$\mathfrak{m}'(i) = \mathfrak{m}(i) \text{ for all } i \in [nd] \backslash B, \quad \dot{\sigma}(\dot{e}_i) = \dot{\zeta}(\hat{e}_{\mathfrak{m}'(i)}) \text{ for all } i \in B. \tag{49}$$

   Sample $\mathfrak{M}$ uniformly at random from $\mathcal{M}$. Let $\mathscr{G}'$ be the new graph formed from $\mathscr{G}$ by replacing $\underline{\mathtt{L}}_{\mathcal{N}}$ with $\underline{\mathtt{L}}'_{\mathcal{N}}$, and replacing $\mathfrak{m}$ with $\mathfrak{M}$.
3. For each $e \in \delta\mathcal{N}$, let $\boldsymbol{n}(e)$ denote the depth-$2T$ neighborhood of $v(e)$ in the graph $\mathscr{G} \backslash \{a(e)\}$, including the edge e which we regard as the root of $\boldsymbol{n}(e)$. Let

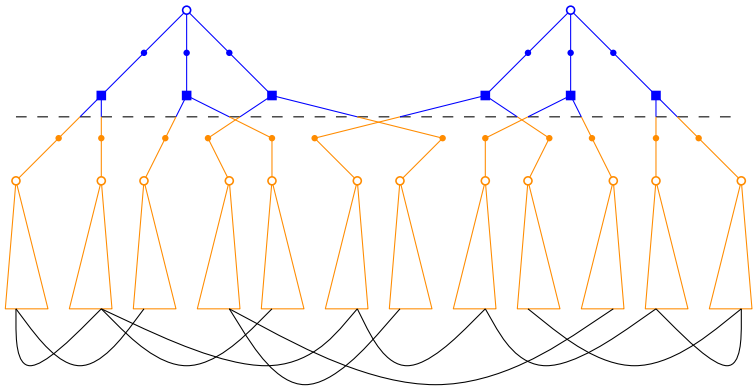
$$\underline{\zeta}_{\boldsymbol{n}(e)} \equiv \mathsf{update}(\underline{\sigma}_{\boldsymbol{n}(e)}, \zeta_e; \boldsymbol{n}(e));$$

   note that, since $\underline{\sigma}$ is a valid T-coloring, $\underline{\zeta}_{\boldsymbol{n}(e)}$ and $\underline{\sigma}_{\boldsymbol{n}(e)}$ must agree at the boundary of $\boldsymbol{n}(e)$. For any edge e' which does not appear in $\mathcal{N}$ or any of the trees $\boldsymbol{n}(e)$, define $\zeta_{e'} = \sigma_{e'}$.

The state of the Markov chain after one step is $(\mathscr{G}', \underline{\zeta}, Y)$. See Figure 2.



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



(B) $(\mathscr{G}', Y, \underline{\sigma}')$

FIGURE 2. $(\mathscr{G}, Y, \underline{\sigma})$ to $(\mathscr{G}', Y, \underline{\sigma}')$.

Lemma 5.7. *Suppose we have a measure $\mathbb{P}(Y|\mathscr{G})$ such that, whenever the tuples $(\mathscr{G}, Y, \underline{\sigma})$ and $(\mathscr{G}', Y, \underline{\zeta})$ belong to the same orbit of the Markov chain, it holds that*

$$\mathbb{P}(Y|\mathscr{G}) = \mathbb{P}(Y|\mathscr{G}'). \tag{50}$$

A reversing measure for the Markov chain is then given by

$$\mu(\mathscr{G}, \underline{\sigma}, Y) = \mathbb{P}(\mathscr{G})\mathbb{P}(Y|\mathscr{G})\boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G}}(\underline{\sigma})^{\lambda}$$

Proof. Started from $\mathrm{A} = (\mathscr{G}, \underline{\sigma}, Y)$, let $\pi(\mathrm{A}, \mathrm{B})$ denote the chance to reach state $\mathrm{B} = (\mathscr{G}', \underline{\zeta}, Y)$ in one step of the Markov chain:

$$\pi(\mathrm{A}, \mathrm{B}) = \frac{p((\underline{\mathrm{L}}'_{\mathcal{N}}, \underline{\zeta}_{\mathcal{N}}) \,|\, (\underline{\mathrm{L}}_{\mathcal{N}}, \underline{\sigma}_{\mathcal{N}}))}{|}\mathcal{M}(\mathrm{A}, \mathrm{B})|$$

for $\mathcal{M}$ as defined in (49). The size of $\mathcal{M}$ can be expressed as a function of $(|Y|, \dot{h})$ only, so $|\mathcal{M}(\mathrm{A}, \mathrm{B})| = |\mathcal{M}(\mathrm{B}, \mathrm{A})|$. It follows that

$$\begin{aligned}\frac{\mu(\mathrm{A})\pi(\mathrm{A}, \mathrm{B})}{\mu(\mathrm{B})\pi(\mathrm{B}, \mathrm{A})} &= \frac{\boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G}}(\underline{\sigma})^{\lambda}}{\boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G}}(\underline{\zeta})^{\lambda}} \frac{p((\underline{\mathrm{L}}'_{\mathcal{N}}, \underline{\zeta}_{\mathcal{N}}) \,|\, (\underline{\mathrm{L}}_{\mathcal{N}}, \underline{\sigma}_{\mathcal{N}}))}{p((\underline{\mathrm{L}}_{\mathcal{N}}, \underline{\sigma}_{\mathcal{N}}) \,|\, (\underline{\mathrm{L}}'_{\mathcal{N}}, \underline{\zeta}_{\mathcal{N}}))} \\ &= \frac{\boldsymbol{w}^{\mathrm{lit}}_{\partial}(\underline{\sigma}_{\partial})^{\lambda}\boldsymbol{w}^{\mathrm{lit}}_{\mathcal{N}}(\underline{\sigma}_{\mathcal{N}}|\underline{\mathrm{L}}_{\mathcal{N}})^{\lambda}\boldsymbol{w}^{\mathrm{lit}}_{\mathcal{N}}(\underline{\zeta}_{\mathcal{N}}|\underline{\mathrm{L}}'_{\mathcal{N}})^{\lambda}}{\boldsymbol{w}^{\mathrm{lit}}_{\partial}(\underline{\zeta}_{\partial})^{\lambda}\boldsymbol{w}^{\mathrm{lit}}_{\mathcal{N}}(\underline{\zeta}_{\mathcal{N}}|\underline{\mathrm{L}}'_{\mathcal{N}})^{\lambda}\boldsymbol{w}^{\mathrm{lit}}_{\mathcal{N}}(\underline{\sigma}_{\mathcal{N}}|\underline{\mathrm{L}}_{\mathcal{N}})^{\lambda}} = \frac{\boldsymbol{w}^{\mathrm{lit}}_{\partial}(\underline{\sigma}_{\partial})^{\lambda}}{\boldsymbol{w}^{\mathrm{lit}}_{\partial}(\underline{\zeta}_{\partial})^{\lambda}},\end{aligned}$$

using (48). It follows from Lemma 5.5 that this ratio equals one, which proves reversibility. (We remark that since the Markov chain breaks up into many disjoint orbits, the reversing measure is not unique.) □

Let A be any subset of the state space, and let B denote the set of states reachable from A in one step of the chain. Then reversibility implies

$$\mu(A) = \sum_{\mathrm{A}\in A}\sum_{\mathrm{B}\in B}\mu(\mathrm{A})\pi(\mathrm{A},\mathrm{B}) = \sum_{\mathrm{A}\in A}\sum_{\mathrm{B}\in B}\mu(\mathrm{B})\pi(\mathrm{B},\mathrm{A}) \leqslant \mu(B)\max_{\mathrm{B}\in B}\pi(\mathrm{B},A). \tag{51}$$

5.3. **From graph to tree optimizations.** Given $\mathscr{G} = (V,F,E)$, a valid coloring $\underline{\sigma}$ on $\mathscr{G}$, and a nonempty subset of variables $Y \subseteq V$, we define

$$H^{\mathrm{samp}} \equiv H^{\mathrm{samp}}(\mathscr{G},\underline{\sigma},Y) \equiv H^{\mathrm{samp}}(\mathcal{G},\underline{\sigma},Y) \equiv (\dot{H}^{\mathrm{samp}},\hat{H}^{\mathrm{samp}},\bar{H}^{\mathrm{samp}})$$

which records the empirical distribution of $\underline{\sigma}$ near Y, as follows. For $v \in Y$ and $e \in \delta v$, let $1 \leqslant j(e) \leqslant k$ denote the index of e in $\delta a(e)$. Let

$$\begin{aligned}
\dot{H}^{\mathrm{samp}}(\zeta) &= |\{v \in Y : \underline{\sigma}_{\delta v} = \zeta\}|/|Y|, && \zeta \in \Omega^d,\\
\hat{H}^{\mathrm{samp}}(\xi) &= |\{(v,e) : v \in Y, e \in \delta v, (\underline{\sigma}_{\delta a(e)})^{(j(e))} = \xi\}|/(d|Y|), && \xi \in \Omega^k,\\
\bar{H}^{\mathrm{samp}}(\zeta) &= |\{(v,e) : v \in Y, e \in \delta v, \sigma_e = \sigma\}|/(d|Y|), && \sigma \in \Omega,
\end{aligned} \tag{52}$$

where $(\underline{\sigma}_{\delta a})^{(j)}$ is the rotation of $\underline{\sigma}_{\delta a}$ in which the j-th entry appears first (Definition 5.2). We then define $\dot{h} = \dot{h}^{\mathrm{tree}}(H^{\mathrm{samp}})$ as the empirical measure of $\dot{\sigma}$ on the edges $\delta a(e)\backslash e$, for $e \in \delta v$:

$$\dot{h}^{\mathrm{tree}}(\dot{\sigma}) = (k-1)^{-1}\sum_{\zeta\in\Omega^k}\hat{H}(\zeta)\sum_{i=2}^{k}\mathbf{1}\{\dot{\zeta}_i = \dot{\sigma}\}, \quad \dot{\sigma} \in \dot{\Omega}.$$

It is clear that $H^{\mathrm{samp}}(\mathscr{G},\underline{\sigma},Y)$ can be expressed as a function of $\underline{\sigma}_{\mathcal{N}}$, and from now on we indicate this relation by

$$H^{\mathrm{samp}}(\mathscr{G},\underline{\sigma},Y) = H(\underline{\sigma}_{\mathcal{N}}).$$

Let $\mathbb{E}Z_{\mathcal{N}}(H^{\mathrm{samp}})$ denote the total weight of pairs $(\underline{\mathtt{L}}_{\mathcal{N}},\underline{\sigma}_{\mathcal{N}})$ which are consistent with H^{samp}, normalized by the number of literal assignments:

$$\mathbb{E}Z_{\mathcal{N}}(H^{\mathrm{samp}}) \equiv \frac{1}{2^{|\mathcal{N}|dk}}\sum_{\underline{\sigma}_{\mathcal{N}}}\mathbf{1}\{H(\underline{\sigma}_{\mathcal{N}}) = H^{\mathrm{samp}}\}\sum_{\underline{\mathtt{L}}_{\mathcal{N}}}\boldsymbol{w}^{\mathrm{lit}}_{\mathcal{N}}(\underline{\sigma}_{\mathcal{N}}|\underline{\mathtt{L}}_{\mathcal{N}})^{\lambda}.$$

Clearly, this depends on $\mathcal{N}$ only through $s = |\mathcal{N}|$, so we denote $\mathbb{E}Z_s(H^{\mathrm{samp}}) \equiv \mathbb{E}Z_{\mathcal{N}}(H^{\mathrm{samp}})$. The following lemma gives an explicit calculation of $\mathbb{E}Z_s(H^{\mathrm{samp}})$.

Lemma 5.8. *With $\hat{\Phi}$ as in Remark 3.1,*

$$\mathbb{E}Z_s(H^{\mathrm{samp}}) = \frac{\binom{s}{s\dot{H}^{\mathrm{samp}}}\binom{ds}{ds\hat{H}^{\mathrm{samp}}}}{\binom{ds}{ds\bar{H}^{\mathrm{samp}}}}\dot{\Phi}^{\lambda s\dot{H}^{\mathrm{samp}}}\hat{\Phi}^{\lambda ds\hat{H}^{\mathrm{samp}}}\bar{\Phi}^{\lambda ds\bar{H}^{\mathrm{samp}}}.$$

This equals $s^{O(1)}\exp\{s\boldsymbol{\Lambda}(H^{\mathrm{samp}})\}$ where $\boldsymbol{\Lambda}$ is given by Definition 5.1, and is concave in H.

Proof. The first assertion follows by a straightforward combinatorial calculation (cf. (34)). Stirling's formula yields the asymptotic expansion $\mathbb{E}Z_s(H^{\mathrm{samp}}) = s^{O(1)}\exp\{s\boldsymbol{\Lambda}(H^{\mathrm{samp}})\}$. The function $\boldsymbol{\Lambda}(H)$ is the sum of $\boldsymbol{\Theta}(H)$ and the linear function $\boldsymbol{s}(H)\lambda$, and we claim that $\boldsymbol{\Theta}$ is concave. To see this, recall that $H = H^{\mathrm{samp}}$ must satisfy

$$\bar{H}(\sigma) = \sum_{\zeta\in\Omega^k}\hat{H}(\zeta)\mathbf{1}\{\zeta_1 = \sigma\}.$$

Let $\hat{H}_\sigma(\zeta)$ denote the probability of ζ under $\hat{H}$, conditioned on $\zeta_1 = \sigma$. Then

$$\boldsymbol{\Theta}(H) = \mathcal{H}(\dot{H}) + d\sum_\sigma \bar{H}(\sigma)\mathcal{H}(\hat{H}_\sigma) + \boldsymbol{v}(H).$$

The entropy function $\mathcal{H}$ is concave, so this proves that $\boldsymbol{\Theta}$ is indeed concave. □

Remark 5.9. An equivalent characterization of $\boldsymbol{\Lambda}$ is as follows. Recall that $\mathcal{N}$ consists of s disjoint trees $\mathcal{N}(v_1), \dots, \mathcal{N}(v_s)$ where each $\mathcal{N}(v_s)$ is a copy of the depth-one tree $\mathcal{D}$ depicted in Figure 3. Both $\mathcal{N}$ and $\mathcal{D}$ do not include edge literals. The natural weight function on colorings of $\mathcal{D}$ is defined by

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

where $\hat{\Phi}$ is as in Remark 3.1. If ν is a probability measure over colorings $\underline{\sigma}_{\mathcal{D}}$, then we denote $H(\nu) = (\dot{H}, \hat{H}, \bar{H})$ where (cf. (52))

$$\dot{H}(\zeta) = \nu(\underline{\sigma}_{\delta v} = \zeta),\ \hat{H}(\xi) = d^{-1}\sum_{e \in \delta v} \nu((\underline{\sigma}_{\delta a(e)})^{j(e)} = \xi),\ \bar{H}(\sigma) = d^{-1}\sum_{e \in \delta v} \nu(\sigma_e = \sigma).$$

Let $Z_s(\nu)$ be the contribution to $Z_s(H^{\mathrm{samp}})$ from colorings $\underline{\sigma}_{\mathcal{N}}$ with empirical measure ν — that is, colorings $\underline{\sigma}_{\mathcal{N}}$ satisfying $s\nu(\underline{\sigma}_{\mathcal{D}}) = |\{i \in [s] : \underline{\sigma}_{\mathcal{N}(v_i)} = \underline{\sigma}_{\mathcal{D}}\}|$ for all $\underline{\sigma}_{\mathcal{D}}$. Using multi-index notation as before, we have

$$\mathbb{E}Z_s(\nu) = \binom{s}{s\nu}(\boldsymbol{w}_{\mathcal{D}})^{\lambda\nu} = s^{O(1)}\exp\{\mathcal{H}(\nu) + \lambda\langle \ln \boldsymbol{w}_{\mathcal{D}}, \nu\rangle\}.$$

Summing over all ν such that $s\nu$ is integer-valued and $H(\nu) = H^{\mathrm{samp}}$ gives

$$\mathbb{E}Z_s(H^{\mathrm{samp}}) = \sum_\nu \mathbb{E}Z_s(\nu) = s^{O(1)}\exp\{n\boldsymbol{\Lambda}(H^{\mathrm{samp}})\}$$

for the following alternative equivalent of $\boldsymbol{\Lambda}$:

$$\boldsymbol{\Lambda}(H^{\mathrm{samp}}) = \sup\{\mathcal{H}(\nu) + \lambda\langle \ln \boldsymbol{w}_{\mathcal{D}}, \nu\rangle : H(\nu) = H^{\mathrm{samp}}\}.$$

This representation also explains clearly why $\boldsymbol{\Lambda}$ is concave. Lastly, note we can express $\boldsymbol{\Lambda}^{\mathrm{opt}}$ similarly as $\boldsymbol{\Lambda}^{\mathrm{opt}}(\dot{h}) = \sup\{\mathcal{H}(\nu) + \lambda\langle \ln \boldsymbol{w}_{\mathcal{D}}, \nu\rangle : \dot{h}^{\mathrm{tree}}(H(\nu)) = H^{\mathrm{samp}}\}$.

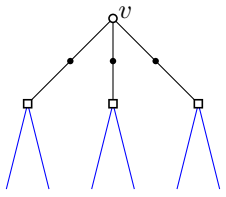


FIGURE 3. The depth-one tree $\mathcal{D}$, rooted at variable v. We use $\mathcal{L}(\mathcal{D})$ to denote the set of boundary edges $e \in \delta a \backslash (av)$, $a \in \partial v$, so $|\mathcal{L}(\mathcal{D})| = d(k-1)$.

Proposition 5.10. *Let $A(H)$ be the set of tuples $(\mathscr{G}, \underline{\sigma}, Y)$ such that $\underline{\sigma}$ is a valid T-coloring on $\mathscr{G}$ with empirical measure H, and Y is a subset of V satisfying (47) as well as*

$$n\epsilon \leqslant |Y| \leqslant 6n\epsilon \text{ and } \|H^{\mathrm{samp}}(\mathscr{G}, \underline{\sigma}, Y) - H^{\mathrm{sym}}\| \leqslant (\ln \ln n)^{-1/2}.$$

Suppose we define an exceptional set of graphs $\mathscr{B}$ with $\mathbb{P}(\mathscr{B}) \leqslant \exp\{-n(\ln n)^{1/2}\}$, and a law $\mathbb{P}(Y|\mathscr{G})$ such that for all $\mathscr{G} \notin \mathscr{B}$ and all $\underline{\sigma}$ with $H(\mathscr{G}, \underline{\sigma}) = H$, we have

$$\mathbb{P}(A(H) \,|\, (\mathscr{G}, \underline{\sigma})) = \sum_{Y} \mathbb{P}(Y|\mathscr{G}) \mathbf{1}\{(\mathscr{G}, \underline{\sigma}, Y) \in A(H)\} \geqslant \tfrac{1}{2}. \tag{53}$$

Then the expected weight of colorings with empirical measure H satisfies

$$\mathbb{E}\boldsymbol{Z}(H) \leqslant \frac{e^{o_n(1)} \max\{\mathbb{E}\boldsymbol{Z}(H') : \|H' - H\|_1 \leqslant \epsilon(dk)^{2T}\}}{\exp\{n\epsilon \min\{\boldsymbol{\Xi}(H'') : \|H'' - H^{\mathrm{sym}}\|_1 \leqslant (\ln\ln n)^{-1/2}\}\}}.$$

Proof. Since $\boldsymbol{Z}(H) \leqslant 2^n$,

$$\mathbb{E}\boldsymbol{Z}(H) \leqslant \mathbb{E}[\boldsymbol{Z}(H); \mathscr{G} \notin \mathscr{B}] + \exp\{-\Omega(n(\ln n)^{1/2})\}.$$

Since we only consider measures H for which $\boldsymbol{F}(H) > -\infty$, the right-hand side above is dominated by the contribution from $\mathscr{G} \notin \mathscr{B}$. Next recall from Lemma 5.7 the reversing measure $\mu(\mathscr{G}, \underline{\sigma}, Y)$. Applying assumption (53),

$$\mathbb{E}\boldsymbol{Z}(H) \leqslant 2\mathbb{E}[\boldsymbol{Z}(H); \mathscr{G} \notin \mathscr{B}] = 2 \sum_{\mathscr{G} \notin \mathscr{B}} \mathbb{P}(\mathscr{G}) \sum_{\underline{\sigma}} \boldsymbol{w}^{\mathrm{lit}}_{\mathscr{G}}(\underline{\sigma})^{\lambda} \leqslant 4\mu(A(H)).$$

We now apply (51), writing $B(H)$ for the set of states $\mathrm{B} = (\mathscr{G}', \underline{\sigma}', Y')$ reachable from $A(H)$ in one step of the Markov chain. First note that if $\mathrm{B} \in B(H)$ then $H' = H(\mathscr{G}', \underline{\sigma}', Y')$ must satisfy (crudely) $\|H' - H\| \leqslant \epsilon(dk)^{2T}$, so summing over the $\epsilon(dk)^{2T}$-neighborhood of H,

$$\mu(B(H)) \leqslant s^{O(1)} \max\{\mathbb{E}\boldsymbol{Z}(H') : \|H' - H\| \leqslant \epsilon(dk)^{2T}\}.$$

Next, writing $s = |Y'|$, we have

$$\pi(\mathrm{B}, A(H)) = \frac{Z_s(H^{\mathrm{samp}})}{\sum_{H''} Z_s(H'') \mathbf{1}\{\dot{h}^{\mathrm{tree}}(H'') = \dot{h}^{\mathrm{tree}}(H^{\mathrm{samp}})\}},$$

where H'' represents $H^{\mathrm{samp}}(\mathscr{G}', \underline{\sigma}', Y')$. Applying Lemma 5.8 gives

$$\pi(\mathrm{B}, A(H)) \leqslant s^{O(1)} \exp\{s[\boldsymbol{\Lambda}(H^{\mathrm{samp}}) - \boldsymbol{\Lambda}^{\mathrm{opt}}(\dot{h}^{\mathrm{tree}}(H^{\mathrm{samp}}))]\}.$$

Recalling $\|H^{\mathrm{samp}} - H^{\mathrm{sym}}\| \leqslant (\ln\ln n)^{-1/2}$ and $n\epsilon \leqslant s \leqslant 6n\epsilon$, the result follows. □

5.4. **Sampling.** We now define the law $\mathbb{P}(Y|\mathscr{G})$ and verify condition (53). To this end, given $\mathscr{G} = (V, F, E)$, 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 μ, then for all $t \geqslant 1$ we have

$$\mathbb{P}(X \geqslant t\mu) \leqslant \exp\{-t\mu \ln(t/e)\}. \tag{54}$$

Lemma 5.11. *If $\mathscr{G} = (V, F, E)$ is sampled from the (d, k)-regular configuration model, then for any fixed t it holds for $n \geqslant n_\circ(t)$ that*

$$\mathbb{P}(|V \backslash V_t| \geqslant n(\ln\ln n)^{-1}) \leqslant \exp\{-n(\ln n)^{1/2}\}.$$

Proof. Let γ count the total number of cycles in $\mathscr{G}$ 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

$$|V \backslash V_t| \leqslant 2t(dk)^t \gamma. \tag{55}$$

Consider breadth-first search exploration in $\mathscr{G}$ 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 $\leqslant 2t$ is upper bounded by

$$\frac{(dk)^{2t}}{nd-t}.$$

It follows that the total number of cycles revealed up to time $nd(1-\delta)$ is stochastically dominated by a binomial random variable

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

The final $nd\delta$ exploration steps can form at most $nd\delta$ new cycles, so $\gamma \leqslant \gamma' + nd\delta$. Applying (54) with $\delta = (\ln\ln n)^{-2}$,

$$\mathbb{P}(\gamma \geqslant 2nd\delta) \leqslant \mathbb{P}(\gamma' \geqslant nd\delta) \leqslant \exp\Big\{ - nd\delta \ln \frac{nd\delta^2}{e(dk)^{2t}}\Big\} \leqslant \exp\{-n(\ln n)^{1/2}\}$$

for large enough n. Recalling (55) gives the claimed bound. □

Recalling Proposition 5.10, let $\mathscr{B}$ be the set of graphs $\mathscr{G}$ for which $|V\backslash V_t| \geqslant n/2$. For $\mathscr{G} \notin \mathscr{B}$, take i.i.d. random variables $I_v \sim \mathrm{Ber}(\epsilon')$ indexed by $v \in V_t$ for some ϵ' to be determined, and let

$$Y_v \equiv \mathbf{1}\{I_v = 1, \text{ and } I_u = 0 \text{ for all } u \in B_t(v)\backslash\{v\}\}, \quad \epsilon \equiv \tfrac{1}{2}\mathbb{E}Y_v. \tag{56}$$

We define $\mathbb{P}(Y|\mathscr{G})$ to be the law of the set $Y = \{v \in V_t : Y_v = 1\}$, with $t = 4T$. Given a valid coloring $\underline{\sigma}$ on $\mathscr{G} = (V, F, E)$, define (cf. (52))

$$\begin{aligned}
\dot{X}_v(\zeta) &\equiv \mathbf{1}\{\underline{\sigma}_{\delta v} = \zeta\}, && \zeta \in \Omega^d,\\
\hat{X}_v(\xi) &\equiv |\{e \in \delta v : (\underline{\sigma}_{\delta a(e)})^{j(e)} = \xi\}|, && \xi \in \Omega^k,\\
\bar{X}_v(\sigma) &\equiv |\{(a,e) : a \in \partial v, e \in \delta a\backslash(av), \sigma_e = \sigma\}|, && \sigma \in \Omega.
\end{aligned}$$

Lemma 5.12. *Fix $(\mathscr{G}, \underline{\sigma})$ and let $n' = |V_t| < n/2$. Then for all $x > 4|n - n'|$ we have the concentration bounds*

$$\mathbb{P}\Big(\Big|\sum_{v\in V_t} Y_v - n'\epsilon\Big| \geqslant x\Big) \leqslant \exp\Big\{-\frac{x^2}{8n'(dk)^{2t}}\Big\}$$

$$\mathbb{P}\Big(\Big|\sum_{v\in V_t} Y_v\dot{X}_v(\zeta) - n'\epsilon\dot{H}(\zeta)\Big| \geqslant x\Big) \leqslant \exp\Big\{-\frac{x^2}{8n'(dk)^{2t}}\Big\}$$

$$\mathbb{P}\Big(\Big|\frac{1}{d}\sum_{v\in V_t} Y_v\hat{X}_v(\xi) - n'\epsilon\hat{H}^{\mathrm{sym}}(\xi)\Big| \geqslant x\Big) \leqslant \exp\Big\{-\frac{x^2}{8n'(dk)^{2t+1}}\Big\}$$

$$\mathbb{P}\Big(\Big|\frac{1}{d(k-1)}\sum_{v\in V_t} Y_v\bar{X}_v(\dot{\sigma}) - n'\epsilon\bar{H}(\sigma)\Big| \geqslant x\Big) \leqslant \exp\Big\{-\frac{x^2}{8n'(dk)^{2t+1}}\Big\}$$

Proof. Assume without loss that $V_t = [n'] \equiv \{1, \dots, n'\}$, and for $0 \leqslant s \leqslant n'$ let $\mathscr{F}_s$ denote the sigma-field generated by $Y_1, \dots, Y_s$. Let

$$S \equiv \sum_{v\leqslant n'} A_v Y_v, \quad M_s \equiv \mathbb{E}[S|\mathscr{F}_s]$$

where we take different values of A_v for the various bounds:

$$A_v = 1, \ A_v = \dot{X}_v(\zeta), \ A_v = \hat{X}_v(\xi), \ A_v = \bar{X}_v(\sigma).$$

We emphasize that $\mathscr{G}$ and $\underline{\sigma}$ are fixed, so the only randomness is in the Y's:

$$M_s = \sum_{v \leqslant n'} A_v \mathbb{E}[Y_v | \mathscr{F}_s].$$

If v lies at distance greater than $2t$ from any variable in $[s] \equiv \{1, \dots, s\}$, then

$$\mathbb{E}[Y_v | \mathscr{F}_s] = \mathbb{E}[Y_v] = 2\epsilon.$$

More generally, $\mathbb{E}[Y_v|\mathscr{F}_s]$ is a measurable function of all the Y_w values for $w \in [s] \cap B_{2t}(v)$. Therefore the only possibility for $\mathbb{E}[Y_v|\mathscr{F}_{s+1}] \neq \mathbb{E}[Y_v|\mathscr{F}_s]$ is that $[s+1] \cap B_{2t}(v)$ differs from $[s] \cap B_{2t}(v)$, which implies in particular that $v \in B_{2t}(s+1)$. The number of such v is at most $(dk)^t$, so we conclude

$$|M_{s+1} - M_s| \leqslant (dk)^t \|A\|_\infty.$$

It follows by the Azuma–Hoeffding martingale inequality that

$$\mathbb{P}(|S - \mathbb{E}S| \geqslant x) \leqslant \exp\Big\{ - \frac{x^2}{2n'(dk)^{2t}\|A\|_\infty} \Big\},$$

and the claimed bounds follow from the fact that removing $n - n' < n/2$ vertices from a graph can change the empirical measure by at most $2(n-n')/n$. □

Proof of Theorem 5.3. Take $\epsilon' > 0$ small enough such that the resulting ϵ defined by (56) satisfies $6\epsilon(dk)^{2T} < 1$. It then follows from Lemmas 5.11 and 5.12 that the conditions of Proposition 5.10 are satisfied by taking $\mathscr{B}$ to be the set of graphs with $|V \backslash V_t| \geqslant n(\ln \ln n)^{-1}$, and $\mathbb{P}(Y|\mathscr{G})$ to be the law of $Y = \{v \in V_t : Y_v = 1\}$, for Y_v as given by (56). □

6. Tree optimization problem

In this section we give the analysis of $\Xi(H)$ (Definition 5.1 and Theorem 5.3). Recall from (37) the definition of $\mathbf{N}_\circ$, and from (38) the definition of $H_\star$.

Proposition 6.1. *For Ξ, Ξ_2 as defined by (35) and (40), we have*
(a) On $\{H \in \mathbf{N}_\circ : H = H^{\mathrm{sym}}\}$, Ξ is uniquely minimized at $H = H_\star$, with $\Xi(H_\star) = 0$.
(b) On $\{H \in \mathbf{N}_{\mathrm{sep}} : H = H^{\mathrm{sym}}\}$, Ξ_2 is uniquely minimized at $H = H_\otimes$, with $\Xi_2(H_\otimes) = 0$.

Proposition 6.2. *There exists a positive constant $\epsilon = \epsilon(k)$ such that*

$$\Xi(H) \geqslant \epsilon \|H - H_\star\|^2 \quad \text{for all } \|H - H_\star\| \leqslant \epsilon,$$
$$\Xi_2(H) \geqslant \epsilon \|H - H_\otimes\|^2 \quad \text{for all } \|H - H_\otimes\| \leqslant \epsilon.$$

6.1. Uniqueness of minimizer.

We now outline the proof of Proposition 6.1. Let ν be any probability measure over colorings of the depth-one $\mathcal{D}$ (Figure 3). Recall from Remark 5.9 that

$$\mathbf{\Lambda}(H) = \sup\{\mathcal{H}(\nu) + \lambda \langle \ln \boldsymbol{w}_{\mathcal{D}}, \nu \rangle : H(\nu) = H\},$$
$$\mathbf{\Lambda}^{\mathrm{opt}}(\dot{h}) = \sup\{\mathcal{H}(\nu) + \lambda \langle \ln \boldsymbol{w}_{\mathcal{D}}, \nu \rangle : \dot{h}^{\mathrm{tree}}(H(\nu)) = \dot{h}\}.$$

The mappings $\nu \mapsto H(\nu)$ and $\nu \mapsto \dot{h}^{\mathrm{tree}}(H(\nu))$ are linear, so we are in the setting of Section A. The discussion in that section (see in particular Remark A.7) implies the following: there is a unique measure $\nu = \nu^{\mathrm{opt}}(\dot{h})$ achieving the maximum in $\mathbf{\Lambda}^{\mathrm{opt}}(\dot{h})$, and there exists a probability measure $\dot{q}$ on $\dot{\Omega}$ such that

$$\nu(\underline{\sigma}) = [\nu^{\mathrm{bd}}(\dot{q})](\underline{\sigma}) \equiv \frac{1}{Z} \Big\{ \dot{\Phi}(\underline{\sigma}_{\delta v}) \prod_{a \in \partial v} [\bar{\Phi}(\sigma_{av}) \hat{\Phi}(\underline{\sigma}_{\delta a})] \Big\}^\lambda \prod_{e \in \mathcal{L}(\mathcal{D})} \dot{q}(\dot{\sigma}_e), \tag{57}$$

with Z the normalizing constant. Likewise, in the second moment there is a unique measure $\nu = \nu_2^{\mathrm{opt}}(\dot{h})$ achieving the maximum in $\mathbf{\Lambda}_2^{\mathrm{opt}}(\dot{h})$, and there exists a probability measure $\dot{q}$ on $\dot{\Omega}^2$ such that

$$\nu(\underline{\sigma}) = [\nu_2^{\mathrm{bd}}(\dot{q})](\underline{\sigma}) \equiv \frac{1}{Z}\left\{\dot{\Phi}_2(\underline{\sigma}_{\delta v})\prod_{a\in\partial v}[\bar{\Phi}_2(\sigma_{av})\hat{\Phi}_2(\underline{\sigma}_{\delta a})]\right\}^{\lambda}\prod_{e\in\mathcal{L}(\mathcal{D})}\dot{q}(\dot{\sigma}_e). \tag{58}$$

In each case, although $\nu^{\mathrm{opt}}(\dot{h})$ is uniquely determined by $\dot{h}$, $\dot{q}$ need not be if the constraints are rank-deficient. Nevertheless we shall proceed simply from the existence of some $\dot{q}$.

Lemma 6.3. *$\mathbf{\Xi}(H_\star) = 0$ and $\mathbf{\Xi}_2(H_\otimes) = 0$.*

Lemma 6.4. *Zeroes of $\mathbf{\Xi}, \mathbf{\Xi}_2$ correspond to* BP *fixed points, as follows:*

(a) Suppose $\nu = \nu^{\mathrm{opt}}(\dot{h}^{\mathrm{tree}}(H)) = \nu^{\mathrm{bd}}(\dot{q})$, and let $\mu = \mu^{\mathrm{opt}}(H)$ be the optimizer for $\mathbf{\Lambda}(H)$. If $H \in \mathbf{N}_\circ$ with $H = H^{\mathrm{sym}}$ and $\mathbf{\Xi}(H) = 0$, then $\mu = \nu$ and $\mathtt{BP}\dot{q} = \dot{q}$.

(b) Suppose $\nu = \nu_2^{\mathrm{opt}}(\dot{h}^{\mathrm{tree}}(H)) = \nu^{\mathrm{bd}}(\dot{q})$, and let $\mu = \mu_2^{\mathrm{opt}}(H)$ be the optimizer for $\mathbf{\Lambda}_2(H)$. If $H \in \mathbf{N}_{\mathrm{sep}}$ with $H = H^{\mathrm{sym}}$ and $\mathbf{\Xi}_2(H) = 0$, then $\mu = \nu$ and $\mathtt{BP}\dot{q} = \dot{q}$.

Lemma 6.5. *The fixed points of Lemma 6.4 correpond to $\dot{q}_\star$:*

(a) If $H \in \mathbf{N}_\circ$ and $\nu = \nu^{\mathrm{opt}}(\dot{h}^{\mathrm{tree}}(H)) = \nu^{\mathrm{bd}}(\dot{q})$ with $\dot{q} = \mathtt{BP}\dot{q}$, then $\dot{q} = \dot{q}_\star$.

(b) If $H \in \mathbf{N}_{\mathrm{sep}}$ and $\nu = \nu_2^{\mathrm{opt}}(\dot{h}^{\mathrm{tree}}(H)) = \nu_2^{\mathrm{bd}}(\dot{q})$ with $\dot{q} = \mathtt{BP}\dot{q}$, then $\dot{q} = \dot{q}_\star \otimes \dot{q}_\star$.

Proof of Proposition 6.1. From Lemma 6.3, it suffices to show that if $H \in \mathbf{N}_\circ$ with $H = H^{\mathrm{sym}}$ and $\mathbf{\Xi}(H) = 0$, then $H = H_\star$. From Lemmas 6.4 and 6.5, $\nu = \nu^{\mathrm{opt}}(\dot{h}^{\mathrm{tree}}(H))$ and $\mu = \mu^{\mathrm{opt}}(H)$ are equal, and can be expressed via (57) in terms of $\dot{q} = \dot{q}_\star$. It follows that $H = H(\mu) = H_\star$ as claimed. □

Proof of Lemma 6.3. As we noted above, $\nu^{\mathrm{opt}}(\dot{h})$ can be expressed via (57) in terms of $\dot{q}$, but $\dot{q}$ is not uniquely determined by $\dot{h}$ if the constraints are rank-deficient. However, if $\dot{h}$ is a strictly positive measure on $\dot{\Omega}$, then it is straightforward to check that the constraints are of full rank, so $\dot{q}$ is unique. Let $\nu_\star$ denote the measure given by (57) with $\dot{q} = \dot{q}_\star$. It is easy to check, from the proof of Proposition 4.2, that $\dot{q}_\star$ is fully supported on $\dot{\Omega}_T$. Therefore $H(\nu_\star) = H_\star$ and $\dot{h}^{\mathrm{tree}}(H(\nu_\star)) = \dot{h}^{\mathrm{tree}}(H_\star)$, and these are strictly positive. It follows that $\nu_\star$ is the (unique) optimizer for both $\mathbf{\Lambda}(H_\star)$ and $\mathbf{\Lambda}^{\mathrm{opt}}(\dot{h}^{\mathrm{tree}}(H_\star))$, which proves $\mathbf{\Xi}(H_\star) = 0$. □

Proof of Lemma 6.4. Note that $\mathbf{\Lambda}(H)$ is an optimum over a subset of the measures ν which are considered for $\mathbf{\Lambda}^{\mathrm{opt}}(\dot{h}^{\mathrm{tree}}(H))$. Let $\mu = \mu^{\mathrm{opt}}(H)$ be the (unique) optimizer for $\mathbf{\Lambda}(H)$, and write $\nu = \nu^{\mathrm{opt}}(\dot{h}^{\mathrm{tree}}(H))$. Since ν is the unique optimizer in $\mathbf{\Lambda}^{\mathrm{opt}}(\dot{h}^{\mathrm{tree}}(H))$, we have $\mathbf{\Xi}(H) = 0$ if and only if $\mu = \nu$. In this case, since $H(\mu) = H$ with $H = H^{\mathrm{sym}}$, the same must hold for $H(\nu)$. Recall $H = H^{\mathrm{sym}}$ means that $\hat{H}$ is rotationally symmetric. We can take a marginal of (57) to obtain an expression for $\hat{H}$: in the first-moment calculation,

$$\hat{H}(\underline{\sigma}) = (\hat{\boldsymbol{z}})^{-1}\hat{\Phi}(\underline{\sigma})((\mathtt{BP}\dot{q})(\dot{\sigma}_1))\prod_{i=2}^{k}\dot{q}(\dot{\sigma}_i), \quad \underline{\sigma} \in \Omega^k.$$

The analogous expression holds in the second moment. We now claim that for the above measure $\hat{H}$ to be symmetric, we must have $\mathtt{BP}\dot{q} = \dot{q}$. Note that if $\hat{\Phi}$ were fully supported on Ω^k, and both $\dot{q}$ and $\mathtt{BP}\dot{q}$ were fully supported on $\dot{\Omega}$, 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:

1. In the first moment, Lemma 3.6 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
$$\hat{\Phi}((\dot{\sigma}, \hat{\sigma}), \mathtt{b}_0, \dots, \mathtt{b}_0) > 0. \tag{59}$$
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.
2. In the second moment, since we restrict to $H \in \mathbf{N}_{\mathrm{sep}}$, $\dot{q}(\dot{\sigma})$ 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 (59) 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. □

Proof of Lemma 6.5. It suffices to prove that $\dot{q} \in \mathbf{\Gamma}$. Since by assumption $\dot{q} = \mathtt{BP}\dot{q}$, we must have $\dot{q} = \dot{q}^{\mathrm{avg}}$. We then argue separately for the first and second moment:

1. For the first moment, we must verify (45). It follows directly from the relation $\dot{q} = \mathtt{BP}\dot{q}$ that $\dot{q}(\mathtt{r}) \geqslant \dot{q}(\mathtt{b})$. Since $H \in \mathbf{N}_\circ$, the majority of clauses have all `blue` edges, so
$$1/2 \leqslant \hat{H}(\mathtt{b}^k) \leqslant (\hat{\boldsymbol{z}})^{-1}\dot{q}(\mathtt{b})^k.$$
Next, for any $\underline{\sigma} \in \Omega^k$ which has exactly one entry `free` and the remaining $k-1$ entries `blue`, we must have $\hat{\Phi}(\underline{\sigma}) \geqslant 1/2$. It follows that
$$1 \gtrsim 2^k(\bar{H}(\mathtt{r}) + \bar{H}(\mathtt{f})) \gtrsim (\hat{\boldsymbol{z}})^{-1}[\dot{q}(\mathtt{r}) + 2^k\dot{q}(\mathtt{f})]\dot{q}(\mathtt{b})^{k-1}.$$
Comparing the two displays above gives $\dot{q}(\mathtt{r}) + 2^k\dot{q}(\mathtt{f}) \lesssim \dot{q}(\mathtt{b})$, proving $\dot{q} \in \mathbf{\Gamma}$.
2. For the second moment, we must verify (46). Condition (C) is immediate from the relation $\dot{q} = \mathtt{BP}\dot{q}$. From Lemma 4.4 it suffices to verify the condition with $\kappa = 0$, in which case condition (B) follows from (A). It therefore remains to verify (A). Denote
$$\mathtt{B} \equiv \{\mathtt{b}_0, \mathtt{b}_1\}^2, \quad \mathtt{B}_= \equiv \{\mathtt{b}_0\mathtt{b}_0, \mathtt{b}_1\mathtt{b}_1\} \subseteq \mathtt{B}, \quad \mathtt{B}_= \equiv \{\mathtt{b}_0\mathtt{b}_1, \mathtt{b}_1\mathtt{b}_0\} \subseteq \mathtt{B}.$$
Since the total density of `red` and `free` edges is small, the majority of clauses must have all colors in $\mathtt{B}$: $\hat{H}(\mathtt{B}^k) = 1 - O(k/2^k)$. For any $\underline{\sigma} \in \mathtt{B}^k$, $\hat{\Phi}(\underline{\sigma}) = 1 - O(k/2^k)$. Therefore
$$1 \asymp \hat{H}(\mathtt{B}^k) \asymp \dot{q}(\mathtt{B})^k/\hat{\boldsymbol{z}}. \tag{60}$$

For $H \in \mathbf{N}_{\mathrm{sep}}$, we have $|\bar{H}(\mathtt{B}_{=}) - \bar{H}(\mathtt{B}_{\neq})| \lesssim k^4/2^{k/2}$. We can obtain $\bar{H}$ as a marginal of $\hat{H}$: using the rotational symmetry of $\hat{H}$, we can express

$$\begin{aligned}&\bar{H}(\mathtt{B}_{=}) - \bar{H}(\mathtt{B}_{\neq}) - \mathsf{err}(H)\\ &= \sum_{\underline{\xi}=(\xi_2,\dots,\xi_k)\in \mathtt{B}^{k-1}} \prod_{i=2}^{k} \dot{q}(\xi_i) \Big[\sum_{\sigma\in\mathtt{B}_{=}} \dot{q}(\sigma)\hat{\Phi}(\sigma,\underline{\xi}) - \sum_{\sigma'\in\mathtt{B}_{\neq}} \dot{q}(\sigma')\hat{\Phi}(\sigma',\underline{\xi}) \Big]\end{aligned}$$

where $\mathsf{err}(H)$ is the contribution from the clauses which are not all $\mathtt{B}$, and is bounded by $O(k/2^k)$. Recalling $\hat{\Phi}(\underline{\sigma}) = 1 - O(k/2^k)$ for $\underline{\sigma} \in \mathtt{B}^k$, the right-hand side above equals

$$\frac{\dot{q}(\mathtt{B})^k}{\hat{\boldsymbol{z}}} \Big[O(k/2^k) + \frac{\dot{q}(\mathtt{B}_{=}) - \dot{q}(\mathtt{B}_{\neq})}{\dot{q}(\mathtt{B})} \Big].$$

Applying (60) and rearranging gives

$$\frac{k^4}{2^{k/2}} \gtrsim \frac{|\dot{q}(\mathtt{B}_{=}) - \dot{q}(\mathtt{B}_{\neq})|}{\dot{q}(\mathtt{B})}.$$

It remains to show that $\sum_{\dot{\sigma}\notin\mathtt{B}} (2^{-k})^{\mathtt{r}[\dot{\sigma}]} \dot{q}(\dot{\sigma}) = O(2^{-k})\dot{q}(\mathtt{B})$. We will deduce this from the fact that the total fraction of clauses where $\sigma_i \notin \mathtt{B}$ for some $i \in [k]$ is $O(k/2^k)$. By rotational symmetry of $\hat{H}$, the fraction with $\sigma_1 \notin \mathtt{B}$ is $O(2^{-k})$. Take $\dot{\sigma} = (\dot{\sigma}^1, \dot{\sigma}^2) \in \dot{\Omega}^2 \backslash \mathtt{B}$. For $j = 1, 2$, let

$$\sigma^j = \begin{cases} \sigma^j & \sigma^j \in \{\mathtt{r}, \mathtt{b}\}, \\ (\dot{\sigma}^j, \square) & \text{otherwise.} \end{cases}$$

Denote $\sigma \equiv (\sigma^1, \sigma^2)$. We now consider separately the cases $\mathtt{r}[\dot{\sigma}] = 0, 1, 2$:

(a) If $\mathtt{r}[\dot{\sigma}] = 0$, then note that for any $\underline{\xi} \in \mathtt{B}^{k-1}$ we have $\hat{\Phi}(\sigma, \underline{\xi}) \asymp 1$. On the other hand, using the rotational symmetry of $\hat{H}$, the total fraction of clauses where the first incident edge has a color in $\dot{\Omega}^2 \backslash \mathtt{B}$ is $O(2^{-k})$. Thus

$$2^{-k} \gtrsim \frac{\dot{q}(\mathtt{B})^k}{\hat{\boldsymbol{z}}} \sum_{\dot{\sigma}\notin\mathtt{B}} \mathbf{1}\{\mathtt{r}[\dot{\sigma}] = 0\} \frac{\dot{q}(\dot{\sigma})}{\dot{q}(\mathtt{B})} \asymp \frac{\dot{q}(\dot{\sigma} \notin \mathtt{B} : \mathtt{r}[\dot{\sigma}] = 0)}{\dot{q}(\mathtt{B})}.$$

(b) If $\mathtt{r}[\dot{\sigma}] = 1$, then for any $\underline{\xi} \in \mathtt{B}^{k-1}$ with at least two indices each in $\mathtt{B}_{=}$ and $\mathtt{B}_{\neq}$, we have $\hat{\Phi}(\sigma, \underline{\xi}) \asymp 2^{-k}$. Thus

$$2^{-k} \gtrsim \frac{\dot{q}(\mathtt{B})^k}{\hat{\boldsymbol{z}}} \sum_{\dot{\sigma}\notin\mathtt{B}} \mathbf{1}\{\mathtt{r}[\dot{\sigma}] = 1\} 2^{-k} \frac{\dot{q}(\dot{\sigma})}{\dot{q}(\mathtt{B})} \asymp \frac{\dot{q}(\dot{\sigma} \notin \mathtt{B} : \mathtt{r}[\dot{\sigma}] = 1)}{2^k \dot{q}(\mathtt{B})}.$$

(c) If $\mathtt{r}[\dot{\sigma}] = 2$, then

$$2^{-k} \gtrsim \frac{1}{\hat{\boldsymbol{z}}} \sum_{\dot{\sigma}\notin\mathtt{B}} \mathbf{1}\{\mathtt{r}[\dot{\sigma}] = 2\} 2^{-k} \min\{\dot{q}(\mathtt{B}_{=}), \dot{q}(\mathtt{B}_{\neq})\}^{k-1} \asymp \frac{\dot{q}(\dot{\sigma} \notin \mathtt{B} : \mathtt{r}[\dot{\sigma}] = 2)}{4^k \dot{q}(\mathtt{B})}$$

Combining the above estimates verifies $\sum_{\dot{\sigma}\notin\mathtt{B}} (2^{-k})^{\mathtt{r}[\dot{\sigma}]} \dot{q}(\dot{\sigma}) = O(2^{-k})\dot{q}(\mathtt{B})$.

Altogether this verifies, in both the first and second moment, that $\dot{q}$ lies in the regime for BP contraction, and consequently must equal $\dot{q}_\star$ as claimed. □

6.2. **Non-degeneracy around minimizer.**

Proof of Proposition 6.2. Consider H near $H_\star$, and let $\nu = \nu^{\mathrm{opt}}(\dot{h}^{\mathrm{tree}}(H))$ and $\mu = \mu^{\mathrm{opt}}(H)$. It follows from Proposition A.6 that

$$\boldsymbol{\Xi}(H) = \mathcal{H}(\mu|\nu) \gtrsim \|\mu - \nu\|^2,$$

so it suffices to show that $\|\mu - \nu\| \gtrsim \|H - H_\star\|$. To this end, recall ν can be expressed via (57) in terms of some $\dot{q}$, while $\nu_\star$ can be expressed in terms of $\dot{q}_\star$. Thus

$$\|\nu - \nu_\star\|_1 \lesssim \|\dot{q} - \dot{q}_\star\|_1.$$

For H in a small enough neighborhood of $H_\star$, the constraints are of full rank, so $\nu^{\mathrm{opt}}(\dot{h}^{\mathrm{tree}}(H))$ is expressible in terms of $\dot{q}$ for $\dot{q}$ *uniquely* determined by $\dot{h}^{\mathrm{tree}}(H)$, hereafter denoted $\dot{q} = \dot{q}^{\mathrm{opt}}(H)$. In fact, we see further from (133) that $\dot{q}^{\mathrm{opt}}$ is differentiable in a neighborhood of $H_\star$. Then, since $\dot{q}^{\mathrm{opt}}(H_\star) = \dot{q}_\star$ which lies in the interior of $\boldsymbol{\Gamma}$, we must have $\dot{q}^{\mathrm{opt}}(H) \in \boldsymbol{\Gamma}$ for H in some neighborhood of $H_\star$. It then follows by Proposition 4.2 in the first moment, and by Proposition 4.3 in the second moment, that

$$(1-c)\|\dot{q} - \dot{q}_\star\|_1 \leqslant \|\dot{q} - \dot{q}_\star\|_1 - \|\mathtt{BP}\dot{q} - \dot{q}_\star\|_1 \leqslant \|\dot{q} - \mathtt{BP}\dot{q}\|_1.$$

To compare $\dot{q}$ with $\mathtt{BP}\dot{q}$, consider

$$\sup\{\mathcal{H}(\hat{\nu}) + \lambda\langle \ln \hat{\Phi}, \hat{\nu}\rangle : \hat{\nu}(\dot{\sigma}_i = \dot{\sigma}) = \hat{H}(\dot{\sigma}_i = \dot{\sigma}) \text{ for each } i\}.$$

There is a unique optimizer $\hat{\nu} = \hat{\nu}^{\mathrm{opt}}(\hat{H})$ which can be expressed as

$$\hat{\nu}(\underline{\sigma}) \cong \hat{\Phi}(\underline{\sigma})^\lambda \prod_{i=1}^{k} \widetilde{\gamma}_i(\dot{\sigma}_i).$$

In a neighborhood of $\hat{H}_\star$, the vector $\widetilde{\gamma} \equiv (\widetilde{\gamma}_i)_i$ is uniquely determined as a smooth function of $\hat{H}$, which we denote $\widetilde{\gamma}^{\mathrm{opt}}(\hat{H})$. Consequently, if we denote $\hat{H}^{\mathrm{rot}}(\underline{\sigma}) = \hat{H}(\sigma_2, \dots, \sigma_k, \sigma_1)$, then

$$\begin{aligned}
\|\dot{q} - \mathtt{BP}\dot{q}\|_1 &\leqslant \|(\mathtt{BP}\dot{q}, \dot{q}, \dots, \dot{q}) - (\dot{q}, \mathtt{BP}\dot{q}, \dot{q}, \dots, \dot{q})\|_1 = \|\widetilde{\gamma}^{\mathrm{opt}}(\hat{H}(\nu)) - \widetilde{\gamma}^{\mathrm{opt}}(\hat{H}(\nu)^{\mathrm{rot}})\|_1 \\
&\lesssim \|\hat{H}(\nu) - \hat{H}(\nu)^{\mathrm{rot}}\| \leqslant \|\hat{H}(\nu) - \hat{H}(\mu)\| + \|\hat{H}(\mu) - \hat{H}(\nu)^{\mathrm{rot}}\| \\
&= 2\|\hat{H}(\nu) - \hat{H}(\mu)\| \lesssim \|\mu - \nu\|.
\end{aligned}$$

where in the last line we used that $\hat{H}(\mu) = \hat{H}(\mu)^{\mathrm{rot}}$. Combining the above inequalities gives $\|H - H_\star\| \lesssim \|\mu - \nu\|$ as claimed. □

Proof of Propositions 3.7 and 3.14. Follows from Proposition 6.1 and 6.2. □

7. Conclusion of lower bound

In this section we prove Propositions 3.10 and 3.17.

7.1. **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 7.1. *Consider random regular* NAE-SAT *at clause density* $\alpha \geqslant 2^{k-1}\ln 2 - O(1)$. *On* $\mathscr{G} = (V, F, E)$, *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* ρ *fraction of variables. Then*

$$\mathbb{E}Z^2[\rho] \leqslant (\mathbb{E}Z)\exp\Big\{n\Big[H(\rho) - (\ln 2)\pi(\rho) + O(1/2^k)\Big]\Big\},$$

for $\pi(\rho) \equiv 1 - \rho^k - (1-\rho)^k$.

Proof. For $\underline{u} \in \{0,1\}^V$, let $I^{\mathrm{NAE}}_{\mathscr{G}}(\underline{u})$ be the indicator that $\underline{u}$ is a valid NAE-SAT solution on $\mathscr{G}$. Fix any pair of vectors $\underline{x}, \acute{\underline{x}} \in \{0,1\}^V$ which agree on ρ fraction of variables:

$$\mathbb{E}Z^2[\rho] = 2^n \binom{n}{n\rho} \mathbb{E}[I^{\mathrm{NAE}}_{\mathscr{G}}(\underline{x}) I^{\mathrm{NAE}}_{\mathscr{G}}(\acute{\underline{x}})] = (\mathbb{E}Z) \binom{n}{n\rho} \mathbb{E}[I^{\mathrm{NAE}}_{\mathscr{G}}(\acute{\underline{x}}) \,|\, I^{\mathrm{NAE}}_{\mathscr{G}}(\underline{x}) = 1].$$

Given $\underline{x}, \acute{\underline{x}}$, let $M \equiv M(\underline{x}, \acute{\underline{x}})$ 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 $\acute{\underline{x}}$. If we define i.i.d. binomial random variables $D_a \sim \mathrm{Bin}(k, \rho)$, indexed by $a \in F$, then

$$\mathbb{P}(M = m\gamma) = \mathbb{P}\bigg(\sum_{a \in F} \mathbf{1}\{D_a \notin \{0,k\}\} \,\bigg|\, \sum_{a \in F} D_a = mk\rho \bigg).$$

The $(D_a)_{a \in F}$ sum to $mk\rho$ with probability which is polynomial in n, so

$$\mathbb{P}(M = m\gamma) \leqslant n^{O(1)} \mathbb{P}(\mathrm{Bin}(m, \pi) = m\gamma)$$

with $\pi = \pi(\rho)$ as in the statement of the lemma. Therefore

$$\mathbb{E}[I^{\mathrm{NAE}}_{\mathscr{G}}(\acute{\underline{x}}) \,|\, I^{\mathrm{NAE}}_{\mathscr{G}}(\underline{x}) = 1] \leqslant n^{O(1)} \mathbb{E}\bigg[\bigg(\frac{2^k - 4}{2^k - 2} \bigg)^X \bigg]$$

for $X \sim \mathrm{Bin}(m, \rho)$. It is easily seen that the above is $\leqslant \exp\{-m\pi/2^{k-1}\}$, and the claimed bound follows, using the lower bound on $\alpha = m/n$. □

Corollary 7.2. *Let* $\psi(\rho) = H(\rho) - (\ln 2)\pi(\rho)$. *Then* $\psi(\rho) \leqslant -2k/2^k$ *for all* ρ *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 $\alpha = m/n \geqslant 2^{k-1} \ln 2 - O(1)$, $\mathbb{E}Z^2[\rho] \leqslant \exp\{-nk/2^k\}$ *for all such* ρ.

Proof. Note that $H(\frac{1+\epsilon}{2}) \leqslant \ln 2 - \epsilon^2/2$. If $(k \ln k)/2^k \leqslant \epsilon \leqslant 1/k$, then

$$\psi(\tfrac{1+\epsilon}{2}) \leqslant -\epsilon^2/2 + O(k\epsilon/2^k) \leqslant -\epsilon^2/3.$$

Both $H(\frac{1+\epsilon}{2})$ and $\pi(\frac{1+\epsilon}{2})$ are symmetric about $\epsilon = 0$, and decreasing on the interval $0 \leqslant \epsilon \leqslant 1$. It follows that for any $0 \leqslant a \leqslant b \leqslant 1$,

$$\max_{a \leqslant \epsilon \leqslant b} \psi(\tfrac{1+\epsilon}{2}) \leqslant H(\tfrac{1+a}{2}) - (\ln 2)\pi(\tfrac{1+b}{2}).$$

With this in mind, if $1/k \leqslant \epsilon \leqslant 1 - 5(\ln k)/k$,

$$\psi(\tfrac{1+\epsilon}{2}) \leqslant -(2k^2)^{-1} + O(k^{-5/2}) \leqslant -(4k^2)^{-1}.$$

If $1 - 5(\ln k)/k \leqslant \epsilon \leqslant 1 - (\ln k)^3/k^2$,

$$\psi(\tfrac{1+\epsilon}{2}) \leqslant O(1)(\ln k)^2/k - \Omega(1)(\ln k)^3/k \leqslant -\Omega(1)(\ln k)^3/k.$$

Finally, if $1 - (\ln k)^3/k^2 \leqslant \epsilon \leqslant 1 - \exp\{-2k/(\ln k)\}$, then

$$\psi(\tfrac{1+\epsilon}{2}) \leqslant O(1)\epsilon k/(\ln k) - \Omega(1)\epsilon k \leqslant -\Omega(1)\epsilon k.$$

Combining these estimates proves the claimed bound on $\psi(\rho)$. The assertion for $\mathbb{E}[Z^2(\rho)]$ then follows by substituting into Lemma 7.1, and noting that $\mathbb{E}Z \leqslant \exp\{O(n/2^k)\}$. □

7.2. **Large overlap.** In what follows, we restrict consideration to a small neighborhood $\mathbf{N}$ of $H_\star$. We abbreviate $\underline{\sigma} \in H$ if $H(\mathscr{G}, \underline{\sigma}) = H$, and $\underline{\sigma} \in \mathbf{N}$ if $H(\mathscr{G}, \underline{\sigma}) \in \mathbf{N}$. 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 $\boldsymbol{Z}^{\mathrm{ns}}(H, H')$ count the colorings $\underline{\sigma} \in H$ such that

$$\left|\left\{\underline{\sigma}' \in H' : \delta(\underline{\sigma}, \underline{\sigma}') \leqslant \exp\{-k/(\ln k)\}\right\}\right| \geqslant \omega(n),$$

for $\omega(n) = \exp\{(\ln n)^4\}$. (Although we will not write it explicitly, it should be understood that $\boldsymbol{Z}^{\mathrm{ns}}(H, H')$ depends on $\mathscr{G}$, since both $\underline{\sigma}, \underline{\sigma}'$ are required to be valid colorings of $\mathscr{G}$.) Let $\boldsymbol{Z}^{\mathrm{ns}}(\mathbf{N})$ denote the sum of $\boldsymbol{Z}^{\mathrm{ns}}(H; H')$ over all pairs $H, H' \in \mathbf{N}$ with $H' \succcurlyeq H$. Let $\boldsymbol{Z}(\mathbf{N})$ denote the sum of $\boldsymbol{Z}(H)$ over all $H \in \mathbf{N}$.

Proposition 7.3. *There exists a small enough positive constant $\epsilon_{\max}(k)$ such that, if $\mathbf{N}$ is the ϵ-neighborhood of $H_\star$ for any $\epsilon \leqslant \epsilon_{\max}$, then*

$$\mathbb{E}\boldsymbol{Z}^{\mathrm{ns}}(\mathbf{N}) \leqslant \mathbb{E}\boldsymbol{Z}(\mathbf{N}) \exp\{-(\ln n)^2\}.$$

Proof. By definition,

$$\boldsymbol{Z}^{\mathrm{ns}}(\mathbf{N}) = \sum_{H \in \mathbf{N}} \boldsymbol{Z}^{\succcurlyeq}(H), \quad \boldsymbol{Z}^{\succcurlyeq}(H) \equiv \sum_{H' \in \mathbf{N}} \mathbf{1}\{H' \succcurlyeq H\} \boldsymbol{Z}^{\mathrm{ns}}(H, H').$$

It suffices to show that for every $H \in \mathbf{N}$, $\mathbb{E}\boldsymbol{Z}^{\succcurlyeq}(H) \leqslant \mathbb{E}\boldsymbol{Z}(H) \exp\{-2(\ln n)^2\}$. 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

$$\left|\left\{\underline{\sigma}' \in \mathbf{N} : \underline{\sigma}' \succcurlyeq \underline{\sigma} \text{ and } \delta(\underline{\sigma}, \underline{\sigma}') \leqslant \exp\{-k/(\ln k)\}\right\}\right| \geqslant \omega(n).$$

(Again, it is understood that both $\underline{\sigma}, \underline{\sigma}'$ must be valid colorings of $\mathscr{G}$.) Then

$$\boldsymbol{Z}^{\succcurlyeq}(H) \leqslant n^c \sum_{\underline{\sigma} \in H} \mathbf{1}\{(\mathscr{G}, \underline{\sigma}) \in \boldsymbol{E}\}.$$

Consequently, in order to show the required bound on $\mathbb{E}\boldsymbol{Z}^{\succcurlyeq}(H)$, it suffices to show

$$\mathbb{P}^H(\boldsymbol{E}) \leqslant n^{-c} \exp\{-2(\ln n)^2\}, \tag{61}$$

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 $\delta(\underline{\sigma}, \underline{\sigma}') \leqslant \delta$ is at most $(cn)^{n\delta}$, which is $\ll \omega(n)$ if $\delta \leqslant n^{-1}(\ln n)^2$. This implies that there must exist $\underline{\sigma}' \in \mathbf{N}$ such that $\underline{\sigma}' \succcurlyeq \underline{\sigma}$ and $n^{-1}(\ln n)^2 \leqslant \delta(\underline{\sigma}, \underline{\sigma}') \leqslant \exp\{-k/(\ln k)\}$. It follows that

$$S \equiv \{v \in V : x_v(\underline{\sigma}) \in \{\mathtt{0}, \mathtt{1}\} \text{ and } x_v(\underline{\sigma}') \neq x_v(\underline{\sigma})\}$$

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 `red` (resp. `blue`) edges incident to a subset of vertices U. Let I_S be the indicator that all variables in S are forced. For any fixed $S \subseteq V$,

$$\mathbb{P}^H(S \text{ internally forced}) \leqslant \mathbb{E}_{\mathbb{P}^H}\left[I_S k^{\mathtt{R}_S} \frac{(\mathtt{B}_S)_{\mathtt{R}_S}}{(\mathtt{B}_F)_{\mathtt{R}_S}} \right] \leqslant \mathbb{E}_{\mathbb{P}^H}[I_S (4ks)^{\mathtt{R}_S}].$$

In the first inequality, the factor $k^{\mathtt{R}_S}$ accounts for the choice, for each S-incident `red` 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 be `blue`) will also be S-incident. The second inequality follows by noting that we certainly have $\mathtt{B}_S \leqslant nsd$, and for H near $H_\star$ we also clearly have $\mathtt{B}_F \geqslant nd/4$.

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 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 $\mathbb{Q}^H(\mathsf{MARG}) \geqslant n^{-c}$ (adjusting c as needed), we have

$$\mathbb{P}^H((\mathscr{G}, \underline{\sigma})) = \mathbb{Q}^H((\mathscr{G}, \underline{\sigma}) \,|\, \mathsf{MARG}) \leqslant n^c \, \mathbb{Q}^H((\mathscr{G}, \underline{\sigma}); \mathsf{MARG}).$$

Let us abbreviate $\dot{H}(\ell)$ for the probability under $\dot{H}$ that $\underline{\sigma}$ has ℓ `red` entries: then

$$\mathbb{E}_{\mathbb{P}^H}[I_S(4ks)^{\mathtt{R}_S}] \leqslant n^c \, \mathbb{E}_{\mathbb{Q}^H}[I_S(4ks)^{\mathtt{R}_S}; \mathsf{MARG}] \leqslant n^c \left(\sum_{\ell \geqslant 1} \dot{H}(\ell)(4ks)^\ell \right)^{ns}. \tag{62}$$

For H sufficiently close to $H_\star$, we will have

$$\dot{H}(\ell) \leqslant 2\dot{H}_\star(\ell) \leqslant 2\binom{d}{\ell} \frac{\hat{q}_\star(\mathtt{r}_1)^\ell \hat{q}_\star(\mathtt{b}_1)^{d-\ell}}{[\hat{q}_\star(\mathtt{r}_1) + \hat{q}_\star(\mathtt{b}_1)]^d - \hat{q}_\star(\mathtt{b}_1)^d}.$$

It follows that the right-hand side of (62) is (for some absolute constant δ)

$$\leqslant n^c \, 2^{ns} \left(\frac{[\hat{q}_\star(\mathtt{r}_1) \cdot 4ks + \hat{q}_\star(\mathtt{b}_1)]^d - \hat{q}_\star(\mathtt{b}_1)^d}{[\hat{q}_\star(\mathtt{r}_1) + \hat{q}_\star(\mathtt{b}_1)]^d - \hat{q}_\star(\mathtt{b}_1)^d} \right)^{ns} \leqslant n^c s^{ns} 2^{-\delta kns},$$

where the last inequality uses that $s \leqslant \exp\{-k/(\ln k)\}$. Summing over S gives

$$\mathbb{P}^H(\boldsymbol{E}) \leqslant \max_{s \geqslant (2n)^{-1}(\ln n)^2} n^c 2^{-\delta kns/2} \leqslant \exp\{-\Omega(1) k (\ln n)^2\}.$$

This implies (61); and the claimed result follows as previously explained. □

Proof of Proposition 3.10. Follows by combining Corollary 7.2 and Proposition 7.3. □

8. Upper bound

In this section we prove the upper bound, Proposition 3.19.

8.1. Interpolation bound for regular graphs.

For a certain family of spin systems that includes NAE-SAT, an interpolative calculation gives an upper bound for the free energy on Erdős-Rényi graphs ([FL03, PT04], cf. [Gue03]). These bounds build on earlier work [GT03] concerning the subadditivity of the free energy in the Sherrington–Kirkpatrick model, which was later generalized to a broad class of models [BGT13, Gam14]. (Although these results are closely related, we remark that interpolation gives quantitative bounds whereas subadditivity does not.) To prove our main result, we establish the analogue of [FL03, PT04] 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.

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 $\mathcal{X} \to \mathbb{R}_{\geqslant 0}$. Independently of b, let $f_1, \dots, f_k$ be i.i.d. copies of f, and define the random function

$$\theta(\underline{x}) \equiv \sum_{s \in S} b_s \prod_{j=1}^{k} f_{s,j}(x_j). \tag{63}$$

Let h be another (random) element of $\mathcal{F}(\mathcal{X})$. Assume there is a constant $\epsilon > 0$ so that

$$\epsilon \leqslant \{h, 1 - \theta\} \leqslant 1/\epsilon \quad \text{almost surely.} \tag{64}$$

Note we do not require the b_s to be non-negative; however, we assume that

$$b^p(\underline{s}) \equiv \mathbb{E}\Big[\prod_{\ell=1}^{p} b_{s_\ell}\Big] \geqslant 0 \quad \text{for any } p \geqslant 1,\ \underline{s} \equiv (s_1, \dots, s_p) \in S^p. \tag{65}$$

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 θ_a are i.i.d. copies of θ. 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

$$\mu_{\mathscr{G}}(\underline{x}) \equiv \frac{1}{Z(\mathscr{G})} \prod_{v \in V} h_v(x_v) \prod_{a \in F} [1 - \theta_a(\underline{x}_{\delta a})] \tag{66}$$

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

$$F_n \equiv n^{-1} \mathbb{E}_n \ln Z(\mathscr{G}).$$

Example 8.1 (positive temperature NAE-SAT)**.** Let $\mathcal{X} = \{\mathtt{0}, \mathtt{1}\}$, and let $\underline{\mathtt{L}} \equiv (\mathtt{L}_i)_{i \leqslant k}$ 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

$$\theta(\underline{x}) \equiv (1 - e^{-\beta}) \bigg(\prod_{i=1}^{k} \frac{\mathtt{L}_1 \oplus x_i}{2} + \prod_{i=1}^{k} \frac{\mathtt{1} \oplus \mathtt{L}_i \oplus x_i}{2} \bigg)$$

where $\beta \in (0, \infty)$ is the inverse temperature. In this model, each violated clause incurs a multiplicative penalty $e^{-\beta}$.

Example 8.2 (positive-temperature coloring)**.** Let $\mathcal{X} = [q]$. The positive-temperature coloring (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 bound is a random regular graph analog of [PT04, Thm. 3]. (We have stated our result for a more general class of models than considered in [PT04]; however the main result of [PT04] extends to these models with minor modifications.)

Theorem 8.3. *Consider a (random) Gibbs measure (66) satisfying assumptions (63)-(65), and let $F_n \equiv n^{-1}\mathbb{E}_n \ln Z(\mathscr{G})$. 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 $\underline{\eta} \equiv (\eta_{a,j})_{a\geqslant 0, j\geqslant 0}$ be an array of i.i.d. samples from ζ. For each index (a,j) let $\rho_{a,j}$ be a conditionally independent sample from $\eta_{a,j}$, and denote $\underline{\rho} \equiv (\rho_{a,j})_{a\geqslant 0, j\geqslant 0}$. 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 $\zeta \in \mathcal{M}_2$,

$$F_n \leqslant \lambda^{-1}\mathbb{E}\ln\mathbb{E}'\bigg[\Big(\sum_{x\in\mathcal{X}} h(x)\prod_{a=1}^{d}\boldsymbol{u}_a(x)\Big)^{\lambda}\bigg] - (k-1)\alpha\lambda^{-1}\mathbb{E}\ln\mathbb{E}'[(\boldsymbol{u}_0)^{\lambda}] + O_\epsilon(n^{-1/3})$$

where $\mathbb{E}'$ denotes the expectation over $\underline{\rho}$ conditioned on all else, and $\mathbb{E}$ denotes the overall expectation.

Remark 8.4. 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 [MM09, Ch. 19]), which has state space $\mathcal{M}_0$. The informal picture is that the η associated to variable x is determined by the geometry of the local neighborhood of x — that is to say, the randomness of ζ 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 8.3 can be achieved with ζ a point mass.

8.2. **Replica symmetric bound.** Along the lines of [PT04], we first prove a weaker "replica symmetric" version of Theorem 8.3. Afterwards we will apply it to obtain the full result.

Theorem 8.5. *In the setting of Theorem 8.3, 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 $F_n \leqslant \Phi_V - \Phi_F - O_\epsilon(n^{-1/3})$.

Inspired by the proof of [BGT13], we prove Theorem 8.5 by a combinatorial interpolation between two graphs, $\mathscr{G}_{-1}$ and $\mathscr{G}_{nd+1}$. The initial graph $\mathscr{G}_{-1}$ will have free energy Φ_V, and the final graph $\mathscr{G}_{nd+1}$ will have free energy $F_n + \Phi_F$. We will show that, up to $O_\epsilon(n^{1/3})$ error, the free energy of $\mathscr{G}_{-1}$ will be larger than that of $\mathscr{G}_{nd+1}$, from which the bound of Theorem 8.5 follows.

To begin, we take $\mathscr{G}_{-1}$ to be a factor graph consisting of n disjoint trees (Figure 4a). 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 η_u is a sample from ζ, and ρ_u is a sample from η_u.[2] As before, the variables and clauses in $\mathscr{G}_{-1}$ are labelled independently with functions h_v and θ_a. We now additionally assign to each $u \in U$ the label (η_u, ρ_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 $\mathscr{G}_{-1}, \mathscr{G}_0, \dots, \mathscr{G}_{nd+1}$. Fix $m' \equiv 2n^{2/3}$. The construction proceeds by adding and removing clauses. Whenever we remove a clause a, the edges δa 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 θ_b of θ. 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 COUP_r. Let $\mathsf{COUP}_{\leqslant r}$ be the event that COUP_s occurs for all $0 \leqslant s \leqslant r$. The event $\mathsf{COUP}_{\leqslant -1}$ occurs vacuously, so $\mathbb{P}(\mathsf{COUP}_{\leqslant -1}) = 1$. 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 $0 \leqslant r \leqslant nd - m' - 1$, we start from $\mathscr{G}_r$ and form $\mathscr{G}_{r+1}$ as follows.
   a. If $\mathsf{COUP}_{\leqslant r-1}$ 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
   $$\mathsf{COUP}_r \equiv \{\min\{\delta' U_{r,\circ}, \delta' V_{r,\circ}\} \geqslant k\}.$$
   If instead $\mathsf{COUP}_{\leqslant r-1}$ fails, then $\mathsf{COUP}_{\leqslant r}$ fails by definition.
   b. If $\mathsf{COUP}_{\leqslant r}$ fails, let $\mathscr{G}_{r+1} = \mathscr{G}_r$. If $\mathsf{COUP}_{\leqslant r}$ 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.
3. For $nd - m' \leqslant r \leqslant nd - 1$ 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 $\mathscr{G}_{nd+1}$.

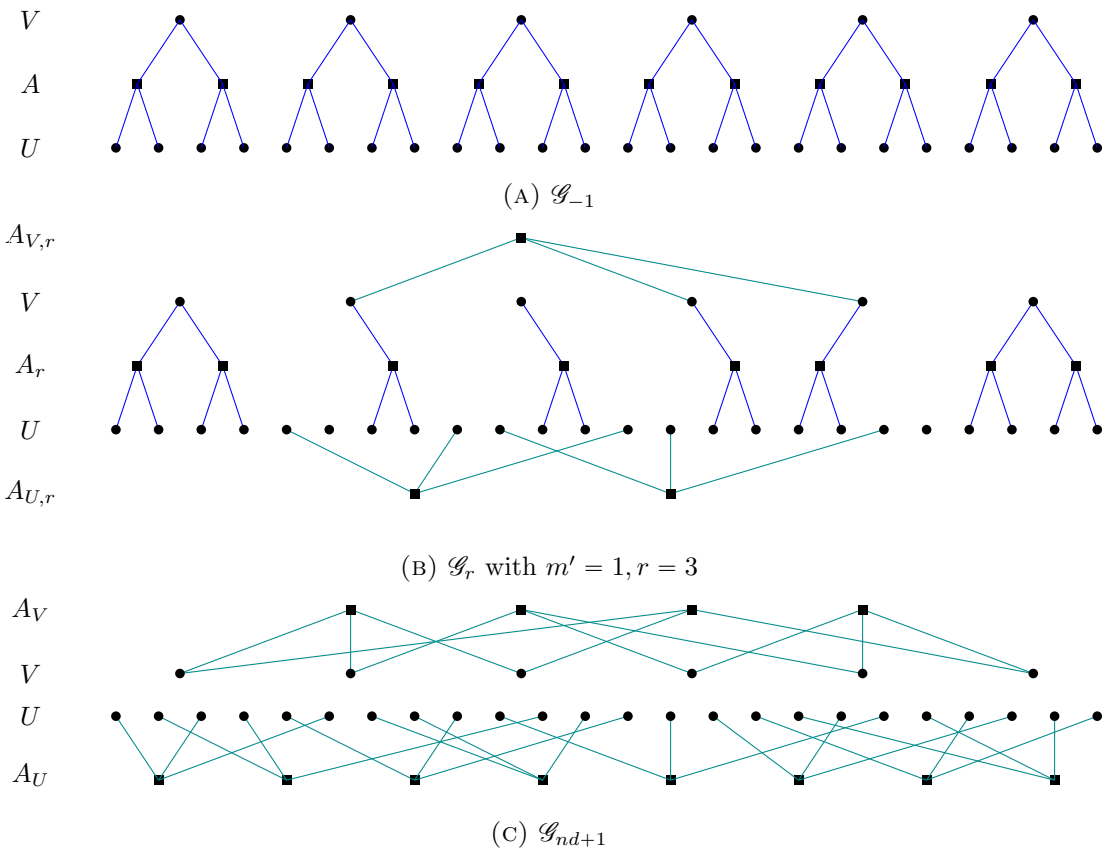
By construction, $\mathscr{G}_{nd+1}$ consists of two disjoint subgraphs, which are the induced subgraphs $\mathscr{G}_U, \mathscr{G}_V$ of U, V respectively. Note that $\mathscr{G}_V$ is distributed as the random graph $\mathscr{G}$ of interest, while $\mathscr{G}_U$ consists of a collection of $nd(k-1)/k = n\alpha(k-1)$ disjoint trees.

Lemma 8.6. *Under the construction above,*

$$\mathbb{E}\ln Z(\mathscr{G}_0) \geqslant \mathbb{E}\ln Z(\mathscr{G}_{nd}) - O_\epsilon(n^{1/3}), \tag{67}$$

where the expectation $\mathbb{E}$ is over the sequence of random graphs $(\mathscr{G}_r)_{-1\leqslant r\leqslant nd+1}$.

[2]For the proof of Theorem 8.5 it is equivalent to sample ρ from $\eta^{\mathrm{avg}} \equiv \int \eta\, d\zeta$.



(A) $\mathscr{G}_{-1}$

(B) $\mathscr{G}_r$ with $m' = 1, r = 3$

(C) $\mathscr{G}_{nd+1}$

FIGURE 4. Interpolation with $d = 2, k = 3$, $n = 6$.

Proof. Let $\mathscr{F}_{r,\circ}$ be the σ-field generated by $\mathscr{G}_{r,\circ}$, and write $\mathbb{E}_{r,\circ}$ for expectation conditioned on $\mathscr{F}_{r,\circ}$. One can rewrite (67) as

$$\mathbb{E} \ln \frac{Z(\mathscr{G}_0)}{Z(\mathscr{G}_{nd})} = \sum_{r=0}^{nd-1} \mathbb{E}\Delta_r, \quad \Delta_r \equiv \mathbb{E}_{r,\circ} \ln \frac{Z(\mathscr{G}_r)}{Z(\mathscr{G}_{r,\circ})} - \mathbb{E}_{r,\circ} \ln \frac{Z(\mathscr{G}_{r+1})}{Z(\mathscr{G}_{r,\circ})}. \tag{68}$$

In particular, $\Delta_r = 0$ if the coupling fails. Therefore it suffices to show that Δ_r is positive conditioned on $\mathsf{COUP}_{\leqslant r}$.[3] 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 with a random label θ_a which is a fresh copy of θ. 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

$$\mathbb{E}_{r,\circ} \ln \frac{Z(\mathscr{G}_r)}{Z(\mathscr{G}_{r,\circ})} = \mathbb{E}_{r,\circ} \ln(1 - \langle \theta(\underline{x}_{\delta a}) \rangle_{r,\circ}) = \sum_{p \geqslant 1} \frac{1}{p} \mathscr{A}_p, \quad \mathscr{A}_p \equiv \mathbb{E}_{r,\circ} \Big[\Big\langle \prod_{\ell=1}^{p} \theta(\underline{x}^\ell_{\delta a}) \Big\rangle_{r,\circ} \Big].$$

We have $\mathbb{E}_{r,\circ} = \mathbb{E}_a \mathbb{E}_\theta$ where $\mathbb{E}_a$ is expectation over the choice of δa, and $\mathbb{E}_\theta$ is expectation over the choice of θ. Under $\mathbb{E}_a$, the edges $(e_2, \dots, e_k)$ are weakly dependent, since they

[3]The event $\mathsf{COUP}_{\leqslant r}$ 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.

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 δa. Under $\mathbb{E}_{a,\mathrm{ind}}$ the chance of a collision $e_i = e_j$ $(i \leqslant j)$ is $O(k^2/|\delta' U_{r,\circ}|)$. Recalling $1 - \theta \geqslant \epsilon$ almost surely, we have

$$\mathscr{A}_{p,\mathrm{ind}} \equiv \mathbb{E}_{a,\mathrm{ind}} \mathbb{E}_\theta \biggl[\Bigl\langle \prod_{\ell=1}^{p} \theta(\underline{x}^\ell_{\delta a}) \Bigr\rangle_{r,\circ} \biggr] = \mathscr{A}_p + O(1)(1-\epsilon)^p \min\biggl\{ \frac{k^2}{|\delta' U_{r,\circ}|}, 1 \biggr\}.$$

Recall from (63) the product form of θ, 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 (65), we have

$$\begin{aligned} \mathscr{A}_{p,\mathrm{ind}} &= \sum_{\underline{s} \in S^p} b^p(\underline{s}) \biggl\langle \mathbb{E}_{a,\mathrm{ind}} \biggl\{ \prod_{j=1}^{k} \mathbb{E}_f \biggl[\prod_{\ell=1}^{p} f_{s_\ell}(x^\ell_{e_j}) \biggr] \biggr\} \biggr\rangle_{r,\circ} \\ &= \sum_{\underline{s} \in S^p} b^p(\underline{s}) \langle I_{V,\underline{s}}(\underline{\underline{x}}) I_{U,\underline{s}}(\underline{\underline{x}})^{k-1} \rangle_{r,\circ}, \end{aligned}$$

where, for $W = U$ or $W = V$, we define

$$I_{W,\underline{s}}(\underline{\underline{x}}) \equiv \frac{1}{|\delta' W_{r,\circ}|} \sum_{e \in \delta' W_{r\,\circ}} \mathbb{E}_f \biggl[\prod_{\ell=1}^{p} f_{s_\ell}(x^\ell_e) \biggr].$$

Summing over $p \geqslant 1$ gives that, on the event $\mathsf{COUP}_{\leqslant r}$,

$$\mathbb{E}_{r,\circ} \ln \frac{Z(\mathscr{G}_r)}{Z(\mathscr{G}_{r,\circ})} = \sum_{p \geqslant 1} \frac{1}{p} \sum_{\underline{s} \in S^p} b^p(\underline{s}) \mathbb{E}_{r,\circ} \langle I_{V,\underline{s}}(\underline{\underline{x}}) I_{U,\underline{s}}(\underline{\underline{x}})^{k-1} \rangle_{r,\circ} + \mathsf{err}_{r,1},$$

$$\text{where } |\mathsf{err}_{r,1}| \leqslant O_\epsilon(1) \min\biggl\{ \frac{k^2}{|\delta' U_{r,\circ}|}, 1 \biggr\}.$$

A similar comparison between $\mathscr{G}_{r+1}$ and $\mathscr{G}_{r,\circ}$ gives

$$\mathbb{E}_{r,\circ} \ln \frac{Z(\mathscr{G}_r)}{Z(\mathscr{G}_{r,\circ})} = \sum_{p \geqslant 1} \frac{1}{p} \mathbb{E}_{r,\circ} \biggl[\sum_{\underline{s} \in S^p} b^p(\underline{s}) \biggl\langle \frac{k-1}{k} I_{U,\underline{s}}(\underline{\underline{x}})^k + \frac{1}{k} I_{V,\underline{s}}(\underline{\underline{x}})^k \biggr\rangle_{r,\circ} \biggr] + \mathsf{err}_{r,2},$$

$$|\mathsf{err}_{r,2}| \leqslant O_\epsilon(1) \min\biggl\{ \frac{k^2}{\min\{|\delta' U_{r,\circ}|, |\delta' V_{r,\circ}|\}}, 1 \biggr\}.$$

We now argue that the sum of the error terms $\mathsf{err}_{r,1}, \mathsf{err}_{r,2}$, over $0 \leqslant r \leqslant nd - 1$, is small in expectation. First note that for a constant $C = C(k, \epsilon)$,

$$\sum_{r=0}^{nd-1} \mathbb{E}[\mathsf{err}_{r,1} + \mathsf{err}_{r,2}] \leqslant Cn \biggl[n^{-2/3} + \mathbb{P}\Bigl(\min\{|\delta' V_{r,\circ}|, |\delta' V_{r,\circ}|\} \leqslant n^{2/3} \text{ for some } r \leqslant nd \Bigr) \biggr].$$

The process $(|\delta' V_{r,\circ}|)_{r \geqslant 0}$ 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, $(|\delta' U|_{r,\circ})_{r \geqslant 0}$ 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

$$\mathbb{P}(|\delta' V_{r,\circ}| \leqslant |\delta' V_{0,\circ}| - n^{2/3}) + \mathbb{P}(|\delta' U_{r,\circ}| \leqslant |\delta' U_{0,\circ}| - n^{2/3}) \leqslant \exp\{-cn^{1/3}\}$$

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)

$$\sum_{r=0}^{nd-1} \mathbb{E}[\mathsf{err}_{r,1} + \mathsf{err}_{r,2}] \leqslant Cn^{1/3}.$$

Altogether this gives

$$\begin{aligned}&\mathbb{E} \ln \frac{Z(\mathscr{G}_0)}{Z(\mathscr{G}_{nd})} - O_\epsilon(n^{1/3}) \\ &= \sum_{r=0}^{nd-1} \sum_{p \geqslant 1} \frac{1}{p} \sum_{\underline{s}} b^p(\underline{s}) \mathbb{E}_{r,\circ} \left\langle I_{V,\underline{s}}(\underline{x}) I_{U,\underline{s}}(\underline{\underline{x}})^{k-1} - \frac{k-1}{k} I_{U,\underline{s}}(\underline{\underline{x}})^{k-1} - \frac{1}{k} I_{V,\underline{s}}(\underline{\underline{x}})^{k-1} \right\rangle_{r,\circ}.\end{aligned}$$

Using the fact that $x^k - kxy^{k-1} + (k-1)y^k \geqslant 0$ for all $x, y \in \mathbb{R}$ and even $k \geqslant 2$, or $x, y \geqslant 0$ and odd $k \geqslant 3$ finishes the proof. □

Corollary 8.7. *In the setting of Lemma 8.6,*

$$\mathbb{E} \ln Z(\mathscr{G}_{-1}) \geqslant \mathbb{E} \ln Z(\mathscr{G}_{nd+1}) - O_\epsilon(n^{2/3}),$$

where the expectation $\mathbb{E}$ is over the sequence of random graphs $(\mathscr{G}_r)_{-1 \leqslant r \leqslant nd+1}$.

Proof. Adding or removing a clause can change the partition function by at most a multiplicative constant (depending on ϵ). 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 8.6 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 8.6 proves the claim. □

Proof of Theorem 8.5. 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 8.7. □

8.3. **1RSB bound.** For the proof of Theorem 8.3, 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, γ_u) where γ goes over the positive integers. Let $\underline{\gamma} \equiv (\gamma_u)_u$. Next, instead of labeling u with (h_u, η_u, ρ_u) as before, we now label it with $(h_u, \eta_u, (\rho_u^\gamma)_{\gamma \geqslant 1})$ where $(\rho_u^\gamma)_{\gamma \geqslant 1}$ is an infinite sequence of i.i.d. samples from η_u. Lastly, we join all variables in U to a new clause a_* (Figure 5), which is labelled with the function

$$\varphi_{a*}(\underline{\gamma}) = \sum_{\gamma \geqslant 1} z_\gamma \prod_{u \in U} \mathbf{1}\{\gamma_u = \gamma\}$$

for some sequence of (random) weights $(z_\gamma)_{\gamma \geqslant 1}$. Let $\mathscr{H}_{-1}$ denote the resulting graph.

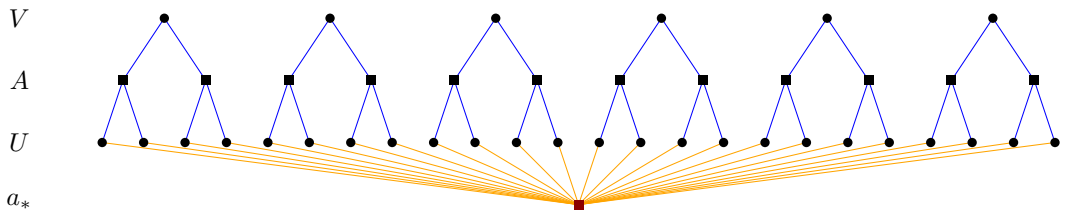


FIGURE 5. $\mathscr{H}_{-1}$

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 φ_{a*}, the support of $\mu_{\mathscr{H}_{-1}}$ contains only those configurations where all the γ_u share a common value γ, 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 $\mathscr{H}_{-1}, \dots, \mathscr{H}_{nd+1}$ precisely as in the proof of Theorem 8.5, 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 $0 \leqslant r \leqslant nd + 1$,

$$Z(\mathscr{H}_r) = \sum_\gamma z_\gamma Z_\gamma(\mathscr{G}_r).$$

The proofs of Lemma 8.6 and Corollary 8.7 carry over to this setting with essentially no changes, giving

Corollary 8.8. *Under the assumptions above,*

$$\mathbb{E} \ln Z(\mathscr{H}_{-1}) \geqslant \mathbb{E} \ln Z(\mathscr{H}_{nd+1}) - O_\epsilon(n^{2/3}),$$

where the expectation $\mathbb{E}$ is over the sequence of random graphs $(\mathscr{H}_r)_{-1 \leqslant r \leqslant nd+1}$.

The result of Corollary 8.8 applies for any choice of $(z_\gamma)_{\gamma \geqslant 1}$. Let us now take $(z_\gamma)_{\gamma \geqslant 1}$ to be a Poisson–Dirichlet process with parameter $\lambda \in (0, 1)$.[4] The process has the following invariance property (see e.g. [Pan13, Ch. 2]):

[4]That is to say, let $(w_\gamma)_{\gamma \geqslant 1}$ 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_γ are arranged in decreasing order, and write $z_\gamma \equiv w_\gamma / W$. Then $(z_\gamma)_{\gamma \geqslant 1}$ is distributed as a Poisson–Dirichlet process with parameter λ.

Proposition 8.9. *Let $(z_\gamma)_{\gamma\geqslant 1}$ be a Poisson–Dirichlet process with parameter $\lambda\in(0,1)$. Independently, let $(\xi_\gamma)_{\gamma\geqslant 1}$ be a sequence of i.i.d. positive random variables with finite second moment. Then the two sequences $(z_\gamma\xi_\gamma)_{\gamma\geqslant 1}$ and $(z_\gamma(\mathbb{E}\xi_1^\lambda)^{1/\lambda})_{\gamma\geqslant 1}$ have the same distribution, and consequently*

$$\mathbb{E}\ln\sum_{\gamma\geqslant 1} z_\gamma\xi_\gamma = \frac{1}{\lambda}\ln\mathbb{E}\xi^\lambda.$$

Proof of Theorem 8.3. Consider

$$\underline{Z}(\gamma)\equiv(Z_\gamma(\mathscr{G}_r))_{-1\leqslant r\leqslant nd+1}.$$

If we condition on everything else except for the ρ's, then $(\underline{Z}(\gamma))_{\gamma\geqslant 1}$ is an i.i.d. sequence indexed by γ. Let $\mathbb{E}_{z,\rho}$ denote expectation over the z's and ρ's, conditioned on all else: then applying Proposition 8.9 gives

$$n^{-1}\mathbb{E}\ln Z(\mathscr{H}_{-1}) = (n\lambda)^{-1}\mathbb{E}\ln\mathbb{E}_{z,\rho}[Z(\mathscr{G}_{-1})^\lambda] = \lambda^{-1}\mathbb{E}\ln\mathbb{E}_{z,\rho}\Big[\Big(\sum_{x\in\mathcal{X}}h(x)\prod_{a=1}^d \boldsymbol{u}_a(x)\Big)^\lambda\Big],$$

$$n^{-1}\mathbb{E}\ln Z(\mathscr{H}_{nd+1}) = F_n + \lambda^{-1}\mathbb{E}\ln\mathbb{E}_{z,\rho}[(\boldsymbol{u}_0)^\lambda].$$

Combining with Corollary 8.8 proves the result. □

8.4. **Conclusion of upper bound.** We now apply Theorem 8.3 to prove the upper bound for the NAE-SAT model, Proposition 3.19. Following Example 8.1, let $F_n(\beta)\equiv n^{-1}\mathbb{E}\ln Z_n(\beta)$ be the expected free energy for NAE-SAT at inverse temperature β. (The expectation is with respect to the law of the random (d,k)-regular graph.)

Let $\dot\mu_\lambda$ be the fixed point specified by Proposition 1.2, and let $(\rho_{aj})_{a,j\geqslant 0}$ be an array of i.i.d. samples from $\dot\mu_\lambda$. For each $\rho=\rho_{aj}$ we can define a (random) measure on $\mathcal{X}=\{0,1\}$ by giving mass ρ to 1, and giving the remaining mass $1-\rho$ to 0. Let $\eta\equiv\eta_\lambda$ be the law of this measure, and let $\zeta\equiv\zeta_\lambda$ denote the Dirac mass at η (cf. Remark 8.4). Recall from Proposition 1.2 that ρ has the same distribution as $1-\rho$. Using this symmetry, the quantities $\boldsymbol{u}_0$ and $\boldsymbol{u}_a(x)$ in Theorem 8.3 are equidistributed with $\boldsymbol{v}_0$ and $\boldsymbol{v}_a(x)$ where

$$\boldsymbol{v}_0\equiv 1-(1-e^{-\beta})\Big\{\prod_{j=1}^k\rho_{0j}+\prod_{j=1}^k(1-\rho_{0j})\Big\},$$

$$(\boldsymbol{v}_a(0),\boldsymbol{v}_a(1))\equiv\Big(1-(1-e^{-\beta})\prod_{j=1}^{k-1}\rho_{0j},\,1-(1-e^{-\beta})\prod_{j=1}^{k-1}(1-\rho_{0j})\Big).$$

In the following calculation we will accumulate some error terms of size $O(e^{-\beta})$, which we will eventually take care of by sending $\beta\to\infty$. It is useful to recall that for any $a,b\geqslant 0$ and $\lambda\in[0,1]$ we have $(a+b)^\lambda\leqslant a^\lambda+b^\lambda$. It follows that for any $x\geqslant 0$ and any $\epsilon\in[-x,\infty)$,

$$|(x+\epsilon)^\lambda - x^\lambda|\leqslant|\epsilon|^\lambda. \tag{69}$$

(Note this bound is not useful for $\lambda=0$, but in that case $(x+\epsilon)^\lambda=1=x^\lambda$.)

Lemma 8.10. *Let $\dot\mu_\lambda$ be the fixed point of Proposition 1.2. With $\dot{\mathfrak{Z}}_\lambda$ and $\hat{\mathfrak{Z}}_\lambda$ as in (8),*

$$\mathbb{E}[(\boldsymbol{u}_0)^\lambda] = \hat{\mathfrak{Z}}_\lambda + O(e^{-\lambda\beta}),$$

$$\mathbb{E}\Big[\Big(\sum_{x\in\{0,1\}^k}\prod_{a=1}^d\boldsymbol{u}_a(x)\Big)^\lambda\Big] = (\hat{\mathfrak{Z}}_\lambda/\bar{\mathfrak{Z}}_\lambda)^d\,\dot{\mathfrak{Z}}_\lambda + O(e^{-\lambda\beta}).$$

Proof. We assume $\lambda \in (0,1]$, since for $\lambda = 0$ there is nothing to prove. It follows straightforwardly from the above definitions that

$$\mathbb{E}[(\boldsymbol{u}_0)^\lambda] = \mathbb{E}[(\boldsymbol{v}_0)^\lambda] = \hat{\mathfrak{Z}}_\lambda + O(e^{-\lambda\beta}),$$

where the $O(e^{-\lambda\beta})$ error is by an application of (69). Next, let

$$\boldsymbol{z}_a \equiv \boldsymbol{v}_a(0) + \boldsymbol{v}_a(1), \quad \boldsymbol{r}_a \equiv \boldsymbol{v}_a(1)/\boldsymbol{z}_a.$$

Recall from (7) the definition of the distributional recursion $\hat{\mathscr{R}}_\lambda : \dot{\mu}_\lambda \mapsto \hat{\mu}_\lambda$, and the associated normalizing constant $\hat{\mathscr{Z}}(\dot{\mu}_\lambda)$. For any continuous bounded function $f : [0,1]^d \to \mathbb{R}$,

$$\begin{aligned}&\int f(\boldsymbol{r}_1,\dots,\boldsymbol{r}_d)\Big(\prod_{a=1}^d \boldsymbol{z}_a\Big)^\lambda \prod_{a=1}^d \Big\{\prod_{j=1}^{k-1} \dot{\mu}_\lambda(d\rho_{aj})\Big\}\\ &= \hat{\mathscr{Z}}(\dot{\mu}_\lambda)^d \int f(\hat{\rho}_1,\dots,\hat{\rho}_d)\prod_{a=1}^d \hat{\mu}_\lambda(d\hat{\rho}_a) + O(e^{-\lambda\beta}).\end{aligned}$$

It follows from this that

$$\begin{aligned}&\mathbb{E}\Big[\Big(\sum_{x\in\{0,1\}^k}\prod_{a=1}^d \boldsymbol{u}_a(x)\Big)^\lambda\Big] + O(e^{-\lambda\beta})\\ &= \hat{\mathscr{Z}}(\dot{\mu}_\lambda)^d \int \Big(\prod_{a=1}^d \hat{\rho}_a + \prod_{a=1}^d (1-\hat{\rho}_a)\Big)^\lambda \prod_{a=1}^d \hat{\mu}_\lambda(d\hat{\rho}_a) = \hat{\mathscr{Z}}(\dot{\mu}_\lambda)^d \dot{\mathfrak{Z}}_\lambda.\end{aligned}$$

Finally, it is straightforward to check that for the fixed point $\dot{\mu}_\lambda$ we have

$$\hat{\mathscr{Z}}(\dot{\mu}_\lambda)\bar{\mathfrak{Z}}_\lambda = \hat{\mathfrak{Z}}_\lambda, \tag{70}$$

so the lemma follows. □

Proof of Proposition 3.19. Applying Lemma 8.10 to the bound of Theorem 8.3, we have

$$F_n(\beta) \leqslant \lambda^{-1}\Big(\ln \dot{\mathfrak{Z}}_\lambda + \alpha \ln \hat{\mathfrak{Z}}_\lambda - d\ln \bar{\mathfrak{Z}}_\lambda + O(e^{-\lambda\beta})\Big) = \lambda^{-1}\Big(\mathfrak{F}(\lambda) + O(e^{-\lambda\beta})\Big).$$

A standard argument gives that for any finite β, $n^{-1}\ln Z_n(\beta)$ is well-concentrated around its expected value $F_n(\beta)$.[5] Thus, for any fixed $\lambda \in (0,1]$ and $\epsilon > 0$, we can choose $\beta = \beta(\lambda,\epsilon)$ sufficiently large so that

$$\limsup_{n\to\infty} \mathbb{P}\Big((Z_n(\beta))^{1/n} \geqslant \exp\{(1+\epsilon)\lambda^{-1}\mathfrak{F}(\lambda)\}\Big) = 0.$$

Since $Z_n \leqslant Z_n(\beta)$ for any finite β, we conclude

$$\mathsf{f}(\alpha) \leqslant \inf\{\lambda^{-1}\mathfrak{F}(\lambda) : \lambda \in (0,1]\}.$$

For $\alpha < \alpha_{\mathrm{sat}}$, if $\lambda = \lambda_\star \in (0,1)$ then $\lambda^{-1}\mathfrak{F}(\lambda) \leqslant s_\star = \mathsf{f}^{\mathrm{1RSB}}(\alpha)$. If instead $\lambda = \lambda_\star = 1$ then again $\lambda^{-1}\mathfrak{F}(\lambda) = s_\star + \Sigma(s_\star) = \mathsf{f}^{\mathrm{1RSB}}(\alpha)$. In any case this proves $\mathsf{f}(\alpha) \leqslant \mathsf{f}^{\mathrm{1RSB}}(\alpha)$. □

9. Contraction estimates

In this section we prove Propositions 4.2 and 4.3, as well as Lemma 4.4.

[5] Take the Doob martingale of $\ln Z_n(\beta)$ with respect to the clause-revealing filtration for the random NAE-SAT instance, then apply the Azuma–Hoeffding concentration bound.

9.1. **Single-copy coloring recursions.** We first analyze the BP recursions for the single-copy coloring model, and prove Proposition 4.2. We first consider the BP recursion with fixed parameters $\lambda \in [0,1]$ and $1 \leqslant T \leqslant \infty$. Recall that we have restricted our attention to measures $\dot{Q}, \hat{Q}$ such that

$$\dot{Q}(\sigma) \cong \dot{q}(\dot{\sigma})\mathbf{1}\{|\sigma| \leqslant T\},$$
$$\hat{Q}(\sigma) \cong \hat{q}(\hat{\sigma})\mathbf{1}\{|\sigma| \leqslant T\}$$

for some probability measures $\dot{q}, \hat{q}$ defined on $\dot{\Omega}_T, \hat{\Omega}_T$. Recall further that we can assume $\dot{q} = \dot{q}^{\mathrm{avg}}$ and $\hat{q} = \hat{q}^{\mathrm{avg}}$. For measures of this type we can give a fairly explicit description of the BP recursion. In what follows it will be convenient to take the convention

$$\dot{m}(\mathtt{r}_1) = \hat{m}(\mathtt{b}_1) = 1, \quad \dot{m}(\mathtt{r}_0) = \hat{m}(\mathtt{b}_0) = 0. \tag{71}$$

For $x \in \{\mathtt{0}, \mathtt{1}\}$ we abbreviate

$$\mathtt{g} \equiv \mathtt{b} \cup \mathtt{f}, \ \mathtt{g}_x \equiv \mathtt{b}_x \cup \mathtt{f}, \ \mathtt{y} \equiv \mathtt{r} \cup \mathtt{f}, \ \mathtt{p}_x \equiv \mathtt{b}_x \cup \mathtt{r}_x.$$

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

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

where $\cong$ indicates the normalization which makes $\dot{\mathtt{BP}}\hat{q}$ a probability measure on $\dot{\Omega}_T$.

For the clause BP recursion, by symmetry it suffices to consider a clause a with all incident edge literals $\mathtt{L}_{aj} = \mathtt{0}$. We 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 there is a valid coloring $\underline{\sigma}$ of δa with

$$\underline{\sigma} = ((\dot{\sigma}, \hat{\sigma}), (\dot{\sigma}_2, \hat{\sigma}_2), \ldots (\dot{\sigma}_k, \hat{\sigma}_k)) \in (\Omega_T)^k. \tag{72}$$

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

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

where the last line uses the convention (71). Recall that $\mathtt{BP} \equiv \dot{\mathtt{BP}} \circ \hat{\mathtt{BP}} \equiv \mathtt{BP}_{\lambda,T}$. We will show the following contraction result.

Proposition 9.1. *Suppose $\dot{q}_1, \dot{q}_2$ belong to $\mathbf{\Gamma}$, as defined by (45). Let $\mathtt{BP} \equiv \mathtt{BP}_{\lambda,T}$ for $\lambda \in [0,1]$ and $1 \leqslant T \leqslant \infty$. Then $\mathtt{BP}\dot{q}_1, \mathtt{BP}\dot{q}_2 \in \mathbf{\Gamma}$ and $\|\mathtt{BP}\dot{q}_1 - \mathtt{BP}\dot{q}_2\|_1 = O(k^2/2^k)\|\dot{q}_1 - \dot{q}_2\|_1$.*

Before the proof of Proposition 9.1 we deduce the following consequences:

Proof of Proposition 4.2. Let $\dot{q}^{(0)}$ be the uniform measure on $\{\mathtt{b}_0, \mathtt{b}_1, \mathtt{r}_1, \mathtt{r}_0\}$, and recursively define $\dot{q}^{(l)} \equiv \mathtt{BP}(\dot{q}^{(l-1)})$. It is clear that $\dot{q}^{(0)} \in \mathbf{\Gamma}$, so Proposition 9.1 implies $\dot{q}^{(l)} \in \mathbf{\Gamma}$ for all $l \geqslant 1$, and furthermore that $(\dot{q}^{(l)})_{l\geqslant 1}$ forms an ℓ^1 Cauchy sequence. By completeness of ℓ^1 we conclude that there exists $\dot{q}^{(\infty)} = \dot{q}_\star \in \mathbf{\Gamma}$ satisfying

$$\lim_{l\to\infty} \|\dot{q}^{(l)} - \dot{q}_\star\|_1 = 0, \quad \mathtt{BP}\dot{q}_\star = \dot{q}_\star.$$

Applying Proposition 9.1 again gives $\|\mathtt{BP}\dot{q} - \dot{q}_\star\|_1 = O(k^2/2^k)\|\dot{q} - \dot{q}_\star\|_1$ for any $\dot{q} \in \mathbf{\Gamma}$, from which it follows that $\dot{q}_\star$ is the unique fixed point of $\mathtt{BP}$ in $\mathbf{\Gamma}$. □

Corollary 9.2. *For $\lambda \in [0,1]$ and $1 \leqslant T \leqslant \infty$, let $\dot{q}_{\lambda,T}$ be the fixed point of $\mathtt{BP}_{\lambda,T}$ given by Proposition 4.2. Then $\|\dot{q}_{\lambda,T} - \dot{q}_{\lambda,\infty}\|_1 \to 0$ in the limit $T \to \infty$.*

Proof. For each $1 \leqslant T \leqslant \infty$, let $(\dot{q}_{\lambda,T})^{(l)}$ $(l \geqslant 0)$ be defined in the same way as $\dot{q}^{(l)}$ in the proof of Proposition 9.1. It follows from the definition that $(\dot{q}_{\lambda,T})^{(l)} = (\dot{q}_{\lambda,\infty})^{(l)}$ for all $l \leqslant l_T$, where $l_T \equiv \ln T / \ln(dk)$. By the triangle inequality and Proposition 4.2,

$$\|\dot{q}_{\lambda,T} - \dot{q}_{\lambda,\infty}\|_1 \leqslant \|\dot{q}_{\lambda,T} - (\dot{q}_{\lambda,\infty})^{(l_T)}\|_1 + \|(\dot{q}_{\lambda,\infty})^{(l_T)} - \dot{q}_{\lambda,\infty}\|_1 \leqslant (C/2^k)^{l_T}$$

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

We now turn to the proof of Proposition 9.1. We work with the non-normalized BP recursions $\mathtt{N\dot{B}P} \equiv \mathtt{N\dot{B}P}_{\lambda,T}$ and $\mathtt{N\hat{B}P} \equiv \mathtt{N\hat{B}P}_{\lambda,T}$, 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 $\mathtt{N\dot{B}P}, \mathtt{N\hat{B}P}$ via

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

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

$$\dot{p}(\dot{\sigma}) \equiv [\dot{p}(\dot{q})](\dot{\sigma}) \equiv \frac{\dot{q}(\dot{\sigma})}{1 - \dot{q}(\mathtt{r})}. \tag{73}$$

In the above we have assumed that the inputs to $\dot{\mathtt{BP}}, \hat{\mathtt{BP}}, \mathtt{N\dot{B}P}, \mathtt{N\hat{B}P}$ are probability measures; we now extend them in the obvious manner to non-negative 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 Δr as a non-negative measure on $\mathcal{X}$: for any subset $S \subseteq \mathcal{X}$,

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

where the inequality may be strict. For any non-negative 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 $\mathtt{N\hat{B}P}$ and $\mathtt{N\dot{B}P}$ be the non-normalized single-copy BP recursions at parameters λ, T. Starting from $\dot{q}_i \in \mathbf{\Gamma}$ $(i = 1, 2)$, denote

$$\dot{p}_i \equiv \dot{p}(\dot{q}_i) \text{ (as defined by (73))},$$
$$\hat{p}_i \equiv \mathtt{N\hat{B}P}(\dot{p}_i) \text{ and } \hat{p}_{i,\infty} \equiv \mathtt{N\hat{B}P}_{\lambda,\infty}(\dot{p}_i),$$
$$\dot{p}_i^{\mathrm{u}} \equiv \mathtt{N\dot{B}P}(\hat{p}_i) \text{ and } \tilde{q}_i \equiv \dot{\mathtt{BP}}\hat{p}_i = \mathtt{BP}\dot{q}_i.$$

With this notation in mind, the proof of Proposition 9.1 is divided into four lemmas.

Lemma 9.3 (effect of reweighting)**.** *Assuming $\dot{q}_1, \dot{q}_2 \in \mathbf{\Gamma}$, $\|\Delta\dot{p}\|_1 = O(1)\|\dot{q}_1 - \dot{q}_2\|_1$, where $O(1)$ indicates a constant depending on the constant appearing in (45).*

Lemma 9.4 (clause BP)**.** *Assuming $\dot{q}_1, \dot{q}_2 \in \mathbf{\Gamma}$,*

$$\begin{aligned} \hat{m}^\lambda \hat{p}_i(\square) &= 1 - 4/2^k + O(k/4^k), \\ \hat{m}^\lambda \hat{p}_i(\mathtt{f}) &= \hat{m}^\lambda \hat{p}_i(\square) + O(k/4^k), \\ \hat{m}^\lambda \hat{p}_i(\mathtt{b}_1) &= 1 + O(k/2^k), \\ \hat{m}^\lambda \hat{p}_i(\mathtt{r}_1) &= (2/2^k)[1 + O(k/2^k)]. \end{aligned} \tag{74}$$

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

$$\begin{aligned} \Delta \hat{m}^\lambda \hat{p}(\mathtt{f}) + \Delta \hat{m}^\lambda \hat{p}(\mathtt{r}) &= O(k/2^k) \Delta \dot{p}(\mathtt{f}), \\ \|\Delta \hat{m}^\lambda \hat{p}\|_1 &= O(k^2/2^k) \|\Delta \dot{p}\|_1. \end{aligned} \tag{75}$$

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

Lemma 9.5 (variable BP, non-normalized)**.** *Assuming $\dot{q}_1, \dot{q}_2 \in \mathbf{\Gamma}$,*

$$\begin{bmatrix} \dot{p}_i^{\mathrm{u}}(\mathtt{f}) \\ \dot{p}_i^{\mathrm{u}}(\mathtt{r}) \end{bmatrix} = \begin{bmatrix} O(2^{-k}) \\ 1 + O(2^{-k}) \end{bmatrix} \dot{p}_i^{\mathrm{u}}(\mathtt{b}), \qquad \begin{bmatrix} \Delta \dot{p}^{\mathrm{u}}(\mathtt{f}) \\ \Delta \dot{p}^{\mathrm{u}}(\mathtt{b}) \\ \Delta \dot{p}^{\mathrm{u}}(\mathtt{r}) \end{bmatrix} = \begin{bmatrix} O(k) \\ O(k2^k) \\ O(k2^k) \end{bmatrix} \|\Delta \hat{m}^\lambda \hat{p}\|_1 \max_{i=1,2} \Big\{ \dot{p}_i^{\mathrm{u}}(\mathtt{b}) \Big\}. \tag{76}$$

Lemma 9.6 (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 $\|\tilde{q}_1 - \tilde{q}_2\|_1 \lesssim k \|\Delta \hat{m}^\lambda \hat{p}\|_1$.*

Proof of Proposition 9.1. Follows by combining the four preceding lemmas 9.3–9.6. □

We now prove the four lemmas.

Proof of Lemma 9.3. This follows from the elementary identity

$$\frac{a_1}{b_1} - \frac{a_2}{b_2} = \frac{1}{b_1}(a_1 - a_2) + \frac{b_2 - b_1}{b_1 b_2} a_2. \tag{77}$$

together with (45). □

In the proof of the next two lemmas, the following elementary fact will be used repeatedly: suppose for $1 \leqslant l \leqslant m$ that we have non-negative 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

$$\begin{aligned} \sum_{\underline{x} \in \underline{\mathcal{X}}} \Big| \prod_{l=1}^m a^l(x^l) - \prod_{l=1}^m b^l(x^l) \Big| &\leqslant \sum_{l=1}^m \sum_{\underline{x} \in \underline{\mathcal{X}}} \Big\{ \prod_{1 \leqslant j < l} b^j(x^j) \Big\} \Big\{ \prod_{l < j \leqslant m} a^j(x^j) \Big\} \Big| a^l(x^l) - b^l(x^l) \Big| \\ &\leqslant \sum_{l=1}^m \|a^l - b^l\|_1 \prod_{j \neq l} \Big(\|a^j\|_1 + \|a^j - b^j\|_1 \Big). \end{aligned} \tag{78}$$

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

$$\sum_{x_1, \ldots, x_m \in \mathcal{X}} \Big| \prod_{i=1}^m a(x_i) - \prod_{i=1}^m b(x_i) \Big| \leqslant m \|a - b\|_1 \Big(\|a\|_1 + \|a - b\|_1 \Big)^{m-1}. \tag{79}$$

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

$$\mathtt{a}_x \equiv \Big\{ \hat{\sigma} \in \hat{\Omega}_T : \underline{\dot{\sigma}} \in (\mathtt{g}_x)^{k-1} \text{ for all } \underline{\dot{\sigma}} \sim \hat{\sigma} \Big\}. \tag{80}$$

Proof of Lemma 9.4. 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

$$\Delta\dot{p}(\mathtt{b}) \leqslant \Delta\dot{p}(\mathtt{f}) \leqslant \dot{p}_1(\mathtt{f}) + \dot{p}_2(\mathtt{f}) = O(2^{-k}).$$

Another consequence of the definition of $\mathbf{\Gamma}$ is that $\|\Delta\dot{p}\|_1 = O(1)$. We now control $\Delta\hat{m}^\lambda\hat{p}(\hat{\sigma})$, distinguishing a few cases:

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

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

   and it is easy to check that

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

   Note moreover that 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 (79) we have

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

   Recalling that $\dot{p}_1, \dot{p}_2 \in \mathbf{\Gamma}$, in the above we have $\dot{p}_1(\mathtt{g}_0) + \Delta\dot{p}(\mathtt{f}) \leqslant [1 + O(2^{-k})]/2$, 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 $\hat{\sigma} = \square$, for which we have $\hat{m}(\hat{\sigma}) = 1/2$ and $\hat{z}(\hat{\sigma}) = 2$. Thus

$$\hat{m}^\lambda\hat{p}(\square) = 1 - (\dot{p}(\mathtt{g}_0))^{k-1} - (\dot{p}(\mathtt{g}_1))^{k-1} + \dot{p}(\mathtt{f})^{k-1}. \tag{81}$$

   Arguing as above gives $\Delta\hat{m}^\lambda\hat{p}(\square) = O(k/2^k)\Delta\dot{p}(\mathtt{f})$, proving the first half of (75).
3. Lastly consider $\hat{\sigma} \in \{\mathtt{b}_0, \mathtt{b}_1\}$. Recalling (71) 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 $\mathtt{red}$ spin, then we also have $\underline{\dot{\sigma}} \sim \hat{\sigma}'$ for some $\hat{\sigma}' \in \{\mathtt{f}, \mathtt{r}\}$. Again making use of (71), 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

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

   The first term on the right-hand side captures the contribution from those $\underline{\dot{\sigma}}$ with no $\mathtt{red}$ spin, 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 $O(k^2/2^k)\|\Delta\dot{p}\|_1$, which finishes the second part of (75).

Combining the above estimates proves (75). We next prove (74). For this purpose we first introduce the notation $\mathtt{f}_{\geqslant s}$ which has the following meaning. In the context of clause-to-variable messages, it refers to the set of $\hat{\sigma} \in \hat{\Omega}_\mathtt{f}$ with $|\hat{\sigma}| \geqslant s$. In the context of variable-to-clause messages, it refers to the set of $\dot{\sigma} \in \dot{\Omega}_\mathtt{f}$ with $|\dot{\sigma}| \geqslant s$. In what follows we will write $\mathtt{f}_{\geqslant s} \subseteq \hat{\Omega}_\mathtt{f}$ or $\mathtt{f}_{\geqslant s} \subseteq \dot{\Omega}_\mathtt{f}$ to make the meaning clear. In particular $\mathtt{f}_{\geqslant 1} \subseteq \hat{\Omega}_\mathtt{f}$ is given by

$$\{\mathtt{f}\}\backslash\{\square\} \subseteq \mathtt{a}_0 \cup \mathtt{a}_1 \subseteq \hat{\Omega}.$$

Since $\dot{q}_i \in \mathbf{\Gamma}$, we must have from (45) that

$$\hat{m}^\lambda \hat{p}_i(\mathtt{f}_{\geqslant 1}) \leqslant 2\sum_{l=1}^{k-1} \binom{k-1}{l} \dot{p}_i(\mathtt{f})^l \dot{p}_i(\mathtt{b}_0)^{k-1-l} \leqslant 2\dot{p}_i(\mathtt{b}_0)^{k-1} \sum_{l=1}^{k-1} \left(\frac{k\dot{p}_i(\mathtt{f})}{\dot{p}_i(\mathtt{b}_0)}\right)^l = O(k/4^k). \tag{82}$$

On the other hand, we see from (81) that

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

If $\underline{\dot{\sigma}} \sim \mathtt{b}_1$ has no $\mathtt{red}$ spin, then there must exist some $\hat{\sigma} \in \hat{\Omega}_\mathtt{f}$ such that $\underline{\dot{\sigma}} \sim \hat{\sigma}$ as well. Conversely, if $\hat{\sigma} \in \hat{\Omega}_\mathtt{f} \cap \hat{\Omega}_T$ 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 $\mathtt{b}_1$.[6] It follows that

$$\begin{aligned} \hat{m}^\lambda \hat{p}_i(\mathtt{b}_1) = \hat{p}_i(\mathtt{b}_1) &= \hat{m}^\lambda \hat{p}_{i,\infty}(\mathtt{f}) + (k-1)\Big[\dot{p}_i(\mathtt{r}_0) - \dot{p}_i(\mathtt{g}_0)\Big]\dot{p}_i(\mathtt{b}_1)^{k-2} \\ &\leqslant \hat{m}^\lambda \hat{p}_{i,\infty}(\mathtt{f}) + (k-1)\dot{p}_i(\mathtt{r}_0)\dot{p}_i(\mathtt{b}_1)^{k-2} = 1 + O(k/2^k). \end{aligned} \tag{83}$$

For a lower bound it suffices to consider the contribution from clauses with all k incident edges colored $\mathtt{blue}$:

$$\hat{m}^\lambda \hat{p}_i(\mathtt{b}_1) = \hat{p}_i(\mathtt{b}_1) \geqslant \dot{p}_i(\mathtt{b})^{k-1}[1 - O(k/2^k)] = 1 - O(k/2^k). \tag{84}$$

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 (74). □

Proof of Lemma 9.5. We control $\dot{p}^{\mathrm{u}}$ and $\Delta\dot{p}^{\mathrm{u}}$ in two cases.

1. First consider $\dot{\sigma} \in \dot{\Omega}_\mathtt{f}$. Up to permutation there is a unique $\underline{\hat{\sigma}} \in (\hat{\Omega}_\mathtt{f})^{d-1}$ such that $\dot{\sigma} = \hat{\mathtt{T}}(\underline{\hat{\sigma}})$. 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

$$\prod_{j=2}^{d} \hat{m}(\hat{\sigma}_j)^\lambda \leqslant \dot{z}(\dot{\sigma})^\lambda \leqslant \prod_{j=2}^{d} \hat{m}(\hat{\sigma}_j)^\lambda + \prod_{j=2}^{d}(1 - \hat{m}(\hat{\sigma}_j))^\lambda, \tag{85}$$

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

$$\mathsf{comb}(\dot{\sigma}) \prod_{j=2}^{d} \hat{m}\hat{p}_i(\hat{\sigma}_j) \leqslant \dot{p}_i^{\mathrm{u}}(\dot{\sigma}) \leqslant \mathsf{comb}(\dot{\sigma}) \left\{ \prod_{j=2}^{d} \hat{m}^\lambda \hat{p}_i(\hat{\sigma}_j) + \prod_{j=2}^{d}(1-\hat{m})^\lambda \hat{p}_i(\hat{\sigma}_j) \right\}.$$

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

$$[\hat{m}^\lambda \hat{p}_i(\square)]^{d-1} \leqslant \dot{p}_i^{\mathrm{u}}(\mathtt{f}) \leqslant [\hat{m}^\lambda \hat{p}_i(\mathtt{f})]^{d-1} + [(1-\hat{m})^\lambda \hat{p}_i(\mathtt{f})]^{d-1} = 2[\hat{m}^\lambda \hat{p}_i(\mathtt{f})]^{d-1}. \tag{86}$$

   Making use of the symmetry together with (85) gives

$$\Delta\dot{p}^{\mathrm{u}}(\mathtt{f}) \leqslant 2 \sum_{\underline{\hat{\sigma}} \in (\hat{\Omega}_\mathtt{f})^{d-1}} \left| \prod_{j=2}^{d-1} \hat{m}^\lambda \hat{p}_1(\hat{\sigma}_j) - \prod_{j=2}^{d-1} \hat{m}^\lambda \hat{p}_2(\hat{\sigma}_j) \right|,$$

   and applying (79) gives

$$\Delta\dot{p}^{\mathrm{u}}(\mathtt{f}) \lesssim d\|\Delta\hat{m}^\lambda \hat{p}\|_1 \Big(\hat{m}^\lambda \hat{p}_1(\mathtt{f}) + \Delta\hat{m}^\lambda \hat{p}_1(\mathtt{f})\Big)^{d-2}.$$

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

Combining (74) with the lower bound from (85) then gives

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

2. Next consider $\dot{\sigma} \in \{\mathtt{red}, \mathtt{blue}\}$: note that $\dot{p}_i^{\mathrm{u}}(\mathtt{r}_x) = \hat{p}_i(\mathtt{p}_x)^{d-1}$, and

$$\frac{\dot{p}_i^{\mathrm{u}}(\mathtt{r}_x) - \dot{p}_i^{\mathrm{u}}(\mathtt{b}_x)}{\dot{p}_i^{\mathrm{u}}(\mathtt{r}_x)} = \frac{\hat{p}_i(\mathtt{b}_x)^{d-1}}{\hat{p}_i(\mathtt{p}_x)^{d-1}} = \left(1 - \frac{\hat{p}_i(\mathtt{r}_x)}{\hat{p}_i(\mathtt{p}_x)}\right)^{d-1} = O(2^{-k}), \tag{87}$$

where the last estimate uses (74) and $d/k = 2^{k-1}\ln 2 + O(1)$. Applying (79) gives

$$\Delta\dot{p}^{\mathrm{u}}(\mathtt{p}_1) \lesssim d\|\hat{m}^{\lambda}\hat{p}\|_1\Big(\min_{i=1,2}\Big\{\hat{m}^{\lambda}\hat{p}_i(\mathtt{p}_1)\Big\} + \Delta\hat{m}^{\lambda}\hat{p}(\mathtt{p}_1)\Big)^{d-2}.$$

Suppose without loss that $\hat{m}^{\lambda}\hat{p}_1(\mathtt{b}_1) \leqslant \hat{m}^{\lambda}\hat{p}_2(\mathtt{b}_1)$: then

$$\begin{aligned}\hat{m}^{\lambda}\hat{p}_1(\mathtt{p}_1) + \Delta\hat{m}^{\lambda}\hat{p}(\mathtt{p}_1) &= \hat{m}^{\lambda}\hat{p}_2(\mathtt{b}_1) + \hat{m}^{\lambda}\hat{p}_1(\mathtt{r}_1) + \Delta\hat{m}^{\lambda}\hat{p}(\mathtt{r}_1)\\ &\leqslant \hat{m}^{\lambda}\hat{p}_2(\mathtt{p}_1) + 2\Delta\hat{m}^{\lambda}\hat{p}(\mathtt{r}_1),\end{aligned}$$

and substituting into the above gives

$$\Delta\dot{p}^{\mathrm{u}}(\mathtt{p}_1) \lesssim d\|\hat{m}^{\lambda}\hat{p}\|_1\Big(\max_{i=1,2}\Big\{\hat{m}^{\lambda}\hat{p}_i(\mathtt{p}_1)\Big\} + \Delta\hat{m}^{\lambda}\hat{p}(\mathtt{r}_1)\Big)^{d-2}.$$

From (75) and the definition (45) 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 (87) that

$$\Delta\dot{p}^{\mathrm{u}}(\mathtt{p}_1) \lesssim d\|\Delta\hat{m}^{\lambda}\hat{p}\|_1 \max_{i=1,2}\Big\{\dot{p}_i^{\mathrm{u}}(\mathtt{b}_1)\Big\}. \tag{88}$$

It remains to show $\dot{p}^{\mathrm{u}}(\mathtt{f})/\dot{p}^{\mathrm{u}}(\mathtt{b}) = O(2^{-k})$. From (83),

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

and from the definition of $\mathbf{\Gamma}$ the right-hand side is $O(k/4^k)\dot{p}_i(\mathtt{b})^{k-1}$. Now recall from (84) that $\hat{m}^{\lambda}\hat{p}_i(\mathtt{b}_1) \gtrsim \dot{p}_i(\mathtt{b})^{k-1}$. Combining these gives

$$\hat{m}^{\lambda}\hat{p}_i(\mathtt{f}) \leqslant [1 + O(k/4^k)]\hat{m}^{\lambda}\hat{p}_i(\mathtt{b}_1). \tag{89}$$

Recalling (85), 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 (74). This concludes the proof. □

Proof of Lemma 9.6. 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 (45) follows from (76), 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 $\tilde{q}_i(\mathtt{R}) \equiv \tilde{q}_i(\mathtt{r}) - \tilde{q}_i(\mathtt{b})$ and

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

Let $\dot{p}_i^{\mathrm{u}}(\mathtt{R})$ and $\Delta\dot{p}^{\mathrm{u}}(\mathtt{R})$ be similarly defined. Arguing similarly as in the derivation of (88),

$$\Delta\dot{p}^{\mathrm{u}}(\mathtt{R}) = 2|\hat{p}_1(\mathtt{b}_1)^{d-1} - \hat{p}_2(\mathtt{b}_1)^{d-1}| \lesssim k\|\Delta\hat{m}^{\lambda}\hat{p}\|_1 \max_{i=1,2}\Big\{\dot{p}_i^{\mathrm{u}}(\mathtt{b})\Big\} \tag{90}$$

Recalling $\|\tilde{q}_i\|_1 = 1$, we have

$$\begin{aligned} 2\tilde{q}_i(\mathtt{r}) &= [1 - \tilde{q}_i(\mathtt{f})] + [\tilde{q}_i(\mathtt{r}) - \tilde{q}_i(\mathtt{b})] \text{ and} \\ 2\tilde{q}_i(\mathtt{b}) &= [1 - \tilde{q}_i(\mathtt{f})] - [\tilde{q}_i(\mathtt{r}) - \tilde{q}_i(\mathtt{b})], \text{ so} \\ \|\Delta\tilde{q}\|_1 &\lesssim \Delta\tilde{q}(\mathtt{f}) + \Delta\tilde{q}(\mathtt{R}). \end{aligned}$$

If we take $a \in \{1, 2\}$ and $b = 2 - a$, and write $\dot{Z}_i \equiv \|\dot{p}_i^{\mathrm{u}}\|_1$, then

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

If we take $a \in \arg\max_i \dot{p}_i^{\mathrm{u}}(\mathtt{b})$, then, by (76) and (90), the first term on the right-hand side is

$$\lesssim \frac{k\|\Delta\hat{m}^{\lambda}\hat{p}\|_1 \dot{p}_a^{\mathrm{u}}(\mathtt{b})}{\dot{Z}_a} \lesssim k\|\Delta\hat{m}^{\lambda}\hat{p}\|_1,$$

where the rightmost inequality uses $\dot{Z}_i \geqslant \dot{p}_i^{\mathrm{u}}(\mathtt{b})$. As for the second term, (76) gives

$$\frac{|\dot{Z}_a - \dot{Z}_b|}{\dot{Z}_a} \lesssim d\|\Delta\hat{m}^{\lambda}\hat{p}\|_1 \quad \text{and} \quad \frac{[\dot{p}_b^{\mathrm{u}}(\mathtt{f}) + \dot{p}_b^{\mathrm{u}}(\mathtt{r}) - \dot{p}_b^{\mathrm{u}}(\mathtt{b})]}{\dot{Z}_b} \lesssim 2^{-k}.$$

Combining these estimates yields the claimed bound. □

9.2. **Pair coloring recursions.** In this section we analyze the BP recursions for the pair coloring model and prove Proposition 4.3 and Lemma 4.4. Recall that we have restricted our attention to measures $\dot{Q}, \hat{Q}$ such that

$$\begin{aligned} \dot{Q}(\sigma^1, \sigma^2) &\cong \dot{q}(\dot{\sigma}^1, \dot{\sigma}^2)\mathbf{1}\{|\sigma^1|, |\sigma^2| \leqslant T\}, \\ \hat{Q}(\sigma^1, \sigma^2) &\cong \hat{q}(\hat{\sigma}^1, \hat{\sigma}^2)\mathbf{1}\{|\sigma^1|, |\sigma^2| \leqslant T\} \end{aligned}$$

for probability measures $\dot{q}, \hat{q}$ defined on $(\dot{\Omega}_T)^2, (\hat{\Omega}_T)^2$. Recall further that we assume $\dot{q} = \dot{q}^{\mathrm{avg}}$ and $\hat{q} = \hat{q}^{\mathrm{avg}}$. For any measure $p(x)$ defined on $x \equiv (x^1, x^2)$ in $(\dot{\Omega}_T)^2$ or $(\hat{\Omega}_T)^2$, define

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

Recall the definition (46) of $\mathbf{\Gamma}(c, \kappa)$. We will prove that

Proposition 9.7. *For any constant $c \in (0, 1]$ and probability measures $\dot{q}_1, \dot{q}_2 \in \mathbf{\Gamma}(c, 1)$, we have $\mathtt{BP}\dot{q}_1, \mathtt{BP}\dot{q}_2 \in \mathbf{\Gamma}(1, 1)$ and*

$$\|\mathtt{BP}\dot{q}_1 - \mathtt{BP}\dot{q}_2\|_1 = O(k^4/2^k)\|\dot{q}_1 - \dot{q}_2\|_1 + O(k^4/2^k) \sum_{i=1,2} \|\dot{q}_i - \mathtt{F}\dot{q}_i\|_1. \tag{91}$$

Assuming this result, it is straightforward to deduce Proposition 4.3:

Proof of Proposition 4.3. 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 $l \geqslant 1$. It is clear that $\dot{q}^{(0)} \in \mathbf{\Gamma}(1, 1)$ and $\dot{q}^{(0)} = \mathtt{F}\dot{q}^{(0)}$. Since $\dot{q}^{(l)} = \mathtt{F}\dot{q}^{(l)}$ for all $l \geqslant 1$, it follows from (91) that $(\dot{q}^{(l)})_{l \geqslant 1}$ forms an ℓ^1 Cauchy sequence. It follows by completeness of ℓ^1 that $\dot{q}^{(l)}$ converges to a limit $\dot{q}^{(\infty)} = \dot{q}_\star \in \mathbf{\Gamma}(1, 1)$, satisfying $\dot{q}_\star = \mathtt{F}\dot{q}_\star = \mathtt{BP}\dot{q}_\star$. This implies that for any probability measure $\dot{q}$,

$$\|\dot{q} - \mathtt{F}\dot{q}\|_1 \leqslant \|\dot{q} - \dot{q}_\star\|_1 + \|\dot{q}_\star - \mathtt{F}\dot{q}\|_1 = 2\|\dot{q} - \dot{q}_\star\|_1.$$

Applying (91) again gives

$$\|\mathtt{BP}\dot{q} - \dot{q}_\star\|_1 = O(k^4/2^k)\|\dot{q} - \dot{q}_\star\|_1 + O(k^4/2^k)\|\dot{q} - \mathtt{F}\dot{q}\|_1 = O(k^4/2^k)\|\dot{q} - \dot{q}_\star\|_1,$$

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

We now turn to the proof of Proposition 9.7. The proof of Lemma 4.4 is given after the proof of Proposition 9.7. Let $\dot{\mathtt{NBP}}, \hat{\mathtt{NBP}}$ now denote the non-normalized BP recursions for the pair model. Let $\dot{p} \equiv \dot{p}(\dot{q})$ be the reweighted measure

$$\dot{p}(\dot{\sigma}) \equiv \frac{\dot{q}(\dot{\sigma})}{1 - \dot{q}(\mathtt{r}[\dot{\sigma}] > 0)}. \tag{92}$$

Recalling convention (71), 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 $\hat{\mathtt{NBP}}$ and $\dot{\mathtt{NBP}}$ be the non-normalized pair BP recursions at parameters λ, T. Starting from $\dot{q}_i \in \mathbf{\Gamma}(c, \kappa)$ $(i = 1, 2)$, we denote

$$\begin{aligned} \dot{p}_i &\equiv \dot{p}(\dot{q}_i) \text{ (as defined by (92))}, \\ \hat{p}_i &\equiv \hat{\mathtt{NBP}}(\dot{p}_i) \text{ and } \hat{p}_{i,\infty} \equiv \hat{\mathtt{NBP}}_{\lambda,\infty}(\dot{p}_i), \\ \dot{p}_i^{\mathrm{u}} &\equiv \dot{\mathtt{NBP}}(\hat{p}_i) \text{ and } \tilde{q}_i \equiv \dot{\mathtt{BP}}\hat{p}_i = \mathtt{BP}\dot{q}_i. \end{aligned}$$

With this notation in mind, the proof of Proposition 9.7 is divided into the following lemmas.

Lemma 9.8 (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*

$$\begin{aligned} \|\Delta\dot{p}\|_1 &\equiv O(2^{2(1-\kappa)k})\|\Delta\dot{q}\|_1, \\ \|\dot{p}_i - \mathtt{F}\dot{p}_i\|_1 &\equiv O(2^{(1-\kappa)k})\|\dot{q}_i - \mathtt{F}\dot{q}_i\|_1. \end{aligned}$$

Lemma 9.9 (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*

$$\begin{aligned} \Delta\hat{m}^{\lambda}\hat{p}(\mathtt{yy}) &= O(k^3/2^k)\Delta\dot{p}(\mathtt{gg}) = O(k^3/2^{(1+c)k}), \\ \Delta\hat{m}^{\lambda}\hat{p}(\{\mathtt{br}, \mathtt{bf}_{\geqslant 1}\}) &= O(k^2/2^k)[\Delta\dot{p}(\mathtt{gg}) + 2^{-k}\Delta\dot{p}(\dot{\Omega}^2 \backslash \{\mathtt{rr}\})] = O(k^3/2^{(1+c)k}), \\ \|\Delta\hat{m}^{\lambda}\hat{p}\|_1 &= O(k^3/2^k)\|\Delta\dot{p}\|_1 = O(k^3 2^{(1-2\kappa)k}), \end{aligned} \tag{93}$$

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

$$\|\hat{m}^{\lambda}\hat{p}_i - \hat{m}^{\lambda}\mathtt{F}\hat{p}_i\|_1 = O(k^3/2^{(1+\kappa)k})\|\dot{p}_i - \mathtt{F}\dot{p}_i\|_1 = O(k^3/2^{2\kappa k})\|\dot{q}_i - \mathtt{F}\dot{q}_i\|_1. \tag{94}$$

Lemma 9.10 (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 $s, t \in \{\mathtt{r}_1, \mathtt{b}_1, \mathtt{f}, \square\}$ that*

$$\frac{\hat{m}^{\lambda}\hat{p}_i(s, t)}{(2/2^k)^{\mathtt{r}[s]+\mathtt{r}[t]}} = \begin{cases} 1 + O(k^2/2^k) & \mathtt{r}[s] + \mathtt{r}[t] \leqslant 1, \\ 1 + O(k^2/2^{ck}) & \mathtt{r}[s] + \mathtt{r}[t] = 2. \end{cases} \tag{95}$$

Furthermore we have the bounds

$$\begin{aligned} \hat{m}^{\lambda}\hat{p}_i(\mathtt{f}_{\geqslant 1}t) + \hat{m}^{\lambda}\hat{p}_i(t\mathtt{f}_{\geqslant 1}) &\leqslant O(k/4^k) \text{ for all } t \in \{\mathtt{r}_1, \mathtt{b}_1, \mathtt{f}, \square\}, \\ \hat{m}^{\lambda}\hat{p}_i(\{\mathtt{f}\} \times \hat{\Omega}) - \hat{m}^{\lambda}\hat{p}_i(\{\mathtt{b}_1\} \times \hat{\Omega}) &\leqslant O(k/4^k). \end{aligned} \tag{96}$$

The same estimates hold with $\mathtt{F}\hat{p}_i$ in place of $\hat{p}_i$.

Lemma 9.11 (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 we have $\mathtt{BP}\dot{q}_1, \mathtt{BP}\dot{q}_2 \in \mathbf{\Gamma}(c', 1)$ with $c' = \max\{0, 2\kappa - 1\}$, and*

$$\|\mathtt{BP}\dot{q}_1 - \mathtt{BP}\dot{q}_2\|_1 = O(k)\big(\|\Delta\hat{m}^{\lambda}\hat{p} + \Delta\hat{m}^{\lambda}\mathtt{F}\hat{p}\|_1\big) + O(k2^k)\sum_{i=1,2}\|\hat{m}^{\lambda}\hat{p}_i - \hat{m}^{\lambda}\mathtt{F}\hat{p}_i\|_1.$$

Proof of Proposition 9.7. Follows by combining the preceding lemmas 9.8–9.11. □

Proof of Lemma 4.4. If $\dot{q} \in \mathbf{\Gamma}(c, 0)$ is a fixed point of BP, then it follows from the preceeding lemmas 9.9–9.11 that $\dot{q} \in \mathbf{\Gamma}(c, 0) \cap \mathbf{\Gamma}(0, 1) = \mathbf{\Gamma}(c, 1)$. □

We now prove the three lemmas leading to Proposition 9.7.

Proof of Lemma 9.8. Applying (77) we have

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

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

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

Since $\dot{q}_i \in \mathbf{\Gamma}$, we have

$$\begin{aligned} \dot{p}_i(\dot{\Omega}^2\backslash\{\mathtt{rr}\}) &= O(1) && \text{by part (A) of (46)},\\ \text{and } \dot{p}_i(\mathtt{rr}) &= O(2^{(1-\kappa)k}) && \text{by part (B) of (46)}. \end{aligned} \tag{97}$$

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

Proof of Lemma 9.9. We will prove (93) for $\hat{p}_i$; the proof for $\mathtt{F}\hat{p}_i$ is entirely similar. It follows from the symmetry $\dot{p}_i = (\dot{p}_i)^{\mathrm{avg}}$ that for any $x, y \in \{\mathtt{0}, \mathtt{1}\}$,

$$\Big|\dot{p}_i(\mathtt{bb}) - 4\dot{p}_i(\mathtt{b}_x\mathtt{b}_y)\Big| = 2\Big|\dot{p}_i(\mathtt{b}_x\mathtt{b}_{y\oplus 1}) - \dot{p}_i(\mathtt{b}_x\mathtt{b}_y)\Big| = 2\Big|\dot{p}_i(\mathtt{b}_0\mathtt{b}_0) - \dot{p}_i(\mathtt{b}_0\mathtt{b}_1)\Big|,$$

from which we obtain that

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

Recall $\mathtt{g} = \{\mathtt{b}, \mathtt{f}\}$ and $\dot{p}_i(\mathtt{gg}) = 1$. Combining the above with (46) gives

$$\begin{aligned} \Delta\dot{p}(\mathtt{gg}) &\leqslant \Delta\dot{p}(\mathtt{bb}) + \Delta\dot{p}(\mathtt{gf}) + \Delta\dot{p}(\mathtt{fg})\\ &\leqslant \sum_{i=1,2}\left\{\Big|\dot{p}_i(\mathtt{b}_0\mathtt{b}_0) - \dot{p}_i(\mathtt{b}_0\mathtt{b}_1)\Big| + \dot{p}_i(\mathtt{gf}) + \dot{p}(\mathtt{fg})\right\} = O(2^{-ck}). \end{aligned} \tag{98}$$

Step I. We first control $\Delta\hat{m}^\lambda\hat{p}(\hat{\sigma})$. As before, 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 $\{\mathtt{y}\}\backslash\{\square\} \subseteq \mathtt{a}_0 \cup \mathtt{a}_1$ (as defined by (80)). Thus

$$\Delta\hat{m}^\lambda\hat{p}(\{\mathtt{yy}\}\backslash\{\square\square\}) \leqslant \sum_{x\in\{0,1\}}\left\{\Delta\hat{m}^\lambda\hat{p}(\mathtt{a}_x \times \{\mathtt{y}\}) + \Delta\hat{m}^\lambda\hat{p}(\{\mathtt{y}\} \times \mathtt{a}_x)\right\}. \tag{99}$$

Consider $\hat{\sigma} \in \mathtt{a}_x \times \{\mathtt{y}\}$: in order for $\underline{\dot{\sigma}} \in (\dot{\Omega}^2)^{k-1}$ to be compatible with $\hat{\sigma}$, it is necessary that $\dot{\sigma}_j \in A \equiv \{\mathtt{g}_x\} \times \{\mathtt{g}\}$ for all $2 \leqslant j \leqslant k$. Combining with (79) gives

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

It follows from (46) that $\dot{p}_1(A) + \Delta\dot{p}(\mathtt{gg}) = \frac{1}{2} + O(2^{-ck})$, so we conclude

$$\Delta\hat{m}^\lambda\hat{p}(\{\mathtt{yy}\}\backslash\{\square\square\}) = O(k/2^k)\Delta\dot{p}(\mathtt{gg}). \tag{100}$$

2. Now take $\hat{\sigma} = \square\square$: 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 $x \in \{0,1\}$ we define $A_x \equiv \{\mathtt{b}_x\mathtt{g}\} \cup \{\mathtt{fg}\}$ and $B_x \equiv \{\mathtt{gb}_x\} \cup \{\mathtt{gf}\}$. Therefore
$$\Delta\hat{m}^\lambda\hat{p}(\square\square) \leqslant \sum_{x\in\{0,1\}} \sum_{\underline{\dot{\sigma}}\in(A_x)^{k-1}\cup(B_x)^{k-1}} \Big|\prod_{j=2}^k \dot{p}_1(\dot{\sigma}_j) - \prod_{j=2}^k \dot{p}_2(\dot{\sigma}_j)\Big| = O(k/2^k)\Delta\dot{p}(\mathtt{gg}),$$
where the last estimate follows by the same argument that led to (100). This concludes the proof of the first line of (93).
3. Now consider $\hat{\sigma}$ with exactly one coordinate in $\{\mathtt{b}\}$, meaning the other must be in $\{\mathtt{y}\}$. Recalling convention (71), 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 $\underline{\dot{\sigma}}$ contains no $\mathtt{red}$ spin, 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
   $$\Delta\hat{m}^\lambda\hat{p}_\infty(\mathtt{yy}) = O(k/2^k)\Delta\dot{p}(\mathtt{gg}) \tag{101}$$
   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
   $$\leqslant (k-1) \sum_{\underline{\dot{\sigma}}\in A\times B^{k-2}} \Big|\prod_{j=2}^k \dot{p}_1(\dot{\sigma}_j) - \prod_{j=2}^k \dot{p}_2(\dot{\sigma}_j)\Big| = O(k^2/2^k)\|\Delta\dot{p}\|_1, \tag{102}$$
   where the last estimate follows using (78) and (97).

   Combining the above estimates (and using the symmetry between $\mathtt{b}_1\mathtt{y}$ and $\mathtt{yb}_1$) gives
   $$\Delta\hat{m}^\lambda\hat{p}(\mathtt{b}_1\mathtt{y}) + \Delta\hat{m}^\lambda\hat{p}(\mathtt{yb}_1) = O(k^2/2^k)\|\Delta\dot{p}\|_1. \tag{103}$$
   If we further assume $\hat{\sigma} \in \{\mathtt{b}_1\} \times \{\mathtt{r}, \mathtt{f}_{\geqslant 1}\}$, then, arguing as above, $\underline{\dot{\sigma}}$ either contributes to $\Delta\hat{m}^\lambda\hat{p}_\infty(\mathtt{y} \times \{\mathtt{r}, \mathtt{f}_{\geqslant 1}\})$, or else belongs to $A_x \times B_x^{k-2}$ for $A_x = \{\mathtt{r}_0\mathtt{g}_x\}$, $B_x = \{\mathtt{b}_1\mathtt{g}_x\}$ and $x \in \{0,1\}$. The contribution from first case is bounded by (100). The contribution from the second case, using (78) and (97), is
   $$\lesssim k\Delta\dot{p}(\dot{\Omega}^2\backslash\{\mathtt{rr}\})\Big(\max_{x\in\{0,1\}} \dot{p}_1(B_x) + \Delta\dot{p}(\mathtt{gg})\Big)^{k-2} = O(k^2/4^k)\Delta\dot{p}(\dot{\Omega}^2\backslash\{\mathtt{rr}\}).$$
   The second claim of (93) follows by combining these estimates and recalling (98).
4. Lastly we 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 $i \in \{1,2\}$, $\underline{\dot{\sigma}}^i$ contains no $\mathtt{red}$ spin. In this case $\underline{\dot{\sigma}}$ is also compatible with some $\hat{\sigma}' \in \{\mathtt{b}_1\mathtt{y}\} \cup \{\mathtt{yb}_1\}$, 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
   $$\Delta\hat{m}^\lambda\hat{p}_\infty(\mathtt{b}_1\mathtt{y}) + \Delta\hat{m}^\lambda\hat{p}_\infty(\mathtt{yb}_1) = O(k^2/2^k)\|\Delta\dot{p}\|_1, \tag{104}$$
   where the last step is by the same argument as for (103).

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

$$(k-1)\Big|\dot p_1(\mathtt{r}_0\mathtt{r}_0)\dot p_1(\mathtt{b}_1\mathtt{b}_1)^{k-2} - \dot p_2(\mathtt{r}_0\mathtt{r}_0)\dot p_2(\mathtt{b}_1\mathtt{b}_1)^{k-2}\Big|.$$

Using (78) and (46), this is at most

$$\begin{aligned} &O(k^2/4^k)\Big(\Delta\dot p(\mathtt{r}_0\mathtt{r}_0) + \dot p(\mathtt{r}_0\mathtt{r}_0)\cdot\Delta\dot p(\mathtt{b}_1\mathtt{b}_1)\Big) \\ &= O(k^2/4^k)\|\dot p\|_1\|\Delta\dot p\|_1 = O(k^2/2^{(1+\kappa)k})\|\Delta\dot p\|_1. \end{aligned} \tag{105}$$

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 $O(k^2/4^k)\|\Delta\dot p\|_1$ by another application of (78) and (46).

The above estimates together give

$$\Delta\hat m^\lambda \hat p(\mathtt{b}_1\mathtt{b}_1) = O(k^2/2^k)\|\Delta\dot p\|_1, \tag{106}$$

where the main contribution comes from (104). Combining with the previous bound (103) yields the last part of (93).

Step II. Next we prove (94) by improving the preceding bounds in the special case that $\dot p_1 = \dot p$ and $\dot p_2 \equiv \mathtt{F}\dot p$. Recall $\hat p_i \equiv \hat{\mathtt{NBP}}(\dot p_i)$; it follows that $\hat p_2 = \mathtt{F}\hat p_1$. Thus, for any $\hat\sigma \in \hat\Omega^2$ with $\hat\sigma^2 = \square$, we have $\hat\sigma = \mathtt{F}\hat\sigma$, so $\hat p_2(\hat\sigma) = \hat p_1(\mathtt{F}\hat\sigma) = \hat p_1(\hat\sigma)$. For $\hat\sigma \in \hat\Omega^2$ with $\hat\sigma^1 = \square$, we have $\hat\sigma = (\mathtt{F}\hat\sigma)\oplus\mathtt{1}$, so $\hat p_2(\hat\sigma) = \hat p_1(\mathtt{F}\hat\sigma) = \hat p_1(\hat\sigma)$, where the last step uses that $\hat p_1 = (\hat p_1)^{\mathrm{avg}}$. It follows that instead of (99) and (101) we have the improved bound

$$\begin{aligned} \Delta\hat m^\lambda\hat p_\infty(\mathtt{yy}) &= \Delta\hat m^\lambda\hat p_\infty(\{\mathtt{yy}\}\backslash(\{\square\mathtt{y}\}\cup\{\mathtt{y}\square\})) \leqslant \sum_{x\in\{0,1\}}\Delta\hat m^\lambda\hat p_\infty(\mathtt{a}_x\times\mathtt{a}_y) \\ &= O(k)\|\Delta\dot p\|_1\sum_{x\in\{0,1\}}\Big(\dot p_1(\mathtt{g}_x,\mathtt{g}_y) + \Delta\dot p(\mathtt{gg})\Big)^{k-2} = O(k/4^k)\|\dot p - \mathtt{F}\dot p\|_1. \end{aligned}$$

Similarly, instead of (102) 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 $A_x = \{\mathtt{r}_0\mathtt{g}_x\}$ and $B_x = \{\mathtt{b}_1\mathtt{g}_x\}$. It follows that instead of (103) and (104) we have the improved bound

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

Previously the main contribution in (106) came from (104), but now it comes instead from (105). 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 (94). The second part follows by applying Lemma 9.8. □

Proof of Lemma 9.10. We first prove (95). Assume $s,t\in\{\mathtt{b}_1,\mathtt{f},\square\}$, and write $st \equiv s\times t\subseteq\hat\Omega^2$. Then for a lower bound we have

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

for an upper bound we have

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

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

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

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 (95), and we turn next to the proof of (96). Arguing similarly as for (82) gives

$$\hat{m}^\lambda \hat{p}_i(\{\mathtt{ff}\} \backslash \{\square\square\}) \leqslant \hat{m}^\lambda \hat{p}_i(\mathtt{f}_{\geqslant 1}\mathtt{f}) + \hat{m}^\lambda \hat{p}_i(\mathtt{ff}_{\geqslant 1}) = O(k/4^k).$$

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

$$\hat{m}^\lambda \hat{p}_i(\mathtt{b}_1\mathtt{f}_{\geqslant 1}) - \hat{m}^\lambda \hat{p}_{i,\infty}(\mathtt{ff}_{\geqslant 1})$$
$$\leqslant \sum_{y \in \{0,1\}} \Big[k\dot{p}_i(\mathtt{r}_0\mathtt{f})\dot{p}_i(\mathtt{b}_1\mathtt{g}_y)^{k-2} + k^2\dot{p}_i(\mathtt{r}_0\mathtt{b}_y)\dot{p}_i(\mathtt{b}_1\mathtt{f})\dot{p}_i(\mathtt{b}_1\mathtt{g}_y)^{k-3} \Big],$$

and applying (46) this is $O(k/4^k)$. Finally,

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

which proves the first part of (96). For the second part, arguing similarly as for (89), we have for any $\xi \in \hat{\Omega}$ that

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

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

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

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

$$\dot{p}_i(\mathtt{g}_0\zeta) - \dot{p}_i(\mathtt{r}_0\zeta) \leqslant O(2^{-k})\dot{p}_i(\mathtt{b}_0\zeta) + \dot{p}_i(\mathtt{f}\zeta) \leqslant \begin{cases} O(1) & \zeta = \mathtt{r}_0, \\ O(2^{-k}) & \text{if } \zeta \in \dot{\Omega} \backslash \{\mathtt{r}_0\}. \end{cases}$$

Combining these estimates and summing over ξ proves the second part of (96). □

An immediate application of (95), which will be useful in the next proof, is that

$$\frac{\hat{m}^\lambda \hat{p}_i(\mathtt{r}_x t)}{\hat{m}^\lambda \hat{p}_i(\mathtt{b}_x t)} \geqslant [1 + O(k^2/2^k)] \cdot (2/2^k). \tag{107}$$

for all $t \in \{\mathtt{b}_0, \mathtt{b}_1, \mathtt{f}, \square\}$.

Proof of Lemma 9.11. We divide the proof in two parts.

Step I. Non-normalized messages.

1. First consider $\dot\sigma \in \{\mathtt{ff}\}$. Recalling $(a+b)^\lambda \leqslant a^\lambda + b^\lambda$ for $a, b \geqslant 0$ and $\lambda \in [0,1]$,
$$\Delta\dot p^{\mathrm{u}}(\mathtt{ff}) \leqslant 2 \sum_{\hat r \in \{\hat p, \mathtt{F}\hat p\}} \sum_{\underline{\hat\sigma} \in \{\mathtt{ff}\}^{k-1}} \left| \prod_{j=2}^{d} \hat m^\lambda \hat r_1(\hat\sigma_j) - \prod_{j=2}^{d} \hat m^\lambda \hat r_2(\hat\sigma_j) \right|$$
where the $\hat r = \mathtt{F}\hat p$ term arises from the fact that
$$\hat m(\hat\sigma^1)^\lambda [1 - \hat m(\hat\sigma^2)]^\lambda \hat p(\hat\sigma) = \hat m(\hat\sigma^1)^\lambda \hat m(\hat\sigma^2 \oplus \mathtt{1})^\lambda (\mathtt{F}\hat p)(\mathtt{F}\hat\sigma) = \hat m^\lambda \mathtt{F}\hat p(\mathtt{F}\hat\sigma).$$
Applying (79) 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 (93) and (95) 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
$$\Delta\dot p^{\mathrm{u}}(\mathtt{ff}) = O(d) \|\Delta\hat m^\lambda \hat p + \Delta\hat m^\lambda \mathtt{F}\hat p\|_1 \cdot \dot p_1^{\mathrm{u}}(\mathtt{ff}). \tag{108}$$
2. Next consider $\dot\sigma \in \{\mathtt{p_1 f}\}$. Let $\hat r_{\max}(\hat\sigma) \equiv \max_{i=1,2} \hat r_i(\hat\sigma)$ — in this notation,
$$\hat r_{\max}(\hat\Omega) = \sum_{\hat\sigma \in \hat\Omega} \max_{i=1,2} \hat r_i(\hat\sigma) \geqslant \max_{i=1,2} \sum_{\hat\sigma \in \hat\Omega} \hat r_i(\hat\sigma) = \max_{i=1,2} \hat r_i(\hat\Omega)$$
where the inequality may be strict. Then
$$\Delta\dot p^{\mathrm{u}}(\mathtt{p_1 f}) = O(d) \sum_{\hat r \in \{\hat p, \mathtt{F}\hat p\}} \Delta\hat m^\lambda \hat r(\mathtt{p_1 f}) [\hat m^\lambda \hat r_{\max}(\mathtt{p_1 f})]^{d-2}.$$
Let $a \in \arg\max_i \hat r_i(\mathtt{b_1}\square)$, so that
$$0 \leqslant \hat m^\lambda \hat r_{\max}(\mathtt{p_1 f}) - \hat m^\lambda \hat r_a(\mathtt{p_1 f}) \leqslant \Delta\hat m^\lambda \hat r(\mathtt{r_1 f}) + \Delta\hat m^\lambda \hat r(\mathtt{b_1 f}_{\geqslant 1}) = O(2^{-(1+c)k}),$$
where the last estimate is by (93) and (96). On the other hand, we have from (95) that $\hat m^\lambda \hat p(\mathtt{p_1 f}) \geqslant \hat m^\lambda \hat p(\mathtt{b_1 f}) \asymp 1$, and it follows that
$$[\hat m^\lambda \hat r_{\max}(\mathtt{p_1 f})]^{d-2} \asymp [\hat m^\lambda \hat r_a(\mathtt{p_1 f})]^{d-1}. \tag{109}$$
Applying (95) and (96) again, we have (for $i = 1, 2$)
$$[\hat m^\lambda \hat r_i(\mathtt{p_1 f})]^{d-1} \asymp [\hat m^\lambda \hat r_i(\mathtt{p_1}\square)]^{d-1}.$$
On the other hand, assuming $T \geqslant 1$, we have
$$\dot p_i^{\mathrm{u}}(\mathtt{r_1 f}) \geqslant [\hat m^\lambda \hat r_i(\mathtt{p_1}\square)]^{d-1} - [\hat m^\lambda \hat r_i(\mathtt{b_1}\square)]^{d-1} \asymp [\hat m^\lambda \hat r_i(\mathtt{p_1}\square)]^{d-1}$$
where the last step follows by (107). Similarly,
$$\begin{aligned} \dot p_i^{\mathrm{u}}(\mathtt{r_1 f}) - \dot p_i^{\mathrm{u}}(\mathtt{b_1 f}) &= O(1) \sum_{\hat r \in \{\hat p, \mathtt{F}\hat p\}} \hat m^\lambda \hat r_i(\mathtt{b_1 f})^{d-1} = O(2^{-k}) \sum_{\hat r \in \{\hat p, \mathtt{F}\hat p\}} \hat m^\lambda \hat r_i(\mathtt{p_1 f})^{d-1} \\ &= O(2^{-k}) \dot p_i^{\mathrm{u}}(\mathtt{r_1 f}) = O(2^{-k}) \dot p_i^{\mathrm{u}}(\mathtt{b_1 f}), \end{aligned} \tag{110}$$
where the last step follows by rearranging the terms. Combining the above gives
$$\Delta\dot p^{\mathrm{u}}(\mathtt{p_1 f}) \leqslant O(d) \|\Delta\hat m^\lambda \hat p + \Delta\hat m^\lambda \mathtt{F}\hat p\|_1 \max_{i=1,2} \dot p_i^{\mathrm{u}}(\mathtt{b_1 f}). \tag{111}$$
Clearly, similar bounds hold if we replace $\mathtt{p_1 f}$ with any of $\mathtt{p_0 f}$, $\mathtt{f p_1}$, or $\mathtt{f p_0}$.

3. Lastly we bound $\Delta\dot{p}^{\mathrm{u}}(\mathtt{p}_x\mathtt{p}_y)$. As in the single-copy recursion, for $x, y \in \{\mathtt{0}, \mathtt{1}\}$ we denote

$$\begin{aligned}\dot{r}(\mathtt{R}_x\dot{\sigma}) &\equiv \dot{r}(\mathtt{r}_x\dot{\sigma}) - \dot{r}(\mathtt{b}_x\dot{\sigma}),\\ \dot{r}(\dot{\sigma}\mathtt{R}_x) &\equiv \dot{r}(\dot{\sigma}\mathtt{r}_x) - \dot{r}(\dot{\sigma}\mathtt{b}_x),\\ \dot{r}(\mathtt{R}_x\mathtt{R}_y) &\equiv \dot{r}(\mathtt{r}_x\mathtt{r}_y) - \dot{r}(\mathtt{r}_x\mathtt{b}_y) - \dot{r}(\mathtt{b}_x\mathtt{r}_y) + \dot{r}(\mathtt{b}_x\mathtt{b}_y).\end{aligned}$$

Applying (107) gives

$$\begin{aligned}\dot{p}_i^{\mathrm{u}}(\mathtt{R}_x\mathtt{r}_y) &= [\hat{p}_i(\mathtt{b}_x\mathtt{p}_y)]^{d-1} = O(2^{-k})[\hat{p}_i(\mathtt{p}_x\mathtt{p}_y)]^{d-1} = O(2^{-k})\dot{p}_i^{\mathrm{u}}(\mathtt{r}_x\mathtt{r}_y),\\ \dot{p}_i^{\mathrm{u}}(\mathtt{R}_x\mathtt{R}_y) &= [\hat{p}_i(\mathtt{b}_x\mathtt{b}_y)]^{d-1} = O(2^{-k})\dot{p}_i^{\mathrm{u}}(\mathtt{r}_x\mathtt{r}_y).\end{aligned}$$

Combining the above estimates gives

$$\dot{p}_i^{\mathrm{u}}(\mathtt{r}_x\mathtt{r}_y) - \dot{p}_i^{\mathrm{u}}(\mathtt{b}_x\mathtt{b}_y) = \dot{p}_i^{\mathrm{u}}(\mathtt{R}_x\mathtt{r}_y) + \dot{p}_i^{\mathrm{u}}(\mathtt{r}_x\mathtt{R}_y) - \dot{p}_i^{\mathrm{u}}(\mathtt{R}_x\mathtt{R}_y) = O(2^{-k})\dot{p}_i^{\mathrm{u}}(\mathtt{r}_x\mathtt{r}_y).$$

Further, it follows from the BP equations that

$$\begin{aligned}&\max\{\dot{p}_i^{\mathrm{u}}(\mathtt{r}_x\mathtt{R}_y), \dot{p}_i^{\mathrm{u}}(\mathtt{b}_x\mathtt{R}_y), \dot{p}_i^{\mathrm{u}}(\mathtt{R}_x\mathtt{r}_y), \dot{p}_i^{\mathrm{u}}(\mathtt{R}_x\mathtt{b}_y)\} \leqslant \dot{p}_i^{\mathrm{u}}(\mathtt{r}_x\mathtt{r}_y) - \dot{p}_i^{\mathrm{u}}(\mathtt{b}_x\mathtt{b}_y),\\ &\text{so } \dot{p}_i^{\mathrm{u}}(st) = [1 + O(2^{-k})]\dot{p}_i^{\mathrm{u}}(\mathtt{b}_x\mathtt{b}_y) \text{ for all } s \in \{\mathtt{r}_x, \mathtt{b}_x\},\ t \in \{\mathtt{r}_y, \mathtt{b}_y\}.\end{aligned} \tag{112}$$

Similarly, we can upper bound

$$\begin{aligned}\Delta\dot{p}^{\mathrm{u}}(\mathtt{p}_x\mathtt{p}_y) &\leqslant 4[\Delta\dot{p}^{\mathrm{u}}(\mathtt{r}_x\mathtt{r}_y) + \Delta\dot{p}^{\mathrm{u}}(\mathtt{R}_x\mathtt{r}_y) + \Delta\dot{p}^{\mathrm{u}}(\mathtt{r}_x\mathtt{R}_y) + \Delta\dot{p}^{\mathrm{u}}(\mathtt{R}_x\mathtt{R}_y)].\\ &\leqslant O(d)\sum_{\hat{r}\in\{\hat{p},\mathtt{F}\hat{p}\}}\sum_{\substack{s\in\{\mathtt{p}_x,\mathtt{b}_x\}\\ t\in\{\mathtt{p}_y,\mathtt{b}_y\}}}\|\Delta\hat{m}^{\lambda}\hat{r}\|_1[\hat{m}^{\lambda}\hat{r}_{\max}(st)]^{d-2}.\end{aligned} \tag{113}$$

For $\hat{r} \in \{\hat{p}, \mathtt{F}\hat{p}\}$, let $a = \arg\max_{i=1,2}\hat{m}^{\lambda}\hat{r}_i(\mathtt{b}_1\mathtt{b}_1)$: then, for any $s \in \{\mathtt{p}_x, \mathtt{b}_x\}$, $t \in \{\mathtt{p}_y, \mathtt{b}_y\}$,

$$\begin{aligned}0 &\leqslant \hat{m}^{\lambda}\hat{r}_{\max}(st) - \max_{i=1,2}\hat{m}^{\lambda}\hat{r}_i(st) \leqslant \hat{m}^{\lambda}\hat{r}_{\max}(st) - \hat{m}^{\lambda}\hat{r}_a(st)\\ &\leqslant O(1)\Delta\hat{m}^{\lambda}\hat{r}(\{\mathtt{pp}\}\backslash\{\mathtt{bb}\}) \leqslant O(1/2^{(1+c)k}),\end{aligned}$$

where the last estimate is by (93). Combining with (74) and (112) gives

$$\sum_{\substack{s\in\{\mathtt{p}_x,\mathtt{b}_x\}\\ t\in\{\mathtt{p}_y,\mathtt{b}_y\}}}[\hat{m}^{\lambda}\hat{r}_{\max}(st)]^{d-2} = O(1)\Big[\max_{i=1,2}\hat{r}_i(\mathtt{p}_x\mathtt{p}_y)\Big]^{d-1} = O(1)\max_{i=1,2}\dot{p}_i^{\mathrm{u}}(\mathtt{bb}).$$

Substituting into (113) gives

$$\Delta\dot{p}^{\mathrm{u}}(\mathtt{p}_x\mathtt{p}_y) \leqslant O(d)\|\Delta\hat{m}^{\lambda}\hat{p} + \Delta\hat{m}^{\lambda}\mathtt{F}\hat{p}\|_1\max_{i=1,2}\dot{p}_i^{\mathrm{u}}(\mathtt{bb}). \tag{114}$$

Further, for any $st \in \{\mathtt{r}_x\mathtt{R}_y, \mathtt{R}_x\mathtt{r}_y, \mathtt{R}_x\mathtt{R}_y\}$, we have

$$\Delta\dot{p}^{\mathrm{u}}(st) \leqslant O(k)\|\Delta\hat{m}^{\lambda}\hat{p} + \Delta\hat{m}^{\lambda}\mathtt{F}\hat{p}\|_1\max_{i=1,2}\dot{p}_i^{\mathrm{u}}(\mathtt{bb}). \tag{115}$$

Lastly, in the special case $\hat{p}_2 = \mathtt{F}\hat{p}_1$, (114) reduces to

$$\begin{aligned}|\dot{p}_1^{\mathrm{u}}(\mathtt{b}_0\mathtt{b}_0) - \dot{p}_1^{\mathrm{u}}(\mathtt{b}_0\mathtt{b}_1)| &\leqslant O(d)\|\hat{m}^{\lambda}\hat{p}_1 - \hat{m}^{\lambda}\mathtt{F}\hat{p}_1\|_1\dot{p}_1^{\mathrm{u}}(\mathtt{bb})\\ &\leqslant k^5 2^{(1-2\kappa)k}\|\dot{p}_i - \mathtt{F}\dot{p}_i\|_1.\end{aligned} \tag{116}$$

where the last estimate is by (94).

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 (46), this means

$$\begin{aligned}
&\text{(A)} \quad \textstyle\sum_{\dot\sigma \notin \{\mathtt{bb}\}} (2^{-k})^{\mathtt{r}[\dot\sigma]} p(\dot\sigma) = O(2^{-k}) p(\mathtt{bb}), \quad |p(\mathtt{b_0b_0}) - p(\mathtt{b_0b_1})| \leqslant (k^9/2^{c'k}) p(\mathtt{bb}),\\
&\text{(B)} \quad p(\mathtt{fr}) = O(2^{-k}) p(\mathtt{bb}), \quad p(\mathtt{rr}) = O(1) p(\mathtt{bb}),\\
&\text{(C)} \quad p(\mathtt{r}_x\dot\sigma) \geqslant [1 - O(2^{-k})] p(\mathtt{b}_x\dot\sigma) \text{ for all } x \in \{\mathtt{0}, \mathtt{1}\} \text{ and } \dot\sigma \in \dot\Omega.
\end{aligned} \tag{117}$$

Condition (C) is automatically satisfied due to the BP equations. The second part of (B) follows from (112). The second part of (A) holds trivially in the case $c' = 0$, and otherwise follows from (116). We claim that

$$\tilde{q}_i(\{\mathtt{rf}, \mathtt{fr}, \mathtt{ff}\}) = O(2^{-k}) \tilde{q}_i(\mathtt{bb}). \tag{118}$$

This immediately implies the first part of (B). Further, the BP equations give $\tilde{q}_i(\mathtt{bf}) \leqslant \tilde{q}_i(\mathtt{rf})$ and $\tilde{q}_i(\mathtt{fb}) \leqslant \tilde{q}_i(\mathtt{fr})$, so the first part of (A) also follows. To see that (118) holds, note that the second part of (96) gives

$$\begin{aligned}
\dot{p}_i^{\mathrm{u}}(\mathtt{ff}) &\leqslant O(1) \sum_{\hat{r} \in \{\hat{p}, \mathtt{F}\hat{p}\}} [\hat{m}^\lambda \hat{r}_i(\mathtt{ff})]^{d-1} \leqslant O(1) \sum_{\hat{r} \in \{\hat{p}, \mathtt{F}\hat{p}\}} [\hat{m}^\lambda \hat{r}_i(\mathtt{b_1b_1})]^{d-1},\\
\dot{p}_i^{\mathrm{u}}(\mathtt{r_1f}) &\leqslant O(1) \sum_{\hat{r} \in \{\hat{p}, \mathtt{F}\hat{p}\}} [\hat{m}^\lambda \hat{r}_i(\mathtt{p_1f})]^{d-1} \leqslant O(1) \sum_{\hat{r} \in \{\hat{p}, \mathtt{F}\hat{p}\}} [\hat{m}^\lambda \hat{r}_i(\mathtt{p_1b_1})]^{d-1}.
\end{aligned}$$

Combining with (107) gives $\dot{p}_i^{\mathrm{u}}(\{\mathtt{r_1f}, \mathtt{ff}\}) = O(2^{-k}) \dot{p}_i^{\mathrm{u}}(\mathtt{r_1r_1})$. Recalling (112) (and making use of symmetry) gives (118). 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 $\mathtt{R}_x$, we have

$$\begin{aligned}
\Delta\tilde{q}(\mathtt{pp}) &\leqslant O(1) \Delta\tilde{q}(\{\mathtt{bb}, \mathtt{rR}, \mathtt{Rr}, \mathtt{RR}\}),\\
\Delta\tilde{q}(\dot\Omega^2 \backslash \{\mathtt{pp}\}) &\leqslant O(1) \Delta\tilde{q}(\{\mathtt{bf}, \mathtt{fb}, \mathtt{ff}, \mathtt{fR}, \mathtt{Rf}\}).
\end{aligned}$$

We next bound $\Delta\tilde{q}(\mathtt{bb})$, which is the sum of $\Delta\tilde{q}(\mathtt{b}_x\mathtt{b}_y)$ over $x, y \in \{\mathtt{0}, \mathtt{1}\}$. By symmetry let us take $x = y = \mathtt{0}$. Since $\tilde{q}_i = (\tilde{q}_i)^{\mathrm{avg}}$, $\tilde{q}_i(\mathtt{b_0b_0}) = \frac{1}{4}\tilde{q}_i(\mathtt{bb}) + \frac{1}{2}[\tilde{q}_i(\mathtt{b_0b_0}) - \tilde{q}_i(\mathtt{b_0b_1})]$, so

$$\Delta\tilde{q}(\mathtt{b_0b_0}) \leqslant \tfrac{1}{4}|\tilde{q}_1(\mathtt{bb}) - \tilde{q}_2(\mathtt{bb})| + \tfrac{1}{2} \sum_{i=1,2} |\tilde{q}_i(\mathtt{b_0b_0}) - \tilde{q}_i(\mathtt{b_0b_1})|.$$

Since the $\tilde{q}_i$ are normalized to be probability measures,

$$1 - \tilde{q}_i(\dot\Omega^2 \backslash \{\mathtt{pp}\}) = \tilde{q}_i(\mathtt{pp}) = 2\tilde{q}_i(\mathtt{rR}) + 2\tilde{q}_i(\mathtt{Rr}) - 3\tilde{q}_i(\mathtt{RR}) + 4\tilde{q}_i(\mathtt{bb}),$$

from which it follows that

$$|\tilde{q}_1(\mathtt{bb}) - \tilde{q}_2(\mathtt{bb})| \lesssim |\tilde{q}_1(\dot\Omega^2 \backslash \{\mathtt{pp}\}) - \tilde{q}_2(\dot\Omega^2 \backslash \{\mathtt{pp}\})| + \Delta\tilde{q}(\{\mathtt{rR}, \mathtt{Rr}, \mathtt{RR}\}).$$

Combining the above estimates gives

$$\|\Delta\tilde{q}\|_1 \lesssim \Delta\tilde{q}(\mathtt{A}) + \sum_{i=1,2} |\tilde{q}_i(\mathtt{b_0b_0}) - \tilde{q}_i(\mathtt{b_0b_1})|, \quad \mathtt{A} \equiv \{\mathtt{bf}, \mathtt{fb}, \mathtt{ff}, \mathtt{fR}, \mathtt{Rf}, \mathtt{rR}, \mathtt{Rr}, \mathtt{RR}\}.$$

Write $\dot{Z}_i \equiv \|\dot{p}_i^{\mathrm{u}}\|_1$. Taking $a \in \{1, 2\}$ and $b = 2 - a$,

$$\|\Delta\tilde{q}\|_1 \leqslant e_1 + e_2e_3 + e_4 \quad \text{with } e_1 \equiv \frac{\Delta\dot{p}^{\mathrm{u}}(\mathtt{A})}{\dot{Z}_a},\ e_2 \equiv \frac{|\dot{Z}_1 - \dot{Z}_2|}{\dot{Z}_a} \leqslant \frac{\|\Delta\dot{p}^{\mathrm{u}}\|_1}{\dot{Z}_a},$$

$$e_3 \equiv \frac{\dot{p}_b^{\mathrm{u}}(\mathtt{A})}{\dot{Z}_b},\ e_4 \equiv \sum_{i=1,2} \frac{|\dot{p}_i^{\mathrm{u}}(\mathtt{b_0b_0}) - \dot{p}_i^{\mathrm{u}}(\mathtt{b_0b_1})|}{\dot{Z}_i}.$$

It follows from (108), (111), (115) and (118), and taking $a = \arg\max_i \dot{p}_i^{\mathrm{u}}(\mathtt{bb})$, that

$$e_1 \lesssim \|\Delta\hat{m}^\lambda\hat{p} + \Delta\hat{m}^\lambda \mathtt{F}\hat{p}\|_1 (d/2^k) \max_{i=1,2} \dot{p}_i^{\mathrm{u}}(\mathtt{bb})/\dot{Z}_a \lesssim k\|\Delta\hat{m}^\lambda\hat{p} + \Delta\hat{m}^\lambda\mathtt{F}\hat{p}\|_1.$$

Further, recalling (114) gives

$$e_2 \lesssim k2^k\|\Delta\hat{m}^\lambda\hat{p} + \Delta\hat{m}^\lambda\mathtt{F}\hat{p}\|_1.$$

Combining (110), (112), and (118) gives $e_3 = O(2^{-k})$. Finally, (116) gives

$$e_4 \lesssim k2^k\|\hat{m}^\lambda\hat{p}_i - \hat{m}^\lambda\mathtt{F}\hat{p}_i\|_1.$$

Combining the pieces together finishes the proof. □

10. The 1RSB free energy

10.1. **Equivalence of recursions.** In this section, we relate the coloring recursion (44) to the distributional recursion (7). The main task of this section is to show the following

Proposition 10.1. *Let $\dot{q}_\lambda$ be the fixed point given by Proposition 4.2 for parameters $\lambda \in [0,1]$ and $T = \infty$. Let $H_\lambda \equiv (\dot{H}_\lambda, \hat{H}_\lambda, \bar{H}_\lambda)$ be the associated triple of measures defined by Proposition 3.7. Then $(\boldsymbol{s}(H_\lambda), \boldsymbol{\Sigma}(H_\lambda), \boldsymbol{F}(H_\lambda)) = (s_\lambda, \Sigma(s_\lambda), \mathfrak{F}(\lambda))$.*

In the course of proving Proposition 10.1, 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

$$\dot{\pi}(\dot{\tau}) = \begin{cases} [1-\dot{q}(\mathtt{r})]^{-1}\dot{q}(\dot{\tau}) & \text{if } \dot{\tau} \in \dot{\Omega}_{\mathtt{f}}, \\ [1-\dot{q}(\mathtt{r})]^{-1}\dot{q}(\mathtt{b}_x) & \text{if } \dot{\tau} = x \in \{\mathtt{0},\mathtt{1}\}. \end{cases}$$

Recalling the definition (16) of the mapping $\dot{m} : \dot{\Omega} \to [0,1]$, we 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

$$\hat{\pi}(\hat{\tau}) \equiv \begin{cases} [1-\hat{q}(\mathtt{b})]^{-1}\hat{q}(\hat{\tau}) & \text{if } \hat{\tau} \in \hat{\Omega}_{\mathtt{f}}, \\ [1-\hat{q}(\mathtt{b})]^{-1}\hat{q}(\mathtt{r}_x) & \text{if } \hat{\tau} = x \in \{\mathtt{0},\mathtt{1}\}. \end{cases}$$

Recalling the definition (17) of the mapping $\hat{m} : \hat{\Omega} \to [0,1]$, we denote the pushforward measure $\hat{u} \equiv \hat{u}(\hat{q}) \equiv \hat{\pi} \circ \hat{m}^{-1}$, so that $\hat{u} \in \mathscr{P}$ also. The next two lemmas follow straightforwardly from the above definitions, and we omit their proofs:

Lemma 10.2. *Suppose $\dot{q} \in \mathscr{P}(\dot{\Omega})$ satisfies $\dot{q} = \dot{q}^{\mathrm{avg}}$ and*

$$\dot{m}^\lambda\dot{q}(\mathtt{f}) = \dot{q}(\mathtt{r}_1) - \dot{q}(\mathtt{b}_1) = \dot{q}(\mathtt{r}_0) - \dot{q}(\mathtt{b}_0) = (1-\dot{m})^\lambda\dot{q}(\mathtt{f}) \tag{119}$$

Then $\hat{q} \equiv \hat{\mathtt{BP}}\dot{q} \in \mathscr{P}(\hat{\Omega})$ must satisfy $\hat{q} = \hat{q}^{\mathrm{avg}}$ and

$$\hat{m}^\lambda\hat{q}(\mathtt{f}) = \hat{q}(\mathtt{b}_1) = \hat{q}(\mathtt{b}_0) = (1-\hat{m})^\lambda\hat{q}(\mathtt{f}), \tag{120}$$

Let $\hat{\boldsymbol{z}} \equiv (\hat{\mathtt{N}}\hat{\mathtt{B}}\hat{\mathtt{P}}\dot{q})/(\hat{\mathtt{B}}\hat{\mathtt{P}}\dot{q})$ be the normalizing constant. Then $\dot{u} \equiv \dot{u}(\dot{q})$ and $\hat{u} \equiv \hat{u}(\hat{q})$ satisfy

$$\hat{u} = \hat{\mathscr{R}}_\lambda(\dot{u}), \quad \hat{\mathscr{Z}}_\lambda(\dot{u}) = \frac{\hat{\boldsymbol{z}}(1-\hat{q}(\mathtt{b}))}{(1-\dot{q}(\mathtt{r}))^{k-1}}. \tag{121}$$

Lemma 10.3. *Suppose $\hat{q} \in \mathscr{P}(\hat{\Omega})$ satisfies $\hat{q} = \hat{q}^{\mathrm{avg}}$ and (120). Then $\dot{q} \equiv \hat{\mathtt{BP}}\hat{q} \in \mathscr{P}(\dot{\Omega})$ must satisfy $\dot{q} = \dot{q}^{\mathrm{avg}}$ and (119). Let $\dot{\boldsymbol{z}} \equiv (\dot{\mathtt{NBP}}\hat{q})/(\dot{\mathtt{BP}}\hat{q})$ be the normalizing constant: then*

$$\dot{u} = \dot{\mathscr{R}}_\lambda(\hat{u}), \quad \dot{\mathscr{Z}}_\lambda(\hat{u}) = \frac{\dot{\boldsymbol{z}}(1-\dot{q}(\mathtt{r}))}{(1-\hat{q}(\mathtt{b}))^{d-1}}. \tag{122}$$

Proof of Proposition 1.2. This is simply a rephrasing of the proof of Proposition 4.2, using Lemma 10.2 and Lemma 10.3. □

We next prove Proposition 10.1. 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 4.2. 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

$$\hat{\boldsymbol{z}} \equiv \hat{\boldsymbol{z}}_\lambda \equiv (\hat{\mathtt{NBP}}\dot{q})/(\hat{\mathtt{BP}}\dot{q}), \quad \dot{\boldsymbol{z}} \equiv \dot{\boldsymbol{z}}_\lambda \equiv (\dot{\mathtt{NBP}}\hat{q})/(\dot{\mathtt{BP}}\hat{q}).$$

Let $H_\lambda \equiv (\dot{H}_\lambda, \hat{H}_\lambda, \bar{H}_\lambda)$ be the triple of associated measures defined as in Proposition 3.7, with normalizing constants $(\dot{\mathcal{Z}}_\lambda, \hat{\mathcal{Z}}_\lambda, \bar{\mathcal{Z}}_\lambda)$. Recall from (9) 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 10.4. *Under the above notations, $\boldsymbol{F}(H_\lambda) = \ln \dot{\mathcal{Z}}_\lambda + \alpha \ln \hat{\mathcal{Z}}_\lambda - d \ln \bar{\mathcal{Z}}_\lambda$, and*

$$\bar{\mathfrak{Z}}_\lambda = \frac{\bar{\mathcal{Z}}_\lambda}{(1-\dot{q}_\lambda(\mathtt{r}))(1-\hat{q}_\lambda(\mathtt{b}))}, \quad \dot{\mathfrak{Z}}_\lambda = \frac{\dot{\mathcal{Z}}_\lambda}{(1-\hat{q}_\lambda(\mathtt{b}))^d}, \quad \hat{\mathfrak{Z}}_\lambda = \frac{\hat{\mathcal{Z}}_\lambda}{(1-\dot{q}_\lambda(\mathtt{r}))^k}. \tag{123}$$

Consequently $\mathfrak{F}(\lambda) = \boldsymbol{F}(H_\lambda)$.

Proof. It follows from (35) (and recalling (31) that $\hat{\Phi}(\underline{\sigma})^\lambda = \hat{\Phi}^{\max}(\underline{\sigma})^\lambda \hat{v}(\underline{\sigma})$) 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 (38) and rearranging gives

$$\begin{aligned}
&\boldsymbol{F}(H_\lambda) - \Big(\ln \dot{\mathcal{Z}}_\lambda + \alpha \ln \hat{\mathcal{Z}}_\lambda - d \ln \bar{\mathcal{Z}}_\lambda \Big) \\
&\quad = -\Big\langle \sum_{i=1}^d \ln \hat{q}_\lambda(\hat{\sigma}_i), \dot{H}_\lambda \Big\rangle - \alpha \Big\langle \sum_{i=1}^k \ln \dot{q}_\lambda(\dot{\sigma}_i), \hat{H}_\lambda \Big\rangle + d\langle \ln[\dot{q}_\lambda(\dot{\sigma})\hat{q}_\lambda(\hat{\sigma})], \bar{H}_\lambda\rangle.
\end{aligned}$$

This equals zero by (39). The proof of (123) is straightforward from the preceding definitions, and is omitted. □

Proof of Proposition 10.1. 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. □

10.2. **Large-k asymptotics.** We now evaluate the large-k asymptotics of the free energy, beginning with (9). Let $\dot{\mu}_\lambda$ be as 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

$$\psi_\lambda \equiv \int x^\lambda \mathbf{1}\{x \in (0,1)\} \dot{\mu}_\lambda(dx), \quad \rho_\lambda \equiv \int y^\lambda \mathbf{1}\{y \in (0,1)\backslash\{\tfrac{1}{2}\}\} \hat{\mu}_\lambda(dy).$$

Proposition 10.5. *For $k \geqslant k_0$, $\alpha_{\mathrm{lbd}} \leqslant \alpha = (2^{k-1} - c)\ln 2 \leqslant \alpha_{\mathrm{ubd}}$, and $\lambda \in [0,1]$,*

$$\ln \dot{\mathfrak{Z}}_\lambda = \ln 2 - (1-2^{\lambda-1})/2^k + d\ln\Big(2^{-\lambda}\hat{\mu}_\lambda(\tfrac{1}{2}) + \hat{\mu}_\lambda(1) + \rho_\lambda\Big) + \mathsf{err}, \tag{124}$$

$$-d\ln\bar{\mathfrak{Z}} = -d\ln\Big(2^{-\lambda}\hat{\mu}_\lambda(\tfrac{1}{2}) + \hat{\mu}(1) + \rho_\lambda\Big) - (k\ln 2)[-\dot{\mu}(\mathtt{f}) + 2\psi_\lambda] + \mathsf{err}, \tag{125}$$

$$\alpha\ln\hat{\mathfrak{Z}} = \alpha\ln(1-2/2^k) + (k\ln 2)(-\dot{\mu}(\mathtt{f}) + 2\psi_\lambda) + \mathsf{err}, \tag{126}$$

where err *denotes any error bounded by $k^{O(1)}/4^k$. Altogether this yields*

$$\mathfrak{F}(\lambda) = \mathsf{f}^{\mathrm{RS}}(\alpha) - (1-2^{\lambda-1})/2^k + \mathsf{err} = [(2c-1)\ln 2 - (1-2^{\lambda-1})]/2^k + \mathsf{err}.$$

On the other hand $\lambda s_\lambda = \lambda(\ln 2)2^{\lambda-1}/2^k + \mathsf{err}$.

Proof of Proposition 1.5. Apply Proposition 10.5: setting $\mathfrak{F}(\lambda) = \lambda s_\lambda$ gives

$$\alpha_\lambda = (2^{k-1} - c_\lambda)\ln 2 + \mathsf{err}, \quad c_\lambda = \frac{1}{2} + \frac{1 - 2^{\lambda-1}(1-\lambda\ln 2)}{2\ln 2}.$$

Substituting the special values $\lambda = 1$ and $\lambda = 0$ gives

$$c_{\mathrm{cond}} = c_1 = 1, \quad c_{\mathrm{sat}} = c_0 = \frac{1}{2} + \frac{1}{4\ln 2},$$

as claimed. □

Proof of Proposition 10.5. Throughout the proof we abbreviate ϵ_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, ψ_λ is upper bounded by $\dot{\mu}(\mathtt{f}) \equiv \dot{\mu}((0,1))$, and we will show below that

$$\dot{\mu}_\lambda(\mathtt{f}) = \frac{2^{\lambda-1}}{2^k} + \epsilon_k. \tag{127}$$

Estimate of $\dot{\mathfrak{Z}}_\lambda$. Recall from the definition (8) that

$$\dot{\mathfrak{Z}}_\lambda = \int\Big(\prod_{i=1}^d y_i + \prod_{i=1}^d(1-y_i)\Big)^\lambda \prod_{i=1}^d \hat{\mu}_\lambda(dy_i).$$

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 $\dot{\mathfrak{Z}}_\lambda(\mathtt{f}_{\geqslant 2})$. 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:

$$\begin{aligned}\dot{\mathfrak{Z}}_\lambda(\mathtt{f}_{\geqslant 2}) &= \sum_{j=1}^d \binom{d}{j}\Big(\int y^\lambda \mathbf{1}\{y\in(0,1)\backslash\{\tfrac{1}{2}\}\}\hat{\mu}_\lambda(dy)\Big)^j\Big(2^{-\lambda}\hat{\mu}_\lambda(\tfrac{1}{2})\Big)^{d-j}\\ &\leqslant O(1)d\hat{\mu}_\lambda((0,1)\backslash\{\tfrac{1}{2}\})\Big(2^{-\lambda}\hat{\mu}_\lambda(\tfrac{1}{2})\Big)^d \leqslant \dot{\mathfrak{Z}}_\lambda(\mathtt{f}_1)k^{O(1)}/2^k.\end{aligned}$$

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

$$\frac{\dot{\mathfrak{Z}}_\lambda(\mathtt{f})}{2\dot{\mathfrak{Z}}_\lambda(\mathtt{1})} = \frac{2^\lambda}{2}\Big(\frac{\hat{\mu}_\lambda(\frac{1}{2})}{\hat{\mu}_\lambda(\frac{1}{2}) + 2^\lambda\hat{\mu}_\lambda(1)}\Big)^d + \epsilon_k = \frac{2^{\lambda-1}}{2^k} + \epsilon_k.$$

Combining these yields (124). The proof of (127) is very similar.

Estimate of $\bar{\mathfrak{Z}}_\lambda$. Recall from the definition (8) 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 $x=1$ is given by

$$\bar{\mathfrak{Z}}_\lambda(x=1) = \dot{\mu}_\lambda(1)\Big(2^{-\lambda}\hat{\mu}_\lambda(\tfrac{1}{2}) + \hat{\mu}_\lambda(1) + \rho_\lambda\Big).$$

There is an equal contribution from the case $x=0$. Next, the contribution from $x\in(0,1)$ and $y=1/2$ is given by

$$\bar{\mathfrak{Z}}_\lambda(x\in(0,1), y=1/2) = \dot{\mu}(\mathtt{f})2^{-\lambda}\hat{\mu}_\lambda(\tfrac{1}{2}).$$

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 $\leqslant k^{O(1)}/8^k$. Combining these estimates gives

$$\begin{aligned}d\ln\bar{\mathfrak{Z}}_\lambda &= d\ln\Big(2^{-\lambda}\hat{\mu}_\lambda(\tfrac{1}{2}) + 2\dot{\mu}_\lambda(1)\hat{\mu}(1) + 2\dot{\mu}_\lambda(1)\rho_\lambda + 2\hat{\mu}(1)\psi_\lambda\Big) + \epsilon_k \\ &= d\ln\Big(2^{-\lambda}\hat{\mu}_\lambda(\tfrac{1}{2}) + \hat{\mu}(1) + \rho_\lambda\Big) + d\ln\Big(1 + \frac{\hat{\mu}(1)[-\dot{\mu}_\lambda(\mathtt{f}) + 2\psi_\lambda]}{2^{-\lambda}\hat{\mu}_\lambda(\frac{1}{2})}\Big) + \epsilon_k.\end{aligned}$$

Recalling $\hat{\mu}_\lambda = \hat{\mathscr{R}}\dot{\mu}_\lambda$ gives

$$d\ln\Big(1 + \frac{\hat{\mu}(1)[-\dot{\mu}_\lambda(\mathtt{f}) + 2\psi_\lambda]}{2^{-\lambda}\hat{\mu}_\lambda(\frac{1}{2})}\Big) = d\dot{\mu}(0)^{k-1}(-\dot{\mu}(\mathtt{f}) + 2\psi_\lambda) + \epsilon_k,$$

and (125) follows.

Estimate of $\hat{\mathfrak{Z}}_\lambda$. Recall from the definition (8) that

$$\hat{\mathfrak{Z}}_\lambda = \int\Big(1 - \prod_{i=1}^k x_i - \prod_{i=1}^k(1-x_i)\Big)\prod_{i=1}^k \dot{\mu}_\lambda(x_i).$$

The contribution to $\hat{\mathfrak{Z}}$ from separating clauses is

$$1 - 2\dot{\mu}(\mathtt{0},\mathtt{f})^k + \dot{\mu}(\mathtt{f})^k = 1 - (2/2^k)(1 + k\dot{\mu}(\mathtt{f})) + k^{O(1)}/8^k.$$

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 $\leqslant k^{O(1)}/8^k$, and (126) follows.

Estimate of s_λ. Recall from (10) the definition of s_λ. 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. □

10.3. **Properties of the complexity function.** We conclude by deducing some properties of the complexity function $\Sigma(s)$.

Lemma 10.6. *For fixed $1 \leqslant T < \infty$, the fixed point $\dot{q}_{\lambda,T}$ 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 Ω_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 (45), Proposition 9.1 guarantees that

$$\begin{aligned}\|f_T[\dot{q}, \lambda] - f_T[\dot{q}_\star(\lambda, T), \lambda]\|_1 &\geqslant \|\dot{q} - \dot{q}_\star(\lambda)\|_1 - \|\mathtt{BP}_{\lambda,T}[\dot{q}] - \mathtt{BP}_{\lambda,T}[\dot{q}_\star(\lambda, T)]\|_1 \\ &\geqslant (1 - O(k^2 2^{-k}))\|\dot{q} - \dot{q}_\star(\lambda, T)\|_1.\end{aligned}$$

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 λ. □

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 [KMR+07] takes $\mathcal{S}(s)$ as a conceptual starting point. However, for purposes of explicit calculation the actual starting point is the Legendre dual

$$\mathfrak{F}(\lambda) \equiv (-\mathcal{S})^\star(\lambda) = \sup_{s\in\mathbb{R}} \Big\{\lambda s + \mathcal{S}(s)\Big\} = \sup_H \boldsymbol{F}_\lambda(H),$$

where $\boldsymbol{F}_\lambda(H) \equiv \lambda \boldsymbol{s}(H) + \boldsymbol{\Sigma}(H)$. The replica symmetry breaking heuristic gives an explicit conjecture for $\mathfrak{F}$. One then makes the assuption that $\mathcal{S}(s)$ is *concave* in s: this means it is the same as

$$\mathcal{R}(s) \equiv -\mathfrak{F}^\star(s) = -(-\mathcal{S})^{\star\star}(s),$$

so if $\mathcal{S}$ is concave then it can be recovered from $\mathfrak{F}$.

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_λ which maximizes $\lambda s + \mathcal{S}(s)$, given by

$$s_\lambda = \boldsymbol{s}(H_\lambda).$$

Recall that H_λ and s_λ both depend implicitly on T. We also have from Lemma 10.6 that for any fixed $T < \infty$, s_λ is continuous in λ, so it maps $\lambda \in [0,1]$ onto some compact interval $\mathcal{I} \equiv [s_-, s_+]$. Define the modified function

$$\overline{\mathcal{S}}(s) \equiv \begin{cases} \mathcal{S}(s) & s \in \mathcal{I}, \\ -\infty & \text{otherwise.} \end{cases}$$

Lemma 10.7. *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_λ is nondecreasing in λ.*

Proof. The function $-\mathcal{S}(s)$ has Legendre dual

$$\overline{\mathfrak{F}}(\lambda) = \sup_{s \in \mathbb{R}} \Big\{ \lambda s + \overline{\mathcal{S}}(s) \Big\} = \sup_{s \in \mathcal{I}} \Big\{ \lambda s + \mathcal{S}(s) \Big\} \leqslant \mathfrak{F}(\lambda).$$

For $\lambda \in [0,1]$ it is clear that $\overline{\mathfrak{F}}(\lambda) = \mathfrak{F}(\lambda)$. It is straightforward to check that if $\lambda < 0$ then

$$\overline{\mathfrak{F}}(\lambda) \leqslant \max_{s \in \mathcal{I}} \lambda s + \max_{s \in \mathcal{I}} \mathcal{S}(s) = \lambda s_{\min} + \mathcal{S}(s_0),$$

so if $s < s_{\min}$ then

$$(-\overline{\mathcal{S}})^{\star\star}(s) = (\overline{\mathfrak{F}})^{\star}(s) \geqslant \sup_{\lambda < 0} \Big\{ \lambda s - \overline{\mathfrak{F}}(\lambda) \Big\} \geqslant \sup_{\lambda < 0} \Big\{ \lambda(s - s_{\min}) - \mathcal{S}(s_0) \Big\} = +\infty.$$

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

$$(-\overline{\mathcal{S}})^{\star\star}(s) \geqslant \lambda_\circ s - \mathfrak{F}(\lambda_\circ) = -\mathcal{S}(s).$$

This proves $(-\overline{\mathcal{S}})^{\star\star}(s) \geqslant -\overline{\mathcal{S}}(s)$ for all $s \in \mathbb{R}$. On the other hand, it holds for any function f that $f^{\star\star} \leqslant f$, 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 1.4. We can obtain $\Sigma(s)$ as the limit of $\overline{\mathcal{S}}(s)$ in the limit $T \to \infty$. It follows from Lemma 10.7 together with Corollary 9.2 that it is strictly decreasing in s. □

Appendix A. Constrained entropy maximization

In this section we review some general theory for entropy maximization problems under affine constraints.

A.1. **Constraints and continuity.** We will optimize a functional over non-negative measures ν on a finite space X (with $|X| = s$), subject to some affine constraints $M\nu = b$. We begin by discussing basic continuity properties. Denote

$$\mathbb{H}(b) \equiv \{\nu \geqslant 0\} \cap \{M\nu = b\} \subseteq \mathbb{R}^s.$$

Let $\Delta \equiv \{\nu \geqslant 0\} \cap \{\langle \mathbf{1}, \nu \rangle = 1\}$, 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 Δ under M, so $\boldsymbol{B}$ is a compact subset of $\mathbb{R}^r$.

Proposition A.1. *If $\boldsymbol{F}$ is any continuous function on Δ and*

$$F(b) = \max\{\boldsymbol{F}(\nu) : \nu \in \mathbb{H}(b)\}, \tag{128}$$

then F is (uniformly) continuous over $b \in \boldsymbol{B}$.

Proposition A.1 is a straightforward consequence of the following two lemmas.

Lemma A.2. *For $b \in \boldsymbol{B}$ and any vector u in the unit sphere $\mathbb{S}^{r-1}$, let*

$$d(b,u) \equiv \inf\{t \geqslant 0 : b + tu \notin \boldsymbol{B}\}.$$

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 $H_1, \dots, H_k$, where $H_i = \{x \in \mathbb{R}^r : \langle a_i, x\rangle \leqslant c_i\}$. Consequently there is at least one index $1 \leqslant i \leqslant k$ such that

$$d(b,u) = \inf\{t \geqslant 0 : b + tu \notin H_i\}.$$

It follows that $\langle a_i, u\rangle > 0$ and

$$d(b,u) = \frac{c_i - \langle a_i, b\rangle}{\langle a_i, u\rangle} \geqslant \frac{c_i - \langle a_i, b\rangle}{|a_i|} = d(b, \partial H_i)$$

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

$$\delta = \delta(b) = \min\{d(b, \partial H_i) : d(b, \partial H_i) > 0\};$$

δ is a minimum over finitely many positive numbers so it is also positive. □

Lemma A.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

$$d_{\mathcal{H}}(X,Y) = \inf\{\epsilon \geqslant 0 : X \subseteq Y^\epsilon \text{ and } Y \subseteq X^\epsilon\},$$

where X^ϵ, Y^ϵ are the ϵ-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

$$\mathbb{H}(b_n) \subseteq (\mathbb{H}(b))^\epsilon, \quad n \geqslant n_0(\epsilon).$$

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 ν. For $u \in \mathbb{S}^{r-1}$ let $d(b,u)$ be as in Lemma A.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 $\geqslant \delta(b)$ by Lemma A.2. We can express $b' = b + tu$ as the convex combination

$$b' = (1-\epsilon)b + \epsilon[b + d(b,u)u], \quad \epsilon = \frac{t}{d(b,u)} = \frac{|b'-b|}{d(b,u)} \leqslant \frac{|b'-b|}{\delta}.$$

Then $\nu' = (1-\epsilon)\nu + \epsilon\nu(b,u) \in \mathbb{H}(b')$, so

$$|\nu' - \nu| = \epsilon|\nu(b,u) - \nu| \leqslant \frac{(\operatorname{diam}\Delta)|b-b'|}{\delta}$$

This implies $H(b) \subseteq (H(b_n))^\epsilon$ for large enough n, and the result follows. □

Proof of Proposition A.1. Take $\nu \in \mathbb{H}(b)$ so that $F(b) = \boldsymbol{F}(\nu)$. If $b' = b + tu \in \boldsymbol{B}$ for $u \in \mathbb{S}^{r-1}$, then Lemma A.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 Δ by the Heine–Cantor theorem. Therefore

$$F(b) = \boldsymbol{F}(\nu) = \boldsymbol{F}(\nu') + o_t(1) \leqslant F(b') + o_t(1).$$

By the same argument $F(b') \leqslant F(b) + o_t(1)$, concluding the proof. □

When solving (128) for a *fixed* value of $b \in \boldsymbol{B}$, it will be convenient to make the following reduction:

Remark A.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 $\{\nu \geqslant 0\}$, it must be that

$$X_\circ \equiv \{x \in X : \exists \nu \in \{\nu \geqslant 0\} \cap \{M\nu = b\} \text{ so that } \nu(x) > 0\}$$

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

$$\mathbb{H}_\circ(b) = \{\nu_\circ \geqslant 0\} \cap \{M'\nu_\circ = b\}$$

contains an interior point of $\{\nu_\circ \geqslant 0\}$. 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 (128) 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 $\mathbb{H}(b) = \{\nu \geqslant 0\} \cap \{M\nu = b\}$ contains an interior point of $\{\nu \geqslant 0\}$. It follows that this space has dimension $s - r > 0$, and its boundary is contained in the boundary of $\{\nu \geqslant 0\}$.

A.2. **Entropy maximization.** We now restrict (128) to the case of functionals $\boldsymbol{F}$ which are *concave* on the domain $\{\nu \geqslant 0\}$. 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

$$\mathcal{L}(\gamma; b) = \sup\{\boldsymbol{F}(\nu) + \langle \gamma, M\nu - b\rangle : \nu \geqslant 0\} = -\langle \gamma, b\rangle + G(\gamma).$$

It holds for any $\gamma \in \mathbb{R}^r$ that $\mathcal{L}(\gamma; b) \geqslant F(b)$, so

$$F(b) \leqslant \inf\{\mathcal{L}(\gamma; b) : \gamma \in \mathbb{R}^r\} = -G^\star(b). \tag{129}$$

Now assume ψ is a positive function on X, and consider (128) for the special case

$$\boldsymbol{F}(\nu) = \mathcal{H}(\nu) + \langle \nu, \ln \psi \rangle = \sum_{x \in X} \nu(x) \ln \frac{\psi(x)}{\nu(x)}. \tag{130}$$

We remark that the supremum in $(-\mathcal{H})^\star(\nu^\star) = \sup\{\langle \nu^\star, \nu\rangle + \mathcal{H}(\nu) : \nu \geqslant 0\}$ is uniquely attained by $\nu^{\mathrm{opt}}(x) = \exp\{-1 + \nu^\star(x)\}$, yielding

$$(-\mathcal{H})^\star(\nu^\star) = \langle \nu^{\mathrm{opt}}(\nu^\star), 1\rangle = \sum_x \exp\{-1 + \nu^\star(x)\}.$$

This gives the explicit expression

$$G(\gamma) = (-\boldsymbol{F})^\star(M^t\gamma) = (-\mathcal{H})^\star(\ln\psi + M^t\gamma) = \sum_x \psi(x) \exp\{-1 + (M^t\gamma)(x)\}. \tag{131}$$

Lemma A.5. *Assume ψ 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 (131) is strictly convex in γ.*

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 $H \equiv \operatorname{Hess} G(\gamma)$ 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 $(M_{i,x})_{i\leqslant r}$, 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 ν 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 γ. □

Proposition A.6. *Let $b \in \boldsymbol{B}$ such that $\mathbb{H}(b) = \{\nu \geqslant 0\} \cap \{M\nu = b\}$ contains an interior point of $\{\nu \geqslant 0\}$, and consider the optimization problem (128) for $\boldsymbol{F}$ as in (130). For this problem, the inequality (129) 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 γ, and its infimum is achieved by a unique $\gamma = \gamma(b)$. The optimum value of (128) is uniquely attained by the measure $\nu = \nu^{\mathrm{opt}}(b)$ defined by

$$\nu(x) = \psi(x) \exp\{-1 + (M^t\gamma)(x)\}. \tag{132}$$

For any $\mu \in \mathbb{H}(b)$, $\boldsymbol{F}(\nu) - \boldsymbol{F}(\mu) = \mathcal{H}(\mu|\nu) \gtrsim \|\nu - \mu\|^2$. 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 $\{\nu \geqslant 0\}$. The entropy $\mathcal{H}$ has unbounded gradient at this boundary, so for $\boldsymbol{F}$ as in (130), the optimization problem (128) must be solved by a strictly positive measure $\nu > 0$. Since $\nu > 0$, we can differentiate in the direction of any vector δ 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 A.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 γ so that $M\epsilon = 0$:[7]

$$\gamma = (MM^t)^{-1} M(\ln \nu - \ln \psi + 1).$$

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 (128) is optimized by $\nu = \nu^{\text{opt}}(b)$, as given by (132). The optimal value of (128) is then

$$\begin{aligned} F(b) &= \langle 1, \nu^{\text{opt}}(b) \rangle - \langle M^t \gamma(b), \nu^{\text{opt}}(b) \rangle \\ &= \sum_x \psi(x) \exp\{-1 + (M^t \gamma)(x)\} - \langle \gamma, b \rangle \Big|_{\gamma = \gamma(b)} = \mathcal{L}(\gamma(b), b). \end{aligned}$$

In view of (129), 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 A.5 that $G(\gamma)$ is strictly convex, which implies that $\mathcal{L}(\gamma; b)$ is strictly convex in γ. 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 $\{\nu \geqslant 0\}$, 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 γ gives

$$b'(\gamma) = \operatorname{Hess} G(\gamma), \quad \gamma'(b) = [\operatorname{Hess} G(\gamma(b))]^{-1}. \tag{133}$$

We also find (in this neighborhood) that

$$F'(b) = -\gamma(b), \quad F''(b) = -\gamma'(b) = -[\operatorname{Hess} G(\gamma(b))]^{-1},$$

so F is strictly concave.

It remains to prove that $\boldsymbol{F}(\nu) - \boldsymbol{F}(\mu) = \mathcal{H}(\mu|\nu)$. (The estimate $\mathcal{H}(\mu|\nu) \gtrsim \|\mu - \nu\|^2$ is well known and straightforward to verify.) For any measure μ,

$$-\mathcal{H}(\mu|\nu) = \mathcal{H}(\mu) + \langle \mu, \ln(\psi \exp\{-1 + M^t \gamma\}) \rangle.$$

Applying this with $\mu = \nu$ gives

$$0 = -\mathcal{H}(\nu|\nu) = \mathcal{H}(\nu) + \langle \nu, \ln(\psi \exp\{-1 + M^t \gamma\}) \rangle.$$

Subtracting these two equations gives

$$-\mathcal{H}(\mu|\nu) = \mathcal{H}(\mu) - \mathcal{H}(\nu) + \langle \mu - \nu, \ln \psi \rangle + \langle \mu - \nu, \ln(\exp\{-1 + M^t \gamma\}) \rangle.$$

If $M\nu = M\nu = b$ then the last term vanishes, giving $-\mathcal{H}(\mu|\nu) = \boldsymbol{F}(\mu) - \boldsymbol{F}(\nu)$. □

Remark A.7. Our main application of Proposition A.6 is for the depth-one tree $\mathcal{D}$ as shown in Figure 3. 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

$$\psi(\underline{\sigma}) = \boldsymbol{w}_{\mathcal{D}}(\underline{\sigma})^\lambda = \left\{ \dot{\Phi}(\underline{\sigma}_{\delta v}) \prod_{a \in \partial v} [\bar{\Phi}(\sigma_{av}) \hat{\Phi}(\underline{\sigma}_{\delta a})] \right\}^\lambda.$$

[7] The matrix MM^t is invertible: if $MM^t x = 0$ then $M^t x \in \ker M = (\operatorname{im} M^t)^\perp$. On the other hand clearly $M^t x \in \operatorname{im} M^t$, so $M^t x \in (\operatorname{im} M^t) \cap (\operatorname{im} M^t)^\perp = \{0\}$. Therefore $x \in \ker M^t$, but M^t is injective by assumption.

We then wish to solve the optimization problem (128) for $\boldsymbol{F}(\nu)$ as in (130), under the constraint that ν has marginals $\dot{h}^{\text{tree}}(\dot{\sigma})$ on the boundary edges $\mathcal{L}(\mathcal{D})$. 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 ζ of $\mathcal{D}$: the $(\dot{\sigma}, \zeta)$ entry of M is given by

$$M(\dot{\sigma}, \zeta) = |\mathcal{L}(\mathcal{D})|^{-1} \sum_{e \in \mathcal{L}(\mathcal{D})} \mathbf{1}\{\dot{\zeta}_e = \dot{\sigma}\}.$$

Recall Remark A.4, let $\dot{\Omega}_+ = \{\dot{\sigma} \in \dot{\Omega} : \dot{h}^{\text{tree}}(\dot{\sigma}) > 0\}$, and $X_\circ = \{\zeta \in X : M(\dot{\sigma}, \zeta) = 0 \ \forall \dot{\sigma} \notin \dot{\Omega}\}$. 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 $\dot{\zeta}$ row is a linear combination of other rows, then set $\dot{q}(\dot{\zeta}) = 1$ 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 $\{\nu_\circ \geqslant 0\} \cap \{M_\circ \nu_\circ = b_\circ\}$ where $b_\circ$ denotes the entries of b indexed by $\dot{\Omega}_\circ$. It follows from Proposition A.6 that the unique maximizer of (128) is the measure $\nu = \nu^{\text{opt}}(b)$ given by

$$\nu(\underline{\sigma}) = \frac{1}{Z} \boldsymbol{w}_{\mathcal{D}}(\underline{\sigma})^\lambda = \frac{1}{Z} \left\{ \dot{\Phi}(\underline{\sigma}_{\delta v}) \prod_{a \in \partial v} [\bar{\Phi}(\sigma_{av}) \hat{\Phi}(\underline{\sigma}_{\delta a})] \right\}^\lambda \prod_{e \in \mathcal{L}(\mathcal{D})} \dot{q}(\sigma_e).$$

Note however that if M_+ is not of full rank then $\dot{q}$ need not be unique.

References

[AM06] D. Achlioptas and C. Moore. Random k-SAT: two moments suffice to cross a sharp threshold. *SIAM J. Comput.*, 36(3):740–762 (electronic), 2006.

[BC15a] V. Bapst and A. Coja-Oghlan. The condensation phase transition in the regular k-SAT model. arXiv:1507.03512, 2015.

[BC15b] V. Bapst and A. Coja-Oghlan. Harnessing the Bethe free energy. arXiv:1504.03975, 2015.

[BCH+16] V. Bapst, A. Coja-Oghlan, S. Hetterich, F. Raßmann, and D. Vilenchik. The condensation phase transition in random graph coloring. *Comm. Math. Phys.*, 341(2):543–606, 2016.

[BCR14] V. Bapst, A. Coja-Oghlan, and F. Raßmann. A positive temperature phase transition in random hypergraph 2-coloring. arXiv:1410.2190, 2014.

[BGT13] M. Bayati, D. Gamarnik, and P. Tetali. Combinatorial approach to the interpolation method and scaling limits in sparse random graphs. *Ann. Probab.*, 41(6):4080–4115, 2013.

[BMZ05] A. Braunstein, M. Mézard, and R. Zecchina. Survey propagation: an algorithm for satisfiability. *Random Structures Algorithms*, 27(2):201–226, 2005.

[COZ12] A. Coja-Oghlan and L. Zdeborová. The condensation transition in random hypergraph 2-coloring. In *Proc. 23rd SODA*, pages 241–250. ACM, New York, 2012.

[CP12] A. Coja-Oghlan and K. Panagiotou. Catching the k-NAE-SAT threshold. In *Proc. 44th STOC*, pages 899–907. ACM, New York, 2012.

[CP16a] A. Coja-Oghlan and K. Panagiotou. The asymptotic k-SAT threshold. *Adv. Math.*, 288:985–1068, 2016.

[CP16b] A. Coja-Oghlan and W. Perkins. Belief propagation on replica symmetric random factor graph models. arXiv:1603.08191, 2016.

[CPS15] A. Coja-Oghlan, W. Perkins, and K. Skubch. Limits of discrete distributions and Gibbs measures on random graphs. arXiv:1512.06798, 2015.

[DKMZ11] A. Decelle, F. Krząkała, C. Moore, and L. Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E*, 84(6):066106, 2011.

[DSS13] J. Ding, A. Sly, and N. Sun. Maximum independent sets on random regular graphs. arXiv:1310.4787, 2013.

[DSS14] J. Ding, A. Sly, and N. Sun. Satisfiability threshold for random regular NAE-SAT. In *Proc. 46th STOC*, New York, NY, USA, 2014. ACM.

[DSS15] J. Ding, A. Sly, and N. Sun. Proof of the satisfiability conjecture for large k. In *Proc. 47th STOC*, pages 59–68. ACM, New York, 2015.

[FL03] S. Franz and M. Leone. Replica bounds for optimization problems and diluted spin systems. *J. Statist. Phys.*, 111(3-4):535–564, 2003.

[Gam14] D. Gamarnik. Right-convergence of sparse random graphs. *Probab. Theory Related Fields*, 160(1-2):253–278, 2014.

[GT03] F. Guerra and F. L. Toninelli. Infinite volume limit and spontaneous replica symmetry breaking in mean field spin glass models. *Ann. Henri Poincaré*, 4(suppl. 1):S441–S444, 2003.

[Gue03] F. Guerra. Broken replica symmetry bounds in the mean field spin glass model. *Comm. Math. Phys.*, 233(1):1–12, 2003.

[KMR+07] F. Krząkała, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborová. Gibbs states and the set of solutions of random constraint satisfaction problems. *Proc. Natl. Acad. Sci. USA*, 104(25):10318–10323 (electronic), 2007.

[Mas14] L. Massoulié. Community detection thresholds and the weak Ramanujan property. In *Proc. 46th STOC*, pages 694–703. ACM, New York, 2014.

[MM09] M. Mézard and A. Montanari. *Information, physics, and computation*. Oxford Graduate Texts. Oxford University Press, Oxford, 2009.

[MMW07] E. Maneva, E. Mossel, and M. J. Wainwright. A new look at survey propagation and its generalizations. *J. ACM*, 54(4):Art. 17, 41, 2007.

[MMZ06] S. Mertens, M. Mézard, and R. Zecchina. Threshold values of random k-SAT from the cavity method. *Random Structures Algorithms*, 28(3):340–373, 2006.

[MNS15] E. Mossel, J. Neeman, and A. Sly. Reconstruction and estimation in the planted partition model. *Probab. Theory Related Fields*, 162(3-4):431–461, 2015.

[MRS08] A. Montanari, F. Ricci-Tersenghi, and G. Semerjian. Clusters of solutions and replica symmetry breaking in random k-satisfiability. *J. Stat. Mech.*, 2008(04):P04004, 2008.

[Pan13] D. Panchenko. *The Sherrington-Kirkpatrick model*. Springer Monographs in Mathematics. Springer, New York, 2013.

[Par02] G. Parisi. On local equilibrium equations for clustering states. arXiv:cs/0212047, 2002.

[PT04] D. Panchenko and M. Talagrand. Bounds for diluted mean-fields spin glass models. *Probab. Theory Related Fields*, 130(3):319–336, 2004.

[ZK07] L. Zdeborova and F. Krząkała. Phase transitions in the coloring of random graphs. *Phys. Rev. E*, 76(3):031131, 2007.