Received: 16 March 2020 | Accepted: 21 September 2020 | Published on: 17 January 2021

DOI: 10.1002/rsa.20993

RESEARCH ARTICLE

WILEY

The number of satisfying assignments of random 2-SAT formulas

Dimitris Achlioptas[1] | Amin Coja-Oghlan[2] | Max Hahn-Klimroth[2] | Joon Lee[2] | Noëla Müller[2] | Manuel Penschuck[3] | Guangyan Zhou[4]

[1]Department of Computer Science and Telecommunications, University of Athens, Ilissia, Greece
[2]Mathematics Institute, Goethe University, 10 Robert Mayer St Frankfurt, Germany
[3]Computer Science Institute, Goethe University, 11–15 Robert Mayer St Frankfurt, Germany
[4]School of Mathematics and Statistics, Beijing Technology and Business University, Beijing, China

Correspondence
Amin Coja-Oghlan, Mathematics Institute, Goethe University, 10 Robert Mayer Street, Frankfurt 60325, Germany.
Email: acoghlan@math.uni-frankfurt.de

Funding information
This research was supported by the DFG Stiftung Polytechnische Gesellschaft DFG ME National Natural Science Foundation of China, Grant/Award Numbers: CO 646/4, 2088/3-2, ME 2088/4-2, 61702019.

Abstract

We show that throughout the satisfiable phase the normalized number of satisfying assignments of a random 2-SAT formula converges in probability to an expression predicted by the cavity method from statistical physics. The proof is based on showing that the Belief Propagation algorithm renders the correct marginal probability that a variable is set to "true" under a uniformly random satisfying assignment.

KEYWORDS

2-SAT, Belief Propagation, satisfiability problem

1 | INTRODUCTION

1.1 | Background and motivation

The random 2-SAT problem was the first random constraint satisfaction problem whose satisfiability threshold could be pinpointed precisely, an accomplishment attained independently by Chvátal and Reed [15] and Goerdt [31] in 1992. The proofs evince the link between the 2-SAT threshold and the percolation phase transition of a random digraph. This connection subsequently enabled Bollobás, Borgs, Chayes, Kim, and Wilson [12] to identify the size of the scaling window, which matches that

of the giant component phase transition of the Erdős-Rényi random graph [11, 34]. Ramifications and extensions of these results pertain to random 2-SAT formulas with given literal degrees [20], the random MAX 2-SAT problem [21] and the performance of algorithms [46]. But despite the great attention devoted to random 2-SAT over the years, a fundamental question, mentioned prominently in the survey [29], remained conspicuously open: *how many satisfying assignments does a random 2-SAT formula typically possess?* While percolation-type arguments have been stretched to derive (rough) bounds [13], the exact answer remained beyond the reach of elementary techniques.

In addition to the mathematical literature, the 2-SAT problem attracted the interest of statistical physicists, who brought to bear a canny but nonrigorous approach called the cavity method [37, 38]. Instead of relying on percolation ideas, the physics *ansatz* seizes upon a heuristic message passing scheme called Belief Propagation. Its purpose is to calculate the marginal probabilities that a random satisfying assignment sets specific variables of the 2-SAT formula to "true". According to physics intuition Belief Propagation reveals a far more fine-grained picture than a mere percolation argument possibly could. Indeed, in combination with a functional called the Bethe free entropy, Belief Propagation renders a precise conjecture as to the number of satisfying assignments.

We prove this conjecture. Specifically, we show that for all clause-to-variable densities below the 2-SAT threshold the number of satisfying assignments is determined by the Bethe functional applied to a particular solution of a stochastic fixed point equation that mimics Belief Propagation. The formula that we obtain does not boil down to a simple algebraic expression, which may explain why the problem has confounded classical methods for nearly three decades. Nonetheless, thanks to rapid convergence of the stochastic fixed point iteration, the formula can be evaluated numerically within arbitrary precision. A crucial step towards the main theorem is to verify that Belief Propagation does indeed yield the correct marginals, a fact that may be of independent interest.

By comparison to prior work on Belief Propagation in combinatorics (e.g., [17, 22, 23, 40]), we face the substantial technical challenge of dealing with the "hard" constraints of the 2-SAT problems, which demands that *all* clauses be satisfied. A second novelty is that in order to prove convergence of Belief Propagation to the correct marginals we need to investigate delicately constructed extremal initial conditions for the message passing process. Since these depend on the random 2-SAT formula itself, we need to develop means to confront the ensuing stochastic dependencies between the construction of the initial condition and the subsequent message passing iterations. We proceed to state the main results precisely. An outline of the proofs and a detailed discussion of related work follow in Sections 2 and 3.

1.2 | The main result

Let $n > 1$ be an integer, let $d > 0$ be a positive real and let $\boldsymbol{m} \stackrel{\mathrm{d}}{=} \mathrm{Po}(dn/2)$ be a Poisson random variable. Furthermore, let $\boldsymbol{\Phi} = \boldsymbol{\Phi}_n$ be a random 2-SAT formula with Boolean variables $x_1, \ldots, x_n$ and $\boldsymbol{m}$ clauses, drawn uniformly and independently from the set of all $4n(n-1)$ possible clauses with two distinct variables. Thus, each variable appears in d clauses on the average and the satisfiability threshold occurs at $d = 2$. We aim to estimate the number $Z(\boldsymbol{\Phi})$ of satisfying assignments, the *partition function* in physics jargon. More precisely, since $Z(\boldsymbol{\Phi})$ remains exponentially large for all $d < 2$ w.h.p., in order to obtain a well-behaved limit we compute the normalized logarithm $n^{-1} \log Z(\boldsymbol{\Phi})$.

The result comes in terms of the solution to a stochastic fixed point equation on the unit interval. Hence, let $\mathscr{P}(0,1)$ be the set of all Borel probability measures on $(0,1)$, endowed with the weak topology. Furthermore, define an operator $\mathrm{BP}_d : \mathscr{P}(0,1) \to \mathscr{P}(0,1)$, $\pi \mapsto \hat{\pi}$ as follows. With $\boldsymbol{d}^+, \boldsymbol{d}^-$ Poisson variables with mean $d/2$ and $\boldsymbol{\mu}_{\pi,1}, \boldsymbol{\mu}_{\pi,2}, \ldots$ random variables with distribution π, all mutually

independent, let $\hat{\pi}$ be the distribution of the random variable

$$\frac{\prod_{i=1}^{d^-} \mu_{\pi,i}}{\prod_{i=1}^{d^-} \mu_{\pi,i} + \prod_{i=1}^{d^+} \mu_{\pi,i+d^-}} \in (0,1). \tag{1.1}$$

Let $\delta_{1/2} \in \mathscr{P}(0,1)$ signify the atom at 1/2 and write $\mathrm{BP}_d^{\ell}(\cdot)$ for the ℓ-fold application of the operator BP_d.

Theorem 1.1. *For any $d < 2$ the limit $\pi_d = \lim_{\ell\to\infty} \mathrm{BP}_d^{\ell}(\delta_{1/2})$ exists and*

$$\lim_{n\to\infty} \frac{1}{n} \log Z(\mathbf{\Phi}) = \mathbb{E}\left[\log\left(\prod_{i=1}^{d^-} \mu_{\pi_d,i} + \prod_{i=1}^{d^+} \mu_{\pi_d,i+d^-} \right) - \frac{d}{2} \log\left(1 - \mu_{\pi_d,1}\mu_{\pi_d,2}\right) \right] \textit{in probability.} \tag{1.2}$$

Of course, the fact that the r.h.s. of (1.2) is well-defined is part of the statement of Theorem 1.1. By construction, the distribution π_d is a solution to the stochastic fixed point equation

$$\pi_d = \mathrm{BP}_d(\pi_d). \tag{1.3}$$

Equation (1.3) is known as the *density evolution* equation in physics lore, while the expression on the r.h.s. of (1.2) is called the *Bethe free entropy* [35]. Hence, Theorem 1.1 matches the conjecture from [37]. By comparison, Markov's inequality (specifically, $Z(\mathbf{\Phi}) \le \mathbb{E}[Z(\mathbf{\Phi})] \cdot \exp(o(n))$ w.h.p.) yields the elementary first moment bound

$$\frac{1}{n} \log Z(\mathbf{\Phi}) \le \frac{1}{n} \log \mathbb{E}[Z(\mathbf{\Phi})] + o(1) = (1-d)\log 2 + \frac{d}{2}\log 3 + o(1) \text{ w.h.p.}, \tag{1.4}$$

which, however, fails to be tight for any $0 < d < 2$ [43]. Furthermore, while (1.2) may appear difficult to evaluate, the proof reveals that the fixed point iteration $\mathrm{BP}_d^{\ell}(\delta_{1/2})$ converges geometrically (in an appropriate metric). In effect, decent numerical approximations can be obtained; see Figure 1. A quick remark on the right figure is in order. The jumps on the graph are due to specific small tree components of the formula. For instance, isolated variables can take either $+1$ or -1 independently, which explains the jump at 1/2. In addition, isolated clauses with two variables that are not connected to any other clause render variable marginals 1/3 or 2/3, depending on the signs. This explains the other two noticeable jumps.

For $d < 1$ the random digraph on $\{x_1, \neg x_1, \ldots, x_n, \neg x_n\}$ obtained by inserting for each clause $l_1 \vee l_2$ of $\mathbf{\Phi}$ the two directed edges $\neg l_1 \to l_2$, $\neg l_2 \to l_1$ is sub-critical, thus guaranteeing no weak giant component and the distribution π_d is supported on a countable set. Such description of clause is also known as the implication graph, first introduced in [5]. In effect, for $d < 1$ the formula (1.2) can be obtained via elementary counting arguments. By contrast, the emergence of a weak giant component for $1 < d < 2$ turns the computation of $Z(\mathbf{\Phi})$ into a challenge. Finally, for $d > 2$ the digraph contains a strongly connected giant component w.h.p. Its long directed cycles likely cause contradictions, which is why satisfying assignments cease to exist.

An asymptotically tight upper bound on $n^{-1}\log Z(\mathbf{\Phi})$ could be obtained via the interpolation method from mathematical physics [30, 43]. We will revisit this point in Section 3. Thus, the principal contribution of Theorem 1.1 is the lower bound on $\log Z(\mathbf{\Phi})$. The best prior lower bound was obtained

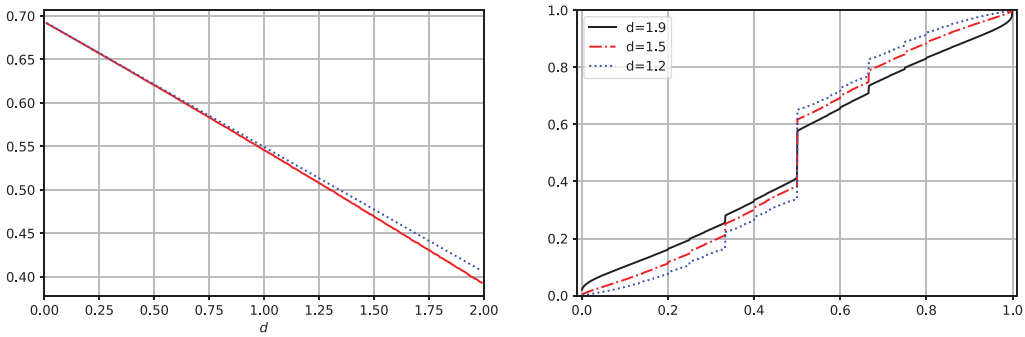

FIGURE 1 *Left:* the red line depicts a numerical approximation to the r.h.s. of (1.2) after 24 iterations of $\mathrm{BP}_d(\cdot)$. The dotted blue line displays the first moment bound. *Right:* the cumulative density functions of numerical approximations to $\mathrm{BP}_d^{24}(\delta_{1/2})$ for various d

by Boufkhad and Dubois [13] in 1999 via percolation arguments. However, this bound drastically undershoots the actual value from Theorem 1.1. For instance, for $d = 1.2$, [13] gives $n^{-1} \log Z(\mathbf{\Phi}) \geq 0.072 \ldots$, while actually $n^{-1} \log Z(\mathbf{\Phi}) = 0.515 \ldots$ w.h.p.

1.3 | Belief propagation

To elaborate on the combinatorial meaning of the distribution π_d, we need to look into the Belief Propagation heuristic. Instantiated to 2-SAT, Belief Propagation is a message passing algorithm designed to approximate the marginal probability that a specific variable takes the value "true" under a random satisfying assignment. While finding satisfying assignments of a given 2-SAT formula is an easy computational task, calculating these marginals is not. In fact, the problem is #P-hard [50]. Nonetheless, we are going to prove that Belief Propagation approximates the marginals well on random formulas w.h.p.

To introduce Belief Propagation, we associate a bipartite graph $G(\mathbf{\Phi})$ with the formula $\mathbf{\Phi}$. One vertex class $V_n = \{x_1, \ldots, x_n\}$ represents the propositional variables, the other class $F_{\boldsymbol{m}} = \{a_1, \ldots, a_{\boldsymbol{m}}\}$ represents the clauses. Each clause a_i is adjacent to the two variables that it contains. Because each clause contains two distinct variables and each of the $\boldsymbol{m}$ clauses is represented by a separate vertex, $G(\mathbf{\Phi})$ is a simple graph. We write $\partial v = \partial(\mathbf{\Phi}, v)$ for the set of neighbors of a vertex v of $G(\mathbf{\Phi})$. Moreover, for $\ell \geq 1$ let $\partial^\ell v$ signify the set of all vertices at distance precisely ℓ from v.

Associated with the edges of $G(\mathbf{\Phi})$, the Belief Propagation messages are probability distributions on the Boolean values "true" and "false." To be precise, any adjacent clause/variable pair a, x comes with two messages, one directed from a to x and a reverse one from x to a. Encoding "true" and "false" by ± 1, we initialize all messages by

$$\nu^{(0)}_{\mathbf{\Phi},a\to x}(\pm 1) = \nu^{(0)}_{\mathbf{\Phi},x\to a}(\pm 1) = 1/2. \tag{1.5}$$

For $\ell \geq 1$ the messages $\nu^{(\ell)}_{\mathbf{\Phi},a\to x}, \nu^{(\ell)}_{\mathbf{\Phi},x\to a}$ are defined inductively. Specifically, suppose that clause a contains the two variables x, y. Let $r, s \in \{\pm 1\}$ indicate whether x, y appear as positive or negative literals in a. Then for $t = \pm 1$ let

$$\nu^{(\ell)}_{\mathbf{\Phi},a\to x}(t) = \frac{1 - \mathbf{1}\{r \neq t\}\, \nu^{(\ell-1)}_{\mathbf{\Phi},y\to a}(-s)}{1 + \nu^{(\ell-1)}_{\mathbf{\Phi},y\to a}(s)},$$

$$\nu^{(\ell)}_{\mathbf{\Phi},x\to a}(t) = \frac{\prod_{b\in\partial x\setminus\{a\}} \nu^{(\ell)}_{\mathbf{\Phi},b\to x}(t)}{\prod_{b\in\partial x\setminus\{a\}} \nu^{(\ell)}_{\mathbf{\Phi},b\to x}(1) + \prod_{b\in\partial x\setminus\{a\}} \nu^{(\ell)}_{\mathbf{\Phi},b\to x}(-1)}. \tag{1.6}$$

The last expression is deemed to equal 1/2 if the denominator vanishes (which does not happen if $\mathbf{\Phi}$ is satisfiable). Finally, the Belief Propagation estimate of the marginal of a variable x after ℓ iterations reads

$$\nu^{(\ell)}_{\mathbf{\Phi},x}(t) = \frac{\prod_{a\in\partial x} \nu^{(\ell)}_{\mathbf{\Phi},a\to x}(t)}{\prod_{a\in\partial x} \nu^{(\ell)}_{\mathbf{\Phi},a\to x}(1) + \prod_{a\in\partial x} \nu^{(\ell)}_{\mathbf{\Phi},a\to x}(-1)}, \tag{1.7}$$

again interpreted to yield 1/2 if the denominator vanishes. For an excellent exposition of Belief Propagation, including the derivation of (1.6)–(1.7), we point to [35, Chapter 14].

The next theorem establishes that (1.7) approximates the true marginals well for large ℓ. In fact, we prove a significantly stronger result. To set the stage, let $S(\mathbf{\Phi})$ be the set of all satisfying assignments of $\mathbf{\Phi}$. Assuming $S(\mathbf{\Phi}) \neq \emptyset$, let

$$\mu_{\mathbf{\Phi}}(\sigma) = \mathbf{1}\{\sigma \in S(\mathbf{\Phi})\}/Z(\mathbf{\Phi}) \ (\sigma \in \{\pm 1\}^{\{x_1,\ldots,x_n\}}) \tag{1.8}$$

be the uniform distribution on $S(\mathbf{\Phi})$. Furthermore, write $\boldsymbol{\sigma}$ for a sample from $\mu_{\mathbf{\Phi}}$. Then for a satisfying assignment $\tau \in S(\mathbf{\Phi})$ and $\ell \geq 1$ the conditional distribution $\mu_{\mathbf{\Phi}}(\cdot\,|\boldsymbol{\sigma}_{\partial^{2\ell}x_1} = \tau_{\partial^{2\ell}x_1}) = \mu_{\mathbf{\Phi}}(\cdot\,|\forall y \in \partial^{2\ell}x_1 : \boldsymbol{\sigma}_y = \tau_y)$ imposes the "boundary condition" τ on all variables y at distance 2ℓ from x_1. The following theorem shows that Belief Propagation does not just approximate the plain, unconditional marginals well w.h.p., but even the conditional marginals given any conceivable boundary condition. Recall that $\mathbb{P}[Z(\mathbf{\Phi}) > 0] = 1 - o(1)$ for $d < 2$.

Theorem 1.2. *If $d < 2$, then*

$$\lim_{\ell\to\infty}\limsup_{n\to\infty} \mathbb{E}\left[\max_{\tau\in S(\mathbf{\Phi})} \left|\mu_{\mathbf{\Phi}}(\boldsymbol{\sigma}_{x_1} = 1|\boldsymbol{\sigma}_{\partial^{2\ell}x_1} = \tau_{\partial^{2\ell}x_1}) - \nu^{(\ell)}_{\mathbf{\Phi},x_1}(1)\right| \;\middle|\; Z(\mathbf{\Phi}) > 0\right] = 0. \tag{1.9}$$

Since $\nu^{(\ell)}_{\mathbf{\Phi},x_1}$ does not depend on τ, averaging (1.9) on the boundary condition $\tau \in S(\mathbf{\Phi})$ yields

$$\lim_{\ell\to\infty}\limsup_{n\to\infty} \mathbb{E}\left[\left|\mu_{\mathbf{\Phi}}(\boldsymbol{\sigma}_{x_1} = \pm 1) - \nu^{(\ell)}_{\mathbf{\Phi},x_1}(\pm 1)\right| | Z(\mathbf{\Phi}) > 0\right] = 0. \tag{1.10}$$

Thus, Belief Propagation approximates the unconditional marginal of x_1 well in the limit of large n and ℓ. Indeed, because the distribution of $\mathbf{\Phi}$ is invariant under permutations of the variables $x_1, \ldots, x_n$, (1.10) implies that the marginals of all but $o(n)$ variables x_i are within $\pm o(1)$ of the Belief Propagation approximation w.h.p.

But thanks to the presence of the boundary condition τ, Theorem 1.2 leads to further discoveries. For a start, applying the triangle inequality to (1.9) and (1.10), we obtain

$$\lim_{\ell\to\infty}\limsup_{n\to\infty} \mathbb{E}\left[\max_{\tau\in S(\mathbf{\Phi})} \left|\mu_{\mathbf{\Phi}}(\boldsymbol{\sigma}_{x_1} = 1|\boldsymbol{\sigma}_{\partial^{2\ell}x_1} = \tau_{\partial^{2\ell}x_1}) - \mu_{\mathbf{\Phi}}(\boldsymbol{\sigma}_{x_1} = 1)\right| \;\middle|\; Z(\mathbf{\Phi}) > 0\right] = 0. \tag{1.11}$$

Thus, no discernible shift of the marginal of x_1 is likely to ensue upon imposition of any possible boundary condition τ. The spatial mixing property (1.11) is colloquially known as *Gibbs uniqueness* [33].

Furthermore, (1.11) rules out extensive long-range correlations. Specifically, for any fixed ℓ the first two variables x_1, x_2 with high probability have distance greater than 4ℓ in $G(\mathbf{\Phi})$. Therefore, (1.11) implies that for all $d < 2$,

$$\lim_{n\to\infty} \sum_{s,t\in\{\pm 1\}} \mathbb{E}\left[\left|\mu_{\mathbf{\Phi}}(\boldsymbol{\sigma}_{x_1} = s, \boldsymbol{\sigma}_{x_2} = t) - \mu_{\mathbf{\Phi}}(\boldsymbol{\sigma}_{x_1} = s)\cdot \mu_{\mathbf{\Phi}}(\boldsymbol{\sigma}_{x_2} = t)\right| \;\middle|\; Z(\mathbf{\Phi}) > 0\right] = 0. \tag{1.12}$$

Thus, the truth values $\boldsymbol{\sigma}_{x_1}, \boldsymbol{\sigma}_{x_2}$ are asymptotically independent. Of course, once again by permutation invariance, (1.12) implies that asymptotic independence extends to all but $o(n^2)$ pairs of variables x_i, x_j w.h.p. The decorrelation property (1.12) is called *replica symmetry* in the physics literature [33].

Finally, we can clarify the combinatorial meaning of the distribution π_d from Theorem 1.1. Namely, π_d is the limit of the empirical distribution of the marginal probabilities $\mu_{\mathbf{\Phi}}(\boldsymbol{\sigma}_{x_i} = 1)$.

Corollary 1.3. *For any $0 < d < 2$ the random probability measure*

$$\pi_{\mathbf{\Phi}} = \frac{1}{n}\sum_{i=1}^{n} \delta_{\mu_{\mathbf{\Phi}}(\boldsymbol{\sigma}_{x_i}=1)} \tag{1.13}$$

converges to π_d weakly in probability. 1

Thus, the stochastic fixed point equation (1.3) that characterizes π_d simply expresses that the marginal probabilities $\mu_{\mathbf{\Phi}}(\boldsymbol{\sigma}_{x_i} = 1)$ result from the Belief Propagation recurrence (1.6).

1.4 | Preliminaries and notation

Throughout we denote by $V_n = \{x_1, \ldots, x_n\}$ the variable set of $\mathbf{\Phi}_n$. Generally, given a 2-SAT formula Φ we write $V(\Phi)$ for the set of variables and $F(\Phi)$ for the set of clauses. The bipartite clause/variable-graph $G(\Phi)$ is defined as in Section 1.3. For a vertex v of $G(\Phi)$ we let $\partial(\Phi, v)$ be the set of neighbors. Where Φ is apparent we just write ∂v. Moreover, $\partial^\ell(\Phi, v)$ or briefly $\partial^\ell v$ stands for the set of vertices at distance exactly ℓ from v. Additionally, $\nabla^\ell(\Phi, v)$ denotes the sub-formula obtained from Φ by deleting all clauses and variables at distance greater than ℓ from v. This subformula may contain clauses of length less than two. Furthermore, for a clause a and a variable x of Φ we let $sign(x, a) = sign_\Phi(x, a) \in \{\pm 1\}$ be the sign with which x appears in a. In addition, we let $S(\Phi)$ be the set of all satisfying assignments of Φ, $Z(\Phi) = |S(\Phi)|$ and, assuming $Z(\Phi) > 0$, we let μ_Φ be the probability distribution on $\{\pm 1\}^{V(\Phi)}$ that induces the uniform distribution on $S(\Phi)$ as in (1.8). Moreover, $\boldsymbol{\sigma}_\Phi = (\boldsymbol{\sigma}_{\Phi,x})_{x\in V(\Phi)}$ signifies a uniformly random satisfying assignment; we drop Φ where the reference is apparent.

For any Φ we set up Belief Propagation as in (1.5)–(1.7). It is well known that Belief Propagation yields the correct marginals if $G(\Phi)$ is a tree. To be precise, the *depth* of $x \in V(\Phi)$ is the maximum distance between x and a leaf of $G(\Phi)$.

Proposition 1.4 ([35, Theorem 14.1])**.** *If $G(\Phi)$ is a tree and $x \in V(\Phi)$, then for any ℓ greater than or equal to the depth of x we have $\mu_\Phi(\boldsymbol{\sigma}_x = \pm 1) = \nu^{(\ell)}_{\Phi,x}(\pm 1)$.*

We will encounter the following functions repeatedly. For $\varepsilon > 0$ let

$$\Lambda_\varepsilon(z) = \log(z \vee \varepsilon) \tag{1.14}$$

1That is, for any continuous function $f : [0, 1] \to \mathbb{R}$ we have $\lim_{n\to\infty}\mathbb{E}\left|\int_0^1 f(z)\mathrm{d}\pi_d(z) - \int_0^1 f(z)\mathrm{d}\pi_{\mathbf{\Phi}}(z)\right| = 0$.

be the log function truncated at $\log \varepsilon$. Moreover, we need the continuous and mutually inverse functions

$$\psi : \mathbb{R} \to (0,1), \quad z \mapsto (1+\tanh(z/2))/2, \quad \varphi : (0,1) \to \mathbb{R}, \quad p \mapsto \log(p/(1-p)). \tag{1.15}$$

Let $\mathscr{P}(\mathbb{R})$ be the set of all Borel probability measures on $\mathbb{R}$ with the weak topology. Moreover, for a real $q \geq 1$ let $\mathscr{W}_q(\mathbb{R})$ be the set of all $\rho \in \mathscr{P}(\mathbb{R})$ such that $\int_{\mathbb{R}} |x|^q \mathrm{d}\rho(x) < \infty$. We equip this space with the Wasserstein metric

$$W_q(\rho,\rho') = \inf\left\{ \left(\int_{\mathbb{R}^2} |x-y|^q \mathrm{d}\gamma(x,y) \right)^{1/q} : \gamma \text{ is a coupling of } \rho,\rho' \right\}, \tag{1.16}$$

thereby turning $\mathscr{W}_q(\mathbb{R})$ into a complete separable space [10].

For $\rho \in \mathscr{P}(\mathbb{R})$ we denote by $\boldsymbol{\eta}_\rho, \boldsymbol{\eta}_{\rho,1}, \boldsymbol{\eta}_{\rho,2}, \ldots$ random variables with distribution ρ. Similarly, for $\pi \in \mathscr{P}(0,1)$ we let $\boldsymbol{\mu}_\pi, \boldsymbol{\mu}_{\pi,1}, \boldsymbol{\mu}_{\pi,2}, \ldots$ be a sequence of random variables with distribution π. We continue to let $\boldsymbol{d}^+, \boldsymbol{d}^-$ Poisson variables with mean $d/2$ and also define $\boldsymbol{d}$ be a Poisson variable with mean d. Moreover, $\boldsymbol{s}_1, \boldsymbol{s}_1', \boldsymbol{s}_2, \boldsymbol{s}_2', \ldots \in \{\pm 1\}$ always denote uniformly distributed random variables. All of these random variables are mutually independent as well as independent of any other sources of randomness.

Finally, from here on we tacitly assume that $0 < d < 2$.

2 | OVERVIEW

The proof of Theorem 1.1 proceeds in four steps. First we show that the limit π_d from Theorem 1.1 exists. Subsequently we establish the fact (1.9) that Belief Propagation approximates the conditional marginals well. This will easily imply the convergence of the empirical marginals (1.13) to π_d. Third, building upon these preparations, we will prove that the truncated mean $n^{-1}\mathbb{E}[\log(Z(\boldsymbol{\Phi}) \vee 1)]$ converges to the r.h.s. of (1.2). The truncation is necessary to deal with the unlikely event that $Z(\boldsymbol{\Phi}) = 0$ which occurs with probability tending to 0. Finally, we will show that $\log(Z(\boldsymbol{\Phi}) \vee 1)$ concentrates about its mean to obtain convergence in probability, thus completing the proof of Theorem 1.1.

2.1 | Step 1: Density evolution

We begin by verifying that the distribution π_d from Theorem 1.1 is well-defined and that π_d satisfies a tail bound.

Proposition 2.1. *The weak limit $\pi_d = \lim_{\ell\to\infty} \mathrm{BP}_d^\ell(\delta_{1/2})$ exists and*

$$\mathbb{E}\left[\log^2 \frac{\boldsymbol{\mu}_{\pi_d}}{1-\boldsymbol{\mu}_{\pi_d}}\right] < \infty. \tag{2.1}$$

Moreover, $\boldsymbol{\mu}_{\pi_d}$ and $1-\boldsymbol{\mu}_{\pi_d}$ are identically distributed and

$$\mathbb{E}\left|\log\left(\prod_{i=1}^{\boldsymbol{d}^-} \boldsymbol{\mu}_{\pi_d,i} + \prod_{i=1}^{\boldsymbol{d}^+} \boldsymbol{\mu}_{\pi_d,i+\boldsymbol{d}^-}\right)\right| < \infty, \qquad \mathbb{E}\left|\log\left(1-\boldsymbol{\mu}_{\pi_d,1}\boldsymbol{\mu}_{\pi_d,2}\right)\right| < \infty. \tag{2.2}$$

The proof of Proposition 2.1, which we carry out in Section 4, is based on a contraction argument. This argument implies that the fixed point iteration converges rapidly to π_d, a fact that can be exploited to obtain numerical estimates. The bounds (2.2) ensure that the expectation on the r.h.s. of (1.2) is well-defined.

2.2 | Step 2: Gibbs uniqueness

As a next step we verify the Gibbs uniqueness property (1.11). We proceed by way of analyzing a multi-type Galton–Watson tree $\boldsymbol{T}$ that mimics the local structure of the graph $G(\boldsymbol{\Phi})$ upon exploration from variable x_1. The Galton–Watson process has five types: variable nodes and four types of clause nodes $(+1,+1),(+1,-1),(-1,+1),(-1,-1)$. The root is a variable node o. Moreover, each variable node spawns independent Po(d/4) numbers of clauses nodes of each of the four types. Additionally, each clause has a single offspring, which is a variable. The semantics of the clause types is that the first component indicates whether the parent variable appears in the clause positively or negatively. The second component indicates whether the child variable appears as a positive or as a negative literal. Clearly, for $d \leq 1$ the tree $\boldsymbol{T}$ is finite with probability one, while infinite trees appear with positive probability for $d > 1$.

Let $\boldsymbol{T}^{(\ell)}$ be the finite tree obtained from $\boldsymbol{T}$ by dropping all nodes at distance greater than ℓ from the root. For even ℓ it will be convenient to view $\boldsymbol{T}^{(\ell)}$ interchangeably as a tree or as a 2-SAT formula. In particular, we write $\partial^{2\ell} o = \partial^{2\ell}(\boldsymbol{T}, o)$ for the set of all variables at distance exactly 2ℓ from o. The following proposition, which is the linchpin of the entire proof strategy, establishes the Gibbs uniqueness property for the tree formula $\boldsymbol{T}^{(2\ell)}$.

Proposition 2.2. *We have*

$$\lim_{\ell\to\infty} \mathbb{E}\left[\max_{\tau\in S(\boldsymbol{T}^{(2\ell)})} \left|\mu_{\boldsymbol{T}^{(2\ell)}}(\boldsymbol{\sigma}_o = 1 | \boldsymbol{\sigma}_{\partial^{2\ell} o} = \tau_{\partial^{2\ell} o}) - \mu_{\boldsymbol{T}^{(2\ell)}}(\boldsymbol{\sigma}_o = 1)\right|\right] = 0. \tag{2.3}$$

Thus, w.h.p. no conceivable boundary condition is apt to significantly shift the marginal of the root.

We prove Proposition 2.2 by a subtle contraction argument in combination with a construction of extremal boundary conditions of the tree formula $\boldsymbol{T}^{(2\ell)}$. More specifically, we will construct boundary conditions $\boldsymbol{\sigma}^{\pm}$ that maximize or minimize the conditional probability

$$\mu_{\boldsymbol{T}^{(2\ell)}}(\boldsymbol{\sigma}_o = 1 | \boldsymbol{\sigma}_{\partial^{2\ell} o} = \boldsymbol{\sigma}^{\pm}_{\partial^{2\ell} o}), \tag{2.4}$$

respectively. Then we will show that the difference of the conditional marginals induced by both these extremal boundary conditions vanishes with probability tending to one as $\ell \to \infty$. The delicate point is that the extremal boundary conditions $\boldsymbol{\sigma}^{\pm}$ depend on the tree $\boldsymbol{T}^{(2\ell)}$. Thus, at first glance it seems that we need to pass the tree twice, once top–down to construct $\boldsymbol{\sigma}^{\pm}$ and then bottom–up to calculate the conditional marginals (2.4). But such an analysis seems untenable because after the top–down pass the tree is exposed and "no randomness remains" to facilitate the bottom–up phase. Fortunately, we will see that a single stochastic fixed point equation captures both the top–down and the bottom–up phase. This discovery reduces the proof of Proposition 2.2 to showing that the fixed point iteration contracts. The details of this delicate argument can be found in Section 5.

Proposition 2.2 easily implies the Gibbs uniqueness condition (1.11) and thereby Theorem 1.2. A further consequence is the asymptotic independence of the joint truth values of bounded numbers of variables.

Corollary 2.3. *The statement (1.9) is true and for any integer $k \geq 2$ we have*

$$\lim_{n\to\infty} \sum_{\sigma\in\{\pm 1\}^k} \mathbb{E}\left[\left|\mu_{\mathbf{\Phi}}(\boldsymbol{\sigma}_{x_1} = \sigma_1, \ldots, \boldsymbol{\sigma}_{x_k} = \sigma_k) - \prod_{i=1}^{k} \mu_{\mathbf{\Phi}}(\boldsymbol{\sigma}_{x_i} = \sigma_i)\right| \,|Z(\mathbf{\Phi}) > 0\right] = 0.$$

2.3 | Step 3: The Aizenman–Sims–Starr scheme

The aforementioned results pave the way for deriving an expression for the conditional expectation of $\log Z(\mathbf{\Phi})$ given that $\mathbf{\Phi}$ is satisfiable. Since $\mathbf{\Phi}$ is satisfiable w.h.p. for all $d < 2$, an equivalent task is to calculate $\mathbb{E}[\log(Z(\mathbf{\Phi}) \vee 1)]$. To this end we seize upon a simple but powerful strategy colloquially called the Aizenman-Sims-Starr scheme [6]. Originally proposed in the context of the Sherrington-Kirkpatrick spin glass model, this proof strategy suggests to compute the asymptotic mean of a random variable on a "system" of size n by carefully estimating the change of that mean upon going to a "system" of size $n+1$. This difference is calculated by coupling the systems of size n and $n+1$ such that the latter is obtained from the former by a small expected number of local changes.

We apply this idea to the random 2-SAT problem by coupling the random formula $\mathbf{\Phi}_n$ with n variables and $\mathrm{Po}(dn/2)$ clauses and the random formula $\mathbf{\Phi}_{n+1}$ with $n+1$ variables and $\mathrm{Po}(d(n+1)/2)$ clauses. Roughly speaking, we obtain $\mathbf{\Phi}_{n+1}$ from $\mathbf{\Phi}_n$ by adding a new variable x_{n+1} along with a few random adjacent clauses that connect x_{n+1} with the variables $x_1, \ldots, x_n$ of $\mathbf{\Phi}_n$. Then the information about the joint distribution of the truth values of bounded numbers of variables furnished by Corollaries 1.3 and 2.3 and the tail bound (2.1) will enable us to accurately estimate $\mathbb{E}\left[\log(Z(\mathbf{\Phi}_{n+1}) \vee 1) - \log(Z(\mathbf{\Phi}_n) \vee 1)\right]$.

Needless to say, upon closer inspection matters will emerge to be rather subtle. The main source of complications is that, in contrast to other models in mathematical physics such as the Sherrington-Kirkpatrick model or the Ising model, the 2-SAT problem has hard constraints. Thus, the addition of a single clause could trigger a dramatic drop in the partition function. In fact, in the worst case a single awkward clause could wipe out all satisfying assignments. In Section 6 we will iron out all these difficulties and prove the following.

Proposition 2.4. *We have*

$$\begin{aligned}&\lim_{n\to\infty} \mathbb{E}[\log(Z(\mathbf{\Phi}_{n+1}) \vee 1)] - \mathbb{E}[\log(Z(\mathbf{\Phi}_n) \vee 1)] \\ &\quad = \mathbb{E}\left[\log\left(\prod_{i=1}^{d^-} \boldsymbol{\mu}_{\pi_d,i} + \prod_{i=1}^{d^+} \boldsymbol{\mu}_{\pi_d,i+d^-}\right) - \frac{d}{2}\log\left(1 - \boldsymbol{\mu}_{\pi_d,1}\boldsymbol{\mu}_{\pi_d,2}\right)\right]. \end{aligned} \tag{2.5}$$

We notice that (2.2) guarantees that the r.h.s. of (2.5) is well-defined. As an immediate consequence of Proposition 2.4 we obtain a formula for $\mathbb{E}[\log(Z(\mathbf{\Phi}) \vee 1)]$.

Corollary 2.5. *For any $d < 2$ we have*

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}[\log(Z(\mathbf{\Phi}) \vee 1)] = \mathbb{E}\left[\log\left(\prod_{i=1}^{d^-} \boldsymbol{\mu}_{\pi_d,i} + \prod_{i=1}^{d^+} \boldsymbol{\mu}_{\pi_d,i+d^-}\right) - \frac{d}{2}\log\left(1 - \boldsymbol{\mu}_{\pi_d,1}\boldsymbol{\mu}_{\pi_d,2}\right)\right].$$

Proof. Writing $\mathbb{E}[\log(Z(\mathbf{\Phi}) \vee 1)]$ as a telescoping sum and applying Proposition 2.4, we obtain

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\log(Z(\mathbf{\Phi}_n)\vee 1)] = \lim_{n\to\infty}\frac{1}{n}\sum_{N=2}^{n-1}\mathbb{E}[\log(Z(\mathbf{\Phi}_{N+1})\vee 1)] - \mathbb{E}[\log(Z(\mathbf{\Phi}_N)\vee 1)]$$
$$= \mathbb{E}\left[\log\left(\prod_{i=1}^{d^-}\boldsymbol{\mu}_{\pi_d,i} + \prod_{i=1}^{d^+}\boldsymbol{\mu}_{\pi_d,i+d^-}\right) - \frac{d}{2}\log\left(1-\boldsymbol{\mu}_{\pi_d,1}\boldsymbol{\mu}_{\pi_d,2}\right)\right],$$

as desired. ∎

2.4 | Step 4: Concentration

The final step towards Theorem 1.1 is to show that $\log(Z(\mathbf{\Phi})\vee 1)$ concentrates about its mean.

Proposition 2.6. *We have* $\lim_{n\to\infty} n^{-1}\mathbb{E}\,|\log(Z(\mathbf{\Phi})\vee 1) - \mathbb{E}[\log(Z(\mathbf{\Phi})\vee 1)]| = 0$.

Proposition 2.6 does not easily follow from routine arguments such as the Azuma–Hoeffding inequality. Once more the issue is that changing a single clause could alter $\log(Z(\mathbf{\Phi})\vee 1)$ by as much as $\Theta(n)$. Instead we will resort to another technique from mathematical physics called the interpolation method. The details can be found in Section 7.

Proof of Theorem 1.1. The theorem follows from Proposition 2.1, Corollary 2.5 and Proposition 2.6. ∎

3 | DISCUSSION

The random 2-SAT satisfiability threshold was established mathematically shortly after the experimental work of Cheeseman, Kanefsky, and Taylor [14] that triggered the quest for satisfiability thresholds appeared. The second successful example, nearly a decade later, was the random 1-in-k-SAT threshold (to satisfy exactly one literal in each clause), which Achlioptas, Chtcherba, Istrate, and Moore pinpointed by analyzing the Unit Clause algorithm [2]. In a subsequent landmark contribution Dubois and Mandler determined the 3-XORSAT threshold via the second moment method [28]. Subsequent work extended this result to random k-XORSAT [24, 44]. Finally, the most notable success thus far has been the verification of the "1RSB cavity method" prediction [36] of the random k-SAT threshold for large k due to Ding, Sly and Sun [26], the culmination of a line of work that refined the use of the second moment method [3, 4, 18].

Over the past two decades the general theme of estimating the partition functions of discrete structures has received a great deal of attention; e.g., [9]. With respect to random 2-SAT (and, more generally, k-SAT), Montanari and Shah [40], Panchenko [42] and Talagrand [49] investigated "soft" versions of the partition function. To be precise, introducing a parameter $\beta > 0$ called the "inverse temperature", these articles study the random variable

$$Z_\beta(\mathbf{\Phi}) = \sum_{\sigma\in\{\pm 1\}^n}\prod_{i=1}^{m}\exp\left(-\beta\mathbf{1}\left\{\sigma \text{ violates clause } a_i\right\}\right). \tag{3.1}$$

Thus, instead of dismissing assignments that fail to satisfy all clauses outright, there is an $\exp(-\beta)$ penalty factor for each violated clause. Talagrand [49] computes $\lim_{n\to\infty} n^{-1}\mathbb{E}[\log Z_\beta(\mathbf{\Phi})]$ for β not exceeding a small but unspecified $\beta_0 > 0$. Panchenko [42] calculates this limit under the assumption $(4\beta \wedge 1)d < 1$. Thus, for $\beta > 1/4$ the result is confined to $d < 1$, in which case the random graph $G(\mathbf{\Phi})$ is sub-critical and both $Z_\beta(\mathbf{\Phi})$ and the actual number $Z(\mathbf{\Phi})$ of satisfying assignments could be calculated via elementary methods. Furthermore, Montanari and Shah [40] obtain $\lim_{n\to\infty} n^{-1}\mathbb{E}[\log Z_\beta(\mathbf{\Phi})]$ for all finite β under the assumption $d < 1.16\ldots$. Although for any fixed formula $\mathbf{\Phi}$ the limit $\lim_{\beta\to\infty} Z_\beta(\mathbf{\Phi})$ is equal to the number of satisfying assignments, it is not possible to interchange the limits $\beta \to \infty$ and $n \to \infty$. Thus, [40, 42] do not yield the number of actual satisfying assignments even for $d < 1.16\ldots$ or $d < 1$, respectively. Apart from estimating $\mathbb{E}\log Z_\beta(\mathbf{\Phi})$, Montanari and Shah [40] also show that the Belief Propagation message passing scheme approximates the marginals of the Boltzmann distribution that goes with $Z_\beta(\mathbf{\Phi})$ well, that is, they obtain a "soft" version of Theorem 1.2 for $d < 1.16\ldots$.

In terms of proof techniques, all three contributions [40, 42, 49] are based on establishing the Gibbs uniqueness property. So is the present paper. But while [40, 42, 49] rely on relatively straightforward contraction arguments, a key distinction is that here we develop a more accurate (and delicate) method for verifying the Gibbs uniqueness property based on the explicit construction of an extremal boundary condition. This is the key to pushing the range of d all the way up to the satisfiability threshold $d = 2$.

Specifically, in order to construct a boundary condition of the random tree $\boldsymbol{T}^{(2\ell)}$ for large ℓ that maximizes the conditional probability of observing the truth value $+1$ at the root we will work our way top–down from the root to level 2ℓ. Exposing the degrees and the signs with which the variables appear, the construction assigns a "desired" truth value to each variable of the tree so as to nudge the parent variable towards its desired value as much as possible. Subsequently, once this process reaches the bottom level of the tree, we go into reverse gear and study the Belief Propagation messages bottom–up to calculate the conditional marginal of the root. Clearly, analyzing this upwards process seems like a tall order because the tree was already exposed during the top-down phase, a challenge that is exacerbated by the presence of hard constraints. Fortunately, in Section 5 we will see how this problem can be transformed into the study of another stochastic fixed point equation that captures the effect of the children's "nudging" their parents. This fixed point problem is amenable to the contraction method. A spatial mixing analysis from an extremal boundary condition was previously conducted in by Dembo and Montanari [22] for the Ising model on random graphs. But of course a crucial difference is that in the Ising model the extremal boundary conditions are constant (all-$+1$ and all--1, respectively).

A second novelty of the present work is that we directly deal with the "hard" 2-SAT problem. Montanari and Shah [40] interpolate on the "inverse temperature" parameter $\beta > 0$, effectively working their way from smaller to larger β. Because the limits $\beta \to \infty$ and $n \to \infty$ do not commute, this approach does not seem applicable to problems with hard constraints. Furthermore, while Panchenko [41, 42] applies the Aizenman-Sims-Starr scheme to the soft constraint version, the hard problem of counting actual satisfying assignments requires a far more careful analysis. Indeed, adding one clause can shift $\log Z_\beta(\mathbf{\Phi})$ merely by $\pm\beta$. By contrast, a single additional clause could very well reduce the logarithm $\log Z(\mathbf{\Phi})$ of the number of satisfying assignments by as much as $\Omega(n)$, or even render the formula unsatisfiable. A few prior applications of the Aizenman–Sims–Starr scheme to problems with hard constraints exist [8, 16, 17], but these hinge on peculiar symmetry properties that enable an indirect approach via a "planted" version of the problem in question. The required symmetries for this approach are absent in several important problems, with random satisfiability the most prominent example. Thus, a significant technical contribution of the present work is that we show how to apply the Aizenman–Sims–Starr scheme directly to problems with hard constraints. Among other things, this requires a careful quantification of the probabilities of rare, potentially cataclysmic events in comparison to their impact on $\log Z(\mathbf{\Phi})$. That said, we should point out that [40, 42, 49] actually also deal with

the (soft) k-SAT partition function for $k > 2$ for certain regimes of clause/variable densities, while the technique that we develop here does not seem to extend beyond binary problems.

A mathematical physics technique called the interpolation method, first proposed by Guerra for the study of the Sherrington–Kirkpatrick model [32], can be applied to the random k-SAT problem [30, 43] to bound the number of satisfying assignments from above. For $k = 2$ the interpolation method yields the upper bound

$$\frac{1}{n}\log Z(\mathbf{\Phi}) \leq \inf_{\pi\in\mathscr{P}(0,1)} \mathbb{E}\left[\log\left(\prod_{i=1}^{d^-}\boldsymbol{\mu}_{\pi,i} + \prod_{i=1}^{d^+}\boldsymbol{\mu}_{\pi,i+d^-}\right) - \frac{d}{2}\log\left(1-\boldsymbol{\mu}_{\pi,1}\boldsymbol{\mu}_{\pi,2}\right)\right] + o(1) \quad \text{w.h.p.,} \quad (3.2)$$

for all $0 < d < 2$; we will revisit this bound in Section 7. Since the expression on the r.h.s. coincides with (1.2) for $\pi = \pi_d$, the main contribution of Theorem 1.1 is the matching lower bound on $\log Z(\mathbf{\Phi})$. We emphasize that Theorem 1.1 and (3.2) imply that π_d from Theorem 1.1 minimizes the functional on the r.h.s. of (3.2). Furthermore, Abbe and Montanari [1] used the interpolation method to establish the *existence* of a function ϕ such that

$$\lim_{n\to\infty} n^{-1}\log(Z(\mathbf{\Phi}) \vee 1) = \phi(d) \qquad \text{in probability} \quad (3.3)$$

for all but a countable number of $d \in (0, 2)$. Theorem 1.1 actually determines $\phi(d)$ and shows that convergence holds for *all* $d \in (0, 2)$. Clearly, (3.3) implies the concentration bound from Proposition 2.6 for all d outside the countable set. But of course we need concentration for all d, and in Section 7 we will use the upper bound (3.2) to prove this concentration result. As an aside, a conditional concentration inequality for $\log Z(\mathbf{\Phi})$, quoted in [29], was obtained by Sharell [47] (unpublished). But the necessary conditions appear to be difficult to check.

In addition, several prior contributions deal with the combinatorial problem of counting solutions to random CSPs. For problems such as k-NAESAT, k-XORSAT or graph coloring where the first moment provides the correct answer due to inherent symmetry properties, the second moment method and small subgraph conditioning yield very precise information as to the number of solutions [16, 19, 45]. Verifying that the number of solutions is determined by the physicists" 1RSB formula [35], the contribution of Sly, Sun and Zhang [48] on the random regular k-NAESAT problem near its satisfiability threshold [25] deals with an even more intricate scenario.

Finally, returning to random 2-SAT, as an intriguing question for future work determining the precise limiting distribution of $\log Z(\mathbf{\Phi})$ stands out. This random variable has standard deviation $\Omega(\sqrt{n})$ for all $0 < d < 2$ even once we condition on $\boldsymbol{m}$, as is easily seen by re-randomizing the signs of the literals in small components. In effect, $\log Z(\mathbf{\Phi})$ is far less concentrated than the partition functions of symmetric random constraint satisfaction problems [16]. May $n^{-1/2}(\log Z(\mathbf{\Phi}) - \mathbb{E}[\log Z(\mathbf{\Phi})])$ be asymptotically normal?

4 | PROOF OF PROPOSITION 2.1

We prove Proposition 2.1 by means of a contraction argument. The starting point is the following observation. For $\ell \geq 0$ let $\pi_d^{(\ell)} = \mathrm{BP}_d^{\ell}(\delta_{1/2})$ be the probability measure obtained after ℓ iterations of the operator $\mathrm{BP}_d(\cdot)$.

Fact 4.1. For all $\ell \geq 0$ the random variables $\boldsymbol{\mu}_{\pi_d^{(\ell)}}$ and $1 - \boldsymbol{\mu}_{\pi_d^{(\ell)}}$ are identically distributed.

Proof. This is because $\boldsymbol{d}^-, \boldsymbol{d}^+$ and hence the random variables

$$\left(\prod_{i=1}^{\boldsymbol{d}^-}\boldsymbol{\mu}_{\pi_d^{(\ell-1)},i}, \prod_{i=1}^{\boldsymbol{d}^-}\boldsymbol{\mu}_{\pi_d^{(\ell-1)},i} + \prod_{i=1}^{\boldsymbol{d}^+}\boldsymbol{\mu}_{\pi_d^{(\ell-1)},i+\boldsymbol{d}^-}\right) \quad \text{and} \quad \left(\prod_{i=1}^{\boldsymbol{d}^+}\boldsymbol{\mu}_{\pi_d^{(\ell-1)},i+\boldsymbol{d}^-}, \prod_{i=1}^{\boldsymbol{d}^-}\boldsymbol{\mu}_{\pi_d^{(\ell-1)},i} + \prod_{i=1}^{\boldsymbol{d}^+}\boldsymbol{\mu}_{\pi_d^{(\ell-1)},i+\boldsymbol{d}^-}\right)$$

from (1.1) are identically distributed. ∎

Due to Fact 4.1 we can rewrite the construction of the sequence $\pi_d^{(\ell)}$ in terms of another operator that is easier to analyze. This operator describes the expression (1.1) in terms of log-likelihood ratios, a simple reformulation that proved useful in the context of Belief Propagation for random satisfiability before [39]. Thus, we define an operator $\mathrm{LL}_d : \mathscr{P}(\mathbb{R}) \to \mathscr{P}(\mathbb{R})$, $\rho \mapsto \hat{\rho}$ by letting $\hat{\rho}$ be the distribution of the random variable

$$\sum_{i=1}^{\boldsymbol{d}} \boldsymbol{s}_i \log \frac{1 + \boldsymbol{s}_i' \tanh(\boldsymbol{\eta}_{\rho,i}/2)}{2}. \tag{4.1}$$

Furthermore, let $\rho_d^{(\ell)} = \mathrm{LL}_d^{\ell}(\delta_0) \in \mathscr{P}(\mathbb{R})$ be the result of ℓ iterations of LL_d launched from the atom at zero. We recall the functions ψ, φ from (1.15). For a measure $\rho \in \mathscr{P}(\mathbb{R})$ and a measurable $f : \mathbb{R} \to \mathbb{R}$ let $f(\rho)$ denote the pushforward measure of ρ that assigns mass $\rho(f^{-1}(A))$ to Borel sets $A \subseteq \mathbb{R}$.

Lemma 4.2. *For all $\ell \geq 0$ we have $\pi_d^{(\ell)} = \psi(\rho_d^{(\ell)})$.*

Proof. Since $\psi(\delta_0) = \delta_{1/2}$, the assertion is true for $\ell = 0$. Proceeding by induction, we obtain

$$\begin{aligned}
\boldsymbol{\mu}_{\pi_d^{(\ell+1)}} &\stackrel{\mathrm{d}}{=} \frac{\prod_{i=1}^{\boldsymbol{d}^+}\boldsymbol{\mu}_{\pi_d^{(\ell)},i}}{\prod_{i=1}^{\boldsymbol{d}^-}\boldsymbol{\mu}_{\pi_d^{(\ell)},i} + \prod_{i=1}^{\boldsymbol{d}^+}\boldsymbol{\mu}_{\pi_d^{(\ell)},i+\boldsymbol{d}^-}} = \psi\left(\log\frac{\prod_{i=1}^{\boldsymbol{d}^-}\boldsymbol{\mu}_{\pi_d^{(\ell)},i}}{\prod_{i=1}^{\boldsymbol{d}^+}\boldsymbol{\mu}_{\pi_d^{(\ell)},i+\boldsymbol{d}^-}}\right) \\
&= \psi\left(\sum_{i=1}^{\boldsymbol{d}^-}\log\left(\boldsymbol{\mu}_{\pi_d^{(\ell)},i}\right) - \sum_{i=1}^{\boldsymbol{d}^+}\log\left(\boldsymbol{\mu}_{\pi_d^{(\ell)},i+\boldsymbol{d}^-}\right)\right) \\
&\stackrel{\mathrm{d}}{=} \psi\left(\sum_{i=1}^{\boldsymbol{d}}\boldsymbol{s}_i\log\boldsymbol{\mu}_{\pi_d^{(\ell)},i}\right) \stackrel{\mathrm{d}}{=} \psi\left(\sum_{i=1}^{\boldsymbol{d}}\boldsymbol{s}_i\log\left(\psi(\boldsymbol{\eta}_{\rho_d^{(\ell)},i})\right)\right).
\end{aligned} \tag{4.2}$$

Moreover, since $\boldsymbol{s}_i \in \{\pm 1\}$ is random, it is immediate from (4.1) that $\boldsymbol{\eta}_{\rho_d^{(\ell)},i} \stackrel{\mathrm{d}}{=} -\boldsymbol{\eta}_{\rho_d^{(\ell)},i}$. Consequently, (4.2) yields

$$\boldsymbol{\mu}_{\pi_d^{(\ell+1)}} \stackrel{\mathrm{d}}{=} \psi\left(\sum_{i=1}^{\boldsymbol{d}}\boldsymbol{s}_i\log\left(\psi(\boldsymbol{s}_i'\boldsymbol{\eta}_{\rho_d^{(\ell)},i})\right)\right) \stackrel{\mathrm{d}}{=} \psi(\boldsymbol{\eta}_{\rho_d^{(\ell+1)}}),$$

which completes the induction. ∎

Due to the continuous mapping theorem, to establish convergence of $(\pi_d^{(\ell)})_{\ell\geq 0}$ it suffices to show that $(\rho_d^{(\ell)})_{\ell\geq 0}$ converges weakly. To this end, we will prove that the operator $\mathrm{LL}_d(\cdot)$ is a contraction.

Lemma 4.3. *If $d < 2$, then LL_d is a contraction on the space $\mathscr{W}_2(\mathbb{R})$.*

Proof. The operator LL_d maps the space $\mathscr{W}_2(\mathbb{R})$ into itself because the derivative of $x \mapsto \log((1+\tanh(x/2))/2)$ is bounded by one in absolute value for all $x \in \mathbb{R}$. To show contraction let $\rho, \rho' \in \mathscr{W}_2(\mathbb{R})$ and consider a sequence of independent random pairs $(\boldsymbol{\eta}_i, \boldsymbol{\eta}'_i)_{i\geq 1}$ such that the $\boldsymbol{\eta}_i$ have distribution ρ and the $\boldsymbol{\eta}'_i$ have distribution ρ'. Because the signs $\boldsymbol{s}_i$ are uniform and independent, we obtain

$$\begin{aligned} W_2(\mathrm{LL}(\rho), \mathrm{LL}(\rho'))^2 &\leq \mathbb{E}\left[\left(\sum_{i=1}^{d} \boldsymbol{s}_i \log \frac{1+\boldsymbol{s}'_i \tanh(\boldsymbol{\eta}_i/2)}{1+\boldsymbol{s}'_i \tanh(\boldsymbol{\eta}'_i/2)}\right)^2\right] \\ &= \mathbb{E}\left[\sum_{h,i=1}^{d} \boldsymbol{s}_h \boldsymbol{s}_i \log \frac{1+\boldsymbol{s}'_h \tanh(\boldsymbol{\eta}_h/2)}{1+\boldsymbol{s}'_h \tanh(\boldsymbol{\eta}'_h/2)} \log \frac{1+\boldsymbol{s}'_i \tanh(\boldsymbol{\eta}_i/2)}{1+\boldsymbol{s}'_i \tanh(\boldsymbol{\eta}'_i/2)}\right] \\ &= \mathbb{E}\left[\sum_{i=1}^{d} \log^2 \frac{1+\boldsymbol{s}'_i \tanh(\boldsymbol{\eta}_i/2)}{1+\boldsymbol{s}'_i \tanh(\boldsymbol{\eta}'_i/2)}\right] \\ &= d\mathbb{E}\left[\log^2 \frac{1+\boldsymbol{s}_1 \tanh(\boldsymbol{\eta}_1/2)}{1+\boldsymbol{s}_1 \tanh(\boldsymbol{\eta}'_1/2)}\right]. \end{aligned} \tag{4.3}$$

Furthermore,

$$\log^2 \frac{1+\tanh(\boldsymbol{\eta}_1/2)}{1+\tanh(\boldsymbol{\eta}'_1/2)} = \left[\int_{\boldsymbol{\eta}'_1}^{\boldsymbol{\eta}_1} \frac{\partial \log(1+\tanh(z/2))}{\partial z} \mathrm{d}z\right]^2 = \left[\int_{\boldsymbol{\eta}_1 \wedge \boldsymbol{\eta}'_1}^{\boldsymbol{\eta}_1 \vee \boldsymbol{\eta}'_1} \frac{1-\tanh(z/2)}{2} \mathrm{d}z\right]^2, \tag{4.4}$$

$$\log^2 \frac{1-\tanh(\boldsymbol{\eta}_1/2)}{1-\tanh(\boldsymbol{\eta}'_1/2)} = \left[\int_{\boldsymbol{\eta}'_1}^{\boldsymbol{\eta}_1} \frac{\partial \log(1-\tanh(z/2))}{\partial z} \mathrm{d}z\right]^2 = \left[\int_{\boldsymbol{\eta}_1 \wedge \boldsymbol{\eta}'_1}^{\boldsymbol{\eta}_1 \vee \boldsymbol{\eta}'_1} \frac{1+\tanh(z/2)}{2} \mathrm{d}z\right]^2. \tag{4.5}$$

Combining (4.4)–(4.5) and applying the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned} &\mathbb{E}\left[\log^2 \frac{1+\boldsymbol{s}_1 \tanh(\boldsymbol{\eta}_1/2)}{1+\boldsymbol{s}_1 \tanh(\boldsymbol{\eta}'_1/2)}\right] \\ &= \frac{1}{2}\mathbb{E}\left[\left[\int_{\boldsymbol{\eta}_1 \wedge \boldsymbol{\eta}'_1}^{\boldsymbol{\eta}_1 \vee \boldsymbol{\eta}'_1} \frac{1-\tanh(z/2)}{2} \mathrm{d}z\right]^2 + \left[\int_{\boldsymbol{\eta}_1 \wedge \boldsymbol{\eta}'_1}^{\boldsymbol{\eta}_1 \vee \boldsymbol{\eta}'_1} \frac{1+\tanh(z/2)}{2} \mathrm{d}z\right]^2\right] \\ &\leq \frac{1}{2}\mathbb{E}\left[|\boldsymbol{\eta}_1 - \boldsymbol{\eta}'_1| \int_{\boldsymbol{\eta}_1 \wedge \boldsymbol{\eta}'_1}^{\boldsymbol{\eta}_1 \vee \boldsymbol{\eta}'_1} \left(\frac{1-\tanh(z/2)}{2}\right)^2 + \left(\frac{1+\tanh(z/2)}{2}\right)^2 \mathrm{d}z\right] \\ &\leq \frac{1}{2}\mathbb{E}\left[(\boldsymbol{\eta}_1 - \boldsymbol{\eta}'_1)^2\right]. \end{aligned} \tag{4.6}$$

Finally, (4.3) and (4.6) yield $W_2(\mathrm{LL}(\rho), \mathrm{LL}(\rho'))^2 \leq d\mathbb{E}[(\boldsymbol{\eta}_1 - \boldsymbol{\eta}'_1)^2]/2$, which implies contraction because $d < 2$. ∎

Proof of Proposition 2.1. Together with the Banach fixed point theorem Lemma 4.3 ensures that the W_2-limit $\rho_d = \lim_{\ell\to\infty} \mathrm{LL}_d^{\ell}(\delta_0)$ exists. Therefore, Lemma 4.2 implies that the sequence $(\pi_d^{(\ell)})_{\ell\geq 0}$ converges weakly. In addition, since $\rho_d \in \mathscr{W}_2(\mathbb{R})$, Lemma 4.2 also implies the bound (2.1). Finally, to prove (2.2) we apply (2.1) to obtain

$$\mathbb{E}\left|\log\left(\prod_{i=1}^{d^-} \boldsymbol{\mu}_{\pi_d,i} + \prod_{i=1}^{d^+} \boldsymbol{\mu}_{\pi_d,i+d^-}\right)\right| \leq \log(2) - \mathbb{E}\log \prod_{i=1}^{d^-} \boldsymbol{\mu}_{\pi_d,i}$$

$$\leq \log(2) - \frac{d}{2}\mathbb{E}\log \boldsymbol{\mu}_{\pi_d,1}$$
$$\leq 2\log(2) + d\mathbb{E}\left|\log\frac{\boldsymbol{\mu}_{\pi_d}}{1-\boldsymbol{\mu}_{\pi_d}}\right| < \infty,$$
$$\mathbb{E}\left|\log(1-\boldsymbol{\mu}_{\pi_d,1}\boldsymbol{\mu}_{\pi_d,2})\right| \leq \mathbb{E}\left|\log(1-\boldsymbol{\mu}_{\pi_d})\right|$$
$$\leq \mathbb{E}\left|\log\frac{\boldsymbol{\mu}_{\pi_d}}{1-\boldsymbol{\mu}_{\pi_d}}\right| + \log 2 < \infty,$$

thereby completing the proof. ∎

5 | PROOF OF PROPOSITION 2.2

5.1 | Outline

The goal is to prove that the marginal of the root variable o of $\boldsymbol{T}^{(2\ell)}$ remains asymptotically invariant even upon imposition of an arbitrary (feasible) boundary condition on the variables at distance 2ℓ from the root o. A priori, a proof of this statement seems challenging because of the very large number of possible boundary conditions. Indeed, we expect about d^ℓ variables at distance 2ℓ. But a crucial feature of the 2-SAT problem is that we can construct a pair of extremal boundary conditions. One of these maximizes the probability that the root is set to one. The other one minimizes that probability. As a consequence, instead of inspecting all possible boundary conditions, it suffices to show that the marginals on the root o that these two extremal boundary induce asymptotically coincide with the unconditional marginals. Of course, due to symmetry it actually suffices to consider the "positive" extremal boundary condition that maximally nudges the root towards +1.

To construct this extremal boundary condition we define a satisfying assignment $\boldsymbol{\sigma}^+$ by working our way down the tree $\boldsymbol{T}^{(2\ell)}$. We begin by defining $\boldsymbol{\sigma}^+_o = 1$. Furthermore, suppose for $\ell \geq 1$ the values of the variables at distance $2(\ell-1)$ from o have been defined already. Consider a variable $v \in \partial^{2\ell} o$, its parent clause a and the parent variable u of a. Our aim is to choose $\boldsymbol{\sigma}^+_v$ so as to "nudge" u towards $\boldsymbol{\sigma}^+_u$ as much as possible. To this end we set $\boldsymbol{\sigma}^+_v$ so as to not satisfy a if setting u to $\boldsymbol{\sigma}^+_u$ satisfies a. Otherwise we pick the value that satisfies a; see Figure 2. In formulas,

$$\boldsymbol{\sigma}^+_v = sign(a,v)\mathbf{1}\{sign(a,u) \neq \boldsymbol{\sigma}^+_u\} - sign(a,v)\mathbf{1}\{sign(a,u) = \boldsymbol{\sigma}^+_u\}.$$

The following lemma verifies that $\boldsymbol{\sigma}^+$ is extremal, that is, that imposing the values provided by $\boldsymbol{\sigma}^+$ on the boundary variables $\partial^{2\ell} o$ maximizes the probability of the truth value 1 at the root o. The proof can be found in Section 5.2.

Lemma 5.1. *For any integer $\ell \geq 0$ we have* $\max_{\tau \in S(\boldsymbol{T}^{(2\ell)})} \mu_{\boldsymbol{T}^{(2\ell)}}(\boldsymbol{\sigma}_o = 1 | \boldsymbol{\sigma}_{\partial^{2\ell} o} = \tau_{\partial^{2\ell} o}) = \mu_{\boldsymbol{T}^{(2\ell)}}(\boldsymbol{\sigma}_o = 1 | \boldsymbol{\sigma}_{\partial^{2\ell} o} = \boldsymbol{\sigma}^+_{\partial^{2\ell} o})$.

Lemma 5.1 reduces the task of proving Proposition 2.2 to establishing the following statement.

Proposition 5.2. *We have* $\lim_{\ell\to\infty} \mathbb{E}\left|\mu_{\boldsymbol{T}^{(2\ell)}}(\boldsymbol{\sigma}_o = 1) - \mu_{\boldsymbol{T}^{(2\ell)}}(\boldsymbol{\sigma}_o = 1 | \boldsymbol{\sigma}_{\partial^{2\ell} o} = \boldsymbol{\sigma}^+_{\partial^{2\ell} o})\right| = 0.$

In words, the root marginal given the extremal boundary condition $\boldsymbol{\sigma}^+$ asymptotically coincides with the unconditional marginal.

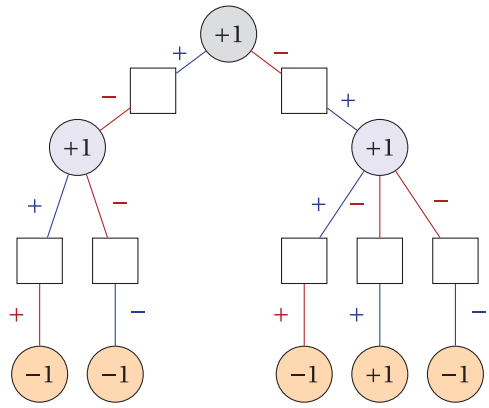

FIGURE 2 The graph $G(\mathbf{\Phi})$ together with extremal boundary condition $\boldsymbol{\sigma}^+$. Variables are indicated by circles and clauses by squares. The labels on the edges illustrate the sign with which variables appears in the clauses. To obtain the extremal boundary condition $\boldsymbol{\sigma}^+$ we proceed top-down. The truth values of the children are chosen so as to nudge the parent variables in the direction provided by $\boldsymbol{\sigma}^+$

The proof of Proposition 5.2 is delicate because the boundary condition $\boldsymbol{\sigma}^+$ depends on the tree $\boldsymbol{T}^{(2\ell)}$. Indeed, it seems hopeless to confront these dependencies head on by first passing down the tree to construct $\boldsymbol{\sigma}^+$ and to subsequently work up the tree to calculate marginals. To sidestep this problem we devise a quantity that recovers the Markov property of the random tree. Specifically, with each variable node $x \in \partial^{2k} o$, $k > 0$, of $\boldsymbol{T}^{(2\ell)}$ we will associate a carefully defined quantity $\boldsymbol{\eta}_x^{(\ell)} \in \mathbb{R} \cup \{\pm\infty\}$ that gauges how strongly x can nudge its (grand-)parent variable y towards the truth value mandated by $\boldsymbol{\sigma}_y^+$. This random variable $\boldsymbol{\eta}_x^{(\ell)}$ will turn out to be essentially independent of the top $2k$ levels of the tree. In effect, we will discover that the distribution of $\boldsymbol{\eta}_o^{(\ell)}$ can be approximated by the k-fold application of a suitable operator that will turn out to be a W_1-contraction. Taking limits $k, \ell \to \infty$ carefully will then complete the proof.

To facilitate this construction we need to count satisfying assignments of sub-formulas of $\boldsymbol{T}^{(2\ell)}$ subject to certain boundary conditions. Specifically, for a variable x we let $\boldsymbol{T}_x^{(2\ell)}$ be the sub-formula of $\boldsymbol{T}^{(2\ell)}$ comprising x and its progeny. Moreover, for a satisfying assignment $\tau \in S(\boldsymbol{T}^{(2\ell)})$ we let

$$S(\boldsymbol{T}_x^{(2\ell)}, \tau) = \left\{ \chi \in S(\boldsymbol{T}_x^{(2\ell)}) : \forall y \in V(\boldsymbol{T}_x^{(2\ell)}) \cap \partial^{2\ell}(\boldsymbol{T}, o) : \chi_y = \tau_y \right\},$$
$$Z(\boldsymbol{T}_x^{(2\ell)}, \tau) = \left| S(\boldsymbol{T}_x^{(2\ell)}, \tau) \right|.$$

In words, $S(\boldsymbol{T}_x^{(2\ell)}, \tau)$ contains all satisfying assignments of $\boldsymbol{T}_x^{(2\ell)}$ that comply with the boundary condition induced by τ. As a final twist, for $t = \pm 1$ we also need the number

$$Z(\boldsymbol{T}_x^{(2\ell)}, \tau, t) = \left| \left\{ \chi \in S(\boldsymbol{T}_x^{(2\ell)}, \tau) : \chi_x = t \right\} \right|$$

of satisfying assignments of $\boldsymbol{T}_x^{(2\ell)}$ that agree with τ on the boundary and assign value t to x.

The protagonist of the proof of Proposition 5.2 is the log-likelihood ratio

$$\boldsymbol{\eta}_x^{(\ell)} = \log \frac{Z(\boldsymbol{T}_x^{(2\ell)}, \boldsymbol{\sigma}^+, \boldsymbol{\sigma}_x^+)}{Z(\boldsymbol{T}_x^{(2\ell)}, \boldsymbol{\sigma}^+, -\boldsymbol{\sigma}_x^+)} \in \mathbb{R} \cup \{\pm\infty\} \qquad (x \in V(\boldsymbol{T}^{(2\ell)})), \tag{5.1}$$

with the conventions $\log 0 = -\infty$, $\log \infty = \infty$. Thus, $\boldsymbol{\eta}_x^{(\ell)}$ gauges how likely a random satisfying assignment $\boldsymbol{\sigma}$ of $\boldsymbol{T}_x^{(2\ell)}$ subject to the $\boldsymbol{\sigma}^+$-boundary condition is to set x to its designated value $\boldsymbol{\sigma}_x^+$.

To get a handle on the $\boldsymbol{\eta}_x^{(\ell)}$, we show that these quantities can be calculated by propagating the extremal boundary condition $\boldsymbol{\sigma}^+$ bottom–up toward the root of the tree. Specifically, we consider the operator

$$\mathrm{LL}^+_{\boldsymbol{T}^{(2\ell)}} : (-\infty, \infty]^{V(\boldsymbol{T}^{(2\ell)})} \to (-\infty, \infty]^{V(\boldsymbol{T}^{(2\ell)})}, \qquad \eta \mapsto \hat{\eta}$$

defined as follows. For all $x \in \partial^{2\ell} o$ we set $\hat{\eta}_x = \infty$. Moreover, for a variable $x \in \partial^{2k} o$ with $k < \ell$ with children $a_1, \ldots, a_j$ and grandchildren $y_1 \in \partial a_1 \setminus \{x\}, \ldots, y_j \in \partial a_j \setminus \{x\}$ we define

$$\hat{\eta}_x = -\sum_{i=1}^{j} \boldsymbol{\sigma}_x^+ sign(x, a_i) \log \frac{1 - \boldsymbol{\sigma}_x^+ sign(x, a_i) \tanh(\eta_{y_i}/2)}{2}. \tag{5.2}$$

It may not be apparent that the above sum is well-defined as a $-\infty$ summand might occur. However, the next lemma rules this out and shows that ℓ-fold iteration of $\mathrm{LL}^+_{\boldsymbol{T}^{(2\ell)}}$ from all-$+\infty$ yields $\boldsymbol{\eta}^{(\ell)} = (\boldsymbol{\eta}_x^{(\ell)})_{x \in V(\boldsymbol{T}^{(2\ell)})}$.

Lemma 5.3. *The operator* $\mathrm{LL}^+_{\boldsymbol{T}^{(2\ell)}}$ *is well-defined and* $\mathrm{LL}^{+(\ell)}_{\boldsymbol{T}^{(2\ell)}}(\infty, \ldots, \infty) = \boldsymbol{\eta}^{(\ell)}$.

We defer the proof of Lemma 5.3 to Section 5.3.

The next aim is to approximate the ℓ-fold iteration of $\mathrm{LL}^+_{\boldsymbol{T}^{(2\ell)}}$, and specifically the distribution of the value $\boldsymbol{\eta}_o^{(\ell)}$ associated with the root, via a nonrandom operator $\mathscr{P}(\mathbb{R}) \to \mathscr{P}(\mathbb{R})$. To this end we need to cope with the $\pm\infty$-entries of the vector $\boldsymbol{\eta}^{(\ell)}$, a task that we solve by bounding $\boldsymbol{\eta}_x^{(\ell)}$ for variables x near the top of the tree.

Lemma 5.4. *There exist* $c = c(d) > 0$ *and a sequence* $(\varepsilon_k)_{k \geq 1}$ *with* $\lim_{k \to \infty} \varepsilon_k = 0$ *such that for any* $k > 0$, $\ell > ck$ *we have* $\mathbb{P}[\max_{x \in \partial^{2k} o} |\boldsymbol{\eta}_x^{(\ell)}| \leq ck] > 1 - \varepsilon_k$.

The proof of Lemma 5.4, based on a percolation argument, can be found in Section 5.4. We continue to denote by c and $(\varepsilon_k)_k$ the number and the sequence supplied by Lemma 5.4.

Guided by Lemma 5.4 we consider the vector $\overline{\boldsymbol{\eta}}^{(\ell,k)}$ of truncated log-likelihood ratios

$$\overline{\boldsymbol{\eta}}_x^{(\ell,k)} = \begin{cases} -ck & \text{if } x \in \partial^{2k} o \text{ and } \boldsymbol{\eta}_x^{(\ell)} < -ck, \\ ck & \text{if } x \in \partial^{2k} o \text{ and } \boldsymbol{\eta}_x^{(\ell)} > ck, \\ \boldsymbol{\eta}_x^{(\ell)} & \text{otherwise.} \end{cases}$$

Furthermore, let

$$\boldsymbol{\eta}^{(\ell,k)} = \mathrm{LL}^{+(k)}_{\boldsymbol{T}^{(2\ell)}}(\overline{\boldsymbol{\eta}}^{(\ell,k)})$$

be the result of k iterations of $\mathrm{LL}^+_{\boldsymbol{T}^{(2\ell)}}(\cdot)$ starting from $\overline{\boldsymbol{\eta}}^{(\ell,k)}$.

Corollary 5.5. *For any* $\ell > ck$ *we have* $d_{\mathrm{TV}}(\boldsymbol{\eta}_o^{(\ell,k)}, \boldsymbol{\eta}_o^{(\ell)}) < \varepsilon_k$.

Proof. This follows from Lemmas 5.3 and 5.4, which shows that the truncation is inconsequential with probability at least $1 - \varepsilon_k$. ∎

We are ready to introduce the operator $\mathscr{P}(\mathbb{R}) \to \mathscr{P}(\mathbb{R})$ that mimics $\mathrm{LL}^{+}_{\boldsymbol{T}^{(2\ell)}}$. Specifically, $\mathrm{LL}^{+}_{d} : \mathscr{P}(\mathbb{R}) \to \mathscr{P}(\mathbb{R})$ maps $\rho \in \mathscr{P}(\mathbb{R})$ to the distribution of

$$-\sum_{i=1}^{d} \boldsymbol{s}_i \log \frac{1 - \boldsymbol{s}_i \tanh(\boldsymbol{\eta}_{\rho,i}/2)}{2}. \tag{5.3}$$

We emphasize the subtle difference between (5.3) and (4.1), which involves two independent signs $\boldsymbol{s}_i, \boldsymbol{s}'_i$. The next lemma establishes the connection between the random operator $\mathrm{LL}^{+}_{\boldsymbol{T}^{(2\ell)}}$ and the operator LL^{+}_{d}. Namely, let $\rho^{(\ell,k)}$ be the distribution of $\boldsymbol{\eta}_o^{(\ell,k)}$. Moreover, let $\overline{\rho}^{(\ell-k)}$ be the distribution of

$$\boldsymbol{\eta}_o^{(\ell-k)} \mathbf{1}\{-ck < \boldsymbol{\eta}_o^{(\ell-k)} < ck\} + ck \mathbf{1}\{ck < \boldsymbol{\eta}_o^{(\ell-k)}\} - ck \mathbf{1}\{\boldsymbol{\eta}_o^{(\ell-k)} < -ck\},$$

i.e., the truncation of $\boldsymbol{\eta}_o^{(\ell-k)}$.

Lemma 5.6. *For $\ell > ck$ we have $\rho^{(\ell,k)} = \mathrm{LL}^{+(k)}_{d}(\overline{\rho}^{(\ell-k)})$.*

We prove Lemma 5.6 in Section 5.5. Recalling φ from (1.15), as in the proof of Proposition 2.1 we let $\rho_d = \varphi(\pi_d)$ be the distribution of the log-likelihood ratio $\log(\boldsymbol{\mu}_{\pi_d}/(1 - \boldsymbol{\mu}_{\pi_d}))$.

Lemma 5.7. *The operator LL^{+}_{d} is a W_1-contraction with unique fixed point ρ_d.*

The proof of Lemma 5.7 can be found in Section 5.6. Let $(\rho^{(\ell)})_\ell$ be the sequence of distributions of $(\boldsymbol{\eta}_o^{(\ell)})_\ell$. As an immediate consequence we obtain the limit of the sequence $(\rho^{(\ell)})_\ell$. We recall ψ from (1.15).

Corollary 5.8. *The sequence $(\psi(\rho^{(\ell)}))_{\ell \geq 0}$ converges weakly to π_d.*

Proof. This follows from Corollary 5.5, Lemma 5.6, Lemma 5.7 and the continuous mapping theorem. ∎

Proof of Proposition 5.2. Set $\boldsymbol{\vartheta}_o^{(\ell)} = (\mathrm{LL}^{+(\ell)}_{\boldsymbol{T}^{(2\ell)}}(0, \dots, 0))_o = \log(\mu_{\boldsymbol{T}^{(2\ell)}}(\boldsymbol{\sigma}_o = 1)/\mu_{\boldsymbol{T}^{(2\ell)}}(\boldsymbol{\sigma}_o = -1))$. Then

$$\mu_{\boldsymbol{T}^{(2\ell)}}(\boldsymbol{\sigma}_o = 1) = \psi(\boldsymbol{\vartheta}_o^{(\ell)}) \qquad \text{and} \qquad \mu_{\boldsymbol{T}^{(2\ell)}}(\boldsymbol{\sigma}_o = 1 | \boldsymbol{\sigma}_{\partial^{2\ell} o} = \boldsymbol{\sigma}^{+}_{\partial^{2\ell} o}) = \psi(\boldsymbol{\eta}_o^{(\ell)}).$$

Moreover, Lemma 5.1 shows that $0 \leq \psi(\boldsymbol{\vartheta}_o^{(\ell)}) \leq \psi(\boldsymbol{\eta}_o^{(\ell)}) \leq 1$. Furthermore, Lemma 5.7 implies that $\psi(\boldsymbol{\vartheta}_o^{(\ell)})$ converges weakly to π_d. Finally, Corollary 5.8 implies that $\psi(\boldsymbol{\eta}_o^{(\ell)})$ also converges weakly to π_d, whence

$$\lim_{\ell \to \infty} \mathbb{E}\left|\psi(\boldsymbol{\eta}_o^{(\ell)}) - \psi(\boldsymbol{\vartheta}_o^{(\ell)})\right| = \lim_{\ell \to \infty} \left|\mathbb{E}[\psi(\boldsymbol{\vartheta}_o^{(\ell)})] - \mathbb{E}[\psi(\boldsymbol{\eta}_o^{(\ell)})]\right| = 0,$$

which directly implies the assertion. ∎

Proof of Proposition 2.2. The proposition follows immediately from Lemma 5.1 and Proposition 5.2. ∎

5.2 | Proof of Lemma 5.1

The proof is by induction on the height of the tree. The following claim summarizes the main step of the induction.

Claim 5.9. For all $\ell \geq 0$, all variables x of $\boldsymbol{T}^{(2\ell)}$ and all satisfying assignments $\tau \in S(\boldsymbol{T}^{(2\ell)})$ we have

$$\frac{Z(\boldsymbol{T}_x^{(2\ell)}, \tau, \boldsymbol{\sigma}_x^+)}{Z(\boldsymbol{T}_x^{(2\ell)}, \tau)} \leq \frac{Z(\boldsymbol{T}_x^{(2\ell)}, \boldsymbol{\sigma}^+, \boldsymbol{\sigma}_x^+)}{Z(\boldsymbol{T}_x^{(2\ell)}, \boldsymbol{\sigma}^+)}. \tag{5.4}$$

Proof. For boundary variables $x \in \partial^{2\ell} o$ there is nothing to show because the r.h.s. of (5.4) equals one. Hence, consider a variable $x \in \partial^{2k} o$ for some $k < \ell$. If $Z(\boldsymbol{T}_x^{(2\ell)}, \tau, \boldsymbol{\sigma}_x^+) = 0$, then (5.4) is trivially satisfied. Hence, assume that $Z(\boldsymbol{T}_x^{(2\ell)}, \tau, \boldsymbol{\sigma}_x^+) > 0$. Let $a_1^+, \ldots, a_g^+$ be the children (clauses) of x with $sign(x, a_i^+) = \boldsymbol{\sigma}_x^+$. Also let $y_1, \ldots, y_g$ be the children (variables) of $a_1^+, \ldots, a_g^+$. Similarly, let $a_1^-, \ldots, a_h^-$ be the children of x with $sign(x, a_i^-) = -\boldsymbol{\sigma}_x^+$ and let $z_1, \ldots, z_h$ be their children. We claim that for all $\tau \in S(\boldsymbol{T}^{(2\ell)})$,

$$\begin{aligned} Z(\boldsymbol{T}_x^{(2\ell)}, \tau, \boldsymbol{\sigma}_x^+) &= \prod_{i=1}^{g} Z(\boldsymbol{T}_{y_i}^{(2\ell)}, \tau) \prod_{i=1}^{h} Z(\boldsymbol{T}_{z_i}^{(2\ell)}, \tau, \boldsymbol{\sigma}_{z_i}^+), \\ Z(\boldsymbol{T}_x^{(2\ell)}, \tau, -\boldsymbol{\sigma}_x^+) &= \prod_{i=1}^{g} Z(\boldsymbol{T}_{y_i}^{(2\ell)}, \tau, -\boldsymbol{\sigma}_{y_i}^+) \prod_{i=1}^{h} Z(\boldsymbol{T}_{z_i}^{(2\ell)}, \tau). \end{aligned} \tag{5.5}$$

For setting x to $\boldsymbol{\sigma}_x^+$ satisfies $a_1^+, \ldots, a_g^+$; hence, arbitrary satisfying assignments of the sub-trees $\boldsymbol{T}_{y_i}^{(2\ell)}$ can be combined, which explains the first product. By contrast, upon assigning x the value $\boldsymbol{\sigma}_x^+$ we need to assign the variables z_i the values $\boldsymbol{\sigma}_{z_i}^+$ so that they satisfy the clauses a_i^-. This leaves us with $Z(\boldsymbol{T}_{z_i}^{(2\ell)}, \tau, \boldsymbol{\sigma}_{z_i}^+)$ possible satisfying assignments of the sub-trees $\boldsymbol{T}_{z_i}^{(2\ell)}$; hence the second product, and we obtain the left equation. A similar argument yields the right one. Dividing the two expressions from (5.5) and invoking the induction hypothesis (for $k+1$), we obtain

$$\begin{aligned} \frac{Z(\boldsymbol{T}_x^{(2\ell)}, \tau, -\boldsymbol{\sigma}_x^+)}{Z(\boldsymbol{T}_x^{(2\ell)}, \tau, \boldsymbol{\sigma}_x^+)} &= \prod_{i=1}^{g} \frac{Z(\boldsymbol{T}_{y_i}^{(2\ell)}, \tau, -\boldsymbol{\sigma}_{y_i}^+)}{Z(\boldsymbol{T}_{y_i}^{(2\ell)}, \tau)} \cdot \prod_{i=1}^{h} \frac{Z(\boldsymbol{T}_{z_i}^{(2\ell)}, \tau)}{Z(\boldsymbol{T}_{z_i}^{(2\ell)}, \tau, \boldsymbol{\sigma}_{z_i}^+)} \\ &\geq \prod_{i=1}^{g} \frac{Z(\boldsymbol{T}_{y_i}^{(2\ell)}, \boldsymbol{\sigma}^+, -\boldsymbol{\sigma}_{y_i}^+)}{Z(\boldsymbol{T}_{y_i}^{(2\ell)}, \boldsymbol{\sigma}^+)} \cdot \prod_{i=1}^{h} \frac{Z(\boldsymbol{T}_{z_i}^{(2\ell)}, \boldsymbol{\sigma}^+)}{Z(\boldsymbol{T}_{z_i}^{(2\ell)}, \boldsymbol{\sigma}^+, \boldsymbol{\sigma}_{z_i}^+)} \\ &= \frac{Z(\boldsymbol{T}_x^{(2\ell)}, \boldsymbol{\sigma}^+, -\boldsymbol{\sigma}_x^+)}{Z(\boldsymbol{T}_x^{(2\ell)}, \boldsymbol{\sigma}^+, \boldsymbol{\sigma}_x^+)}, \end{aligned}$$

completing the induction. ∎

Proof of Lemma 5.1. The assertion follows by applying Claim 5.9 to $x = o$. ∎

5.3 | Proof of Lemma 5.3

To show that $\mathrm{LL}^+_{\boldsymbol{T}^{(2\ell)}}$ is well defined we verify that, in the notation of (5.2), $\hat{\eta}_x \in (-\infty, \infty]$ for all x. Indeed, in the expression on the r.h.s. of (5.2) a $\pm\infty$ summand can arise only from variables y_i with $\eta_{y_i} = \infty$. But the definition of $\boldsymbol{\sigma}^+$ ensures that such y_i either render a zero summand if $\boldsymbol{\sigma}_x^+ sign(x, a_i) = -1$, or a $+\infty$ summand if $\boldsymbol{\sigma}_x^+ sign(x, a_i) = 1$. Thus, the sum is well-defined and $\hat{\eta}_x \in (-\infty, \infty]$.

Furthermore, to verify the identity $\boldsymbol{\eta}^{(\ell)} = \mathrm{LL}^{+(\ell)}_{\boldsymbol{T}^{(2\ell)}}(\infty, \ldots, \infty)$, consider a variable x of $\boldsymbol{T}^{(2\ell)}$. Let $a_1^+, \ldots, a_g^+$ be its children with $sign(a_i^+, x) = \boldsymbol{\sigma}_x^+$, let $y_1, \ldots, y_g$ be their children, let $a_1^-, \ldots, a_h^-$ be the children of x with $sign(a_i^-, x) = -\boldsymbol{\sigma}_x^+$ and let $z_1, \ldots, z_h$ be their children. Then (1.15) and (5.5) yield

$$\begin{aligned}\boldsymbol{\eta}_x^{(\ell)} &= -\sum_{i=1}^{g} \log \frac{Z(\boldsymbol{T}_{y_i}^{(2\ell)}, \boldsymbol{\sigma}^+, -\boldsymbol{\sigma}_{y_i}^+)}{Z(\boldsymbol{T}_{y_i}^{(2\ell)}, \boldsymbol{\sigma}^+)} + \sum_{i=1}^{h} \log \frac{Z(\boldsymbol{T}_{z_i}^{(2\ell)}, \boldsymbol{\sigma}^+, \boldsymbol{\sigma}_{z_i}^+)}{Z(\boldsymbol{T}_{z_i}^{(2\ell)}, \boldsymbol{\sigma}^+)} \\ &= -\sum_{i=1}^{g} \log \frac{1 - \tanh(\boldsymbol{\eta}_{y_i}^{(\ell)}/2)}{2} + \sum_{i=1}^{h} \log \frac{1 + \tanh(\boldsymbol{\eta}_{z_i}^{(\ell)}/2)}{2}. \end{aligned} \tag{5.6}$$

The assertion follows because $sign(x, a_i^+)\boldsymbol{\sigma}_x^+ = 1$ and $sign(x, a_i^-)\boldsymbol{\sigma}_x^+ = -1$.

5.4 | Proof of Lemma 5.4

The goal is to prove that for variables some distance away from level 2ℓ of $\boldsymbol{T}^{(2\ell)}$ the counts $Z(\boldsymbol{T}_x^{(2\ell)}, \boldsymbol{\sigma}^+, \pm 1)$ are roughly of the same order of magnitude. Approaching this task somewhat indirectly, we begin by tracing the logical implications of imposing a specific value $s = \pm 1$ on a variable x of the (possibly infinite) tree $\boldsymbol{T}$. Clearly, upon setting x to the value s a child (clause) a of x will be satisfied iff x appears in a with sign s. In effect, all clauses a with $sign(a, x) \neq s$ need to be satisfied by their second variable y, a grandchild of x. Thus, we impose the value $sign(a, y)$ on y and recurse down the tree. Let $\boldsymbol{T}_{x,s}$ denote the sub-tree of $\boldsymbol{T}$ comprising x and all other variables on which this process imposes specific values as well as all clauses that contain two such variables. Clearly, for every leaf y of $\boldsymbol{T}_{x,s}$ the values imposed on y happens to satisfy all child clauses of y in $\boldsymbol{T}$. Let $\boldsymbol{N}_{x,s} \in [1, \infty]$ be the number of variables in $\boldsymbol{T}_{x,s}$. The next lemma shows that the impact of a boundary condition on the marginal of x can be bounded in terms of $\boldsymbol{N}_{x,s}$.

Claim 5.10. Let $s \in \{\pm 1\}$. If $x \in \partial^{2k} o$ satisfies $\boldsymbol{N}_{x,s} < \ell - k$ then $Z(\boldsymbol{T}_x^{(2\ell)}, \tau) \leq 2^{\boldsymbol{N}_{x,s}} Z(\boldsymbol{T}_x^{(2\ell)}, \tau, s)$.

Proof. The construction of the implication tree $\boldsymbol{T}_{x,s}$ imposes a truth value σ_y on each variable y of the tree that y must inevitably take if x gets assigned s. Thus, $\boldsymbol{T}_{x,s}$ comes with a satisfying assignment $\sigma \in S(\boldsymbol{T}_{x,s})$ with $\sigma_x = s$. For any leaf y of $\boldsymbol{T}_{x,s}$ every child clause a of y in the super-tree $\boldsymbol{T}$ will be automatically satisfied by setting y to σ_y (because otherwise a would have been included in $\boldsymbol{T}_{x,s}$). Hence, all the clauses of $\boldsymbol{T}$ that are children of the leaves of $\boldsymbol{T}_{x,s}$ are satisfied by σ. Moreover, because $\boldsymbol{N}_{x,s} < \ell - k$, any leaf y of $\boldsymbol{T}_{x,s}$ has distance less than 2ℓ from o. Thus, the assignment σ does not clash with the boundary condition τ. As a consequence, for any $\chi \in S(\boldsymbol{T}_x^{(2\ell)}, \tau)$ we obtain another satisfying assignment $\chi' \in S(\boldsymbol{T}_x^{(2\ell)}, \tau)$ by letting

$$\chi'_z = \begin{cases} \sigma_z & \text{if } z \in V(\boldsymbol{T}_{x,s}), \\ \chi_z & \text{otherwise.} \end{cases}$$

Moreover, under the map $\chi \mapsto \chi'$ the number of inverse images of any assignment χ' is bounded by the total number $2^{\boldsymbol{N}_{x,s}}$ of different truth assignments of the variables $V(\boldsymbol{T}_{x,s})$. Therefore, $Z(\boldsymbol{T}_x^{(2\ell)}, \tau) \leq 2^{\boldsymbol{N}_{x,s}} Z(\boldsymbol{T}_x^{(2\ell)}, \tau, s)$. ∎

As a next step we bound the random variable $\boldsymbol{N}_{x,s}$.

Claim 5.11. There exists a number $\alpha = \alpha(d) > 0$ such that $\mathbb{P}\left[\boldsymbol{N}_{o,s} \geq u\right] \leq \exp\left(-u\alpha\right)/\alpha$ for all $u \geq 0$, $s \in \{\pm 1\}$.

Proof. In the construction of $\boldsymbol{T}_{o,s}$ we only propagate along clauses in which the parent variable is forced to take a value that fails to satisfy the clause. Since the signs are uniformly random, the number of such child clauses has distribution Po(d/2). Therefore, $\boldsymbol{N}_{o,s}$ is bounded by the total progeny of a Galton–Watson process with Po(d/2) offspring. The assertion therefore follows from the tail bound for such processes (e.g., [[7], eq. (11.7)]). ∎

As a final preparation toward the proof of Lemma 5.4 we need a bound on the size of the $2k$-th level of $\boldsymbol{T}$.

Claim 5.12. We have $\lim_{k\to\infty}\mathbb{P}\left[|\partial^{2k}o| > (2d)^k + k\right] = 0$.

Proof. Since every clause of $\boldsymbol{T}$ has precisely one child, the size of level $2k$ of $\boldsymbol{T}$ coincides with the size of the k-th level of a Po(d) Galton–Watson tree. Therefore, the assertion follows from standard tail bounds for Galton–Watson processes (e.g., [[7], eq. (11.7)]). ∎

Proof of Lemma 5.4. Claim 5.11 ensures that for a large enough constant $c = c(d) > 0$ and all large enough k,

$$\mathbb{P}\left(\boldsymbol{N}_{o,\pm 1} \geq ck\right) \leq (2d)^{-k}. \tag{5.7}$$

Combining (5.7) with Claim 5.12 and using the union bound, we obtain a sequence $\varepsilon_k \to 0$ such that

$$\mathbb{P}\left(\forall x \in \partial^{2k}o : \boldsymbol{N}_{x,\pm 1} < ck\right) \geq 1 - \varepsilon_k. \tag{5.8}$$

Furthermore, if $x \in \partial^{2k}o$ satisfies $\boldsymbol{N}_{x,\pm 1} < ck$ and $\ell > (1+c)k$, Claim 5.10 ensures that for all $x \in \partial^{2k}o$,

$$\left|\boldsymbol{\eta}_x^{(\ell)}\right| \leq \log\frac{Z(\boldsymbol{T}_x^{(2\ell)}, \boldsymbol{\sigma}^+)}{Z(\boldsymbol{T}_x^{(2\ell)}, \boldsymbol{\sigma}^+, 1)} + \log\frac{Z(\boldsymbol{T}_x^{(2\ell)}, \boldsymbol{\sigma}^+)}{Z(\boldsymbol{T}_x^{(2\ell)}, \boldsymbol{\sigma}^+, -1)} \leq \boldsymbol{N}_{x,1} + \boldsymbol{N}_{x,-1} < 2ck. \tag{5.9}$$

Combining (5.8) and (5.9) completes the proof. ∎

5.5 | Proof of Lemma 5.6

A straightforward induction shows that for any $p \in \mathscr{P}(\mathbb{R})$ the result $p^{(k)} = \mathrm{LL}_d^+(k)(p)$ of the k-fold application of LL_d^+ coincides with the distribution of the root value of the random operator $\mathrm{LL}^{+(k)}_{\boldsymbol{T}^{(2k)}}$ applied to a vector $(\boldsymbol{\eta}_x)_{x\in V(\boldsymbol{T}^{(2k)})}$ of independent samples from p. Indeed, for $k = 1$ the claim is immediate from the definitions. Moreover, for the inductive step we notice that the k-fold application of LL_d^+ comes down to applying LL_d^+ once to the outcome of the $(k-1)$-fold application. By the induction hypothesis,

$$p^{(k-1)} = \left(\mathrm{LL}^{+(k-1)}_{\boldsymbol{T}^{(2(k-1))}}(\boldsymbol{\eta}_x)_x\right)_o.$$

Finally, applying LL_d^+ to $p^{(k-1)}$ implies the assertion because the first layer of $\boldsymbol{T}^{(2k)}$ is independent of the subtrees rooted at the grandchildren $\partial^2 o$ of the root, which are distributed as independent random trees $\boldsymbol{T}^{(2(k-1))}$. The lemma follows from applying this identity to $p = \overline{\rho}^{(\ell-k)}$.

5.6 | Proof of Lemma 5.7

The operator LL_d^+ maps the space $\mathscr{W}_1(\mathbb{R})$ into itself because the derivative of $x \mapsto \log((1-\tanh(x/2))/2)$ is bounded by one in absolute value for all $x \in \mathbb{R}$. We proceed to show that $\mathrm{LL}_d^+ : \mathscr{W}_1(\mathbb{R}) \to \mathscr{W}_1(\mathbb{R})$ is a contraction. Thus, consider a sequence of independent random pairs $(\boldsymbol{\eta}_i, \boldsymbol{\eta}'_i)_{i\geq 1}$ with $\boldsymbol{\eta}_i \stackrel{\mathrm{d}}{=} \rho, \boldsymbol{\eta}'_i \stackrel{\mathrm{d}}{=} \rho'$. Then

$$W_1(\mathrm{LL}_d^+(\rho), \mathrm{LL}_d^+(\rho')) \leq \mathbb{E}\left|\sum_{i=1}^{d} \boldsymbol{s}_i \log\frac{1-\boldsymbol{s}_i\tanh(\boldsymbol{\eta}_i/2)}{1-\boldsymbol{s}_i\tanh(\boldsymbol{\eta}'_i/2)}\right| \leq d\mathbb{E}\left|\log\frac{1-\boldsymbol{s}_1\tanh(\boldsymbol{\eta}_1/2)}{1-\boldsymbol{s}_1\tanh(\boldsymbol{\eta}'_1/2)}\right|.$$

Since the function $z \mapsto \log(1+\tanh(z/2))$ is monotonically increasing, we obtain

$$\left|\log\frac{1+\tanh(\boldsymbol{\eta}_1/2)}{1+\tanh(\boldsymbol{\eta}'_1/2)}\right| = \left|\int_{\boldsymbol{\eta}'_1}^{\boldsymbol{\eta}_1} \frac{\partial\log(1+\tanh(z/2))}{\partial z}\mathrm{d}z\right| = \int_{\boldsymbol{\eta}_1\wedge\boldsymbol{\eta}'_1}^{\boldsymbol{\eta}_1\vee\boldsymbol{\eta}'_1} \frac{1-\tanh(z/2)}{2}\mathrm{d}z,$$
$$\left|\log\frac{1-\tanh(\boldsymbol{\eta}_1/2)}{1-\tanh(\boldsymbol{\eta}'_1/2)}\right| = \left|\int_{\boldsymbol{\eta}'_1}^{\boldsymbol{\eta}_1} \frac{\partial\log(1-\tanh(z/2))}{\partial z}\mathrm{d}z\right| = \int_{\boldsymbol{\eta}_1\wedge\boldsymbol{\eta}'_1}^{\boldsymbol{\eta}_1\vee\boldsymbol{\eta}'_1} \frac{1+\tanh(z/2)}{2}\mathrm{d}z.$$

Hence, $W_1(\mathrm{LL}_d^+(\rho), \mathrm{LL}_d^+(\rho')) \leq d\mathbb{E}\left|\boldsymbol{\eta}_1 - \boldsymbol{\eta}'_1\right|/2$ and therefore $W_1(\mathrm{LL}_d^+(\rho), \mathrm{LL}_d^+(\rho')) \leq dW_1(\rho,\rho')/2$.

Finally, we observe that ρ_d is a fixed point of LL_d^+. Indeed, Proposition 2.1 implies that $\boldsymbol{\eta}^{\rho_d}$ and $-\boldsymbol{\eta}^{\rho_d}$ are identically distributed. Therefore, if $\boldsymbol{s}_i, \boldsymbol{s}'_i \in \{\pm 1\}$ are uniform and independent, we obtain

$$\boldsymbol{s}_i \log\left(\left(1-\boldsymbol{s}_i\tanh(\boldsymbol{\eta}_{\rho_d,i}/2)\right)/2\right) \stackrel{\mathrm{d}}{=} \boldsymbol{s}_i \log\left(\left(1+\boldsymbol{s}'_i\tanh(\boldsymbol{\eta}_{\rho_d,i}/2)\right)/2\right).$$

Hence, recalling the definitions (4.1) and (5.3) of the operators, we see that $\mathrm{LL}_d^+(\rho_d) = \mathrm{LL}_d(\rho_d) = \rho_d$.

5.7 | Proof of Theorem 1.2

Consider the sub-formula $\nabla^{2\ell}(\boldsymbol{\Phi}, x_1)$ of $\boldsymbol{\Phi}$ obtained by deleting all clauses and variables at distance greater than 2ℓ from x_1. By design, we can couple $\nabla^{2\ell}(\boldsymbol{\Phi}, x_1)$ and $\boldsymbol{T}^{(2\ell)}$ such that both coincide w.h.p. Therefore, since any satisfying assignment of $\boldsymbol{\Phi}$ induces a satisfying assignment of $\boldsymbol{T}^{(2\ell)}$, Proposition 2.2 implies the Gibbs uniqueness property (1.11). Furthermore, because Proposition 1.4 shows that Belief Propagation correctly computes the root marginal $\mu_{\boldsymbol{T}^{(2\ell)}}(\boldsymbol{\sigma}_o = 1)$, (1.9) follows from (1.11).

5.8 | Proof of Corollary 1.3

Let $\pi_d^{(\ell)} = \mathrm{BP}^{(\ell)}(\delta_{1/2})$. Thanks to Proposition 2.1 it suffices to prove that

$$\lim_{\ell\to\infty}\limsup_{n\to\infty}\mathbb{E}[W_1(\pi_{\boldsymbol{\Phi}}, \pi_d^{(\ell)})] = 0. \tag{5.10}$$

Hence, fix $\varepsilon > 0$, pick a large $\ell = \ell(\varepsilon) > 0$ and a larger $L = L(\ell) > 0$. A routine second moment calculation shows that for any possible outcome T of $\boldsymbol{T}^{(2\ell)}$ the number X_T of variables x_i of $\boldsymbol{\Phi}$ such that $\nabla^{2\ell}(\boldsymbol{\Phi}, x_i) = T$ satisfies $X_T = n\mathbb{P}\left[\boldsymbol{T}^{(2\ell)} = T\right] + o(n)$ w.h.p. Hence, w.h.p. $\boldsymbol{\Phi}$ admits a coupling $\gamma_{\boldsymbol{\Phi}}$ of $\boldsymbol{T}^{(2\ell)}$ and a uniform variable $\boldsymbol{i}$ on $[n]$ such that $\gamma(\{\nabla^{2\ell}(\boldsymbol{\Phi}, x_{\boldsymbol{i}}) = \boldsymbol{T}^{(2\ell)}\}) = 1 - o(1)$. Furthermore, Theorem 1.2 implies that given $\nabla^{2\ell}(\boldsymbol{\Phi}, x_{\boldsymbol{i}}) = \boldsymbol{T}^{(2\ell)}$ we have

$$\mathbb{P}\left[\left|\mu_{\boldsymbol{\Phi}}(\boldsymbol{\sigma}_{x_{\boldsymbol{i}}} = 1) - \mu_{\boldsymbol{T}^{(2\ell)}}(\boldsymbol{\tau}_o = 1)\right| > \varepsilon\right] < \varepsilon, \tag{5.11}$$

provided ℓ is large enough. Finally, Lemma 1.4 implies together with a straightforward induction on ℓ that $\pi_d^{(\ell)}$ is the distribution of $\mu_{\boldsymbol{T}^{(2\ell)}}(\boldsymbol{\tau}_o = 1)$. Therefore, (5.10) follows from (5.11).

5.9 | Proof of Corollary 2.3

Fix $\varepsilon > 0$ and pick a small $\xi = \xi(\varepsilon) > 0$ and large $\ell = \ell(\xi) > 0$. Since k is fixed independently of n, Theorem 1.2 shows that w.h.p.

$$\sum_{i=1}^{k} \max_{\tau \in S(\boldsymbol{\Phi})} \left| \mu_{\boldsymbol{\Phi}}(\boldsymbol{\sigma}_{x_i} = 1 | \boldsymbol{\sigma}_{\partial^{2\ell} x_i} = \tau_{\partial^{2\ell} x_i}) - \mu^{(\ell)}_{\boldsymbol{\Phi}, x_i}(1) \right| < \xi. \tag{5.12}$$

Furthermore, the smallest pairwise distance between $x_1, \ldots, x_k$ exceeds 4ℓ w.h.p. Therefore, we can draw a sample $\boldsymbol{\sigma}$ from $\mu_{\boldsymbol{\Phi}}$ in two steps. First, draw $\boldsymbol{\sigma}'$ from $\mu_{\boldsymbol{\Phi}}$. Then, independently re-sample assignments of all the variables in $\nabla^{2\ell-2}(\boldsymbol{\Phi}, x_i)$ from $\mu_{\boldsymbol{\Phi}}(\cdot \,| \boldsymbol{\sigma}'_{\partial^{2\ell} x_i})$ for $i = 1, \ldots, k$. The resulting assignment $\boldsymbol{\sigma}''$ has distribution $\mu_{\boldsymbol{\Phi}}$ and the values $\boldsymbol{\sigma}''_{x_i}$, $i \in [k]$, are mutually independent given $\boldsymbol{\sigma}'$. Finally, since (5.12) shows that conditioning on the boundary conditions $\boldsymbol{\sigma}'_{\partial^{2\ell} x_i}$ is inconsequential w.h.p., we obtain the assertion by taking $\varepsilon \to 0$ sufficiently slowly.

6 | PROOF OF PROPOSITION 2.4

6.1 | Outline

The proof is based on a natural coupling of the random formulas $\boldsymbol{\Phi}_n$ and $\boldsymbol{\Phi}_{n+1}$ with n and $n+1$ variables, respectively. Specifically, let

$$\boldsymbol{m}' \stackrel{\mathrm{d}}{=} \mathrm{Po}(dn/2 - d/2), \qquad \boldsymbol{\Delta}'' \stackrel{\mathrm{d}}{=} \mathrm{Po}(d/2), \qquad \boldsymbol{\Delta}''' \stackrel{\mathrm{d}}{=} \mathrm{Po}(d) \tag{6.1}$$

be independent random variables. Moreover, let $\boldsymbol{\Phi}'$ be a random formula with n variables and $\boldsymbol{m}'$ clauses, chosen independently and uniformly from the set of all $4n(n-1)$ possible clauses. Then obtain $\boldsymbol{\Phi}''$ from $\boldsymbol{\Phi}'$ by adding another $\boldsymbol{\Delta}''$ uniformly random and independent clauses. Moreover, obtain $\boldsymbol{\Phi}'''$ from $\boldsymbol{\Phi}'$ by adding one variable x_{n+1} along with $\boldsymbol{\Delta}'''$ clauses, chosen uniformly and independently from the set of all $8n$ possible clauses that contain x_{n+1} and another variable from the set $\{x_1, \ldots, x_n\}$.

Fact 6.1. We have $\boldsymbol{\Phi}'' \stackrel{\mathrm{d}}{=} \boldsymbol{\Phi}_n$ and $\boldsymbol{\Phi}''' \stackrel{\mathrm{d}}{=} \boldsymbol{\Phi}_{n+1}$; therefore,

$$\mathbb{E}[\log(Z(\boldsymbol{\Phi}_{n+1}) \vee 1)] - \mathbb{E}[\log(Z(\boldsymbol{\Phi}_n) \vee 1)] = \mathbb{E}\left[\log \frac{Z(\boldsymbol{\Phi}''') \vee 1}{Z(\boldsymbol{\Phi}') \vee 1}\right] - \mathbb{E}\left[\log \frac{Z(\boldsymbol{\Phi}'') \vee 1}{Z(\boldsymbol{\Phi}') \vee 1}\right]. \tag{6.2}$$

Hence, the proof of Proposition 2.4 boils down to establishing the following two statements.

Proposition 6.2. *We have* $\lim_{n\to\infty} \mathbb{E} \log \frac{Z(\boldsymbol{\Phi}'') \vee 1}{Z(\boldsymbol{\Phi}') \vee 1} = \frac{d}{2} \mathbb{E}\left[\log\left(1 - \boldsymbol{\mu}_{\pi_d,1} \boldsymbol{\mu}_{\pi_d,2}\right)\right]$.

Proposition 6.3. *We have* $\lim_{n\to\infty} \mathbb{E} \log \frac{Z(\boldsymbol{\Phi}''') \vee 1}{Z(\boldsymbol{\Phi}') \vee 1} = \mathbb{E}\left[\log\left(\sum_{\sigma \in \{\pm 1\}} \prod_{i=1}^{\boldsymbol{d}} \left(1 - \mathbf{1}\{\sigma \neq \boldsymbol{s}_i\} \boldsymbol{\mu}_{\pi_d,i}\right)\right)\right]$.

Furthermore, to prove Propositions 6.2 and 6.3 we "just" need to understand the impact of a bounded expected number of "local" changes (such as adding a random clause) on the partition function.

The proof strategy sketched in the previous paragraph is known as the Aizenman-Sims-Starr scheme. The technique was originally deployed to study the Sherrington-Kirkpatrick spin glass model [6], but has since found various applications to models on sparse random graphs (e.g., [17, 41]). By comparison to prior applications, the difficulty here is that we apply this technique to a model with hard constraints. In effect, while typically the addition of a single clause will only reduce the number of satisfying assignments by a bounded factor, occasionally a much larger change might ensue. For instance, for any $0 < d < 2$ there is a small but nonzero probability that a single additional clause might close a "bicycle', that is, a sequence of clauses that induce an implication chain $x_i \to \cdots \to \neg x_i \to \cdots \to x_i$. Thus, a single unlucky clause might wipe out all satisfying assignments.

Suppose we wish to roughly estimate the change in the number of satisfying assignments upon going from $\mathbf{\Phi}'$ to $\mathbf{\Phi}'''$. Clearly $Z(\mathbf{\Phi}''') \leq 2Z(\mathbf{\Phi}')$ because we only add one new variable. But of course $Z(\mathbf{\Phi}''')$ might be much smaller than $Z(\mathbf{\Phi}')$. To obtain a bound, consider the new clauses $b_1, \ldots, b_{\mathbf{\Delta}'''}$ that were added along with x_{n+1} and let $y_1, \ldots, y_{\mathbf{\Delta}'''}$ be the variables of $\mathbf{\Phi}'$ where the new clauses attach. Define an assignment $\chi : Y = \{y_1, \ldots, y_{\mathbf{\Delta}'''}\} \to \{\pm 1\}$ by letting $\chi_{y_i} = sign(y_i, b_i)$; thus, χ satisfies the b_i. Furthermore, let

$$S(\mathbf{\Phi}', \chi) = \left\{\sigma \in S(\mathbf{\Phi}') : \forall y \in Y : \sigma_y = \chi_y\right\}, \qquad Z(\mathbf{\Phi}', \chi) = |S(\mathbf{\Phi}', \chi)|$$

be the set and the number of satisfying assignments of $\mathbf{\Phi}'$ that coincide with χ on Y. Because each $\sigma \in S(\mathbf{\Phi}', \chi)$ already satisfies all the new clauses regardless of the value assigned to x_{n+1}, we obtain $Z(\mathbf{\Phi}''') \geq 2Z(\mathbf{\Phi}', \chi)$. Hence, it seems that we just need to lower bound $Z(\mathbf{\Phi}', \chi)$.

To this end we could employ a process similar to the one that we applied in Section 5.4 to the tree $\boldsymbol{T}$. Generally, let $Y \subseteq \{x_1, \ldots, x_n\}$ be a set of variables and let $\chi \in \{\pm 1\}^Y$ be an assignment. The following process, known as the Unit Clause Propagation algorithm [27], chases the implications of imposing the assignment χ on Y:

> while $\mathbf{\Phi}'$ possesses a clause a that has exactly one neighboring variable $z \in \partial a$ on which the value $-sign(z, a)$ has been imposed, impose the value $sign(a, z')$ on the second variable $z' \in \partial a \setminus \{z\}$ of a.

Let $\mathscr{I}_\chi$ be the set of variables on which the process has imposed a value upon termination (including the initial set Y). Unfortunately, it is possible that $\mathbf{\Phi}'$ contains a clause a on whose both variables z, z' the "wrong" values $-sign(a, z), -sign(a, z')$ got imposed. In other words, Unit Clause might be left with contradictions. If such a clause exists we let $\boldsymbol{I}_\chi = n$. Otherwise we set $\boldsymbol{I}_\chi = |\mathscr{I}_\chi|$. We obtain the following lower bound on $Z(\mathbf{\Phi}', \chi)$.

Fact 6.4. We have $Z(\mathbf{\Phi}') \leq 2^{\boldsymbol{I}_\chi}(Z(\mathbf{\Phi}', \chi) \vee 1)$.

Proof. The inequality is trivially satisfied if $Z(\mathbf{\Phi}') = 0$ or $\boldsymbol{I}_\chi = n$. Hence, we may assume that $Z(\mathbf{\Phi}') > 0$ and that Unit Clause did not run into a contradiction. Consequently, Unit Clause produced an assignment χ^* of the variables $\mathscr{I}_\chi$ that satisfies all clauses a of $\mathbf{\Phi}'$ with $\partial a \cap \mathscr{I}_\chi \neq \emptyset$. Hence, for any satisfying assignment $\sigma \in S(\mathbf{\Phi}')$ we obtain another satisfying assignment $\hat{\sigma} \in S(\mathbf{\Phi}', \chi)$ by letting $\hat{\sigma} = \chi^*_x \mathbf{1}\{x \in \mathscr{I}_\chi\} + \sigma_x \mathbf{1}\{x \notin \mathscr{I}_\chi\}$, i.e., we overwrite the variables in $\mathscr{I}_\chi$ according to χ^*. Clearly, under the map $\sigma \mapsto \hat{\sigma}$ an assignment $\hat{\sigma} \in S(\mathbf{\Phi}', \chi)$ has at most $2^{\boldsymbol{I}_\chi}$ inverse images. ∎

Hence, we need an upper bound on $\boldsymbol{I}_\chi$, which will be proven at the end of Section 6.2.

Lemma 6.5. *There exists $C = C(d) > 0$ such that for every set $Y \subseteq \{x_1, \dots, x_n\}$ of size $|Y| \le \log^2 n$ and any $\chi \in \{\pm 1\}^Y$ we have $\mathbb{E}[\boldsymbol{I}_\chi] \le C|Y|^2$.*

Unfortunately, this first moment bound does not quite suffice for our purposes. Indeed, Lemma 6.5 allows for the possibility that $\boldsymbol{I}_\chi = n$ with probability $\Omega(1/n)$. In combination with Fact 6.4 this rough bound would lead to error terms that eclipse the "main" terms displayed in Propositions 6.2 and 6.3. But we cannot hope for a much better bound on $\boldsymbol{I}_\chi$. Indeed, $\mathbb{P}\left[\boldsymbol{I}_\chi = n\right] = \Omega(1/n)$ because the graph $G(\boldsymbol{\Phi}')$ likely contains a few short cycles and if Y contains a variable on a short cycle, then there is a $\Omega(1)$ probability that Unit Clause will cause a contradiction.

Hence, we need to be more circumspect. Previously we aimed for an assignment χ that satisfied *all* the new clauses $b_1, \dots, b_{\boldsymbol{\Delta}'''}$ added upon going to $\boldsymbol{\Phi}'''$. But we still have the new variable x_{n+1} at our disposal to at least satisfy a single clause b_i. Hence, we can afford to start Unit Clause from an assignment χ' that differs from χ on a single variable. Thus, for a set Y of variables and $\chi \in \{\pm 1\}^Y$ we define

$$\boldsymbol{A}_\chi = \min\left\{\boldsymbol{I}_{\chi'} : \chi' \in \{\pm 1\}^Y, \sum_{y \in Y} \mathbf{1}\{\chi_y \neq \chi'_y\} \le 1\right\}. \tag{6.3}$$

Lemma 6.6. *There exists $C' = C'(d) > 0$ such that for every set $Y \subseteq \{x_1, \dots, x_n\}$ of size $|Y| \le \log^2 n$ and any $\chi \in \{\pm 1\}^Y$ we have $\mathbb{E}[\boldsymbol{A}_\chi^2] \le C'|Y|^4$.*

This second moment bound significantly improves over Lemma 6.5. For instance, Lemma 6.6 implies that the probability of an enormous drop $Z(\boldsymbol{\Phi}''') \le \exp(-\Omega(n))Z(\boldsymbol{\Phi}')$ is bounded by $O(n^{-2})$. Once more this estimate is about tight because there is an $\Omega(n^{-2})$ probability that a single new clause closes a bicycle. As we shall see, with a bit of care the bound from Lemma 6.6 suffices to prove Propositions 6.2 and 6.3. Yet Lemma 6.5 has its uses, too, as it implies the following vital tail bound.

Corollary 6.7. *We have* $\mathrm{limsup}_{n\to\infty} \mathbb{E}\left[n \wedge \left|\log \frac{\mu_{\boldsymbol{\Phi}'}(\sigma_{x_1}=1)}{\mu_{\boldsymbol{\Phi}'}(\sigma_{x_1}=-1)}\right| \,\middle|\, Z(\boldsymbol{\Phi}') > 0\right] < \infty.$

We proceed to study Unit Clause Propagation in order to prove Lemmas 6.5,6.6 and Corollary 6.7. Then we will prove Propositions 6.2 and 6.3, which imply Proposition 2.4.

6.2 | Unit clause propagation

To avoid dependencies we consider a binomial model $\boldsymbol{\Phi}^\dagger$ of a random 2-SAT formula with variables $x_1, \dots, x_n$, where each of the $4\binom{n}{2}$ possible (unordered) 2-clauses is present with probability

$$p = d/(4n) + n^{-4/3} \tag{6.4}$$

independently. We define a random variable $\boldsymbol{A}_\chi^\dagger$ on $\boldsymbol{\Phi}^\dagger$ in perfect analogy to $\boldsymbol{A}_\chi$. Since the choice (6.4) of p ensures that $\boldsymbol{\Phi}^\dagger$ and $\boldsymbol{\Phi}'$ can be coupled so that the former has more clauses than the latter with probability $1 - o(n^{-2})$, it suffices to analyze $\boldsymbol{A}_\chi^\dagger$. Moreover, thanks to symmetry it suffices to prove Lemmas 6.5 and 6.6 under the assumption that the initial set of variables is $Y = \{x_1, \dots, x_\ell\}$, $\ell \le \log^2 n$.

At first glance investigating $\boldsymbol{A}_\chi^\dagger$ appears to be complicated by the fact that (6.3) takes the minimum over all possible χ'. To sidestep this issue we will investigate a "comprehensive" propagation process

whose progeny encompasses all the unit clauses that may result from any χ'. In its first round this process pursues for each variable x_i, $i \le \ell$, the Unit Clauses created by imposing either of the two possible truth values on x_i. The effect will be the imposition of truth values on all variables at distance two from Y. Subsequently we trace Unit Clause Propagation from the values imposed on the variables in $\partial^2 Y$. Hence, the difficulty of considering all χ' as in (6.3) disappears because the first step disregards χ.

To deal with possible contradictions the process will actually operate on literals rather than variables. Throughout each literal will belong to one of three possible categories: unexplored, explored, or finished. Initially the 2ℓ literals $x_1, \neg x_1, \ldots, x_\ell, \neg x_\ell$ qualify as explored and all others as unexplored. Formally, we let $\mathscr{E}_0 = \{x_1, \neg x_1, \ldots, x_\ell, \neg x_\ell\}$, $\mathscr{U}_0 = \{x_{\ell+1}, \neg x_{\ell+1}, \ldots, x_n, \neg x_n\}$ and $\mathscr{F}_0 = \emptyset$. Furthermore, for $t \ge 0$ we construct $\mathscr{E}_{t+1}, \mathscr{U}_{t+1}, \mathscr{F}_{t+1}$ as follows. If $\mathscr{E}_t = \emptyset$, the process has terminated and we set $\mathscr{E}_{t+1} = \mathscr{E}_t, \mathscr{U}_{t+1} = \mathscr{U}_t, \mathscr{F}_{t+1} = \mathscr{F}_t$. Otherwise, pick a literal $l_{t+1} \in \mathscr{E}_t$ and let $\mathscr{E}'_{t+1}$ be the set of all literals $l' \in \mathscr{U}_t$ such that $\mathbf{\Phi}^\dagger$ features the clause $\neg l_{t+1} \vee l'$. Furthermore, let

$$\mathscr{U}_{t+1} = \mathscr{U}_t \setminus \mathscr{E}'_{t+1}, \qquad \mathscr{E}_{t+1} = (\mathscr{E}_t \cup \mathscr{E}'_{t+1}) \setminus \{l_{t+1}\}, \qquad \mathscr{F}_{t+1} = \mathscr{F}_t \cup \{l_{t+1}\}.$$

Finally, the set $\mathscr{F}_\infty = \bigcup_{t \ge 1} \mathscr{F}_t$ contains all literals upon which Unit Clause could impose the value "true" from any initial assignment χ. A contradiction might result only if $x_i, \neg x_i \in \mathscr{F}_\infty$ for some $i > \ell$.

Claim 6.8. For any $0 < d < 2$ there exists $\zeta > 0$ such that for all $T > 8\ell/(2-d)$ we have $\mathbb{P}[|\mathscr{F}_\infty| > T] \le \exp(-\zeta T)$.

Proof. Let $t \ge 0$. Given $|\mathscr{U}_t|$ and $|\mathscr{E}_t|$ we have

$$\boldsymbol{X}_{t+1} = |\mathscr{E}_{t+1}| - |\mathscr{E}_t| + \mathbf{1}\left\{\mathscr{E}_t \neq \emptyset\right\} \stackrel{\mathrm{d}}{=} \mathrm{Bin}\left(|\mathscr{U}_t| \mathbf{1}\left\{|\mathscr{E}_t| \ge 0\right\}, p\right).$$

Moreover, given $|\mathscr{U}_t|$ and $|\mathscr{E}_t|$ let $\boldsymbol{Y}_{t+1} \stackrel{\mathrm{d}}{=} \mathrm{Bin}\left(2n - |\mathscr{U}_t| \mathbf{1}\left\{|\mathscr{E}_t| \ge 0\right\}, p\right)$ be independent of $\boldsymbol{X}_{t+1}$ and everything else, and set $\boldsymbol{X}^{\ge}_{t+1} = \boldsymbol{X}_{t+1} + \boldsymbol{Y}_{t+1}$. Then $(\boldsymbol{X}^{\ge}_t)_{t \ge 1}$ is an i.i.d. sequence of $\mathrm{Bin}(2n, p)$ random variables such that $\boldsymbol{X}^{\ge}_t \ge \boldsymbol{X}_t$ for all t. Hence, for any $T \ge 1$,

$$\begin{aligned}\mathbb{P}[|\mathscr{F}_\infty| > T] = \mathbb{P}[|\mathscr{E}_T| > 0] &\le \mathbb{P}\left[\sum_{t=1}^{T} \boldsymbol{X}_t > T - 2\ell\right] \le \mathbb{P}\left[\sum_{t=1}^{T} \boldsymbol{X}^{\ge}_t > T - 2\ell\right] \\ &= \mathbb{P}\left[\mathrm{Bin}(2nT, p) > T - 2\ell\right]. \end{aligned} \tag{6.5}$$

Furthermore, because $T > 8\ell/(2-d)$, (6.4) shows that for large enough n,

$$\frac{T - 2\ell}{2npT} \ge \frac{(1 - (2-d)/4)T}{2T(d/4 + o(1))} = \frac{d+2}{2d} + o(1) > 1.$$

Hence, the Chernoff bound shows that there exists $\zeta = \zeta(d) > 0$ (which approaches 0 as $d \to 2$) such that

$$\mathbb{P}\left[\mathrm{Bin}(2Tn, p) > T - 2\ell\right] \le \exp(-\zeta T) \qquad \text{for } T > 8\ell/(2-d). \tag{6.6}$$

Combining (6.5) and (6.6) completes the proof. ∎

Let $\mathbf{\Phi}^*$ be the sub-formula of $\mathbf{\Phi}^\dagger$ comprising all variables x such that $x \in \mathscr{F}_\infty$ or $\neg x \in \mathscr{F}_\infty$ along with all clauses a that contain two such variables. Let $\boldsymbol{n}^*$ be the number of variables of $\mathbf{\Phi}^*$ and let $\boldsymbol{m}^*$ be the number of clauses.

Claim 6.9. We have $\mathbb{P}[\boldsymbol{m}^* \geq \boldsymbol{n}^* - \ell + 1] \leq O(\ell^2/n)$ and $\mathbb{P}[\boldsymbol{m}^* > \boldsymbol{n}^* - \ell + 1] \leq O(\ell^4/n^2)$.

Proof. We set up a graph representing the literals involved in the exploration process and the clauses that contain such literals. Specifically, let $\neg\mathcal{F}_\infty = \{\neg l : l \in \mathcal{F}_\infty\}$ contain all negations of literals in $\mathcal{F}_\infty$. Moreover, let $\mathcal{G}$ be the graph whose vertices are the literals $\mathcal{F}_\infty \cup \neg\mathcal{F}_\infty$ as well as all clauses a of $\boldsymbol{\Phi}^\dagger$ that consist of two literals from $\mathcal{F}_\infty \cup \neg\mathcal{F}_\infty$. Let $\mathcal{C}_\infty$ be the set of such clauses a. For each clause $a \in \mathcal{C}_\infty$ the graph $\mathcal{G}$ contains two edges joining a and its two constituent literals. The graph $G(\boldsymbol{\Phi}^*)$ that we are ultimately interested in results from $\mathcal{G}$ by contracting pairs of inverse literals $l, \neg l \in \mathcal{F}_\infty \cup \neg\mathcal{F}_\infty$.

A large excess $\boldsymbol{m}^* - \boldsymbol{n}^*$ can either caused by the presence of atypically many clauses in $\mathcal{G}$ or by excess pairs of inverse litetals that get contracted. We first address the gain in clauses due to inclusion of $\neg\mathcal{F}_\infty$ and all induced clauses. The exploration process discovers each literal $\lambda \in \mathcal{F}_\infty \setminus \{x_1, \neg x_1, \ldots, x_\ell, \neg x_\ell\}$ via a clause $\neg l_t \vee \lambda$, where $\neg l_t \in \mathscr{E}_{t-1}$. Thus, $|\mathcal{C}_\infty| \geq |\mathcal{F}_\infty| - 2\ell$. Hence, the random variable $\boldsymbol{X} = |\mathcal{C}_\infty| - |\mathcal{F}_\infty| + 2\ell$ accounts for the number of excess clauses that are present among the literals $\mathcal{F}_\infty \cup \neg\mathcal{F}_\infty$ but that were not probed by the process. We highlight that $\boldsymbol{X}$ also counts clauses that contain two literals from the seed set $\{x_1, \neg x_1, \ldots, x_\ell, \neg x_\ell\}$. Because clauses appear in $\boldsymbol{\Phi}^\dagger$ independently with probability $p = O(d/n)$, we obtain the bounds

$$\mathbb{P}[\boldsymbol{X} \geq 1 || \mathcal{F}_\infty |] \leq O(|\mathcal{F}_\infty|^2/n), \qquad \mathbb{P}[\boldsymbol{X} \geq 2 || \mathcal{F}_\infty |] \leq O(|\mathcal{F}_\infty|^4/n^2). \tag{6.7}$$

Secondly, we investigate the loss in nodes due to contraction. Hence, $\boldsymbol{n}^* = |\mathcal{F}_\infty \cup \neg\mathcal{F}_\infty|/2$. By construction, the seeds $x_1, \neg x_1, \ldots, x_\ell, \neg x_\ell$ come in pairs. Let $\boldsymbol{X}' = \frac{1}{2}|\mathcal{F}_\infty \cap \neg\mathcal{F}_\infty| - \ell$ count the number of excess inverse literal pairs that we need to contract. Since the process is oblivious to the identities of the variables underlying the literals, given its size the set $\mathcal{F}_\infty \setminus \{x_1, \neg x_1, \ldots, x_\ell, \neg x_\ell\}$ is a uniformly random subset of the set $\{x_i, \neg x_i : \ell < i \leq n\}$ of nonseed literals. Therefore, a routine balls-into-bins argument shows that

$$\mathbb{P}[\boldsymbol{X}' \geq 1 || \mathcal{F}_\infty |] \leq O(|\mathcal{F}_\infty|^2/n), \qquad \mathbb{P}[\boldsymbol{X}' \geq 2 || \mathcal{F}_\infty |] \leq O(|\mathcal{F}_\infty|^4/n^2). \tag{6.8}$$

Finally, in order to estimate $\boldsymbol{m}^* - \boldsymbol{n}^*$ we consider four separate cases.

Case 1: $\boldsymbol{X} = \boldsymbol{X}' = 0$ Since $\boldsymbol{X} = 0$ the graph $\mathcal{G}$ is a forest with 2ℓ components rooted at $x_1, \neg x_1, \ldots, x_\ell, \neg x_\ell$. Moreover, since $\boldsymbol{X}' = 0$ we have $\mathcal{F}_\infty \cap \neg\mathcal{F}_\infty = \{x_1, \neg x_1, \ldots, x_\ell, \neg x_\ell\}$. Therefore, $G(\boldsymbol{\Phi}^*)$ is obtained from $\mathcal{G}$ by identifying the pairs $x_i, \neg x_i$ for $i = 1, \ldots, \ell$. Hence, $G(\boldsymbol{\Phi}^*)$ is a forest with ℓ components, and thus

$$\boldsymbol{m}^* = \boldsymbol{n}^* - \ell. \tag{6.9}$$

Case 2: $\boldsymbol{X} = 1$, $\boldsymbol{X}' = 0$ Obtain $\hat{\mathcal{G}}$ from $\mathcal{G}$ by adding one new vertex r whose neighbors are $x_1, \neg x_1, \ldots, x_\ell, \neg x_\ell$. Then $\hat{\mathcal{G}}$ is unicyclic because $\boldsymbol{X} = 1$. Let $\tilde{\mathcal{G}}$ be the graph obtained from $\hat{\mathcal{G}}$ by deleting the vertex r along with one (arbitrary) clause a from the cycle of $\hat{\mathcal{G}}$. Then $\tilde{\mathcal{G}}$ is a forest with 2ℓ components. Therefore, by the same token as in Case 1, $G(\boldsymbol{\Phi}^* - a)$ is a forest with ℓ components. Hence, $G(\boldsymbol{\Phi}^*)$, obtained by inserting clause a into $G(\boldsymbol{\Phi}^* - a)$, either contains a single cycle or consists of exactly $\ell - 1$ components. Thus, by (6.7)

$$\boldsymbol{m}^* \leq \boldsymbol{n}^* - \ell + 1, \qquad \mathbb{P}\left[\boldsymbol{X} = 1,\ \boldsymbol{X}' = 0 || \mathcal{F}_\infty |\right] = O(|\mathcal{F}_\infty|^2/n). \tag{6.10}$$

Case 3: $X = 0, X' = 1$ The graph $\hat{\mathcal{G}}$, defined as in Case 2, is a tree because $X = 0$. Suppose $(\mathscr{F}_\infty \cap \neg\mathscr{F}_\infty) \setminus \{x_1, \neg x_1, \dots, x_\ell, \neg x_\ell\} = \{y, \neg y\}$. Let a be a clause on the unique path from y to $\neg y$ in $\hat{\mathcal{G}}$. Then the same argument as in Case 1 shows that $G(\mathbf{\Phi}^* - a)$ is a forest with ℓ components. Therefore, $G(\mathbf{\Phi}^*)$ either contains a unique cycle or has precisely $\ell - 1$ components. Consequently, (6.8) yields

$$\boldsymbol{m}^* \le \boldsymbol{n}^* - \ell + 1, \qquad \mathbb{P}\left[\boldsymbol{X} = 0,\ \boldsymbol{X}' = 1 \middle| |\mathscr{F}_\infty|\right] = O(|\mathscr{F}_\infty|^2/n). \tag{6.11}$$

Case 4: $\boldsymbol{X} + \boldsymbol{X}' \ge 2$ In this case we do not have a bound on $\boldsymbol{m}^* - \boldsymbol{n}^*$, but we claim that

$$\mathbb{P}\left[\boldsymbol{X} + \boldsymbol{X}' \ge 2 \middle| |\mathscr{F}_\infty|\right] = O(|\mathscr{F}_\infty|^4/n^2). \tag{6.12}$$

Indeed, (6.7) and (6.8) readily imply that $\mathbb{P}\left[\boldsymbol{X} \vee \boldsymbol{X}' \ge 2 \middle| |\mathscr{F}_\infty|\right] = O(|\mathscr{F}_\infty|^4/n^2)$. Furthermore, since $\boldsymbol{X}$ is independent of $\boldsymbol{X}'$ given $\mathscr{F}_\infty$, (6.7) and (6.8) also yield the bound $\mathbb{P}\left[\boldsymbol{X} = \boldsymbol{X}' = 1 \middle| |\mathscr{F}_\infty|\right] = O(|\mathscr{F}_\infty|^4/n^2)$.

The assertion follows by combining (6.9)–(6.12) with Claim 6.8. ∎

Claim 6.10. For all $\chi \in \{\pm 1\}^{\{x_1, \dots x_\ell\}}$ we have $\boldsymbol{A}^\dagger_\chi \le |\mathscr{F}_\infty| \mathbf{1}\{\boldsymbol{m}^* \le \boldsymbol{n}^* - \ell + 1\} + n\mathbf{1}\{\boldsymbol{m}^* > \boldsymbol{n}^* - \ell + 1\}$.

Proof. The graph $G(\mathbf{\Phi}^*)$ consists of at most ℓ components (one for each of the initial variables $x_1, \dots, x_\ell$). Hence, $\boldsymbol{m}^* \ge \boldsymbol{n}^* - \ell$ and $G(\mathbf{\Phi}^*)$ is acyclic if $\boldsymbol{m}^* = \boldsymbol{n}^* - \ell$. Moreover, if $G(\mathbf{\Phi}^*)$ is acyclic then $\boldsymbol{A}^\dagger_\chi \le |\mathscr{F}_\infty|$ by construction.

Thus, we are left to consider the case $\boldsymbol{m}^* = \boldsymbol{n}^* - \ell + 1$. Then $\mathbf{\Phi}^*$ contains a clause a such that $G(\mathbf{\Phi}^* - a)$ is a forest with ℓ components rooted at $x_1, \dots, x_\ell$. Assume without loss that $a = x_{n-1} \vee x_n$. Then by construction we have $\{x_{n-1}, \neg x_{n-1}\} \cap \mathscr{F}_\infty \ne \emptyset$ and $\{x_n, \neg x_n\} \cap \mathscr{F}_\infty \ne \emptyset$. Furthermore, unless $\neg x_{n-1}, \neg x_n \in \mathscr{F}_\infty$ we have $\boldsymbol{A}_\chi \le \boldsymbol{I}_\chi \le |\mathscr{F}_\infty|$ as in the first case. Hence, assume that $\neg x_{n-1}, \neg x_n \in \mathscr{F}_\infty$. Let $i \in [\ell]$ be such that x_n belongs to the connected component of x_i in $G(\mathbf{\Phi}^* - a)$ and obtain χ' from χ by flipping the value assigned to x_i. Because $G(\mathbf{\Phi}^* - a)$ is a forest, we conclude that $\boldsymbol{A}^\dagger_\chi \le \boldsymbol{I}_\chi \wedge \boldsymbol{I}_{\chi'} \le |\mathscr{F}_\infty|$. ∎

Proof of Lemma 6.6. The choice of the clause probability p ensures that $\boldsymbol{A}^\dagger_\chi$ stochastically dominates $\boldsymbol{A}_\chi$. Therefore, the assertion follows from Claims 6.8–6.10. ∎

Proof of Lemma 6.5. The choice of the clause probability p and the construction of the set $\mathscr{F}_\infty$ guarantee that $\boldsymbol{I}_\chi$ is stochastically dominated by the random variable $|\mathscr{F}_\infty| \mathbf{1}\{\boldsymbol{m}^* \le \boldsymbol{n}^* - \ell\} + n\mathbf{1}\{\boldsymbol{m}^* > \boldsymbol{n}^* - \ell\}$. Hence, Claims 6.8–6.10 imply the desired bound. ∎

Proof of Corollary 6.7. Let $Y = \{x_1\}$ and $\chi^+_{x_1} = 1$, $\chi^-_{x_1} = -1$. Assume that $\mathbf{\Phi}'$ is satisfiable. Then Fact 6.4 implies that

$$n \wedge \left|\log \frac{\mu_{\mathbf{\Phi}'}(\boldsymbol{\sigma}_{x_1} = 1)}{\mu_{\mathbf{\Phi}'}(\boldsymbol{\sigma}_{x_1} = -1)}\right| \le \boldsymbol{I}_{\chi^-} + \boldsymbol{I}_{\chi^+}.$$

Therefore, the assertion follows from Lemma 6.5. ∎

6.3 | Proof of Proposition 6.2

Let $c_1, \ldots, c_{\Delta''}$ be the new clauses added to Φ'' and let $Y = \{y_1, z_1, \ldots, y_{\Delta''}, z_{\Delta''}\}$ be the set of variables that occur in these clauses. We begin by deriving the following rough bound.

Lemma 6.11. *We have* $\mathbb{E}\left[\log^2 \frac{Z(\Phi'')\vee 1}{Z(\Phi')\vee 1}\right] = O(1)$.

Proof. If Φ' is unsatisfiable then so is Φ'' and thus $(Z(\Phi'') \vee 1)/(Z(\Phi') \vee 1) = 1$. Hence, we may assume that $Z(\Phi') \geq 1$. If $|Y| = 2\Delta''$, the new clauses attach to disjoint sets of variables. Consider the truth value assignment $\chi \in \{\pm 1\}^Y$ that satisfies both literals in each of the clauses $c_1, \ldots, c_{\Delta''}$. Fact 6.4 shows that

$$Z(\Phi'') \vee 1 \geq Z(\Phi', \chi) \vee 1 \geq 2^{-A_\chi} Z(\Phi'). \tag{6.13}$$

Combining (6.13) with Lemma 6.6 and recalling that $\Delta'' \stackrel{\mathrm{d}}{=} \mathrm{Po}(d/2)$, we obtain

$$\mathbb{E}\left[\mathbf{1}\left\{Z(\Phi') \geq 1,\ |Y| = 2\Delta''\right\} \log^2 \frac{Z(\Phi'') \vee 1}{Z(\Phi') \vee 1}\right] \leq \mathbb{E}\left[\mathbf{1}\left\{Z(\Phi') \geq 1,\ |Y| = 2\Delta''\right\} A_\chi^2\right] = O(1). \tag{6.14}$$

Next, consider the event $|Y| = 2\Delta'' - 1$. Because $c_1, \ldots, c_{\Delta''}$ are drawn independently, we have

$$\mathbb{P}\left[|Y| = 2\Delta'' - 1 | \Delta''\right] \leq O((\Delta'')^2/n). \tag{6.15}$$

Moreover, because the signs of the clauses $c_1, \ldots, c_{\Delta''}$ are independent of Φ', given $|Y| = 2\Delta - 1$ there exists an assignment $\chi \in \{\pm 1\}^Y$, stochastically independent of Φ', that satisfies $c_1, \ldots, c_{\Delta''}$. Fact 6.4 yields $Z(\Phi'')\vee 1 \geq Z(\Phi', \chi) \geq 2^{-I_\chi} Z(\Phi')$. Therefore, since $\log((Z(\Phi'')\vee 1)/(Z(\Phi')\vee 1)) \leq n$, Lemma 6.5 and (6.15) imply

$$\mathbb{E}\left[\mathbf{1}\left\{Z(\Phi') \geq 1,\ |Y| = 2\Delta'' - 1\right\} \log^2 \frac{Z(\Phi'') \vee 1}{Z(\Phi') \vee 1}\right] \leq n\mathbb{E}\left[\mathbf{1}\left\{|Y| = 2\Delta'' - 1\right\} I_\chi\right] = O(1). \tag{6.16}$$

Finally, consider the event $|Y| < 2\Delta'' - 1$. Due to the independence of $c_1, \ldots, c_{\Delta''}$, this event occurs with probability $O(n^{-2})$. Hence, the deterministic bound $(Z(\Phi'') \vee 1)/(Z(\Phi') \vee 1) \geq 2^{-n}$ implies

$$\mathbb{E}\left[\mathbf{1}\left\{Z(\Phi') \geq 1,\ |Y| < 2\Delta'' - 1\right\} \log^2 \frac{Z(\Phi'') \vee 1}{Z(\Phi') \vee 1}\right] = O(1). \tag{6.17}$$

The assertion follows from (6.14), (6.16) and (6.17). ∎

Lemma 6.12. *There exists a number* $K > 0$ *such that for every* $\varepsilon > 0$ *we have*

$$\limsup_{n\to\infty} \mathbb{E}\left[\left(\sum_{i=1}^{\Delta''} \Lambda_\varepsilon (1 - \mu_{\Phi'}(\sigma_{y_i} = -sign(c_i, y_i))\mu_{\Phi'}(\sigma_{z_i} = -sign(c_i, z_i))\right)^2 | Z(\Phi') > 0\right] \leq K.$$

Proof. We recall the function Λ_ε from (1.14). Since $\boldsymbol{\Delta}'' \stackrel{\mathrm{d}}{=} \mathrm{Po}(d/2)$ and the pair $(\boldsymbol{y}_1, \boldsymbol{z}_1)$ is uniformly random, due to Cauchy-Schwarz it suffices to prove $\limsup_{n\to\infty} \mathbb{E}\left[\Lambda_\varepsilon(1-\mu_{\boldsymbol{\Phi}'}(\boldsymbol{\sigma}_{x_1}=1)\mu_{\boldsymbol{\Phi}'}(\boldsymbol{\sigma}_{x_2}=1))^2 | Z(\boldsymbol{\Phi}')>0\right] \le K$ for every $\varepsilon>0$. We observe that

$$\begin{aligned}
&\limsup_{n\to\infty} \mathbb{E}\left[\Lambda_\varepsilon(1-\mu_{\boldsymbol{\Phi}'}(\boldsymbol{\sigma}_{x_1}=1)\mu_{\boldsymbol{\Phi}'}(\boldsymbol{\sigma}_{x_2}=1))^2 | Z(\boldsymbol{\Phi}')>0\right] \\
&\quad\le \limsup_{n\to\infty} \mathbb{E}\left[\Lambda_\varepsilon(1-\mu_{\boldsymbol{\Phi}'}(\boldsymbol{\sigma}_{x_1}=1))^2 | Z(\boldsymbol{\Phi}')>0\right] \\
&\quad= \limsup_{n\to\infty} \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \Lambda_\varepsilon(1-\mu_{\boldsymbol{\Phi}'}(\boldsymbol{\sigma}_{x_i}=1))^2 | Z(\boldsymbol{\Phi}')>0\right]. \qquad (6.18)
\end{aligned}$$

Moreover, $\boldsymbol{\Phi}'$ has $\boldsymbol{m}' \stackrel{\mathrm{d}}{=} \mathrm{Po}(dn/2 - d/2)$ clauses, while $\boldsymbol{\Phi} = \boldsymbol{\Phi}_n$ has $\boldsymbol{m} \stackrel{\mathrm{d}}{=} \mathrm{Po}(dn/2)$ clauses. Since $d_{\mathrm{TV}}(\boldsymbol{m}', \boldsymbol{m}) = o(1)$, the formulas $\boldsymbol{\Phi}', \boldsymbol{\Phi}$ can be coupled such that both coincide w.h.p. Hence, for any fixed $\varepsilon>0$ we have

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \Lambda_\varepsilon(1-\mu_{\boldsymbol{\Phi}'}(\boldsymbol{\sigma}_{x_i}=1))^2 | Z(\boldsymbol{\Phi}')\right] \\
&\quad= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \Lambda_\varepsilon(1-\mu_{\boldsymbol{\Phi}}(\boldsymbol{\sigma}_{x_i}=1))^2 | Z(\boldsymbol{\Phi}')\right] + o(1). \qquad (6.19)
\end{aligned}$$

Furthermore, since for every $\varepsilon>0$ the function $u\in[0,1] \mapsto \Lambda_\varepsilon(1-u)^2$ is continuous, Corollary 1.3 implies that

$$\frac{1}{n}\sum_{i=1}^n \Lambda_\varepsilon(1-\mu_{\boldsymbol{\Phi}}(\boldsymbol{\sigma}_{x_i}=1))^2 \quad \underline{n\to\infty} \quad \mathbb{E}\left[\Lambda_\varepsilon(1-\boldsymbol{\mu}_{\pi_d})^2\right] \qquad \text{in probability.} \qquad (6.20)$$

Since (2.1) shows that $\mathbb{E}\left[\Lambda_\varepsilon(1-\boldsymbol{\mu}_{\pi_d})^2\right] \le \mathbb{E}\left[\log^2(1-\boldsymbol{\mu}_{\pi_d})\right] < \infty$, the assertion follows from (6.18)–(6.20). ∎

Lemma 6.13. *For any $\delta>0$ there exists $\varepsilon>0$ such that*

$$\operatorname*{limsup}_{n\to\infty} \left|\mathbb{E}\left[\log\frac{Z(\boldsymbol{\Phi}'')\vee 1}{Z(\boldsymbol{\Phi}')\vee 1}\right] - \frac{d}{2}\mathbb{E}\left[\Lambda_\varepsilon\left(1-\mu_{\boldsymbol{\Phi}',x_1}(\boldsymbol{s}_1)\mu_{\boldsymbol{\Phi}',x_2}(\boldsymbol{s}_2)\right)|Z(\boldsymbol{\Phi}')\right]\right| < \delta.$$

Proof. Choose small enough $\xi=\xi(\delta)>\eta=\eta(\xi)>\varepsilon=\varepsilon(\eta)>0$, assume that $n>n_0(\varepsilon)$ is sufficiently large and let $(\gamma_n)_n$ be a sequence of positive reals, depending on ξ and η, that tends to zero sufficiently slowly. Let $\mathscr{E}=\mathscr{E}_n$ be the event that the following five statements hold.

E1 $Z(\boldsymbol{\Phi}')>0$.
E2 $|\boldsymbol{Y}| = 2\boldsymbol{\Delta}''$.
E3 $\boldsymbol{\Delta}'' < \xi^{-1/4}$.
E4 for all $y\in\boldsymbol{Y}$ and all $s\in\{\pm1\}$ we have $\mu_{\boldsymbol{\Phi}'}(\boldsymbol{\sigma}_y=s)<1-2\eta$.
E5 $\sum_{\sigma\in\{\pm1\}^{\boldsymbol{Y}}}\left|\mu_{\boldsymbol{\Phi}}(\forall y\in\boldsymbol{Y}:\boldsymbol{\sigma}_y=\sigma_y) - \prod_{y\in\boldsymbol{Y}}\mu_{\boldsymbol{\Phi}}(\boldsymbol{\sigma}_y=\sigma_y)\right| < \gamma_n$.

The first two events **E1**, **E2** occur with probability $1-o(1)$ as $n\to\infty$. Moreover, $\mathbb{P}[\mathbf{E3}] > 1-\xi$ if ξ is small enough. Furthermore, since Corollary 1.3 shows that $\pi_{\boldsymbol{\Phi}}$ converges to π_d weakly in

probability, the tail bound (2.1) implies that $\mathbb{P}\left[\mathbf{E4}|\boldsymbol{\Delta}'' < \xi^{-1/4}\right] > 1-\xi$, provided that η is small enough. Additionally, Corollary 2.3 implies $\mathbb{P}[\mathbf{E5}|\mathbf{E1–E4}] = 1 - o(1)$ if $\gamma_n \to 0$ slowly enough. Consequently,

$$\mathbb{P}[\mathscr{E}] > 1 - 4\xi. \tag{6.21}$$

Combining Lemma 6.11, (6.21) and the Cauchy-Schwarz inequality, we obtain

$$\left|\mathbb{E}\left[(1-\mathbf{1}_{\mathscr{E}})\log\frac{Z(\boldsymbol{\Phi}'')}{Z(\boldsymbol{\Phi}')}\right]\right| \le \delta/3 + o(1). \tag{6.22}$$

Similarly, by Lemma 6.12, (6.21) and Cauchy-Schwarz,

$$\left|\mathbb{E}\left[(1-\mathbf{1}_{\mathscr{E}})\sum_{i=1}^{\boldsymbol{\Delta}''}\Lambda_\varepsilon\left(1-\mu_{\boldsymbol{\Phi}'}(\boldsymbol{\sigma}_{\mathbf{y}_i}=-sign(\mathbf{y}_i,c_i))\mu_{\boldsymbol{\Phi}'}(\boldsymbol{\sigma}_{\mathbf{z}_i}=-sign(\mathbf{z}_i,c_i))\right)\right]\right| \le \delta/3 + o(1). \tag{6.23}$$

Furthermore, because the distribution of $\boldsymbol{\Phi}'$ is invariant under permutations of the variables $x_1, \ldots, x_n$ and $\mathbb{E}[\boldsymbol{\Delta}''] = d/2$,

$$\begin{aligned}&\mathbb{E}\left[\sum_{i=1}^{\boldsymbol{\Delta}''}\Lambda_\varepsilon\left(1-\mu_{\boldsymbol{\Phi}'}(\boldsymbol{\sigma}_{\mathbf{y}_i}=-sign(\mathbf{y}_i,c_i))\mu_{\boldsymbol{\Phi}'}(\boldsymbol{\sigma}_{\mathbf{z}_i}=-sign(\mathbf{z}_i,c_i))\right)|Z(\boldsymbol{\Phi}')>0\right]\\ &\quad=\frac{d}{2}\mathbb{E}\left[\Lambda_\varepsilon\left(1-\mu_{\boldsymbol{\Phi}'}(\boldsymbol{\sigma}_{x_1}=s_1)\mu_{\boldsymbol{\Phi}'}(\boldsymbol{\sigma}_{x_2}=s_2)\right)|Z(\boldsymbol{\Phi}')>0\right].\end{aligned} \tag{6.24}$$

Moreover, on the event $\mathscr{E}$ we have

$$\begin{aligned}\frac{Z(\boldsymbol{\Phi}'')}{Z(\boldsymbol{\Phi}')} &= \sum_{\sigma\in\{\pm1\}^Y}\mathbf{1}\{\sigma \text{ satisfies } c_1,\ldots,c_{\boldsymbol{\Delta}''}\}\,\mu_{\boldsymbol{\Phi}'}\left(\forall y\in\boldsymbol{Y}:\boldsymbol{\sigma}_y=\sigma_y\right)\\ &= \sum_{\sigma\in\{\pm1\}^Y}\mathbf{1}\{\sigma \text{ satisfies } c_1,\ldots,c_{\boldsymbol{\Delta}''}\}\prod_{y\in\boldsymbol{Y}}\mu_{\boldsymbol{\Phi}'}\left(\boldsymbol{\sigma}_y=\sigma_y\right)+o(1) \qquad [\text{due to } \mathbf{E3,E5}]\\ &= \prod_{i=1}^{\boldsymbol{\Delta}''}\left(1-\mu_{\boldsymbol{\Phi}'}\left(\boldsymbol{\sigma}_{\mathbf{y}_i}=-sign(\mathbf{y}_i,c_i)\right)\mu_{\boldsymbol{\Phi}'}\left(\boldsymbol{\sigma}_{\mathbf{z}_i}=-sign(\mathbf{z}_i,c_i)\right)\right)+o(1).\end{aligned}$$

Therefore, by **E4**

$$\begin{aligned}&\mathbb{E}\left[\mathbf{1}_{\mathscr{E}}\log\frac{Z(\boldsymbol{\Phi}'')}{Z(\boldsymbol{\Phi}')}\right]\\ &\quad=\mathbb{E}\left[\mathbf{1}_{\mathscr{E}}\sum_{i=1}^{\boldsymbol{\Delta}''}\log\left(1-\mu_{\boldsymbol{\Phi}'}\left(\boldsymbol{\sigma}_{\mathbf{y}_i}=-sign(\mathbf{y}_i,c_i)\right)\mu_{\boldsymbol{\Phi}'}\left(\boldsymbol{\sigma}_{\mathbf{z}_i}=-sign(\mathbf{z}_i,c_i)\right)\right)\right]+o(1)\\ &\quad=\mathbb{E}\left[\mathbf{1}_{\mathscr{E}}\sum_{i=1}^{\boldsymbol{\Delta}''}\Lambda_\varepsilon\left(1-\mu_{\boldsymbol{\Phi}'}\left(\boldsymbol{\sigma}_{\mathbf{y}_i}=-sign(\mathbf{y}_i,c_i)\right)\mu_{\boldsymbol{\Phi}'}\left(\boldsymbol{\sigma}_{\mathbf{z}_i}=-sign(\mathbf{z}_i,c_i)\right)\right)\right]+o(1).\end{aligned} \tag{6.25}$$

Finally, the assertion follows from (6.22)–(6.25). ∎

Proof of Proposition 6.2. Proposition 2.1 shows that $\boldsymbol{\mu}_{\pi_d,1}$ and $1-\boldsymbol{\mu}_{\pi_d,1}$ are identically distributed. Since Λ_ε is continuous and bounded, Corollary 1.3 therefore implies that

$$\begin{aligned}\lim_{n\to\infty}\mathbb{E}\left[\Lambda_\varepsilon\left(1-\mu_{\boldsymbol{\Phi}',x_1}(s_1)\mu_{\boldsymbol{\Phi}',x_2}(s_2)\right)\right] &= \mathbb{E}\left[\Lambda_\varepsilon\left(1-\left(\frac{1-s_1}{2}+s_1\boldsymbol{\mu}_{\pi_d,1}\right)\left(\frac{1-s_2}{2}+s_2\boldsymbol{\mu}_{\pi_d,2}\right)\right)\right]\\ &= \mathbb{E}\left[\Lambda_\varepsilon\left(1-\boldsymbol{\mu}_{\pi_d,1}\boldsymbol{\mu}_{\pi_d,2}\right)\right]. \end{aligned} \tag{6.26}$$

for every $\varepsilon>0$. Furthermore, since $\Lambda_\varepsilon(1-\boldsymbol{\mu}_{\pi_d,1}\boldsymbol{\mu}_{\pi_d,2})$ decreases monotonically to $\log(1-\boldsymbol{\mu}_{\pi_d,1}\boldsymbol{\mu}_{\pi_d,2})$ as $\varepsilon\to 0$, the monotone convergence theorem and (2.2) yield

$$\lim_{\varepsilon\to 0}\mathbb{E}\left[\Lambda_\varepsilon\left(1-\boldsymbol{\mu}_{\pi_d,1}\boldsymbol{\mu}_{\pi_d,2}\right)\right] = \mathbb{E}\log\left(1-\boldsymbol{\mu}_{\pi_d,1}\boldsymbol{\mu}_{\pi_d,2}\right). \tag{6.27}$$

Combining (6.26) and (6.27) and Lemma 6.13 completes the proof. ∎

6.4 | Proof of Proposition 6.3

The steps that we follow are analogous to the ones from the proof of Proposition 6.2. Recall that $\boldsymbol{\Phi}'''$ is obtained from $\boldsymbol{\Phi}'$ by adding one variable x_{n+1} along with random adjacent clauses $b_1,\ldots,b_{\boldsymbol{\Delta}'''}$, where $\boldsymbol{\Delta}'''$ is a Poisson variable with mean d. Let $\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{\boldsymbol{\Delta}'''}\in\{x_1,\ldots,x_n\}$ be the variables of $\boldsymbol{\Phi}'$ where the new clauses attach and let $\boldsymbol{Y}=\{\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{\boldsymbol{\Delta}'''}\}$. We begin with the following L_2-bound.

Lemma 6.14. *We have* $\mathit{limsup}_{n\to\infty}\mathbb{E}\left[\log^2\frac{Z(\boldsymbol{\Phi}''')\vee 1}{Z(\boldsymbol{\Phi}')\vee 1}\right]<\infty$.

Proof. If $\boldsymbol{\Phi}'$ is unsatisfiable, then so is $\boldsymbol{\Phi}'''$ and thus $(Z(\boldsymbol{\Phi}''')\vee 1)/(Z(\boldsymbol{\Phi}')\vee 1)=1$. Hence, we may assume that $Z(\boldsymbol{\Phi}')\geq 1$. We now consider three scenarios. First, suppose that $|\boldsymbol{Y}|=\boldsymbol{\Delta}'''$, i.e., the new clauses attach to distinct variables of $\boldsymbol{\Phi}'$. Then define an assignment $\chi\in\{\pm1\}^{\boldsymbol{Y}}$ by setting each $y\in\boldsymbol{Y}$ to the value that satisfies the unique clause among $b_1,\ldots,b_{\boldsymbol{\Delta}'''}$ in which y occurs. We claim that

$$Z(\boldsymbol{\Phi}''')\vee 1\geq 2^{-A_\chi}Z(\boldsymbol{\Phi}'). \tag{6.28}$$

Indeed, if $\chi'\in\{\pm1\}^{\boldsymbol{Y}}$ differs from χ on only one variable, then we can always satisfy all clauses $b_1,\ldots,b_{\boldsymbol{\Delta}'''}$ by setting x_{n+1} appropriately. Therefore, (6.28) follows from Fact 6.4 and the definition (6.3) of $\boldsymbol{A}_\chi$. Combining (6.28) with Lemma 6.6, we obtain

$$\mathbb{E}\left[\mathbf{1}\left\{Z(\boldsymbol{\Phi}')\geq 1,\ |\boldsymbol{Y}|=\boldsymbol{\Delta}'''\right\}\log^2\frac{Z(\boldsymbol{\Phi}''')\vee 1}{Z(\boldsymbol{\Phi}')\vee 1}\right]\leq\mathbb{E}\left[\mathbf{1}\left\{|\boldsymbol{Y}|=\boldsymbol{\Delta}'''\right\}A_\chi^2\right]=O(1). \tag{6.29}$$

Second, consider the case $|\boldsymbol{Y}|=\boldsymbol{\Delta}'''-1$. Because $b_1,\ldots,b_{\boldsymbol{\Delta}'''}$ are drawn independently, we have

$$\mathbb{P}\left[|\boldsymbol{Y}|=\boldsymbol{\Delta}'''-1|\boldsymbol{\Delta}''\right]=O((\boldsymbol{\Delta}''')^2/n). \tag{6.30}$$

Furthermore, there exists an assignment $\chi\in\{\pm1\}^{\boldsymbol{Y}}$ under which all but one of the clauses $b_1,\ldots,b_{\boldsymbol{\Delta}'''}$ are satisfied. This assignment is independent of $\boldsymbol{\Phi}'$ because the signs of $b_1,\ldots,b_{\boldsymbol{\Delta}'''}$ are.

Since we can use the new variable x_{n+1} to satisfy the last clause as well, Fact 6.4 implies the bound $(Z(\mathbf{\Phi}''') \vee 1)/Z(\mathbf{\Phi}') \geq 2^{-\boldsymbol{I}_\chi}$. Therefore, Lemma 6.5 and (6.30) yield

$$\mathbb{E}\left[\mathbf{1}\left\{Z(\mathbf{\Phi}') \geq 1,\ |\boldsymbol{Y}| = \boldsymbol{\Delta}''' - 1\right\} \log^2 \frac{Z(\mathbf{\Phi}'') \vee 1}{Z(\mathbf{\Phi}') \vee 1}\right] \leq \mathbb{E}\left[\mathbf{1}\left\{Z(\mathbf{\Phi}') \geq 1,\ |\boldsymbol{Y}| = \boldsymbol{\Delta}''' - 1\right\} \boldsymbol{I}_\chi^2\right]$$
$$\leq n\mathbb{E}\left[\mathbf{1}\left\{|\boldsymbol{Y}| = \boldsymbol{\Delta}''' - 1\right\} \boldsymbol{I}_\chi\right] = O(1). \quad (6.31)$$

Finally, because $b_1, \ldots, b_{\boldsymbol{\Delta}'''}$ are drawn independently, the event $\{|\boldsymbol{Y}| < \boldsymbol{\Delta}''' - 1\}$ has probability $O(n^{-2})$. Therefore, the deterministic bound $(Z(\mathbf{\Phi}''') \vee 1)/(Z(\mathbf{\Phi}') \vee 1) \geq 2^{-n}$ ensures that

$$\mathbb{E}\left[\mathbf{1}\left\{Z(\mathbf{\Phi}') \geq 1,\ |\boldsymbol{Y}| < \boldsymbol{\Delta}''' - 1\right\} \log^2 \frac{Z(\mathbf{\Phi}''') \vee 1}{Z(\mathbf{\Phi}') \vee 1}\right] = O(1). \quad (6.32)$$

The assertion follows from (6.29), (6.31) and (6.32). ∎

Lemma 6.15. *There exists $K > 0$ such that for every $\varepsilon > 0$ we have*

$$\operatorname*{limsup}_{n\to\infty} \mathbb{E}\left[\Lambda_\varepsilon\left(\sum_{s\in\{\pm1\}} \prod_{i=1}^{\boldsymbol{\Delta}'''} \left(1 - \mathbf{1}\left\{s \neq sign(x_{n+1}, b_i)\right\} \mu_{\mathbf{\Phi}'}(\boldsymbol{\sigma}_{\boldsymbol{y}_i} = -sign(\boldsymbol{y}_i, b_i))\right)\right)^2 |Z(\mathbf{\Phi}') > 0\right] \leq K.$$

Proof. Since $\boldsymbol{\Delta}''' \stackrel{\mathrm{d}}{=} \mathrm{Po}(d/2)$, $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{\boldsymbol{\Delta}'''}$ and the signs $sign(b_i, \boldsymbol{y}_i)$ are uniformly random, we obtain

$$\mathbb{E}\left[\Lambda_\varepsilon\left(\sum_{s\in\{\pm1\}} \prod_{i=1}^{\boldsymbol{\Delta}'''} \left(1 - \mathbf{1}\left\{s \neq sign(x_{n+1}, b_i)\right\} \mu_{\mathbf{\Phi}'}(\boldsymbol{\sigma}_{\boldsymbol{y}_i} = -sign(\boldsymbol{y}_i, b_i))\right)\right)^2 |Z(\mathbf{\Phi}') > 0\right]$$
$$\leq 1 + \mathbb{E}\left[\Lambda_\varepsilon\left(\prod_{i=1}^{\boldsymbol{\Delta}''} \mu_{\mathbf{\Phi}'}(\boldsymbol{\sigma}_{\boldsymbol{y}_i} = 1)\right)^2 |Z(\mathbf{\Phi}') > 0\right]$$
$$\leq 1 + d\mathbb{E}\left[\Lambda_\varepsilon\left(\mu_{\mathbf{\Phi}'}(\boldsymbol{\sigma}_{\boldsymbol{y}_1} = 1)\right)^2 |Z(\mathbf{\Phi}') > 0\right]. \quad (6.33)$$

Furthermore, the formulas $\mathbf{\Phi}'$, $\mathbf{\Phi}$ can be coupled such that both coincide w.h.p. (cf. the proof of Lemma 6.12). Therefore, Corollary 1.3 implies that for every $\varepsilon > 0$,

$$\mathbb{E}\left[\Lambda_\varepsilon\left(\mu_{\mathbf{\Phi}'}(\boldsymbol{\sigma}_{\boldsymbol{y}_1} = 1)\right)^2 |Z(\mathbf{\Phi}') > 0\right] = \mathbb{E}\left[\Lambda_\varepsilon\left(\boldsymbol{\mu}_{\pi_{\mathbf{\Phi}}}\right)^2 |Z(\mathbf{\Phi}) > 0\right] + o(1)$$
$$= \mathbb{E}\left[\Lambda_\varepsilon\left(\boldsymbol{\mu}_{\pi_d}^2\right)\right] + o(1) \leq \mathbb{E}\left[\log^2 \boldsymbol{\mu}_{\pi_d}\right] + o(1). \quad (6.34)$$

Since (2.1) implies that $\mathbb{E}\left[\log^2 \boldsymbol{\mu}_{\pi_d}\right] < \infty$, the assertion follows from (6.33)–(6.34). ∎

Lemma 6.16. *For any $\delta > 0$ there exists $\varepsilon_0 > 0$ such that for every $0 < \varepsilon < \varepsilon_0$,*

$$\left|\mathbb{E}\left[\log \frac{Z(\mathbf{\Phi}''') \vee 1}{Z(\mathbf{\Phi}') \vee 1}\right] - \mathbb{E}\left[\Lambda_\varepsilon\left(\sum_{s\in\{\pm1\}} \prod_{i=1}^{d} \left(1 - \mathbf{1}\left\{s \neq \boldsymbol{s}_i\right\} \mu_{\mathbf{\Phi}'}(\boldsymbol{\sigma}_{x_i} = \boldsymbol{s}_i')\right)\right) |Z(\mathbf{\Phi}') > 0\right]\right|$$
$$< \delta + o(1).$$

Proof. Choose small enough $\xi = \xi(\delta) > \eta = \eta(\xi) > \varepsilon = \varepsilon(\eta) > 0$, assume that $n > n_0(\varepsilon)$ is sufficiently large and let $(\gamma_n)_n$ be a sequence of numbers $\gamma_n > 0$ that tends to zero slowly. Let $\mathscr{E} = \mathscr{E}_n$ be the event that the following five statements are satisfied.

E1 $Z(\mathbf{\Phi}') > 0$.
E2 $|\boldsymbol{Y}| = \mathbf{\Delta}'''$.
E3 $\mathbf{\Delta}''' < \xi^{-1/4}$.
E4 for all $y \in \boldsymbol{Y}$ we have $\mu_{\mathbf{\Phi}'}(\boldsymbol{\sigma}_y = 1) \vee \mu_{\mathbf{\Phi}'}(\boldsymbol{\sigma}_y = -1) < 1 - 2\eta$.
E5 $\sum_{\sigma \in \{\pm 1\}^V} \left| \mu_{\mathbf{\Phi}}(\forall y \in \boldsymbol{Y} : \boldsymbol{\sigma}_y = \sigma_y) - \prod_{y \in \boldsymbol{Y}} \mu_{\mathbf{\Phi}}(\boldsymbol{\sigma}_y = \sigma_y) \right| < \gamma_n$.

As in the proof of Lemma 6.13 we obtain $\mathbb{P}[\mathscr{E}] > 1 - 4\xi$. Hence, Lemmas 6.14 and 6.15 and the Cauchy-Schwarz inequality yield

$$\left| \mathbb{E}\left[(1 - \mathbf{1}_{\mathscr{E}}) \log \frac{Z(\mathbf{\Phi}''')}{Z(\mathbf{\Phi}')} \right] \right| \leq \delta/3 + o(1), \tag{6.35}$$

$$\left| \mathbb{E}\left[(1 - \mathbf{1}_{\mathscr{E}}) \Lambda_\varepsilon \left(\sum_{s \in \{\pm 1\}} \prod_{i=1}^{\mathbf{\Delta}'''} \left(1 - \mathbf{1}\left\{ s \neq sign(x_{n+1}, b_i) \right\} \mu_{\mathbf{\Phi}'}(\boldsymbol{\sigma}_{\mathbf{y}_i} = -sign(\mathbf{y}_i, b_i)) \right) \right) | Z(\mathbf{\Phi}') > 0 \right] \right|$$
$$\leq \delta/3 + o(1). \tag{6.36}$$

Moreover, because the distribution of $\mathbf{\Phi}'$ is invariant under variable permutations,

$$\mathbb{E}\left[\Lambda_\varepsilon \left(\sum_{s \in \{\pm 1\}} \prod_{i=1}^{\mathbf{\Delta}'''} \left(1 - \mathbf{1}\left\{ s \neq sign(x_{n+1}, b_i) \right\} \mu_{\mathbf{\Phi}'}(\boldsymbol{\sigma}_{\mathbf{y}_i} = -sign(\mathbf{y}_i, b_i)) \right) \right) | Z(\mathbf{\Phi}') > 0 \right]$$
$$= \mathbb{E}\left[\Lambda_\varepsilon \left(\sum_{s \in \{\pm 1\}} \prod_{i=1}^{d} \left(1 - \mathbf{1}\left\{ s \neq \boldsymbol{s}_i \right\} \mu_{\mathbf{\Phi}'}(\boldsymbol{\sigma}_{x_i} = \boldsymbol{s}'_i) \right) \right) | Z(\mathbf{\Phi}') > 0 \right] + o(1). \tag{6.37}$$

Furthermore, on $\mathscr{E}$ we obtain

$$\begin{aligned} \frac{Z(\mathbf{\Phi}''')}{Z(\mathbf{\Phi}')} &= \sum_{\sigma \in \{\pm 1\}^{V \cup \{x_{n+1}\}}} \mathbf{1}\left\{ \sigma \text{ satisfies } b_1, \ldots, b_{\mathbf{\Delta}'''} \right\} \mu_{\mathbf{\Phi}'}\left(\forall y \in \boldsymbol{Y} : \boldsymbol{\sigma}_y = \sigma_y \right) \\ &= \sum_{\sigma \in \{\pm 1\}^{V \cup \{x_{n+1}\}}} \mathbf{1}\left\{ \sigma \text{ satisfies } b_1, \ldots, b_{\mathbf{\Delta}'''} \right\} \prod_{y \in \boldsymbol{Y}} \mu_{\mathbf{\Phi}'}\left(\boldsymbol{\sigma}_y = \sigma_y \right) + o(1) \qquad \text{[due to } \mathbf{E3}, \mathbf{E5}\text{]} \\ &= \sum_{s \in \{\pm 1\}} \prod_{\substack{i \in [\mathbf{\Delta}'''] \\ sign(x_{n+1}, b_i) = -s}} \mu_{\mathbf{\Phi}'}(\boldsymbol{\sigma}_{\mathbf{y}_i} = sign(\mathbf{y}_i, b_i)); \end{aligned} \tag{6.38}$$

to elaborate, in the last step s represents the value assigned to x_{n+1} and the product ensures that the clauses b_i in which x_{n+1} occurs with sign $-s$ are satisfied by assigning their second variable $\mathbf{y}_i$ the value $sign(\mathbf{y}_i, b_i)$. Furthermore, (6.38), **E3** and **E4** yield

$$\mathbb{E}\left[\mathbf{1}_{\mathscr{E}} \log \frac{Z(\mathbf{\Phi}''')}{Z(\mathbf{\Phi}')} \right]$$

$$= \mathbb{E}\left[\mathbf{1}\mathscr{E}\log\left(\sum_{s\in\{\pm1\}}\prod_{i=1}^{\Delta'''}\left(1-\mathbf{1}\left\{sign(x_{n+1},b_i)=-s\right\}\mu_{\Phi'}(\sigma_{y_i}=-sign(y_i,b_i))\right)\right)\right]+o(1)$$

$$= \mathbb{E}\left[\mathbf{1}\mathscr{E}\Lambda_\varepsilon\left(\sum_{s\in\{\pm1\}}\prod_{i=1}^{\Delta'''}\left(1-\mathbf{1}\left\{sign(x_{n+1},b_i)=-s\right\}\mu_{\Phi'}(\sigma_{y_i}=-sign(y_i,b_i))\right)\right)\right]+o(1) \quad (6.39)$$

Finally, the assertion follows from (6.35), (6.36), (6.37) and (6.39). ∎

Proof of Proposition 6.3. Because $\boldsymbol{\mu}_{\pi_d,1} \stackrel{\mathrm{d}}{=} 1-\boldsymbol{\mu}_{\pi_d,1}$ by Proposition 2.1, Corollary 1.3 shows that for every $\varepsilon>0$,

$$\lim_{n\to\infty}\mathbb{E}\left[\Lambda_\varepsilon\left(\sum_{s\in\{\pm1\}}\prod_{i=1}^{d}\left(1-\mathbf{1}\left\{s\neq \boldsymbol{s}_i\right\}\mu_{\Phi'}(\sigma_{x_i}=\boldsymbol{s}_i')\right)\,|Z(\boldsymbol{\Phi}')>0\right)\right]$$
$$=\mathbb{E}\left[\Lambda_\varepsilon\left(\sum_{s\in\{\pm1\}}\prod_{i=1}^{d}\left(1-\mathbf{1}\left\{s\neq \boldsymbol{s}_i\right\}\boldsymbol{\mu}_{\pi_d,i}\right)\right)\right]. \quad (6.40)$$

Furthermore, the dominated convergence theorem and (2.2) yield

$$\lim_{\varepsilon\to0}\mathbb{E}\left[\Lambda_\varepsilon\left(\sum_{s\in\{\pm1\}}\prod_{i=1}^{d}\left(1-\mathbf{1}\left\{s\neq \boldsymbol{s}_i\right\}\boldsymbol{\mu}_{\pi_d,i}\right)\right)\right]$$
$$=\mathbb{E}\log\left(\sum_{s\in\{\pm1\}}\prod_{i=1}^{d}\left(1-\mathbf{1}\left\{s\neq \boldsymbol{s}_i\right\}\boldsymbol{\mu}_{\pi_d,i}\right)\right). \quad (6.41)$$

To complete the proof we combine (6.40), (6.41) and Lemma 6.13. ∎

7 | PROOF OF PROPOSITION 2.6

Tools such as Azuma's inequality do not apply to the number $Z(\boldsymbol{\Phi})$ of satisfying assignments because adding or removing even a single clause could change $Z(\boldsymbol{\Phi})$ by an exponential factor. Therefore, we prove Proposition 2.6 by way of a "soft" version of the random 2-SAT problem. Specifically, for a real $\beta>0$ we define $Z_\beta(\boldsymbol{\Phi})$ via (3.1). Thus, instead of dismissing assignments $\sigma\notin S(\boldsymbol{\Phi})$ outright, we charge an $\exp(-\beta)$ penalty factor for each violated clause. Because the constraints are soft, showing that $\log Z_\beta(\boldsymbol{\Phi})$ concentrates is a cinch.

Lemma 7.1. *For all $t,\beta>0$ we have* $\mathbb{P}\left[\left|\log Z_\beta(\boldsymbol{\Phi})-\mathbb{E}[\log Z_\beta(\boldsymbol{\Phi})]\right|>t|\boldsymbol{m}\right]\leq 2\exp\left(-\frac{t^2}{2\boldsymbol{m}\beta^2}\right)$.

Proof. Since adding or removing a single clause can alter $Z_\beta(\boldsymbol{\Phi})$ by at most a factor $\exp(\pm\beta)$, the assertion follows from Azuma's inequality. ∎

The following statement, whose proof relies on the interpolation method from mathematical physics, will enable us to link the random variables $\log Z_\beta(\boldsymbol{\Phi})$ and $\log Z(\boldsymbol{\Phi})$. For a probability measure $p\in\mathscr{P}(0,1)$ and $\beta>0$ let

$$\mathfrak{B}_\beta(p)=\mathbb{E}\left[\log\sum_{s=\pm1}\prod_{i=1}^{d}\left(1-\mathbf{1}\{\boldsymbol{s}_i\neq s\}\frac{1-\exp(-\beta)}{2}\left(1-\boldsymbol{s}_i'+2\boldsymbol{s}_i'\boldsymbol{\mu}_{p,i}\right)\right)\right]$$

$$-\frac{d}{2}\mathbb{E}\left[\log\left(1-\frac{1-\exp(-\beta)}{4}\left(1-\boldsymbol{s}_1+2\boldsymbol{s}_1\boldsymbol{\mu}_{p,1}\right)\left(1-\boldsymbol{s}_2+2\boldsymbol{s}_2\boldsymbol{\mu}_{p,2}\right)\right)\right]. \tag{7.1}$$

These two expectations exist and are finite because $0 \le \beta < \infty$. (More precisely, their absolute values are bounded by $\log 2 + \beta d$ and β, respectively.)

Lemma 7.2 ([43, Theorem 1])**.** *For any $p \in \mathscr{P}(0,1)$, any $0 \le \beta < \infty$ and every $n \ge 1$ we have $\mathbb{E}\log Z_\beta(\boldsymbol{\Phi}) \le n\mathfrak{B}_\beta(p)$.*

Combining Lemmas 7.1 and 7.2, we obtain the following bound for "hard" 2-SAT.

Corollary 7.3. *For any $\beta > 0$ we have $\lim_{n\to\infty}\mathbb{P}\left[\log Z(\boldsymbol{\Phi}) > n\mathfrak{B}_\beta(\pi_d) + n^{2/3}\right] = 0$.*

Proof. We have $Z_\beta(\boldsymbol{\Phi}) \ge Z(\boldsymbol{\Phi})$ and Lemmas 7.1 and 7.2 imply $\lim_{n\to\infty}\mathbb{P}\left[\log Z_\beta(\boldsymbol{\Phi}) > n\mathfrak{B}_\beta(\pi_d) + n^{2/3}\right] = 0$. ∎

Proof of Proposition 2.6. We begin by observing that the limit $\lim_{\beta\to\infty}\mathfrak{B}_\beta(\pi_d)$ exists and is finite. First, there is the pointwise and monotone convergence of the integrands:

$$\log\sum_{s=\pm1}\prod_{i=1}^{d}\left(1-\mathbf{1}\{\boldsymbol{s}_i \ne s\}\frac{1-\exp(-\beta)}{2}\left(1-\boldsymbol{s}_i'+2\boldsymbol{s}_i'\boldsymbol{\mu}_{\pi_d,i}\right)\right)$$
$$\xrightarrow{\beta\to\infty}\log\sum_{s=\pm1}\prod_{i=1}^{d}\left(1-\frac{\mathbf{1}\{\boldsymbol{s}_i \ne s\}}{2}\left(1-\boldsymbol{s}_i'+2\boldsymbol{s}_i'\boldsymbol{\mu}_{\pi_d,i}\right)\right), \tag{7.2}$$

$$\log\left(1-\frac{1-\exp(-\beta)}{4}\left(1-\boldsymbol{s}_1+2\boldsymbol{s}_1\boldsymbol{\mu}_{\pi_d,1}\right)\left(1-\boldsymbol{s}_2+2\boldsymbol{s}_2\boldsymbol{\mu}_{\pi_d,2}\right)\right)$$
$$\xrightarrow{\beta\to\infty}\log\left(1-\frac{1}{4}\left(1-\boldsymbol{s}_1+2\boldsymbol{s}_1\boldsymbol{\mu}_{\pi_d,1}\right)\left(1-\boldsymbol{s}_2+2\boldsymbol{s}_2\boldsymbol{\mu}_{\pi_d,2}\right)\right). \tag{7.3}$$

Furthermore, since $\boldsymbol{\mu}_{\pi_d} \overset{\mathrm{d}}{=} 1-\boldsymbol{\mu}_{\pi_d}$ by Proposition 2.1 and because $1-\boldsymbol{s}+2\boldsymbol{s}\boldsymbol{\mu}_{\pi_d}$ equals either $2\boldsymbol{\mu}_{\pi_d}$ or $2(1-\boldsymbol{\mu}_{\pi_d})$, we obtain

$$\log\sum_{s=\pm1}\prod_{i=1}^{d}\left(1-\frac{\mathbf{1}\{\boldsymbol{s}_i \ne s\}}{2}\left(1-\boldsymbol{s}_i'+2\boldsymbol{s}_i'\boldsymbol{\mu}_{\pi_d,i}\right)\right) \overset{d}{=} \log\left(\prod_{i=1}^{d^-}\boldsymbol{\mu}_{\pi_d,i}+\prod_{i=1}^{d^+}\boldsymbol{\mu}_{\pi_d,i+d^-}\right), \tag{7.4}$$

$$\frac{d}{2}\log\left(1-\frac{1}{4}\left(1-\boldsymbol{s}_1+2\boldsymbol{s}_1\boldsymbol{\mu}_{\pi_d,1}\right)\left(1-\boldsymbol{s}_2+2\boldsymbol{s}_2\boldsymbol{\mu}_{\pi_d,2}\right)\right) \overset{d}{=} \frac{d}{2}\log\left(1-\boldsymbol{\mu}_{\pi_d,1}\boldsymbol{\mu}_{\pi_d,2}\right). \tag{7.5}$$

Moreover, Proposition 2.1 shows that the monotone limits are integrable and therefore an application of the monotone convergence theorem to (7.2) and (7.3), followed by the simplifications (7.4), (7.4), yields the identity

$$\lim_{\beta\to\infty}\mathfrak{B}_\beta(\pi_d) = \mathbb{E}\left[\log\left(\prod_{i=1}^{d^-}\boldsymbol{\mu}_{\pi_d,i}+\prod_{i=1}^{d^+}\boldsymbol{\mu}_{\pi_d,i+d^-}\right)-\frac{d}{2}\log\left(1-\boldsymbol{\mu}_{\pi_d,1}\boldsymbol{\mu}_{\pi_d,2}\right)\right] = \mathfrak{B}_\infty(\pi_d) < \infty.$$

Furthermore, Corollary 2.5 shows that $\mathfrak{B}_\infty(\pi_d) = \lim_{n\to\infty} n^{-1}\mathbb{E}[\log(Z(\mathbf{\Phi}) \vee 1)]$. Therefore, Corollary 7.3 implies that

$$\mathbb{P}\left[n^{-1}\log(Z(\mathbf{\Phi}) \vee 1) > \mathfrak{B}_\infty(\pi_d) + \varepsilon\right] = o(1) \qquad \text{for any } \varepsilon > 0. \tag{7.6}$$

To complete the proof, we upper bound

$$\begin{aligned} &n^{-1}\mathbb{E}\left|\log(Z(\mathbf{\Phi}) \vee 1) - \mathbb{E}[\log(Z(\mathbf{\Phi}) \vee 1)]\right| \\ &\quad \le \mathbb{E}\left|n^{-1}\log(Z(\mathbf{\Phi}) \vee 1) - \mathfrak{B}_\infty(\pi_d)\right| + \left|\mathfrak{B}_\infty(\pi_d) - \mathbb{E}[\log(Z(\mathbf{\Phi}) \vee 1)]\right|. \end{aligned} \tag{7.7}$$

Due to Corollary 2.5, the second term on the r.h.s. of (7.7) tends to zero. On the other hand, (7.6) and Corollary 2.5 yield that for any $\varepsilon > 0$,

$$\mathbb{E}\left|n^{-1}\log(Z(\mathbf{\Phi}) \vee 1) - \mathfrak{B}_\infty(\pi_d)\right| \le \mathbb{E}\left[\mathfrak{B}_\infty(\pi_d) - n^{-1}\log(Z(\mathbf{\Phi}) \vee 1)\right] + 2\varepsilon + o(1) = 2\varepsilon + o(1),$$

as desired. ∎

ACKNOWLEDGMENT

We thank Andreas Galanis and Leslie Goldberg for helpful discussions.

REFERENCES

1. E. Abbe and A. Montanari, *On the concentration of the number of solutions of random satisfiability formulas*, Random Structures Algorithms **45** (2014), 362–382.
2. D. Achlioptas, A. Chtcherba, G. Istrate, and C. Moore, *The phase transition in 1-in-*k *SAT and NAE 3-SAT*, Proc. 12th SODA (2001), 721–722.
3. D. Achlioptas and C. Moore, *Random* k*-SAT: two moments suffice to cross a sharp threshold*, SIAM J. Comput. **36** (2006), 740–762.
4. D. Achlioptas and Y. Peres, *The threshold for random* k*-SAT is* $2^k ln2 - O(k)$, J. AMS **17** (2004), 947–973.
5. B. Aspvall, M. F. Plass, and R. E. Tarjan, *A linear-time algorithm for testing the truth of certain quantified Boolean formulas*, Inform. Process. Lett. **8** (1979), 121–123. ISSN: 0020-0190. https://doi.org/10.1016/0020-0190(79)90002-4.
6. M. Aizenman, R. Sims, and S. Starr, *An extended variational principle for the SK spin-glass model*, Phys. Rev. B **68** (2003), 214403.
7. N. Alon and J. Spencer, *The probabilistic method*, Wiley, Hoboken, NJ, 2016.
8. P. Ayre, A. Coja-Oghlan, P. Gao, and N. Müller, *The satisfiability threshold for random linear equations*, Combinatorica **40** (2020), 179–235.
9. A. Barvinok, *Combinatorics and complexity of partition functions*, Springer, Cham, Switzerland, 2016.
10. V. Bogachev and A. Kolesnikov, *The Monge–Kantorovich problem: Achievements, connections, and perspectives*, Russian Math. Surv. **67** (2012), 785–890.
11. B. Bollobás, *The evolution of random graphs*, Trans. AMS **286** (1984), 257–274.
12. B. Bollobás, C. Borgs, J. Chayes, J. Kim, and D. Wilson, *The scaling window of the 2-SAT transition*, Random Structures Algorithms **18** (2001), 201–256.
13. Y. Boufkhad and O. Dubois, *Length of prime implicants and number of solutions of random CNF formulae*, Theor. Comput. Sci. **215** (1999), 1–30.
14. P. Cheeseman, B. Kanefsky, and W. Taylor, *Where the* really *hard problems are*, Proc. IJCAI **1** (1991), 331–337.
15. V. Chvátal and B. Reed, *Mick gets some (the odds are on his side)*, Proc. 33th FOCS (1992), 620–627.

16. A. Coja-Oghlan, T. Kapetanopoulos, and N. Müller, *The replica symmetric phase of random constraint satisfaction problems*, Combin. Probab. Comput. **29** (2020), 346–422.
17. A. Coja-Oghlan, F. Krzakala, W. Perkins, and L. Zdeborová, *Information-theoretic thresholds from the cavity method*, Adv. Math. **333** (2018), 694–795.
18. A. Coja-Oghlan and K. Panagiotou, *The asymptotic* k*-SAT threshold*, Adv. Math. **288** (2016), 985–1068.
19. A. Coja-Oghlan and N. Wormald, *The number of satisfying assignments of random regular* k*-SAT formulas*, Combin. Probab. Comput. **27** (2018), 496–530.
20. C. Cooper, A. Frieze, and G. Sorkin, *Random 2-SAT with prescribed literal degrees*, Algorithmica **48** (2007), 249–265.
21. D. Coppersmith, D. Gamarnik, M. Hajiaghayi, and G. Sorkin, *Random MAX SAT, random MAX CUT, and their phase transitions*, Random Structures Algorithms **24** (2004), 502–545.
22. A. Dembo and A. Montanari, *Ising models on locally tree-like graphs*, Ann. Appl. Probab. **20** (2010), 565–592.
23. A. Dembo, A. Montanari, and N. Sun, *Factor models on locally tree-like graphs*, Ann. Probab. **41** (2013), 4162–4213.
24. M. Dietzfelbinger, A. Goerdt, M. Mitzenmacher, A. Montanari, R. Pagh, M. Rink: *Tight thresholds for cuckoo hashing via XORSAT*. arXiv:0912.0287 (2009).
25. J. Ding, A. Sly, and N. Sun, *Satisfiability threshold for random regular NAE-SAT*, Commun. Math. Phys. **341** (2016), 435–489.
26. J. Ding, A. Sly, and N. Sun, *Proof of the satisfiability conjecture for large* k, Proc. 47th STOC (2015), 59–68.
27. W. Dowling and J. Gallier, *Linear-time algorithms for testing the satisfiability of propositional Horn formulae*, J. Logic Program. **1** (1984), 267–284.
28. O. Dubois and J. Mandler, *The 3-XORSAT threshold*, Proc. 43rd FOCS (2002), 769–778.
29. W. Fernandez de la Vega, *Random 2-SAT: results and problems*, Theor. Comput. Sci. **265** (2001), 131–146.
30. S. Franz and M. Leone, *Replica bounds for optimization problems and diluted spin systems*, J. Stat. Phys. **111** (2003), 535–564.
31. A. Goerdt, *A threshold for unsatisfiability*, J. Comput. Syst. Sci. **53** (1996), 469–486.
32. F. Guerra, *Broken replica symmetry bounds in the mean field spin glass model*, Commun. Math. Phys. **233** (2003), 1–12.
33. F. Krzakala, 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. U.S.A. **104** (2007), 10318–10323.
34. T. Łuczak, *Component behavior near the critical point of the random graph process*, Random Structures Algorithms **1** (1990), 287–310.
35. M. Mézard and A. Montanari, *Information, physics and computation*, Oxford University Press, Oxford, 2009.
36. M. Mézard, G. Parisi, and R. Zecchina, *Analytic and algorithmic solution of random satisfiability problems*, Science **297** (2002), 812–815.
37. R. Monasson and R. Zecchina, *The entropy of the* k*-satisfiability problem*, Phys. Rev. Lett. **76** (1996), 3881.
38. R. Monasson and R. Zecchina, *Statistical mechanics of the random* K*-SAT model*, Phys. Rev. E **56** (1997), 1357–1370.
39. A. Montanari, F. Ricci-Tersenghi, and G. Semerjian, *Solving constraint satisfaction problems through Belief Propagation-guided decimation*, Proc. 45th Allerton (2007), 352–359.
40. A. Montanari and D. Shah, *Counting good truth assignments of random* k*-SAT formulae*, Proc. 18th SODA (2007), 1255–1264.
41. D. Panchenko, *Spin glass models from the point of view of spin distributions*, Ann. Probab. **41** (2013), 1315–1361.
42. D. Panchenko, *On the replica symmetric solution of the* K*-sat model*, Electron. J. Probab. **19** (2014), 67.
43. D. Panchenko and M. Talagrand, *Bounds for diluted mean-fields spin glass models*, Probab. Theory Relat. Fields **130** (2004), 319–336.
44. B. Pittel and G. Sorkin, *The satisfiability threshold for* k*-XORSAT*, Combin. Probab. Comput. **25** (2016), 236–268.
45. F. Rassmann, *On the number of solutions in random graph* k*-colouring*, Combin. Probab. Comput. **28** (2019), 130–158.
46. A. Scott and G. Sorkin, *Solving sparse random instances of Max Cut and Max 2-CSP in linear expected time*, Combin. Probab. Comput. **15** (2006), 281–315.
47. A. Sharell: *Concentration of the number of solutions to a random 2-CNF formula*. Manuscript, 2000.
48. A. Sly, N. Sun, and Y. Zhang, *The number of solutions for random regular NAE-SAT*, Proc. 57th FOCS (2016), 724–731.

49. M. Talagrand, *The high temperature case for the random* K-*sat problem*, Probab. Theory Relat. Fields **119** (2001), 187–212.
50. L. Valiant, *The complexity of enumeration and reliability problems*, SIAM J. Comput. **8** (1979), 410–421.

How to cite this article: Achlioptas D, Coja-Oghlan A, Hahn-Klimroth M, et al. The number of satisfying assignments of random 2-SAT formulas. *Random Struct Alg* 2021;58:609–647. https://doi.org/10.1002/rsa.20993