The Ising model on trees: Boundary conditions and mixing time[†]

Fabio Martinelli[‡] Alistair Sinclair[§] Dror Weitz[¶]

July 22, 2003

Abstract

We give the first comprehensive analysis of the effect of boundary conditions on the mixing time of the Glauber dynamics for the Ising model. Specifically, we show that the mixing time on an n-vertex regular tree with $(+)$-boundary remains $O(n \log n)$ at all temperatures (in contrast to the free boundary case, where the mixing time is not bounded by any fixed polynomial at low temperatures). We also show that this bound continues to hold in the presence of an arbitrary external field. Our results are actually stronger, and provide tight bounds on the log-Sobolev constant and the spectral gap of the dynamics. In addition, our methods yield simpler proofs and stronger results for the mixing time in the regime where it is insensitive to the boundary condition. Our techniques also apply to a much wider class of models, including those with hard constraints like the anti-ferromagnetic Potts model at zero temperature (colorings) and the hard-core model (independent sets).

[†]An extended abstract of this paper appears in *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, October 2003.

[‡]Department of Mathematics, University of Roma Tre, Largo San Murialdo 1, 00146 Roma, Italy. Email: martin@mat.uniroma3.it. This work was done while this author was visiting the Departments of EECS and Statistics, University of California, Berkeley, supported in part by a Miller Visiting Professorship.

[§]Computer Science Division, University of California, Berkeley, CA 94720-1776, U.S.A. Email: sinclair@cs.berkeley.edu. Supported in part by NSF Grant CCR-0121555 and DARPA cooperative agreement F30602-00-2-0601.

[¶]Computer Science Division, University of California, Berkeley, CA 94720-1776, U.S.A. Email: dror@cs.berkeley.edu. Supported in part by NSF Grant CCR-0121555.

1 Introduction

1.1 Background

Local Markov chains (or "Glauber dynamics") for spin systems on finite graphs have been studied intensively in recent years, and much is known about their mixing time. An important issue left open by these investigations is the effect on the mixing time of the environment in which the system is placed, i.e., when the values of certain boundary spins are fixed. In this paper we investigate this question. We focus for simplicity on the classical Ising model, though our techniques apply to more general spin systems including the antiferromagnetic Potts model (colorings) and the hard-core model (independent sets).

In the Ising model on a finite graph $G = (V, E)$, a configuration consists of an assignment of ± 1-values, or "spins", to each vertex (or "site") of V. We often refer to the spin values ± 1 as $(+)$ and $(-)$. The probability of finding the system in configuration $\sigma \in \{\pm 1\}^V \equiv \Omega_G$ is given by the *Gibbs distribution*

$$\mu_G(\sigma) \propto \exp\Big(\beta \sum\nolimits_{xy \in E} \sigma_x \sigma_y\Big), \tag{1}$$

where $\beta \geq 0$ is the inverse temperature. Thus μ_G assigns higher probability to configurations in which many neighboring spins are aligned. This effect increases with β, so that at high temperatures (low β) the spins behave almost independently, while at low temperatures (high β) there is global order. Frequently one imposes a *boundary condition* on the model, which corresponds to fixing the spin values at some specified "boundary" vertices of G; the term *free boundary* is used to indicate that no boundary condition is specified.

In the classical Ising model, $G = G_n$ is a cube of side $n^{1/d}$ in the d-dimensional Cartesian lattice $\mathbb{Z}^d$, and one studies the properties of the Gibbs distribution as $n \to \infty$ with a specified boundary condition (e.g., the all-$(+)$ or the all-$(-)$ configuration) on the faces of the cube; this limit is referred to as the "(infinite volume) Gibbs measure" for the given boundary condition. It is well known that a *phase transition* occurs at a certain critical inverse temperature $\beta = \beta_c$ (which depends on the dimension d): for $\beta < \beta_c$ (the "high temperature" region) there are no long-range correlations between spins and consequently there is a *unique* Gibbs measure independent of the boundary condition, while for $\beta > \beta_c$ (the "low temperature" region) correlations are present at arbitrary distances and there are (at least) two distinct Gibbs measures (or "phases"), corresponding to the $(+)$ and $(-)$-boundary conditions respectively. See, e.g., [15, 40] for more background.

While the classical theory focused on static properties of the Gibbs measure, in modern statistical physics the emphasis has shifted towards dynamical questions with a computational flavor. The key object here is the *Glauber dynamics*, a Markov chain on the set of spin configurations Ω_G of a finite graph G. For definiteness, we describe the "heat-bath" version of Glauber dynamics: at each step, pick a vertex x of G u.a.r., and replace the spin at x by a random spin drawn from the distribution of σ_x *conditional on* all the neighboring spins. It is easy to check that the Glauber dynamics is an ergodic, reversible Markov chain on Ω_G whose stationary distribution is exactly μ_G. The Glauber dynamics is much studied for two reasons: firstly, it is the basis of Markov chain Monte Carlo algorithms, widely used in computational physics for sampling from the Gibbs distribution; and secondly, it is a plausible model for the actual evolution of the underlying physical system towards equilibrium. In both contexts, the central question is to determine the *mixing time*, i.e., the number of steps until the dynamics is close to its stationary distribution.

Advances in statistical physics over the past decade have led to the following remarkable characterization of the mixing time on finite n-vertex cubes with free boundary in the 2-dimensional lattice $\mathbb{Z}^2$ [42, 30, 29, 28, 10]: when $\beta < \beta_c$ the mixing time is $O(n \log n)$, while for $\beta > \beta_c$ it is $\exp(\Omega(\sqrt{n}))$. Thus the phase transition (a static, spatial phenomenon) has a dramatic computational manifestation in the form of an explosion from optimal to exponential in the running time of a natural algorithm. This result stands as perhaps the most convincing example to date of an intimate connection between phase transitions and computational complexity.

One of the most interesting questions left open by the above result is the *influence of the boundary condition* on the mixing time. It has been conjectured that, in the presence of an all-$(+)$ boundary, the mixing time in $\mathbb{Z}^d$ should remain polynomial in n *at all temperatures* [8, 13]. This captures the intuition that the only obstacle to rapid mixing for $\beta > \beta_c$ is the long time required for the dynamics to get through the "bottleneck" between the $(+)$-phase and the $(-)$-phase; the presence of the $(+)$-boundary eliminates the $(-)$-phase and hence the bottleneck. Formalizing this intuition, however, has proved very elusive.

In this paper we prove a strong version of the above conjecture in what is known in statistical physics as the *Bethe approximation*, namely when the lattice $\mathbb{Z}^d$ is replaced by a regular tree. Specifically, we analyze the mixing time of the Glauber dynamics for the Ising model on a tree with $(+)$-boundary condition on its leaves, and show that it remains $O(n \log n)$ at all temperatures. (With a free boundary, the mixing time on a tree is polynomial at all temperatures, but the exponent grows arbitrarily large at low temperatures as $\beta \to \infty$.) This is apparently the first result that quantifies the effect of boundary conditions on the dynamics in an interesting scenario. We stress that, while the tree is simpler in some respects than $\mathbb{Z}^d$ due to the lack of cycles, in other respects it is more complex: e.g., it exhibits a "double phase transition" (see below). Moreover, the Ising model on trees has recently received a lot of attention as the canonical example of a statistical physics model on a "non-amenable" graph (i.e., one whose boundary is of comparable size to its volume) — see, e.g., [4, 6, 7, 12, 19, 24, 27, 39]. In the next subsection, we briefly describe the Ising model on trees before stating our results in more detail.

1.2 The Ising model on trees

Fix $b \geq 2$ and let $\mathbb{T}^b$ denote the infinite b-ary tree. The Ising model on $\mathbb{T}^b$ is known to have *two* critical inverse temperatures, β_0 and β_1. The first of these, $\beta_0 = \frac{1}{2}\ln(\frac{b+1}{b-1})$, marks the boundary between uniqueness and non-uniqueness of the Gibbs measure: i.e., the "high temperature" region, in which the Gibbs measure is unique, is defined by $\beta \leq \beta_0$ [37]. However, in contrast to the model on $\mathbb{Z}^d$, there is now a second critical point $\beta_1 = \frac{1}{2}\ln(\frac{\sqrt{b}+1}{\sqrt{b}-1})$ [7, 19], which delimits the region where "typical" boundary conditions exert long-range influence on the root. I.e., there is now an "intermediate" region $\beta_0 < \beta \leq \beta_1$ in which the $(+)$- and $(-)$-boundaries exert long-range influence but typical boundaries do not, while in the "low temperature" region $\beta > \beta_1$ long-range influence occurs even for typical boundaries. β_1 has alternative interpretations as the critical value for extremality of the Gibbs measure and the threshold for noisy data transmission on the tree [12].

The Glauber dynamics for the Ising model on trees has also been studied. In a recent paper [4], it is shown that the mixing time with free boundary on a complete b-ary tree with n vertices is $O(n \log n)$ at high and intermediate temperatures (i.e., when $\beta < \beta_1$)†. Moreover, as soon as $\beta > \beta_1$ the mixing time becomes $n^{1+\Omega(1)}$, and the exponent is unbounded as $\beta \to \infty$. Thus the critical point $\beta = \beta_1$ is reflected in a jump in the mixing time from optimal to super-linear.

When one considers the effect of boundary conditions, trees differ greatly from $\mathbb{Z}^d$ because their boundary is very large (of size $O(n)$ rather than $O(n^{1/d})$ as in $\mathbb{Z}^d$). To compensate for this, one introduces an *external field* h that adds to all (non-boundary) spins a bias in the direction of the field. The Gibbs distribution then becomes

$$\mu_G(\sigma) \propto \exp\Big[\beta\big(\sum_{xy \in E} \sigma_x\sigma_y + h\sum_{x \in V} \sigma_x\big)\Big]. \tag{2}$$

Now it is well known [15] that, for all $\beta > \beta_0$, there is a critical value $h = h_c(\beta) > 0$ of the field such that the Gibbs measure is not unique when $|h| \leq h_c$, and is unique when $|h| > h_c$ (see Fig. 1). (When $\beta \leq \beta_0$ the Gibbs measure is unique for all h, and h_c is defined to be zero.) Thus in the

†Actually [4] proves this only for sufficiently high temperatures, but the argument can be extended to all $\beta < \beta_1$ [36].

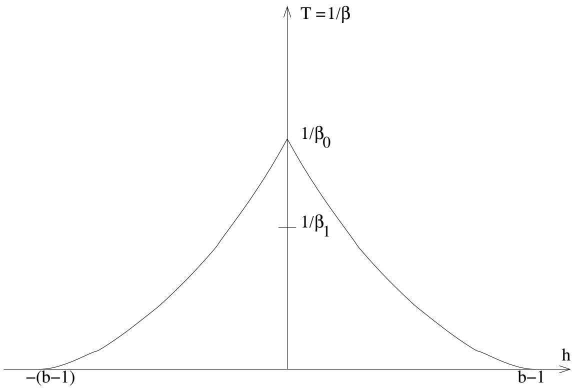

Figure 1: Curve of critical field $h_c(\beta)$. The Gibbs measure is unique above the curve.

presence of a $(+)$-boundary, the tree with an external field of value $h = -h_c$ is the analog of the classical case of $\mathbb{Z}^d$ with zero field. In our results, we analyze the Glauber dynamics over the full range of values of both β and h. The fact that we are able to handle external fields (including the critical value $|h| = h_c$) brings our results for trees rather close to the original conjecture for $\mathbb{Z}^d$.

1.3 Main results and techniques

Translating the conjecture mentioned earlier to the tree setting, we would wish to prove that, in the presence of a $(+)$-boundary, the mixing time on the tree remains bounded by a fixed polynomial at all temperatures, and all values of the external field. This is the content of our first main result; in fact, we are able to prove that the mixing time is $O(n \log n)$, which is optimal:

Theorem A *For any fixed b, the Glauber dynamics on the n-vertex b-ary tree with $(+)$-boundary condition has mixing time $O(n \log n)$ at all inverse temperatures $\beta < \infty$ and all external fields h.*

In our second main result, we give an improved and more general analysis of the mixing time in cases where it is *insensitive* to the boundary condition, i.e., in the high and intermediate temperature region at all fields, and at all temperatures when there is a large external field:

Theorem B *For any fixed b, the Glauber dynamics on the n-vertex b-ary tree with arbitrary boundary conditions has mixing time $O(n \log n)$ both (i) at all inverse temperatures $\beta < \beta_1$ and all external fields h; and (ii) at all inverse temperatures $\beta < \infty$ and all external fields $|h| > h_c(\beta)$.*

This analysis has several advantages over previous ones [4, 36]: it is more direct, applies also when there is an external field, gives a technically stronger result (as explained below), and applies to models more general than the Ising model.

We now proceed to sketch some of our techniques and point out the main technical innovations. We also explain why our results are in fact quite a bit stronger and more general than Theorems A and B stated above.

In the settings of both theorems, we actually prove the stronger property that the Glauber dynamics has *logarithmic Sobolev constant* bounded below by $\Omega(n^{-1})$. The log-Sobolev constant can be viewed as a measure of the rate of decrease of relative entropy; by standard theory, the above bound on it implies not only a mixing time of $O(n \log n)$, but also a number of other properties such as hypercontractivity and Gaussian concentration bounds for the corresponding Gibbs measure (see, e.g., [38]). No analysis of the log-Sobolev constant was known for any of the situations we study here (except at very high temperatures). We warm up for the log-Sobolev constant by first proving that the conceptually simpler *spectral gap* (i.e., the difference between the second-largest eigenvalue and 1) of the Glauber dynamics is $\Omega(n^{-1})$. The spectral gap measures the rate of decay of variance, and the above bound on it leads to a weaker bound of $O(n^2)$ on the mixing time.[‡]

Our analysis of both the log-Sobolev constant and the spectral gap rests on a certain *spatial mixing condition*: if the influence of the spin at the root of the tree on the spins at its leaves decays fast enough with the depth, then we show how to deduce bounds on the spectral gap and the log-Sobolev constant. Our treatment of the two quantities differs only in that influence is measured in terms of the *variance* and *entropy* respectively of functions of the spins. Crucially, in contrast to previous approaches we do not require this decay to hold in arbitrary environments, but only for the measure under consideration. This opens up for the first time the possibility that the condition holds for some boundary conditions and not for others (with the same values of temperature and external field).

The second main ingredient of the paper is establishing the above spatial mixing condition in the scenarios of interest: namely, with a $(+)$-boundary at all temperatures and fields, and arbitrary boundaries at high and intermediate temperatures or large fields. For this purpose we introduce two quantities, κ and γ, that bound the rate at which a spin disagreement at one site (in two copies of the system) can percolate down and up the tree respectively. It is not too hard to see that if the product $\kappa\gamma$ is small enough then the variance mixing condition holds (and hence the spectral gap is bounded); surprisingly, with a bit more work essentially the same condition on $\kappa\gamma$ can be seen to imply entropy mixing and hence a bound on log-Sobolev.

Finally, we mention that our techniques actually apply (with suitable modifications) to a much wider class of spin systems on trees than just the Ising model, including the Potts model and models with hard constraints such as the zero-temperature antiferromagnetic Potts model (colorings) and the hard core model (independent sets). We outline some of these extensions at the end of the paper; full details can be found in a companion paper [32].

The remainder of the paper is organized as follows. In Section 2 we give some basic definitions and notation. Then in Section 3 we define the spatial mixing condition and relate it to the spectral gap and log-Sobolev constant. We go on to verify the mixing condition in the scenarios of interest for the spectral gap and the log-Sobolev constant in Sections 4 and 5 respectively. Finally, in Section 6 we briefly mention some extensions of our results to other models. The proofs of some technical lemmas omitted from the main text are collected in an Appendix.

2 Preliminaries

2.1 Gibbs distributions on trees

For $b \geq 2$, let $\mathbb{T}^b$ denote the infinite b-ary tree (in which every vertex has b children). We will be concerned with (complete) finite subtrees T of $\mathbb{T}^b$; if T has depth m then it has $n = (b^{m+1}-1)/(b-1)$ vertices, and its *boundary* ∂T consists of the children (in $\mathbb{T}^b$) of its leaves, i.e., $|\partial T| = b^{m+1}$. We identify subgraphs of T with their vertex sets, and write $E(A)$ for the edges within a subset A, and ∂A for the boundary of A (i.e., the neighbors of A in $(T \cup \partial T) \setminus A$).

Fix an Ising spin configuration τ on the infinite tree $\mathbb{T}^b$. We denote by Ω_T^τ the set of (finite) spin configurations $\sigma \in \{\pm 1\}^{T\cup\partial T}$ that agree with τ on ∂T; thus τ specifies a *boundary condition* on T. Usually we abbreviate Ω_T^τ to Ω. For any $\eta \in \Omega$ and any subset $A \subseteq T$, we denote by μ_A^η the Gibbs distribution over Ω conditioned on the configuration outside A being η: i.e., if $\sigma \in \Omega$ agrees with η outside A then

$$\mu_A^\eta(\sigma) \propto \exp\Big[\beta\big(\textstyle\sum_{xy\in E(A\cup\partial A)} \sigma_x\sigma_y + h\sum_{x\in A}\sigma_x\big)\Big],$$

‡By a separate argument relating the spectral gap to the log-Sobolev constant for spin systems on trees, which is of independent interest, we are able to improve this bound to $O(n\log^2 n)$ without analyzing the log-Sobolev constant directly: see Theorem 5.7.

where β is the inverse temperature and h the external field. We define $\mu_A^\eta(\sigma) = 0$ otherwise. In particular, when $A = T$, $\mu_T^\eta = \mu_T^\tau$ is simply the Gibbs distribution on the whole of T with boundary condition τ; we abbreviate μ_T^τ to μ.

For a function $f : \Omega \to \mathbb{R}$ we denote by $\mu_A^\eta(f) = \sum_{\sigma\in\Omega} \mu_A^\eta(\sigma) f(\sigma)$ the *expectation* of f w.r.t. the distribution μ_A^η. It will be convenient to view $\mu_A^\eta(f)$ as a function of η, defined by $\mu_A(f)(\eta) = \mu_A^\eta(f)$, the *conditional expectation* of f. Note that $\mu_A(f)$ is a function from Ω to $\mathbb{R}$ but depends only on the configuration outside A. We write $\mathrm{Var}_A^\eta(f) = \mu_A^\eta(f^2) - \mu_A^\eta(f)^2$ and (for $f \geq 0$) $\mathrm{Ent}_A^\eta(f) = \mu_A^\eta(f \log f) - \mu_A^\eta(f) \log \mu_A^\eta(f)$ for the *variance* and *entropy* of f respectively w.r.t. μ_A^η. Note that $\mathrm{Var}_A^\eta(f) = 0$ iff, conditioned on the configuration outside A being η, f does not depend on the configuration inside A. The same holds for $\mathrm{Ent}_A^\eta(f)$. In case $A = T$ we use the abbreviations $\mu(f), \mathrm{Var}(f)$ and $\mathrm{Ent}(f)$.

We record here some basic properties of variance and entropy that we use throughout the paper:
(i) For $B \subseteq A \subseteq T$,

$$\mathrm{Var}_A^\eta(f) = \mu_A^\eta[\mathrm{Var}_B(f)] + \mathrm{Var}_A^\eta[\mu_B(f)]. \tag{3}$$

This equation expresses a decomposition of the variance into the local conditional variance in B and the variance of the projection outside B.
(ii) If $A = \bigcup_i A_i$ for disjoint A_i, and the Gibbs distribution μ_A^η is the product of its marginals over the A_i, then for any function f,

$$\mathrm{Var}_A^\eta(f) \leq \sum_i \mu_A^\eta[\mathrm{Var}_{A_i}(f)]. \tag{4}$$

(iii) For any two subsets $A, B \subseteq T$ such that $(\partial A) \cap B = \emptyset$, and for any function f,

$$\mu[\mathrm{Var}_A(\mu_B(f))] \leq \mu[\mathrm{Var}_A(\mu_{A\cap B}(f))]. \tag{5}$$

Properties (ii) and (iii) are consequences of the fact that variance w.r.t. a fixed measure is a convex functional. Property (ii) can be found, e.g., in [10]; we give proofs of (i) and (iii) in the Appendix. All three properties (i), (ii) and (iii) also hold with Var replaced by Ent.

2.2 The Glauber dynamics

The (*heat-bath*) *Glauber dynamics* is the following Markov chain on $\Omega = \Omega_T^\tau$. In configuration $\sigma \in \Omega$, make a transition as follows:

(i) pick a vertex $x \in T$ u.a.r.

(ii) replace σ by a new configuration drawn from the distribution $\mu_{\{x\}}^\sigma$.

Note that (ii) corresponds simply to replacing the spin σ_x by a new spin σ'_x drawn from the Gibbs distribution conditional on the spins at the neighbors of x. Thus the possible transitions from σ (in addition to self-loops) are to states σ^x in which the spin at vertex x has been flipped, and the transition probability is $P(\sigma, \sigma^x) = \frac{1}{n} \cdot \frac{1}{1+w_x(\sigma)}$, where $w_x(\sigma) = \exp\left[2\beta\sigma_x(\sum_{xy\in E} \sigma_y + h)\right]$.

It is a well-known fact (and easily checked) that the Glauber dynamics is ergodic and reversible w.r.t. the Gibbs distribution $\mu = \mu_T^\tau$, and so converges to the stationary distribution μ. We measure the rate of convergence by the *mixing time*:

$$t_{\mathrm{mix}} = \min\{t : \|P^t(\sigma, \cdot\,) - \mu\| \leq \tfrac{1}{2e} \text{ for all } \sigma \in \Omega\}, \tag{6}$$

where $P^t(\sigma, \cdot\,)$ denotes the distribution of the dynamics after t steps starting from configuration σ, and $\|\cdot\|$ is variation distance. The constant $\frac{1}{2e}$ in this definition is inessential: it is well known [1] that for any $\epsilon > 0$ we have $\max_\sigma \|P^t(\sigma, \cdot\,) - \mu\| \leq \epsilon$ for all $t \geq t_{\mathrm{mix}}\lceil \ln \epsilon^{-1} \rceil$.

To bound the mixing time we will use two standard tools from functional analysis: the *spectral gap* and the *logarithmic Sobolev constant*. For a function $f : \Omega \to \mathbb{R}$, define the *Dirichlet form* of f w.r.t. μ by

$$\frac{1}{2}\sum_{\sigma,\sigma'} \mu(\sigma)P(\sigma,\sigma')(f(\sigma) - f(\sigma'))^2 = \frac{1}{n}\sum_{x} \mu(\mathrm{Var}_{\{x\}}(f)) \stackrel{\mathrm{def}}{=} \frac{1}{n}\mathcal{D}(f). \tag{7}$$

(The l.h.s. here is the general definition for any Markov chain; the equality holds when specializing to the case of the heat-bath dynamics.) Thus $\mathcal{D}(f)$ is the standard Dirichlet form scaled by a factor of n, and can be thought of as the "local variation" of f. Note that $\mathcal{D}(f)$ depends only on the Gibbs distribution μ.

The (scaled) spectral gap and log-Sobolev constant compare the local variation $\mathcal{D}(f)$ (respectively, $\mathcal{D}(\sqrt{f}\,)$) to the variance and entropy respectively of f:

$$c_{\mathrm{gap}}(\mu) = \inf_{f} \frac{\mathcal{D}(f)}{\mathrm{Var}(f)}; \qquad c_{\mathrm{sob}}(\mu) = \inf_{f \geq 0} \frac{\mathcal{D}(\sqrt{f}\,)}{\mathrm{Ent}(f)}, \tag{8}$$

where the infimum in each case is over non-constant functions f. These two quantities measure the rate of decrease of variance and relative entropy respectively (see, e.g., [38]). The quantity c_{gap} also has a natural interpretation as the eigenvalue gap of the Markov transition matrix P (scaled by a factor of n), and is well known in the Computer Science community. For applications in Computer Science of the less familiar log-Sobolev constant, see, e.g., [14, 21, 22].

Specializing to the Glauber dynamics standard results relating the mixing time of a Markov chain to the spectral gap and log-Sobolev constant (see, e.g., [38]) we get:

Theorem 2.1 *The mixing time of the Glauber dynamics on an n-vertex b-ary tree T with boundary condition τ satisfies*

(i) $t_{\mathrm{mix}} \leq c_{\mathrm{gap}}(\mu)^{-1} \times C_1 n^2$;

(ii) $t_{\mathrm{mix}} \leq c_{\mathrm{sob}}(\mu)^{-1} \times C_2 n \log n$,

where $\mu = \mu_T^\tau$ and C_1, C_2 are constants depending only on b, β and h. $\square$

Actually, in the course of this paper (see Theorem 5.7) we will prove that, for the Ising model (or indeed a general spin system) on the tree, $c_{\mathrm{sob}}(\mu) \geq (C_3 \log n)^{-1} \times c_{\mathrm{gap}}(\mu)$, where C_3 is a constant depending only on b, β, h; this means that we can improve the bound in case (i) of Theorem 2.1 to $t_{\mathrm{mix}} \leq c_{\mathrm{gap}}(\mu)^{-1} \times C_2' n \log^2 n$. This result seems to be interesting in its own right.

When discussing the asymptotics of the mixing time as a function of n, the size of T, we fix a boundary condition τ (on the infinite tree $\mathbb{T}^b$) and consider the infinite sequence of Gibbs distributions $\{\mu_T^\tau\}$, where T ranges over all finite complete subtrees of $\mathbb{T}^b$. In particular, when we say that the mixing time for some boundary condition τ is $O(n \log n)$, we mean that there exists a constant C (depending only on b, β, h) such that for all T the mixing time on T with boundary condition τ is $\leq Cn \log n$. Similarly, we will say that $c_{\mathrm{sob}}(\mu) = c_{\mathrm{sob}}(\mu_T^\tau) = \Omega(1)$ to mean that there exists a finite constant $C > 0$ such that, for every T (or equivalently, for every $\mu \in \{\mu_T^\tau\}$), $c_{\mathrm{sob}}(\mu) \geq 1/C$. By Theorem 2.1, this implies that the mixing time is $O(n \log n)$. Note that the foregoing are properties of the boundary condition τ (as well as of the model parameters b, β, h).

Thus to prove Theorems A and B our goal will be to show that, for the stated combinations of b, β, h and boundary condition τ, $c_{\mathrm{sob}}(\mu) = \Omega(1)$, i.e., that $\mathrm{Ent}(f) \leq \mathrm{const} \times \mathcal{D}(\sqrt{f}\,)$. We will in fact first prove that $c_{\mathrm{gap}} = \Omega(1)$, i.e., that $\mathrm{Var}(f) \leq \mathrm{const} \times \mathcal{D}(f)$, because this is conceptually similar and technically easier, and already proves Theorems A and B with the slightly weaker mixing time bound of $O(n^2)$. We will then describe how to extend the analysis to c_{sob}.

Finally, we note that our choice of the heat-bath dynamics is not essential. Since changing to any other reversible update rule (e.g., the Metropolis rule) affects c_{sob} and c_{gap} by at most a constant factor, our analysis applies to any choice of Glauber dynamics.

3 Spatial mixing conditions for spectral gap and log-Sobolev

In this section we define a certain *spatial mixing condition* (i.e., a form of weak dependence between the spin at a site and the configuration far from that site) for a Gibbs distribution μ, and prove that this condition implies that $c_{\text{gap}}(\mu) = \Omega(1)$. An analogous condition implies that $c_{\text{sob}}(\mu) = \Omega(1)$. Our spatial mixing conditions have two main advantages over those used previously: first, the conditions for the spectral gap and the log-Sobolev constant are identical in form, allowing a uniform treatment; second, and more importantly, they are *measure-specific*, i.e., they may hold for the Gibbs distribution induced by some specific boundary configuration while not holding for other boundary configurations. Hence, the conditions are sensitive enough to show rapid mixing for specific boundaries even though the mixing time with other boundaries is slow for the same choice of temperature and external field. We also note that the results of this section hold not just for the Ising model but for any nearest-neighbor interaction model on a tree.

3.1 Reduction to block analysis

Before presenting the main result of this section, we need some more definitions and background. For each site $x \in T$, let $B_{x,\ell} \subseteq T$ denote the subtree (or "block") of height $\ell - 1$ rooted at x, i.e., $B_{x,\ell}$ consists of ℓ levels. (If x is $k < \ell$ levels from the bottom of T then $B_{x,\ell}$ has only k levels.) In what follows we will think of ℓ as a suitably large constant. By analogy with expression (7) for the Dirichlet form, let $\mathcal{D}_\ell(f) \equiv \sum_{x\in T} \mu[\text{Var}_{B_{x,\ell}}(f)]$ denote the local variation of f w.r.t. the blocks $\{B_{x,\ell}\}$. A straightforward manipulation (see, e.g., [28], keeping in mind that each site belongs to at most ℓ blocks) shows that c_{gap} can be bounded as follows:

$$c_{\text{gap}}(\mu) \geq \frac{1}{\ell} \cdot \inf_f \frac{\mathcal{D}_\ell(f)}{\text{Var}(f)} \cdot \min_{\eta,x} c_{\text{gap}}(\mu^\eta_{B_{x,\ell}}). \tag{9}$$

As before, the infimum is taken over non-constant functions (and henceforth we omit explicit mention of this). The importance of (9) is that $\min_{\eta,x} c_{\text{gap}}(\mu^\eta_{B_{x,\ell}})$ depends only on the size of $B_{x,\ell}$ and β, but not on the size of T; in fact, it is at least $\Omega(e^{-c(b,\beta)\cdot\ell})$ [4]. Therefore, in order to show that c_{gap} is bounded by a constant independent of the size of T, it is enough to show that, for some finite ℓ, $\text{Var}(f) \leq \text{const} \times \mathcal{D}_\ell(f)$ for all functions f. This is what we will show below, under the relevant spatial mixing condition. As a side remark, notice that $\inf_f \frac{\mathcal{D}_\ell(f)}{\text{Var}(f)}$ is exactly the (scaled) spectral gap of the Glauber dynamics based on flipping *blocks* $B_{x,\ell}$, rather than single sites x.

An identical manipulation yields an analogous bound for the log-Sobolev constant. For a non-negative function f, let $\mathcal{E}_\ell(f) \equiv \sum_{x\in T} \mu[\text{Ent}_{B_{x,\ell}}(f)]$. Then

$$c_{\text{sob}}(\mu) \geq \frac{1}{\ell} \cdot \inf_{f\geq 0} \frac{\mathcal{E}_\ell(f)}{\text{Ent}(f)} \cdot \min_{\eta,x} c_{\text{sob}}(\mu^\eta_{B_{x,\ell}}). \tag{10}$$

Hence to bound $c_{\text{sob}}(\mu)$ it suffices to show that, for some constant ℓ, $\text{Ent}(f) \leq \text{const} \times \mathcal{E}_\ell(f)$ for all $f \geq 0$.

3.2 Spatial mixing

We are now ready to state our spatial mixing conditions, first for the variance and then for the entropy. For $x \in T$, write T_x for the subtree rooted at x, and $\widetilde{T_x}$ for $T_x \setminus \{x\}$, the subtree T_x excluding its root.

Definition 3.1 [Variance Mixing] *We say that $\mu = \mu^\tau_T$ satisfies $\text{VM}(\ell,\varepsilon)$ if for every $x \in T$, any $\eta \in \Omega^\tau_T$ and any function f that does not depend on $B_{x,\ell}$ the following holds:*

$$\text{Var}^\eta_{T_x}[\mu_{\widetilde{T_x}}(f)] \;\leq\; \varepsilon \cdot \text{Var}^\eta_{T_x}(f).$$

Let us briefly discuss the above condition. Essentially, $\varepsilon = \varepsilon(\ell)$ gives the rate of decay with distance ℓ of point-to-set correlations. To see this, note that the l.h.s. $\mathrm{Var}^{\eta}_{T_x}[\mu_{\widetilde{T_x}}(f)]$ is the variance of the *projection* of f onto the root x of T_x, which is at distance ℓ from the sites on which f depends. It is also worth noting that the required uniformity in η in VM is not very restrictive: since the distribution $\mu^{\eta}_{T_x}$ depends only on the restriction of η to the boundary of T_x, and since $\eta \in \Omega^{\tau}_{T}$ (i.e., η agrees with τ on ∂T and therefore on the bottom boundary of T_x), the only freedom left in choosing η is in choosing the spin of the parent of x. Thus, VM is essentially a property of the distribution induced by the boundary condition τ. It is this lack of uniformity (i.e., the fact that we need not verify VM for other boundary conditions) that makes it flexible enough for our applications.

As the following theorem states, if $\mathrm{VM}(\ell, \varepsilon)$ holds with $\varepsilon \approx \frac{1}{2\ell}$, we get a constant lower bound on c_{gap}:

Theorem 3.2 *For any ℓ and $\delta > 0$, if μ satisfies $\mathrm{VM}(\ell, (1-\delta)/2(\ell+1-\delta))$ then $\mathrm{Var}(f) \leq \frac{3}{\delta} \cdot \mathcal{D}_{\ell}(f)$ for all f.*

Thus in order to show that $c_{\mathrm{gap}}(\mu) = \Omega(1)$ for a particular boundary condition τ, it suffices to show that VM with the above parameters holds for some fixed ℓ and $\delta > 0$, for all $\mu = \mu^{\tau}_{T}$ with T a full subtree.

Remark: In [4] it was shown that for general nearest-neighbor spin systems on any bounded degree graph, if $c_{\mathrm{gap}}(\mu)$ is bounded independently of n then μ exhibits an exponential decay of point-to-set correlations (i.e., $\mathrm{VM}(\ell, \exp(-\Theta(\ell)))$ holds for all ℓ). The authors of [4] posed the question of whether the converse is also true. Theorem 3.2 (which holds for general nearest-neighbor spin systems on a tree) answers this question affirmatively when the graph is a tree. In fact, combining our results with [4] implies that the decay of point-to-set correlations on a tree is either slower than linear or exponentially fast.

The analogous mixing condition for entropy and the log-Sobolev constant is the following:

Definition 3.3 [Entropy Mixing] *We say that $\mu = \mu^{\tau}_{T}$ satisfies $\mathrm{EM}(\ell, \varepsilon)$ if for every $x \in T$, any $\eta \in \Omega^{\tau}_{T}$ and any non-negative function f that does not depend on $B_{x,\ell}$ the following holds:*

$$\mathrm{Ent}^{\eta}_{T_x}[\mu_{\widetilde{T_x}}(f)] \;\leq\; \varepsilon \cdot \mathrm{Ent}^{\eta}_{T_x}(f).$$

Before stating the analog of Theorem 3.2 relating c_{sob} to EM, we need to define one more constant. Let $p_{\min} = \min_{x,s,\eta\in\Omega^{\tau}_{T}} \mu^{\eta}_{T_x}(\sigma_x = s)$, where s ranges over $\{+,-\}$; i.e., $p_{\min}$ is the minimum probability of any spin value at any site with any boundary condition. It is easy to see that $p_{\min} \geq \frac{1}{2}e^{-2\beta(b+|h|)}$, a constant depending only on b, β, h.

Theorem 3.4 *For any ℓ and $\delta > 0$, if μ satisfies $\mathrm{EM}(\ell, [(1-\delta)p_{\min}/(\ell+1-\delta)]^2)$ then $\mathrm{Ent}(f) \leq \frac{2}{\delta} \cdot \mathcal{E}_{\ell}(f)$ for all $f \geq 0$.*

This theorem can be used to show $c_{\mathrm{sob}}(\mu) = \Omega(1)$ for a given boundary condition τ exactly as indicated for c_{gap} immediately following Theorem 3.2 above.

Remark: As in the case of the spectral gap, a strong converse of Theorem 3.4 is also true. In the journal version of this paper [31] we prove that for any nearest-neighbor spin system on any bounded degree graph, if $c_{\mathrm{sob}}(\mu)$ is bounded independently of n then μ satisfies $\mathrm{EM}(\ell, \exp(-\Theta(\ell)))$ for all ℓ.

In order to prove Theorems 3.2 and 3.4 it is convenient to work with spatial mixing conditions that are somewhat more involved than VM and EM. The main difference is that we need to allow for functions that may depend on $B_{x,\ell}$ (the first ℓ levels of T_x) and thus need to introduce a term for this dependency. The modified conditions express the property that the variance (entropy) of the projection of any function f onto the root x of T_x can be bounded up to a constant factor by the local variance (entropy) of f in $B_{x,\ell}$, plus a negligible factor times the local variance (entropy) of f in $\widetilde{T_x}$. As the following lemma states, the modified conditions (with appropriate parameters) can be deduced from VM and EM.

Lemma 3.5 (i) *For any $\varepsilon < \frac{1}{2}$, if $\mu = \mu_T^\tau$ satisfies* $\mathrm{VM}(\ell,\varepsilon)$ *then for every $x \in T$, any $\eta \in \Omega_T^\tau$ and any function f we have $\mathrm{Var}_{T_x}^\eta[\mu_{\widetilde{T_x}}(f)] \le \frac{2-\varepsilon'}{1-\varepsilon'} \cdot \mu_{T_x}^\eta[\mathrm{Var}_{B_{x,\ell}}(f)] + \frac{\varepsilon'}{1-\varepsilon'} \cdot \mu_{T_x}^\eta[\mathrm{Var}_{\widetilde{T_x}}(f)]$, with $\varepsilon' = 2\varepsilon$.*

(ii) *For any $\varepsilon < p_{\min}^2$, if $\mu = \mu_T^\tau$ satisfies* $\mathrm{EM}(\ell,\varepsilon)$ *then for every $x \in T$, any $\eta \in \Omega_T^\tau$ and any function $f \ge 0$ we have $\mathrm{Ent}_{T_x}^\eta[\mu_{\widetilde{T_x}}(f)] \le \frac{1}{1-\varepsilon'} \cdot \mu_{T_x}^\eta[\mathrm{Ent}_{B_{x,\ell}}(f)] + \frac{\varepsilon'}{1-\varepsilon'} \cdot \mu_{T_x}^\eta[\mathrm{Ent}_{\widetilde{T_x}}(f)]$, with $\varepsilon' = \frac{\sqrt{\varepsilon}}{p_{\min}}$.*

Remark: We note that with some extra work, part (ii) of Lemma 3.5 can be improved to hold with $\varepsilon' = c(p_{\min})\varepsilon$. We give the weaker bound because it is simpler to prove while still enough for our applications.

Similar statements to those in Lemma 3.5 appeared in [5]. We defer our proof to the appendix as it only involves the application of standard inequalities from functional analysis.

We can now prove Theorems 3.2 and 3.4 by working with the modified spatial mixing conditions of Lemma 3.5.

Proof of Theorems 3.2 and 3.4: The main step is to show the following claim:

Claim 3.6 *If for every $x \in T$, any $\eta \in \Omega_T^\tau$ and any function f,*

$$\mathrm{Var}_{T_x}^\eta[\mu_{\widetilde{T_x}}(f)] \;\le\; c \cdot \mu_{T_x}^\eta[\mathrm{Var}_{B_{x,\ell}}(f)] \;+\; \left(\frac{1-\delta}{\ell}\right) \cdot \mu_{T_x}^\eta[\mathrm{Var}_{\widetilde{T_x}}(f)],$$

then $\mathrm{Var}(f) \le \frac{c}{\delta} \cdot \mathcal{D}_\ell(f)$ for all f. The same implication holds when Var *is replaced by* Ent*, $\mathcal{D}_\ell$ is replaced by $\mathcal{E}_\ell$ and the function f is restricted to be non-negative.*

Observe that the hypothesis of Theorem 3.2 together with part (i) of Lemma 3.5 establishes the hypothesis of Claim 3.6 with $c \le 3$, and similarly, the hypothesis of Theorem 3.4 together with part (ii) of Lemma 3.5 establishes the hypothesis of Claim 3.6 (after the necessary replacement of symbols) with $c \le 2$.

It therefore suffices to prove Claim 3.6. We prove only the formulation with Var and $\mathcal{D}_\ell$ since the proof for the formulation with Ent and $\mathcal{E}_\ell$ is identical once we make the same replacements in the text of the proof. As will be clear below, the proof uses only properties (3), (4) and (5) from Section 2 which are common to both Var and Ent.

Consider an arbitrary function $f : \Omega \to \mathbb{R}$. Our first goal is to relate $\mathrm{Var}(f)$ to the projections $\mathrm{Var}_{T_x}^\eta[\mu_{\widetilde{T_x}}(f)]$ for $x \in T$, so that we can apply the spatial mixing condition of the hypothesis. Recall that T has $m+1$ levels, and define the increasing sequence $\emptyset = F_0 \subset F_1 \subset \ldots \subset F_{m+1} = T$, where F_i consists of all sites in the lowest i levels of T. Thus F_i is a forest of height $i-1$. Using (3) recursively, and the facts that $\mu_{F_{i+1}}(\mu_{F_i}(f)) = \mu_{F_{i+1}}(f)$ and $\mu_{F_0}(f) = f$, we obtain

$$\begin{aligned}
\mathrm{Var}(f) &= \mu[\mathrm{Var}_{F_1}(f)] + \mathrm{Var}[\mu_{F_1}(f)] \\
&= \mu[\mathrm{Var}_{F_1}(f)] + \mu[\mathrm{Var}_{F_2}(\mu_{F_1}(f))] + \mathrm{Var}[\mu_{F_2}(\mu_{F_1}(f))] \\
&\;\vdots \\
&= \sum_{i=1}^{m+1} \mu[\mathrm{Var}_{F_i}(\mu_{F_{i-1}}(f))].
\end{aligned}$$

Now a fundamental property of nearest-neighbor interaction models on a tree is that, given the configuration on $T \setminus F_i$, the Gibbs distribution on F_i becomes a product of the marginals on the subtrees rooted at the sites $x \in F_i \setminus F_{i-1}$. Using inequality (4) for the variance of a product measure, we therefore have that

$$\mathrm{Var}(f) \le \sum_{i=1}^{m+1} \sum_{x \in F_i \setminus F_{i-1}} \mu[\mathrm{Var}_{T_x}(\mu_{F_{i-1}}(f))] \le \sum_{x \in T} \mu[\mathrm{Var}_{T_x}(\mu_{\widetilde{T_x}}(f))], \tag{11}$$

where in the second inequality we used the convexity of the variance as in (5).

Notice that so far we have not used the spatial mixing condition in the hypothesis of Claim 3.6, but only a natural martingale structure induced by the tree. Let us denote the final sum in (11) by $\mathrm{Pvar}(f)$. In order to bound c_{gap}, we need to compare the projection terms $\mathrm{Var}_{T_x}(\mu_{\widetilde{T_x}}(f))$ in $\mathrm{Pvar}(f)$ with the local conditional variance terms in $\mathcal{D}_\ell(f)$. For example, notice that if μ were the product of its single-site marginals then $\mathrm{Var}_{T_x}(\mu_{\widetilde{T_x}}(f)) \le \mu_{T_x}[\mathrm{Var}_x(f)]$ and $c_{\mathrm{gap}} = 1$. However, in general the variance of the projection on x may also involve terms which depend on other sites, and may lead to a factor that grows with the size of T_x. We will use the spatial mixing condition in order to preclude the latter possibility. Specifically, we show that if for every $x \in T$, any $\eta \in \Omega_T^\tau$ and any function g, $\mathrm{Var}_{T_x}^\eta[\mu_{\widetilde{T_x}}(g)] \;\le\; c \cdot \mu_{T_x}^\eta[\mathrm{Var}_{B_{x,\ell}}(g)] + \varepsilon \cdot \mu_{T_x}^\eta[\mathrm{Var}_{\widetilde{T_x}}(g)]$ then for every $x \in T$ and $\eta \in \Omega$,

$$\mathrm{Var}_{T_x}^\eta[\mu_{\widetilde{T_x}}(f)] \;\le\; c \cdot \mu_{T_x}^\eta[\mathrm{Var}_{B_x}(f)] + \varepsilon \cdot \sum_{y \in B_x \cup \widetilde{\partial} B_x, y \neq x} \mu_{T_x}^\eta[\mathrm{Var}_{T_y}(\mu_{\widetilde{T_y}}(f))], \tag{12}$$

where we have abbreviated $B_{x,\ell}$ to B_x and $\widetilde{\partial} B_x$ stands for the boundary of B_x excluding the parent of x, i.e., the bottom boundary of B_x. Notice that the last term in (12) is relevant only when x is at distance at least ℓ from the bottom of T. When x belongs to one of the ℓ lowest levels of T then $T_x = B_x$, and thus trivially $\mathrm{Var}_{T_x}^\eta[\mu_{\widetilde{T_x}}(f)] \le \mu_{T_x}^\eta[\mathrm{Var}_{B_x}(f)]$.

Let us assume (12) for now and conclude the proof of the theorem. Applying (12) for every x and η, and using the hypothesis that $\varepsilon = \frac{1-\delta}{\ell}$ and the fact that each site appears in at most ℓ blocks, we get

$$\begin{aligned} \mathrm{Pvar}(f) &\le c \cdot \mathcal{D}_\ell(f) + \varepsilon \cdot \sum_{x \in T} \sum_{y \in B_x \cup \widetilde{\partial} B_x, y \neq x} \mu[\mathrm{Var}_{T_y}(\mu_{\widetilde{T_y}}(f))] \\ &\le c \cdot \mathcal{D}_\ell(f) + \varepsilon\ell \cdot \sum_{y \in T} \mu[\mathrm{Var}_{T_y}(\mu_{\widetilde{T_y}}(f))] \\ &= c \cdot \mathcal{D}_\ell(f) + (1-\delta)\mathrm{Pvar}(f), \end{aligned}$$

and hence

$$\mathrm{Var}(f) \;\le\; \mathrm{Pvar}(f) \;\le\; \frac{c}{\delta} \cdot \mathcal{D}_\ell(f),$$

proving Claim 3.6. We now return to proving (12).

Let $g = \mu_{T_x \setminus (B_x \cup \widetilde{\partial} B_x)}(f)$. Once we notice that $\mu_{\widetilde{T_x}}(f) = \mu_{\widetilde{T_x}}(g)$, we can use the spatial mixing assumption that precedes (12) to deduce

$$\begin{aligned} \mathrm{Var}_{T_x}^\eta[\mu_{\widetilde{T_x}}(f)] &\le c \cdot \mu_{T_x}^\eta[\mathrm{Var}_{B_x}(g)] + \varepsilon \cdot \mu_{T_x}^\eta[\mathrm{Var}_{\widetilde{T_x}}(g)] \\ &\le c \cdot \mu_{T_x}^\eta[\mathrm{Var}_{B_x}(f)] + \varepsilon \cdot \mu_{T_x}^\eta[\mathrm{Var}_{\widetilde{T_x}}(g)], \end{aligned}$$

where we used (5) for the second inequality. We will be done once we show that

$$\mu_{T_x}^\eta[\mathrm{Var}_{\widetilde{T_x}}(g)] \;\le\; \sum_{y \in B_x \cup \widetilde{\partial} B_x, y \neq x} \mu_{T_x}^\eta[\mathrm{Var}_{T_y}(\mu_{\widetilde{T_y}}(f))]. \tag{13}$$

But (13) follows from a similar argument to that used earlier to show $\mathrm{Var}(f) \le \mathrm{Pvar}(f)$, starting from the fact that $g = \mu_{F_k'}(f)$, where the forests F_i' are defined analogously to the F_i earlier but restricted to the subtree T_x, and $k = \mathrm{height}(x) - \ell$. We omit the details.

This concludes the proof of Claim 3.6, and thus of Theorems 3.2 and 3.4. $\square$

4 Verifying spatial mixing for the spectral gap

In this section, we will prove that the spectral gap of the Glauber dynamics is bounded in all of the situations covered by Theorems A and B in the Introduction. By Theorem 2.1, this immediately implies that the mixing time is $O(n^2)$ in these situations, thus verifying a weaker version of Theorems A and B. The improvement from $O(n^2)$ to $O(n \log n)$ will follow from our analysis of the log-Sobolev constant in the next section.

In light of Theorem 3.2, to bound the spectral gap it suffices to verify the Variance Mixing condition $\mathrm{VM}(\ell, \varepsilon)$ with $\varepsilon = (1-\delta)/2(\ell+1-\delta)$, for some constants $\ell, \delta > 0$ independent of the size of T. In fact, we will show it with the asymptotically tighter value $\varepsilon = \exp(-\Theta(\ell))$:

Theorem 4.1 *In both of the following situations, there exists a positive constant ϑ (depending only on b, β and h) such that, for all T, the Gibbs distribution $\mu = \mu_T^\tau$ satisfies $\mathrm{VM}(\ell, e^{-\vartheta\ell})$ for all ℓ:*

(i) *τ is arbitrary, and either $\beta < \beta_1$ (with h arbitrary), or $|h| > h_c(\beta)$ (with β arbitrary);*

(ii) *τ is the $(+)$-boundary condition, and β, h are arbitrary.*

As a corollary, in both situations $c_{\mathrm{gap}}(\mu) = \Omega(1)$.

Remark: The validity of VM, i.e, the decay of point-to-set correlations, is of interest independently of its implication for the spectral gap (an implication which is new to this paper): e.g., it is closely related to the purity of the infinite volume Gibbs measure and to bit reconstruction problems on trees [12]. In the special case of the free boundary condition and $h = 0$, part (i) of Theorem 4.1 was first proved in [7] via a lengthy calculation, which was considerably simplified in [19]. It was later reproved in [4] (for arbitrary boundary conditions) as a consequence of the fact that the spectral gap is bounded in this situation. An extension to general trees can be found in [12] and [20]. Our motivation for presenting another proof of part (i) (in addition to handling general fields h) is the simplicity of our argument compared with previous ones. As far as part (ii) is concerned, we are unaware of any previous results for the case of the $(+)$-boundary other than the fact that $\mathrm{VM}(\ell, o(1))$ must hold because the $(+)$-phase is pure (see, e.g., [15]).

The rest of this section is divided into two parts. First, we develop a general framework based on coupling in order to establish the exponential decay of point-to-set correlations. This framework identifies two key quantities, κ and γ, and states that when their product is small enough then VM holds. Then, in the second part, we go back to proving Theorem 4.1 by calculating κ and γ for each of the above two regimes separately.

4.1 A coupling argument for decay of point-to-set correlations

In this section we develop a coupling framework that enables us to verify the exponential decay of point-to-set correlations from a simple calculation involving single-spin distributions.

First we need some additional notation. When x is not the root of T, let $\mu_{T_x}^+$ (respectively, $\mu_{T_x}^-$) denote the Gibbs distribution in which the parent of x has its spin fixed to $(+)$ (respectively, $(-)$) and the configuration on the bottom boundary of T_x is specified by τ (the global boundary condition on T) [§]. For two distributions μ_1 and μ_2, we denote by $\|\mu_1 - \mu_2\|_x$ the variation distance between the projections of μ_1 and μ_2 onto the spin at x. (Since the Ising model has only two spin values, $\|\mu_1 - \mu_2\|_x = |\mu_1(\sigma_x = +) - \mu_2(\sigma_x = +)|$.) Recall also that η^y denotes the configuration η with the spin at site y flipped.

We now identify two constants that are crucial for our coupling argument:

[§] Notice that we do not specify the rest of the configuration outside T_x since it has no influence on the distribution inside T_x once the spin at the parent of x is fixed. However, since our distributions are defined over the whole configuration space, in the discussion below when the configuration outside T_x is relevant it will be understood from the context.

Definition 4.2 *For a sequence of Gibbs distributions $\{\mu_T^\tau\}$ corresponding to a fixed boundary condition τ, define $\kappa \equiv \kappa(\{\mu_T^\tau\})$ and $\gamma \equiv \gamma(\{\mu_T^\tau\})$ by*

(i) $\kappa = \sup_T \max_z \|\mu_{T_z}^+ - \mu_{T_z}^-\|_z$;

(ii) $\gamma = \sup_T \max \|\mu_A^\eta - \mu_A^{\eta^y}\|_z$, *where the maximum is taken over all subsets $A \subseteq T$, all boundary configurations η, all sites y on the boundary of A and all neighbors $z \in A$ of y.*

Note that κ is the same as γ, except that the maximization is restricted to $A = T_z$ and the boundary vertex y being the parent of z; hence always $\kappa \le \gamma$. Since κ involves Gibbs distributions only on maximal subtrees T_z, it may depend on the boundary condition τ at the bottom of the tree. By contrast, γ bounds the worst-case probability of disagreement for an *arbitrary* subset A and arbitrary boundary configuration around A, and hence depends only on (β, h) and not on τ. It is the dependence of κ on τ that opens up the possibility of an analysis that is specific to the boundary condition. For example, at very low temperature and with no external field, κ is close to 1 in the free boundary case, while it is close to zero in the $(+)$-boundary case.

In our arguments κ will be used to bound the probability of a disagreement percolating one level *down* the tree, namely, when we fix a disagreement at x and couple the two resulting marginals on a child z of x. On the other hand, γ will be used in order to bound the probability of a disagreement percolating one level *up* the tree, namely, when we fix a single disagreement on the bottom boundary of a block, say at y (with the rest of the boundary configuration being arbitrary), and couple the marginals on the parent of y.

The novelty of our argument for establishing VM comes from the fact that we identify *two separate* constants κ and γ, and consider their product, rather than working with κ alone:

Theorem 4.3 *Any Gibbs distribution $\mu = \mu_T^\tau$ satisfies $\mathrm{VM}(\ell, (\gamma\kappa b)^\ell)$ for all ℓ, where κ and γ are the constants associated with the sequence $\{\mu_T^\tau\}$ as specified in Definition 4.2. In particular, if $\gamma\kappa b < 1$ then there exists a constant $\vartheta > 0$ such that, for every T, the measure $\mu = \mu_T^\tau$ satisfies $\mathrm{VM}(\ell, e^{-\vartheta\ell})$ for all ℓ, and hence $c_{\mathrm{gap}}(\mu) = \Omega(1)$.*

Proof: Fix arbitrary T, $x \in T$, $\eta \in \Omega_T^\tau$. We need to show that for every function f that does not depend on $B_{x,\ell}$, $\mathrm{Var}_{T_x}^\eta[\mu_{\widetilde{T_x}}(f)] \le \varepsilon \cdot \mathrm{Var}_{T_x}^\eta(f)$ with $\varepsilon = (\kappa\gamma b)^\ell$, i.e., projecting f onto the root (of T_x) causes the variance to shrink by a factor ε. As is well known, it is enough to establish a dual contraction, i.e., to consider an arbitrary function that depends only on the spin at the root and show that, when projecting onto levels ℓ and below, the variance shrinks by a factor of ε. Formally, it is enough to show that for every function g that does not depend on $\widetilde{T_x}$[¶] we have

$$\mathrm{Var}_{T_x}^\eta[\mu_{B_{x,\ell}}(g)] \;\le\; \varepsilon \cdot \mathrm{Var}_{T_x}^\eta(g). \tag{14}$$

This is because for a function f that does not depend on $B_{x,\ell}$, the variance of the projection can be written as

$$\mathrm{Var}_{T_x}^\eta[\mu_{\widetilde{T_x}}(f)] \;=\; \mathrm{Cov}_{T_x}^\eta(f, \mu_{\widetilde{T_x}}(f)) \;=\; \mathrm{Cov}_{T_x}^\eta(f, \mu_{B_{x,\ell}}(\mu_{\widetilde{T_x}}(f))) \;\le\; \sqrt{\mathrm{Var}_{T_x}^\eta(f) \cdot \mathrm{Var}_{T_x}^\eta[\mu_{B_{x,\ell}}(\mu_{\widetilde{T_x}}(f))]}\;,$$

where $\mathrm{Cov}_A^\eta(f, f')$ denotes the covariance $\mu_A^\eta(ff') - \mu_A^\eta(f)\mu_A^\eta(f')$ and the last inequality is an application of Cauchy-Schwartz. We then have

$$\mathrm{Var}_{T_x}^\eta[\mu_{\widetilde{T_x}}(f)] \;\le\; \mathrm{Var}_{T_x}^\eta(f) \cdot \frac{\mathrm{Var}_{T_x}^\eta[\mu_{B_{x,\ell}}(\mu_{\widetilde{T_x}}(f))]}{\mathrm{Var}_{T_x}^\eta[\mu_{\widetilde{T_x}}(f)]}\;.$$

[¶]Effectively this means that, conditioned on the configuration outside T_x being η, g depends only on the spin at the root x.

If we assume (14) then the expression on the r.h.s. is bounded by $\varepsilon \cdot \mathrm{Var}^{\eta}_{T_x}(f)$ since $g = \mu_{\widetilde{T_x}}(f)$ does not depend on $\widetilde{T_x}$.

We therefore proceed with the proof of (14), which goes via a coupling argument. A *coupling* of two distributions μ_1, μ_2 on Ω is any joint distribution ν on Ω^2 whose marginals are μ_1 and μ_2 respectively. For two configurations $\sigma, \sigma' \in \Omega$, let $|\sigma - \sigma'|_{x,\ell}$ denote the *Hamming distance* between the restrictions of σ and σ' to $\widetilde{\partial} B_{x,\ell}$, i.e., the number of sites at distance ℓ below x at which σ and σ' differ. Notice that $|\sigma - \sigma'|_{x,\ell}$ can be at most b^ℓ, the number of sites on the ℓth level below x. Let $\mu^{+}_{\widetilde{T_x}}$ (respectively, $\mu^{-}_{\widetilde{T_x}}$) stand for the Gibbs distribution where the spin at x is set to $(+)$ (respectively, $(-)$) and, as usual, the configuration on the bottom boundary of $\widetilde{T_x}$ is specified by τ. Our goal will be to construct a coupling ν of $\mu^{+}_{\widetilde{T_x}}$ and $\mu^{-}_{\widetilde{T_x}}$ for which the expectation $\mathrm{E}_\nu|\sigma - \sigma'|_{x,\ell} \equiv \sum_{\sigma,\sigma'} \nu(\sigma,\sigma')|\sigma - \sigma'|_{x,\ell}$ is only $(\kappa b)^\ell$.

Claim 4.4 *For every $x \in T$ and all ℓ the following hold*:

(i) *There is a coupling ν of $\mu^{+}_{\widetilde{T_x}}$ and $\mu^{-}_{\widetilde{T_x}}$ for which $\mathrm{E}_\nu|\sigma - \sigma'|_{x,\ell} \le (\kappa b)^\ell$.*

(ii) *For any $\eta, \eta' \in \Omega$ that have the same spin value at the parent of x, $\|\mu^{\eta}_{B_{x,\ell}} - \mu^{\eta'}_{B_{x,\ell}}\|_x \le \gamma^\ell \cdot |\eta - \eta'|_{x,\ell}$.*

Let us assume Claim 4.4 for the moment and complete the proof of (14). Consider an arbitrary g that does not depend on $\widetilde{T_x}$. Let $p = \mu^{\eta}_{T_x}(\sigma_x = +)$ and $q = 1 - p = \mu^{\eta}_{T_x}(\sigma_x = -)$. We also write g^+ for $g(\sigma)$, where σ is any configuration that agrees with η outside T_x and such that $\sigma_x = +$. (This is well defined since g does not depend on $\widetilde{T_x}$). We define g^- similarly. Without loss of generality we may assume that in the coupling ν from Claim 4.4 both the coupled configurations agree with η outside T_x with probability 1. We then have

$$\begin{aligned}
\mathrm{Var}^{\eta}_{T_x}[\mu_{B_{x,\ell}}(g)] &= \mathrm{Cov}^{\eta}_{T_x}[g, \mu_{B_{x,\ell}}(g)] \\
&= \mathrm{Cov}^{\eta}_{T_x}[g, \mu_{\widetilde{T_x}}(\mu_{B_{x,\ell}}(g))] \\
&= pq(g^+ - g^-)[\mu^{+}_{\widetilde{T_x}}(\mu_{B_{x,\ell}}(g)) - \mu^{-}_{\widetilde{T_x}}(\mu_{B_{x,\ell}}(g))] \\
&= pq(g^+ - g^-)\sum_{\sigma,\sigma'} \nu(\sigma,\sigma')[\mu^{\sigma}_{B_{x,\ell}}(g) - \mu^{\sigma'}_{B_{x,\ell}}(g)] \\
&\le pq|g^+ - g^-|\sum_{\sigma,\sigma'} \nu(\sigma,\sigma')\|\mu^{\sigma}_{B_{x,\ell}} - \mu^{\sigma'}_{B_{x,\ell}}\|_x \cdot |g^+ - g^-| \qquad (15) \\
&\le pq(g^+ - g^-)^2\sum_{\sigma,\sigma'} \nu(\sigma,\sigma')|\sigma - \sigma'|_{x,\ell} \cdot \gamma^\ell \\
&= \gamma^\ell \cdot \mathrm{Var}^{\eta}_{T_x}(g) \cdot \mathrm{E}_\nu|\sigma - \sigma'|_{x,\ell} \\
&\le (\gamma\kappa b)^\ell \cdot \mathrm{Var}^{\eta}_{T_x}(g).
\end{aligned}$$

In the sixth line here we have used part (ii) of Claim 4.4, and in the last line we used part (i). This completes the proof of (14), and hence of Theorem 4.3. We thus go back and prove Claim 4.4.

The proof of Claim 4.4 makes use of a standard recursive coupling along paths in the tree (as in, e.g., [4]). We start with part (i), i.e., constructing a coupling ν of $\mu^{+}_{\widetilde{T_x}}$ and $\mu^{-}_{\widetilde{T_x}}$ with the required properties. Since the underlying graph is a tree, we can couple $\mu^{+}_{\widetilde{T_x}}$ and $\mu^{-}_{\widetilde{T_x}}$ recursively. This goes as follows. First, given the spin at x the measures on T_z (where z ranges over the children of x) are all independent of each other, so we can couple the projections on the T_z's independently. Then, we couple the two projections on T_z by first coupling the spin at z using the optimal coupling (the one that achieves the variation distance) of the marginal measures on the spin at z. Thus, the spins at z disagree with probability at most κ. Once a coupled pair of spins at z is chosen, we continue as follows: if the spins at z agree then we can make the configurations in $\widetilde{T_z}$ equal with probability 1 (because the two boundary conditions are the same); if the spins at z differ (i.e., one is $(+)$ and the other $(-)$) then we recursively couple $\mu^{+}_{\widetilde{T}_z}$ and $\mu^{-}_{\widetilde{T}_z}$. We let ν be the resulting coupling of $\mu^{+}_{\widetilde{T}_x}$ and $\mu^{-}_{\widetilde{T}_x}$, and notice that $\mathrm{E}_\nu|\sigma - \sigma'|_{x,l} \le (\kappa b)^\ell$ since for every site y at distance ℓ below x the probability that the two coupled spins at y disagree is at most κ^ℓ.

We go on to prove part (ii) of Claim 4.4. First, by writing a telescopic sum and applying the triangle inequality we get that

$$\|\mu^{\eta}_{B_{x,\ell}} - \mu^{\eta'}_{B_{x,\ell}}\|_x \;\le\; \sum_{i=1}^{k} \|\mu^{\eta^{(i-1)}}_{B_{x,\ell}} - \mu^{\eta^{(i)}}_{B_{x,\ell}}\|_x \ ,$$

where $k = |\eta - \eta'|_{x,\ell}$ and the sequence of configurations $\eta^{(i)}$ is a site-by-site interpolation of the differences between η and η' in $\widetilde{\partial} B_{x,\ell}$. (It suffices to interpolate only over the differences in $\widetilde{\partial} B_{x,\ell}$ since the measure $\mu^{\eta}_{B_{x,\ell}}$ depends only on the configuration in $\partial B_{x,\ell}$ and since η and η' agree on the parent of x.) It is now enough to show that $\|\mu^{\eta}_{B_{x,\ell}} - \mu^{\eta^w}_{B_{x,\ell}}\|_x \le \gamma^\ell$ for all η and $w \in \widetilde{\partial} B_{x,\ell}$. This, however, follows by a coupling argument as before, where this time we couple recursively along the path from w to x (i.e., up the tree). Specifically, suppose by induction that in our coupling there is already a path of disagreement going from w to y, where y is some site on the path from w to x. Let z denote the parent of y. At the next step we choose a coupled pair of spins at z from the two distributions μ^{η}_A and $\mu^{\eta^y}_A$ (using an optimal coupling for the projections onto the spin at z), where the subset A is $B_{x,\ell}$ excluding the path from w to y. The probability of disagreement at z given the disagreement at y is then bounded by γ, by definition. If the resulting spins at z agree then the spins on the rest of the path are coupled to agree with certainty, while if there is a disagreement at z we continue recursively starting from the disagreement at z. We therefore conclude that the probability of disagreement at x in the resulting coupling is γ^ℓ, as required. $\square$

Remark: We stress the fact that Theorem 4.3 is not specific to the Ising model and can be generalized to arbitrary nearest-neighbor models on a tree. Although we used the fact that the Ising model has only two possible spin values, the theorem can be generalized to more than two spin values at the cost of a factor $\frac{1}{p_{\min}}$ in front of $(\gamma\kappa b)^\ell$ in VM, where $p_{\min}$ is the minimum probability of any spin value as defined just before Theorem 3.4. Thus, since Theorem 3.2 also applies to general nearest-neighbor spin systems on a tree, we conclude that the implication from $\gamma\kappa b < 1$ to a bounded $c_{\rm gap}(\mu)$ holds for any such system (with the definitions of κ and γ extended in the obvious way to systems with more than two spin values). The details can be found in the companion paper [32].

4.2 Proof of Theorem 4.1

In this section we go back to proving Theorem 4.1. Using Theorem 4.3, all we need to do for the given choices of the Ising model parameters is to bound κ and γ as in Definition 4.2 such that $\gamma\kappa b < 1$. In contrast to Sections 3 and 4.1, which apply to general nearest-neighbor spin systems on trees, here the calculations are specific to the Ising model.

For both κ and γ, we need to bound a quantity of the form $\|\mu^{\eta}_A - \mu^{\eta^y}_A\|_z$, where $y \in \partial A$ and $z \in A$ is a neighbor of y. The key observation is that this quantity can be expressed very cleanly in terms of the "magnetization" at z, i.e., the ratio of probabilities of a $(-)$-spin and a $(+)$-spin at z. It will actually be convenient to work with the magnetization *without* the influence of the neighbor y: thus we let $\mu^{\eta,y=*}_A$ denote the Gibbs distribution with boundary condition η, except that the spin at y is free (or equivalently, the edge connecting z to y is erased). We then have:

Proposition 4.5 *For any subset $A \subseteq T$, any boundary configuration η, any site $y \in \partial A$ and any neighbor $z \in A$ of y, we have*

$$\|\mu^{\eta}_A - \mu^{\eta^y}_A\|_z = K_\beta(R),$$

where $R = \frac{\mu_A^{\eta,y=*}(\sigma_z=-)}{\mu_A^{\eta,y=*}(\sigma_z=+)}$ *and the function* K_β *is defined by*

$$K_\beta(a) = \frac{1}{e^{-2\beta}a+1} - \frac{1}{e^{2\beta}a+1}.$$

Proof: First, w.l.o.g. we may assume that the edge connecting y and z is the only one connecting y to A; this is because a tree has no cycles, so once the spin at y is fixed A decomposes into disjoint components that are independent. We also assume w.l.o.g. that the spin at y is $(+)$ in η, and we abbreviate μ_A^η and $\mu_A^{\eta^y}$ to μ_A^+ and μ_A^- respectively, and also $\mu_A^{\eta,y=*}$ to μ_A^*. Thus $\|\mu_A^\eta - \mu_A^{\eta^y}\|_z = |\mu_A^+(\sigma_z=+) - \mu_A^-(\sigma_z=+)|$, and $R = \frac{\mu_A^*(\sigma_z=-)}{\mu_A^*(\sigma_z=+)}$. We write R^+ for $\frac{\mu_A^+(\sigma_z=-)}{\mu_A^+(\sigma_z=+)}$ and R^- for $\frac{\mu_A^-(\sigma_z=-)}{\mu_A^-(\sigma_z=+)}$. Since the only influence of y on A is through z, we have $R^+ = e^{-2\beta}R$ and $R^- = e^{2\beta}R$. The proposition now follows once we notice that, by definition of R^+ and R^-, $\mu_A^+(\sigma_z=+) = \frac{1}{R^++1}$ and $\mu_A^-(\sigma_z=+) = \frac{1}{R^-+1}$. □

Now it is easy to check that $K_\beta(a)$ is an increasing function in the interval $[0,1]$, decreasing in the interval $[1,\infty]$, and is maximized at $a=1$. Therefore, we can always bound κ and γ from above by $K_\beta(1) = \frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}}$. Indeed, for γ we must make do with this crude bound because it has to hold for any boundary configuration η and we cannot hope to gain by controlling the magnetization R. However, as we shall see, for κ we can do better in some cases by computing the magnetization at the root; when this differs from 1 we get a better bound than $K_\beta(1)$.

We are now ready to proceed to the proof of Theorem 4.1:

(i) Arbitrary boundary conditions

Here, the boundary condition τ is arbitrary and we first consider the (easy) case when $\beta < \beta_0$ or $|h| > h_c(\beta)$ (i.e., h is super-critical). In this case we do not need to resort to the calculation of κ and γ. As discussed in the Introduction, in this regime there is a unique infinite volume Gibbs measure, so certainly the variation distance at the root $\max_{\eta,\eta'} \|\mu_{B_{x,\ell}}^\eta - \mu_{B_{x,\ell}}^{\eta'}\|_x$ goes to zero as ℓ increases. In fact, it is not too difficult to see that in the above regime this variation distance goes to zero exponentially fast, which directly implies the desired exponential decay of correlations (VM) by plugging the bound on the variation distance into expression (15) in the proof of Theorem 4.3.

We go on to consider the more interesting regime when $\beta_0 \le \beta < \beta_1$ (i.e., intermediate temperatures) and the external field h is arbitrary. Here we use the fact that $\kappa \le \gamma \le K_\beta(1)$. We then certainly have $\gamma\kappa b < 1$ whenever $K_\beta(1) = \frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}} < \sqrt{\frac{1}{b}}$, i.e., whenever $e^{-2\beta} > \frac{\sqrt{b}-1}{\sqrt{b}+1}$. From the definition of β_1 (see Section 1.2), this corresponds precisely to $\beta < \beta_1$. (Observe how this non-trivial result drops out immediately from our machinery, as expressed in the condition $\gamma^2 < \frac{1}{b}$).

This completes the verification of Theorem 4.1 part (i).

(ii) $(+)$**-boundary condition**

We now assume that τ is the all-$(+)$ configuration and consider arbitrary β and h. For convenience, we assume $h \ge -h_c(\beta)$ since the case $|h| > h_c(\beta)$ was covered in part (i) for all boundary conditions τ. The important property of the regime $h \ge -h_c(\beta)$ is that, for the $(+)$-boundary, the spin at the root is at least as likely to be $(+)$ as it is to be $(-)$. We will show that $\gamma\kappa b < 1$ throughout this regime. Recall that we already showed that $\gamma \le K_\beta(1) < 1$ for all finite β. It is therefore enough to show that $\kappa \le \frac{1}{b}$.

To calculate κ, we need to bound the variation distance $\|\mu_{T_z}^+ - \mu_{T_z}^-\|_z$, which by Proposition 4.5 is equal to $K_\beta(R_z)$, where $R_z = \frac{\mu_{T_z}^*(\sigma_z=-)}{\mu_{T_z}^*(\sigma_z=+)}$ and $\mu_{T_z}^*$ is the Gibbs distribution over the subtree T_z when it is disconnected from the rest of T and the spins on its bottom boundary agree with τ. We thus have $\kappa = \sup_T \max_{z\in T} K_\beta(R_z)$.

The final ingredient we need is a recursive computation of the magnetization R_z, the details of which (up to change of variables) can be found in [3] or [6]. Let $y \prec z$ denote that y is a child of z.

A simple direct calculation gives that $R_z = e^{-2\beta h} \prod_{y \prec z} F(R_y)$, where $F(a) \equiv F_\beta(a) = \frac{a+e^{-2\beta}}{e^{-2\beta}a+1}$. In particular, if z is any site on the bottom-most level of T, then since the spins of the children of z are all set deterministically to $(+)$, we get that $R_z = e^{-2\beta h}[F(0)]^b$. We thus define

$$J(a) \equiv J_{\beta,h}(a) = e^{-2\beta h}[F(a)]^b \tag{16}$$

and observe that, for any $z \in T$, $R_z = J^{(\ell)}(0)$, where $J^{(\ell)}$ stands for the ℓ-fold composition of J, and ℓ is the distance of z from the bottom boundary of T.

We now describe some properties of J that we use (refer to Fig. 2): J is continuous and increasing on $[0,\infty)$, with $J(0) = e^{-2\beta(h+b)} > 0$ and $\sup_a J(a) = e^{-2\beta(h-b)} < \infty$. This immediately implies that J has at least one fixed point in $[0,\infty)$; we denote by a_0 the least fixed point. Since a_0 is the least fixed point and $J(0) > 0$ then clearly $J'(a_0) \le 1$, where $J'(a) \equiv \frac{\partial J(a)}{\partial a}$ is the derivative of J. We also note that $a_0 \le 1$ when $h \ge -h_c(\beta)$, which corresponds to the fact that for the all-$(+)$ boundary and the above regime of h, the spin at the root is at least as likely to be $(+)$ as $(-)$.

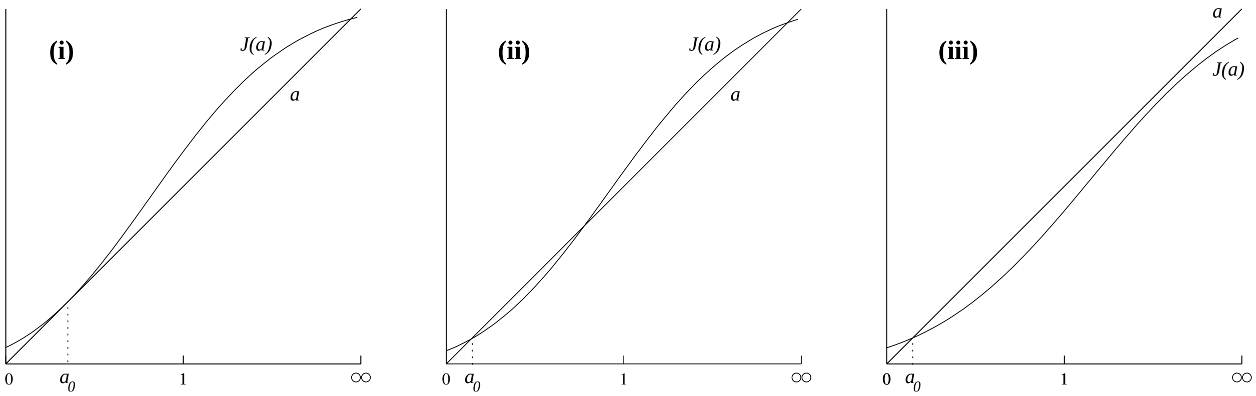

Figure 2: Curve of the function $J(a)$, used in the proof of Theorem 4.1, for $\beta > \beta_0$ and various values of the external field h. (i) $h = -h_c(\beta)$; (ii) $h_c(\beta) > h > -h_c(\beta)$; (iii) $h > h_c(\beta)$. The point a_0 is the smallest fixed point of J.

Now, since J is monotonically increasing and a_0 is the least fixed point of J, then clearly $J^{(\ell)}(0)$ converges to a_0 from below, i.e., $R_z \le a_0$ for every $z \in T$. Thus, since $a_0 \le 1$ for $h \ge -h_c(\beta)$, and the function $K_\beta(a)$ is monotonically increasing in the interval $[0,1]$, $K_\beta(R_z) \le K_\beta(a_0)$ for every $z \in T$.

What remains to be shown is that $K_\beta(a_0) \le \frac{1}{b}$. This follows from the fact that $J'(a_0) \le 1$, together with the following lemma:

Lemma 4.6 *Let a_0 be any fixed point of J. Then $K_\beta(a_0) = \frac{1}{b} \cdot J'(a_0)$.*

Proof: From the definitions of J and F we have:

$$\begin{aligned}
J'(a_0) &= e^{-2\beta h} \cdot b \cdot [F(a_0)]^{b-1} F'(a_0) \\
&= b \cdot J(a_0) \cdot \frac{F'(a_0)}{F(a_0)} \\
&= b \cdot a_0 \cdot \frac{F'(a_0)}{F(a_0)} \\
&= b \cdot a_0 \cdot \Big[\frac{1-e^{-4\beta}}{(a_0+e^{-2\beta})(e^{-2\beta}a_0+1)}\Big] \\
&= b \cdot K_\beta(a_0). \qquad \square
\end{aligned}$$

This completes the verification of Theorem 4.1 part (ii).

5 Verifying spatial mixing for log-Sobolev

In this section we will prove a uniform lower bound (independent of n) on the logarithmic Sobolev constant $c_{\rm sob}(\mu)$ in all the situations covered by Theorems A and B in the Introduction.

In light of Theorem 3.4, to show $c_{\rm sob} = \Omega(1)$ we need only prove the validity of the Entropy Mixing condition $\mathrm{EM}(\ell, [(1-\delta)p_{\min}/2(\ell+1-\delta)]^2)$ for some constants ℓ and δ independent of the size of T. In order to establish EM in the situations covered by Theorem A and B, we extend the coupling framework developed in Section 4.1 so that it can be used to establish EM. As before, we will use a condition on the constants κ and γ, which were defined in Section 4.1. In fact, the condition on κ and γ for establishing EM is practically the same as the one that was used to establish VM, which immediately transfers our $\Omega(1)$ bound on $c_{\rm gap}$ for the relevant parameters to an $\Omega(1)$ bound on $c_{\rm sob}$ for the same choice of parameters. The main result of this section is the following relationship between (κ, γ) and EM.

Theorem 5.1 *Any Gibbs distribution $\mu = \mu_T^\tau$ satisfies $\mathrm{EM}(\ell, c(\gamma\alpha)^{\ell/5})$ for all ℓ, where $\alpha = \max\{\kappa b, 1\}$, κ and γ are the constants associated with the sequence $\{\mu_T^\tau\}$ as specified in Definition 4.2, and c is a constant that depends only on (b, β, h). In particular, if $\max\{\gamma\kappa b, \gamma\} < 1$ then there exists a constant ϑ such that, for every T, the measure $\mu = \mu_T^\tau$ satisfies $\mathrm{EM}(\ell, ce^{-\vartheta\ell})$ for all ℓ, and hence $c_{\rm sob}(\mu) = \Omega(1)$.*

Remark: We should note that the above theorem, like its counterpart for the spectral gap, holds for any spin system on a tree (with the definitions of κ and γ generalized appropriately). See the companion paper [32] for details.

Since in Section 4.2 we have already calculated κ and γ for the regimes in Theorems A and B and shown that in both cases $\max\{\gamma\kappa b, \gamma\} < 1$, we have:

Corollary 5.2 *In both of the following situations, $c_{\rm sob}(\mu) = \Omega(1)$:*

(i) *τ is arbitrary, and either $\beta < \beta_1$ (with h arbitrary), or $|h| > h_c(\beta)$ (with β arbitrary);*

(ii) *τ is the $(+)$-boundary condition and β, h are arbitrary.*

Theorems A and B follow immediately from Corollary 5.2 and Theorem 2.1.

The first step in proving Theorem 5.1 is a reduction of EM to a certain *strong concentration* property of μ, the Gibbs measure under consideration. We believe that this concentration property, as well as its connection to EM, may be of independent interest. The statement of this property and the reduction of EM to it is the content of Section 5.1. Then, in Section 5.2, we complete the proof of Theorem 5.1 by relating the strong concentration property to κ and γ.

It is worth mentioning that we are also able to establish a general (but cruder) bound on $c_{\rm sob}$ as a function of $c_{\rm gap}$. Specifically, we can show that $c_{\rm sob} = \Omega(1/\log n) \times c_{\rm gap}$. Although we do not need this bound in this paper, we present it in Section 5.3 for future reference since its proof is simple and short.

5.1 Establishing EM via a strong concentration property.

In this subsection we reduce EM to a certain strong concentration property of μ. In the next subsection, we will then establish this strong concentration property as a function of κ and γ in order to prove Theorem 5.1. Consider an arbitrary T. For simplicity and without loss of generality, we will analyze the entropy mixing condition only for $T_x = T$, with root r.

Let $\mu_{\widetilde{T}}^+$ and $\mu_{\widetilde{T}}^-$ denote the Gibbs distributions on $\widetilde{T}$ with the spin at the root r set to $(+)$ and $(-)$ respectively (the boundary condition on the leaves of T being specified by τ). Define

$$g_+(\sigma) = \frac{\mu_{\widetilde{T}}^+(\sigma)}{\mu(\sigma)} = \begin{cases} 1/p & \text{if } \sigma_r = (+), \\ 0 & \text{otherwise,} \end{cases}$$

where $p = \mu(\sigma_r = +)$. The key quantity we will work with in the sequel is the following:

$$g_+^{(\ell)} = \mu_{B_{r,\ell}}(g_+).$$

Note that $g_+^{(\ell)}(\sigma)$ depends only on the spins in $\partial B_{r,\ell}$. Indeed, let $\sigma_{r,\ell}$ stand for the restriction of σ to $\partial B_{r,\ell}$, i.e., to the sites at distance ℓ below r. It is easy to verify that $g_+^{(\ell)}(\sigma)$ is equal to $\frac{\mu_{\widetilde{T}}^+(\sigma_{r,\ell})}{\mu(\sigma_{r,\ell})}$. Thus, for a given configuration σ, $g_+^{(\ell)}(\sigma)$ is the ratio of the probabilities of seeing the spins of σ at level ℓ below the root r when the spin at r is $(+)$ and when there is no condition on the spin at r, respectively. We define g_- and $g_-^{(\ell)}$ in an analogous way.

The role played by the functions $g_+^{(\ell)}$ and $g_-^{(\ell)}$ is embodied in the following theorem, which says that if these functions are sufficiently tightly concentrated around their common mean value of 1 then the entropy mixing condition EM holds.

Theorem 5.3 *There exists a constant c (depending only on b, β and h) such that, for any $\delta \geq 0$, if*

$$\mu\big[|g_s^{(\ell)} - 1| > \delta\big] \;\leq\; e^{-2/\delta} \tag{17}$$

for $s \in \{+,-\}$, then we have $\mathrm{Ent}\big[\mu_{\widetilde{T}}(f)\big] \leq c\delta\,\mathrm{Ent}(f)$ for any non-negative function f that does not depend on $B_{r,\ell}$; in particular, $\mathrm{EM}(\ell, c\delta)$ holds.

Proof: Fix $\ell < m$ and a non-negative function f that does not depend on the spins inside the block $B_{r,\ell}$. Since $\mathrm{Ent}(f') \leq \mathrm{Var}(f')/\mu(f')$ for every non-negative function f' (see, e.g., [38]) then

$$\begin{aligned}
\mathrm{Ent}\big[\mu_{\widetilde{T}}(f)\big] \;&\leq\; \frac{\mathrm{Var}\big[\mu_{\widetilde{T}}(f)\big]}{\mu\big[\mu_{\widetilde{T}}(f)\big]} \;=\; \frac{1}{\mu(f)}\cdot\Big[p\big[\mu_{\widetilde{T}}^+(f) - \mu(f)\big]^2 + (1-p)\big[\mu_{\widetilde{T}}^-(f) - \mu(f)\big]^2\Big] \\
&=\; \frac{1}{\mu(f)}\cdot\Big[p\,\mathrm{Cov}\big(g_+, f\big)^2 + (1-p)\,\mathrm{Cov}\big(g_-, f\big)^2\Big] \;\leq\; \max_{s\in\{+,-\}} \frac{\mathrm{Cov}\big(g_s, f\big)^2}{\mu(f)},
\end{aligned} \tag{18}$$

where Cov denotes covariance w.r.t μ. Now observe that, since f does not depend on $B_{r,\ell}$, when computing the covariance term in (18) the function g_s can be replaced by $g_s^{(\ell)}$, which depends only on the spins in $\partial B_{r,\ell}$. Thus, if we can show that (17) implies

$$\mathrm{Cov}\big(g_s^{(\ell)}, f\big)^2 \leq c\delta\mu(f)\,\mathrm{Ent}(f) \tag{19}$$

for some constant c, then by plugging (19) into (18) we will get that $\mathrm{Ent}\big[\mu_{\widetilde{T}}(f)\big] \leq c\delta\,\mathrm{Ent}(f)$, as required.

To establish (19) we make use of the following technical lemma, whose proof can be found in the Appendix.

Lemma 5.4 *Let $\{\Omega, \mathcal{F}, \nu\}$ be a probability space and let f_1 be a mean-zero random variable such that $\|f_1\|_\infty \leq 1$ and $\nu[\,|f_1| > \delta\,] \leq e^{-2/\delta}$ for some $\delta \in (0,1)$. Let f_2 be a probability density w.r.t. ν, i.e. $f_2 \geq 0$ and $\nu(f_2) = 1$. Then there exists a numerical constant $c' > 0$ independent of ν, f_1, f_2 and δ, such that $\nu(f_1 f_2)^2 \leq c'\,\delta\,\mathrm{Ent}_\nu(f_2)$.*

We apply this lemma with $\nu = \mu$ and

$$f_1 = \frac{\big(g_s^{(\ell)} - 1\big)}{\|g_s^{(\ell)}\|_\infty}; \quad f_2 = \frac{f}{\mu(f)},$$

to deduce $\mathrm{Cov}\big(g_s^{(\ell)}, f\big)^2 \;\leq\; c'\delta\|g_s^{(\ell)}\|_\infty^2\mu(f)\,\mathrm{Ent}(f)$. Noting also that $\|g_s^{(\ell)}\|_\infty \leq \|g_s\|_\infty \leq 1/p_{\min}$, where $p_{\min}$ was defined just before Theorem 3.4, this establishes (19) with $c = c'/p_{\min}^2$ and thus completes the proof of the theorem. $\square$

5.2 Proof of Theorem 5.1

In light of Theorem 5.3, to prove Theorem 5.1 it is sufficient to verify the strong concentration property (17) of the functions $g_s^{(\ell)}$ with $\delta = (\gamma\alpha)^{\ell/5}$.

In order to do this we appeal to a strong concentration of the Hamming distance under the coupling ν of $\mu^+_{\widetilde{T}}$ and $\mu^-_{\widetilde{T}}$, as defined in the proof of Claim 4.4. Recall the notation used in that claim, and notice that the Hamming distance is dominated by the size of the population in the ℓth generation of a specific branching process. The following tail bound can be obtained using standard techniques from the analysis of branching processes, and we defer the proof to the end of this section.

Lemma 5.5 *Let $\alpha = \max\{\kappa b, 1\}$. Then for every $C > 0$,*

$$\Pr_\nu\left[|\sigma - \sigma'|_{r,\ell} > C\alpha^\ell\right] \;\leq\; e^{\frac{1}{\ell+1}\left(1-\frac{C}{2e}\right)}.$$

Corollary 5.6 *For every $C > 0$ and $s \in \{+,-\}$,*

$$\Pr_\nu\left[\left|g_s^{(\ell)}(\sigma) - g_s^{(\ell)}(\sigma')\right| > C(\gamma\alpha)^\ell\right] \;\leq\; e^{\frac{1}{\ell+1}\left(1-\frac{p_{\min}C}{2e}\right)}.$$

Proof: It is enough to show that

$$|g_s^{(\ell)}(\sigma) - g_s^{(\ell)}(\sigma')| \leq \frac{\gamma^\ell}{p_{\min}} \cdot |\sigma - \sigma'|_{r,\ell} \tag{20}$$

since we can then apply Lemma 5.5 with C replaced by $p_{\min}C$. On the other hand, (20) follows from part (ii) of Claim 4.4 once we recall that $g_s^{(\ell)}(\sigma) = \mu^\sigma_{B_{r,\ell}}(g_s)$ and that g_s depends only on the spin at the root, implying that $|g_s^{(\ell)}(\sigma) - g_s^{(\ell)}(\sigma')| \leq \|\mu^\sigma_{B_{r,\ell}} - \mu^{\sigma'}_{B_{r,\ell}}\|_r \cdot \|g_s\|_\infty \leq \gamma^\ell\,|\sigma-\sigma'|_{r,\ell}\,/p_{\min}$.
□

Before we go on with the proof of Theorem 5.1, let us compare the way we used the constants κ and γ in the proof of Corollary 5.6 to the way we used them in the proof of Theorem 4.3. In both cases we used κ and γ to get bounds for coupling "down" and "up" the tree respectively. Specifically, we used κ to deduce that the Hamming distance between the coupled configurations at the ℓth level is about $(\kappa b)^\ell$, and we then used γ to bound the effect of each discrepancy at the ℓth level on the spin at the root (or equivalently, on $g_s^{(\ell)}$) by roughly γ^ℓ. While in Theorem 4.3 it was enough that the *average* Hamming distance when coupling down the tree was bounded by $(\kappa b)^\ell$, here we need that this distance is not much larger than $(\kappa b)^\ell$ *with high probability*.

We now return to the proof of Theorem 5.1. W.l.o.g. we may assume that $\gamma\alpha \leq 1$ since $\mathrm{EM}(\ell,1)$ always holds, and also that $\gamma\alpha > 0$ since if $\gamma = 0$ then $\mathrm{EM}(\ell,0)$ holds because then the spin at the root r is independent of the rest of the configuration. Let $a = (\gamma\alpha)^{-1} \geq 1$. Recall that we wish to establish (17) with $\delta = a^{-\ell/5}$ for all large enough ℓ. We will show only that

$$\mu\left[g_s^{(\ell)} - 1 > \delta\right] \leq \frac{1}{2}e^{-2/\delta} \tag{21}$$

since the same bound on the negative tail can be achieved by an analogous argument.

We start by applying Corollary 5.6 with $C = a^{\ell/4}$ to get that, for every $\varepsilon > 0$,

$$\mu^s_{\widetilde{T}}\left[g_s^{(\ell)} - 1 > \varepsilon\right] \;\leq\; \mu\left[g_s^{(\ell)} - 1 > \varepsilon - a^{-3\ell/4}\right] + A, \tag{22}$$

where $A = e^{\frac{1}{\ell+1}\left(1-\frac{p_{\min}a^{\ell/4}}{2e}\right)}$ and we have used the fact that μ is a convex combination of $\mu^+_{\widetilde{T}}$ and $\mu^-_{\widetilde{T}}$.

Next, we notice that by definition of $g_s^{(\ell)}$,

$$\mu^s_{\widetilde{T}}\left[g_s^{(\ell)} - 1 > \varepsilon\right] \;\geq\; (1+\varepsilon)\mu\left[g_s^{(\ell)} - 1 > \varepsilon\right]. \tag{23}$$

Combining (22) and (23) we get that, for every $\varepsilon > 0$,

$$\mu\left[g_s^{(\ell)} - 1 > \varepsilon\right] \;\leq\; \left(\frac{1}{1+\varepsilon}\right)\left(\mu\left[g_s^{(\ell)} - 1 > \varepsilon - a^{-3\ell/4}\right] + A\right). \tag{24}$$

This immediately yields that, for every non-negative integer k and $\varepsilon > 0$,

$$\mu\left[g_s^{(\ell)} - 1 > \varepsilon + ka^{-3\ell/4}\right] \;\leq\; (1+\varepsilon)^{-(k+1)} + A\left(\frac{1+\varepsilon}{\varepsilon}\right), \tag{25}$$

where we applied (24) $k+1$ times, each time increasing ε by $a^{-3\ell/4}$.

Inequality (21) then follows (assuming ℓ is large enough) by applying (25) with $\varepsilon = a^{-\ell/4}$ and $k = \lceil a^{\ell/2} \rceil$. This concludes the proof of Theorem 5.1.

Finally, we supply the missing proof of Lemma 5.5.

Proof of Lemma 5.5: First notice that, by an exponential Markov inequality, it is enough to show that $\mathrm{E}_\nu\left[e^{t|\sigma-\sigma'|_{r,\ell}}\right] \leq e^{2et\alpha^\ell}$ for all $t \leq (2e(\ell+1)\alpha^\ell)^{-1} \leq 1$. We thus fix t as above and let $D_{x,i} = \mathrm{E}_\nu\left[e^{t|\sigma-\sigma'|_{x,i}}\right]$, where ν is the coupling of $\mu^+_{\widetilde{T}_x}$ and $\mu^-_{\widetilde{T}_x}$. Note that $D_{x,i}$ can be calculated recursively as follows. The main observation is that, given a disagreement at x, the random variable $|\sigma-\sigma'|_{x,i}$ is the sum of the b *independent* random variables $|\sigma-\sigma'|_{z,i-1}$ where z ranges over the children of x. In turn, the random variable $e^{t|\sigma-\sigma'|_{z,i-1}}$ takes the value $D_{z,i-1}$ with probability at most κ (the probability of a disagreement at z given a disagreement at x) and the value 1 with the remaining probability (since $|\sigma-\sigma'|_{z,i-1} = 0$ if there is no disagreement at z). Thus, if we let $\delta_i = \max_x D_{x,i} - 1$, then $\delta_{i+1} \leq [1+\kappa\delta_i]^b - 1 \leq e^{\kappa b\delta_i} - 1 \leq e^{\alpha\delta_i} - 1$. We wish to show that, for t in the above range, $\delta_\ell \leq 2et\alpha^\ell$, which implies $\mathrm{E}_\nu\left[e^{t|\sigma-\sigma'|_{r,\ell}}\right] \leq \delta_\ell + 1 \leq e^{2et\alpha^\ell}$, as required. In fact, we show by induction that $\delta_i \leq 2t[\frac{\ell+1}{\ell}\cdot\alpha]^i$ for every $0 \leq i \leq \ell$. For the base case $i = 0$, notice that $|\sigma-\sigma'|_{x,0} = 1$ when starting from a fixed disagreement at x, so $\delta_0 = e^t - 1 \leq 2t$ for t in the given range. For $i+1 > 0$, we use the fact that $\delta_{i+1} \leq e^{\alpha\delta_i} - 1 \leq \frac{\alpha\delta_i}{1-\alpha\delta_i} \leq \frac{\ell+1}{\ell}\cdot\alpha\delta_i$, since by the induction hypothesis $\delta_i \leq \frac{1}{\alpha(\ell+1)}$ for all $0 \leq i \leq \ell-1$ and t in the given range. □

5.3 A crude bound on log-Sobolev via the spectral gap

In this section we state and prove a general bound on c_{sob} using a bound on c_{gap}. Although we do not require this bound for the results in this paper, we believe that it may find applications in the future. We state the bound for the Ising model, but it can be easily verified that it generalizes to any nearest-neighbor spin system on a tree.

Theorem 5.7 *For the Ising model on the b–ary tree, $c_{\mathrm{sob}}(\mu) = c_{\mathrm{gap}}(\mu) \times \Omega(1/\log n)$. In particular, if $c_{\mathrm{gap}}(\mu) = \Omega(1)$ then $c_{\mathrm{sob}}(\mu) = \Omega(1/\log n)$.*

An important consequence of this theorem is that, if one can show that $c_{\mathrm{gap}} = \Omega(1)$ (which may be easier than analyzing the more delicate c_{sob}), then one can immediately deduce that the mixing time is $O(n(\log n)^2)$, which is dramatically better than the $O(n^2)$ obtained from Theorem 2.1. (Of course, Theorem 2.1 is much more general and applies to spin systems on any graph.) In similar vein, we can compare Theorem 5.7 with the general bound $c_{\mathrm{sob}}(\mu) = c_{\mathrm{gap}}(\mu) \times \Omega(1/n)$ (see, e.g.,[38]).

Theorem 5.7 is a consequence of the following lemma.

Lemma 5.8 *For any β and h, there exists a constant $c = c(b,\beta,h)$ such that, for any $x \in T$ and all ℓ,*

$$c_{\rm sob}(\mu^{\tau}_{T_x})^{-1} \le \max_{y \prec x, \eta \in \Omega^{\tau}_{T}} \{c_{\rm sob}(\mu^{\eta}_{T_y})^{-1}\} + c \cdot c_{\rm gap}(\mu^{\tau}_{T_x})^{-1}\,. \tag{26}$$

This lemma immediately implies Theorem 5.7, once we notice that $c_{\rm gap}(\mu^{\eta}_{T_x}) \ge c' \cdot c_{\rm gap}(\mu^{\tau}_{T})$ for a constant $c' = c'(b,\beta,h)$ and every $x \in T$ and $\eta \in \Omega^{\tau}_{T}$, as can easily be checked.

Proof of Lemma 5.8: For simplicity and without loss of generality we will prove the recursive inequality (26) only for $T_x = T$, with root r. Let f be a non–negative function. We then write (using the entropy version of (3))

$$\mathrm{Ent}(f) = \mu\left[\mathrm{Ent}_{\widetilde{T}}(f)\right] + \mathrm{Ent}\left[\mu_{\widetilde{T}}(f)\right]. \tag{27}$$

Using the definition of $c_{\rm sob}$ we have

$$\begin{aligned}\mu\left[\mathrm{Ent}_{\widetilde{T}}(f)\right] &\le \max_{y \prec r, \eta \in \Omega^{\tau}_{T}} \{c_{\rm sob}(\mu^{\eta}_{T_y})^{-1}\} \sum_{x \in \widetilde{T}} \mu\left[\mathrm{Var}_{\{x\}}(\sqrt{f})\right] \\ &\le \max_{y \prec r, \eta \in \Omega^{\tau}_{T}} \{c_{\rm sob}(\mu^{\eta}_{T_y})^{-1}\}\, \mathcal{D}(\sqrt{f})\,. \end{aligned} \tag{28}$$

The second term on the r.h.s. of (27), being the entropy of a Bernoulli random variable, is bounded above by

$$\begin{aligned}\mathrm{Ent}\left[\mu_{\widetilde{T}}(f)\right] &\le \alpha \mathrm{Var}\Big(\sqrt{\mu_{\widetilde{T}}(f)}\Big) && (29)\\ &\le \alpha \mathrm{Var}(\sqrt{f}) \\ &\le \alpha\, c_{\rm gap}(\mu)^{-1}\, \mathcal{D}\big(\sqrt{f}\big), && (30)\end{aligned}$$

where $\alpha \equiv \alpha(p)$ is a constant that depends on $p = \mu(\sigma_r = +)$; specifically $\alpha(p) = \frac{\log(p/1-p)}{2p-1}$ for $p \neq 1/2$, and $\alpha(1/2) = 1/2$ (see [38]).

Putting together (28) and (30), the expression in (27) is bounded above by

$$\Big[\max_{y \prec r, \eta \in \Omega^{\tau}_{T}} \{c_{\rm sob}(\mu^{\eta}_{T_y})^{-1}\} \,+\, \alpha\, c_{\rm gap}(\mu)^{-1}\Big] \mathcal{D}\big(\sqrt{f}\big),$$

so that from the definition of $c_{\rm sob}$ we have

$$c_{\rm sob}(\mu)^{-1} \le \max_{y \prec r, \eta \in \Omega^{\tau}_{T}} \{c_{\rm sob}(\mu^{\eta}_{T_y})^{-1}\} + \alpha\, c_{\rm gap}(\mu)^{-1}. \qquad \square$$

6 Extensions to other models

As we have already indicated, our techniques extend beyond the Ising model to general nearest-neighbor interaction models on trees, including those with hard constraints. In this final section we mention some of these extensions. For a fuller treatment of this material, the reader is referred to the companion paper [32].

A (*nearest neighbor*) *spin system* on a finite graph $G = (V,E)$ is specified by a finite set S of spin values, a symmetric pair potential $U : S \times S \to \mathbb{R} \cup \{\infty\}$, and a singleton potential $W : S \to \mathbb{R}$. A configuration $\sigma \in S^V$ of the system assigns to each vertex (site) $v \in V$ a spin value $\sigma_v \in S$. The Gibbs distribution is given by

$$\mu(\sigma) \propto \exp\Big[-\big(\textstyle\sum_{xy \in E} U(\sigma_x,\sigma_y) + \sum_{x \in V} W(\sigma_x)\big)\Big].$$

Thus the Ising model corresponds to the case $S = \{\pm 1\}$, and $U(s_1, s_2) = -\beta s_1 s_2$, $W(s) = -\beta hs$, where β is the inverse temperature and h is the external field. Note that setting $U(s_1, s_2) = \infty$ corresponds to a *hard constraint*, i.e., spin values s_1, s_2 are forbidden to be adjacent. We denote by Ω the set of all *valid* spin configurations, i.e., those for which $\mu(\sigma) > 0$.

As for the Ising model, we allow boundary conditions which fix the spin values of certain sites. We carry over our notation from the Ising model: thus, e.g., μ_A^τ denotes the Gibbs distribution on a subset $A \subseteq V$ with boundary condition τ on ∂A.

The (heat-bath) Glauber dynamics extends in the obvious way to general spin systems: a site is picked uniformly at random, and its spin value chosen from the conditional distribution determined by the spins of its neighbors. As before, the mixing time can be bounded in terms of the spectral gap c_{gap} and the log-Sobolev constant c_{sob}.

We first note that, as the reader may easily check, neither the spatial mixing conditions in Section 3 nor their proofs made any reference to the details of the Ising model. All of this material therefore carries over without modification to general spin systems on trees.

Theorem 6.1 *The statements of theorems 3.2 and 3.4 hold for general nearest-neighbor spin systems on trees.*

Likewise, the machinery developed in Sections 4 and 5 for verifying the conditions VM and EM also extends to general models, though the details of the calculations are model-specific. In particular, Theorems 4.3 and 5.1 relating VM and EM to the coupling quantities κ and γ of Definition 4.2 still hold (with minor modifications). Thus all we need to do is to carry out the detailed calculations of κ and γ for the model under consideration. We now state without proof the results of these calculations for several models of interest. For the proofs, together with further discussion and extensions, the reader is referred to the companion paper [32].

6.1 The hard-core model (independent sets)

In this model $S = \{0, 1\}$, and we refer to a site as *occupied* if it has spin value 1, and *unoccupied* otherwise. The potentials are

$$U(1,1) = \infty; \quad U(1,0) = U(0,0) = 1; \quad W(1) = L; \quad W(0) = 0,$$

where $L \in \mathbb{R}$. The hard constraint here means that no two adjacent sites may be occupied, so Ω can be identified with the set of all *independent sets* in G. Also, the aggregated potential of a valid configuration is proportional to the number of occupied sites. Hence the Gibbs distribution takes the simple form

$$\mu(\sigma) \propto \lambda^{N(\sigma)},$$

where $N(\sigma)$ is the number of occupied sites and the parameter $\lambda = \exp(-L) \geq 0$, which controls the density of occupation, is referred to as the "activity."

The hard-core model on a b-ary tree undergoes a phase transition at a critical activity $\lambda = \lambda_0 = \frac{b^b}{(b-1)^{b+1}}$ (see, e.g., [41, 25]). For $\lambda \leq \lambda_0$ there is a unique Gibbs measure regardless of the boundary condition on the leaves, while for $\lambda > \lambda_0$ there are (at least) two distinct phases, corresponding to the "odd" and "even" boundary conditions respectively. The even boundary condition is obtained by making the leaves of the tree all occupied if the depth is even, and all unoccupied otherwise. The odd boundary condition is the complement of this. (These boundary conditions are derived from the two maximum-density configurations on the infinite tree $\mathbb{T}^b$ in which alternate levels — either odd or even — are completely occupied.) For $\lambda > \lambda_0$, the probability of occupation of the root in the infinite-volume Gibbs measure differs for odd and even boundary conditions. Relatively little is known about the Glauber dynamics for the hard-core model on trees, beyond the general result of Luby and Vigoda [26, 44] which ensures a mixing time of $O(n \log n)$ when $\lambda < \frac{2}{b-1}$. This result actually holds for any graph G of maximum degree $b+1$.

Our results for the Glauber dynamics in the hard-core model mirror those given earlier for the Ising model. First, for sufficiently small activity λ we show that both $c_{\rm gap}$ and $c_{\rm sob}$ are uniformly bounded away from zero, and hence that the mixing time is $O(n \log n)$, for *arbitrary* boundary conditions. Second, for the *even* (or, symmetrically, odd) boundary condition, we get the same result for *all* activities λ.

Theorem 6.2 *For the hard-core model on the n-vertex b-ary tree with boundary condition τ, the mixing time is $O(n \log n)$ in both of the following situations:*

(i) τ is arbitrary, and $\lambda \le \max\left\{\frac{1}{\sqrt{b}-1}, \lambda_0\right\}$;

(ii) τ is even (or odd), and $\lambda \ge 0$ is arbitrary.

Part (ii) of this theorem is analogous to our earlier result that the mixing time for the Ising model with $(+)$-boundary is $O(n \log n)$ at all temperatures. This is in line with the intuition that the even boundary eliminates the only bottleneck in the dynamics. Part (i) identifies a region in which the mixing time is insensitive to the boundary condition. We would expect this to hold throughout the low-activity region $\lambda \le \lambda_0$, and indeed, by analogy with the Ising model, also in some intermediate region beyond this. Our bound in part (i) confirms this behavior: note that the quantity $\frac{1}{\sqrt{b}-1}$ exceeds λ_0 for all $b \ge 5$, and indeed for large b it grows as $\frac{1}{\sqrt{b}}$ compared to the $\frac{1}{b}$ growth of λ_0. Thus for $b \ge 5$ we establish rapid mixing in a region above the critical value λ_0. To the best of our knowledge this is the first such result. (Note that the result of [26, 44] mentioned earlier establishes rapid mixing for $\lambda < \frac{2}{b-1}$, which is less than λ_0 for all b and so does not even cover the whole uniqueness region.) We should also mention that our coupling analysis of $c_{\rm gap}$ in this region has consequences for the infinite volume Gibbs measure itself, implying that when $\lambda \le \frac{1}{\sqrt{b}-1}$ any $\mu = \lim_{T\to\infty} \mu_T^\tau$ that is the limit of finite Gibbs distributions for some boundary configuration τ is *extremal*, again a new result. We elaborate on these points in the companion paper [32].

6.2 The antiferromagnetic Potts model (colorings)

In this model $S = \{1, 2, \ldots, q\}$, and the potentials are $U(s_1, s_2) = \beta\delta_{s_1,s_2}$, $W(s) = 0$, where β is the inverse temperature. This is the analog of the Ising model except that the interactions are *antiferromagnetic*, i.e., neighbors with unequal spins are favored. The most interesting case of this model is when $\beta = \infty$ (i.e., zero temperature), which introduces hard constraints. Thus if we think of the q spin values as colors, Ω is the set of *proper colorings* of G, i.e., assignments of colors to vertices so that no two adjacent vertices receive the same color. The Gibbs distribution is uniform over proper colorings. In this model it is q that provides the parameterization. For background on the model, see [9].

For colorings on the b-ary tree it is well known that, when $q \le b+1$, there are multiple Gibbs measures; this follows immediately from the existence of "frozen configurations," i.e., colorings in which the color of every internal vertex is forced by the colors of the leaves (see, e.g., [9]). Recently Jonasson [23] proved that, as soon as $q \ge b+2$, the Gibbs measure is unique. Moreover, it is known that there is again an "intermediate" region that includes the value $q = b+1$, in which the Gibbs measure, while not unique, is insensitive to "typical" boundary conditions (chosen from the free measure); see [9].

The sharpest result known for the Glauber dynamics on colorings is due to Vigoda [43], who shows that for arbitrary boundary conditions the mixing time is $O(n \log n)$ provided $q > \frac{11}{6}(b+1)$.

Actually this result holds for *any* n-vertex graph G of maximum degree $b+1$.[‖] Our techniques extend this rapid mixing result all the way down to the critical value $q \geq b+2$, with arbitrary boundary conditions.

Theorem 6.3 *The mixing time of the Glauber dynamics for the colorings model on the n-vertex b-ary tree with $q \geq b+2$ and arbitrary boundary conditions is $O(n \log n)$.*

Again, this theorem follows from a uniform constant lower bound on the log-Sobolev constant c_{sob}.

6.3 The ferromagnetic Potts model

Here we have $S = \{1, 2, \ldots, q\}$ and potentials $U(s_1, s_2) = -\beta\delta_{s_1,s_2}$, $W(s) = 0$. This is a straightforward generalization of the (ferromagnetic) Ising model studied earlier in the paper, in which the spin at each site can take one of q possible values, and the aggregated potential of any configuration depends on the number of adjacent pairs of equal spins. There are no hard constraints.

Qualitatively the behavior of this model is similar to that of the Ising model, though less is known in precise quantitative terms. Again there is a phase transition at a critical $\beta = \beta_0$, which depends on b and q, so that for $\beta > \beta_0$ (and indeed for $\beta \geq \beta_0$ when $q > 2$) there are multiple phases. This value β_0 does not in general have a closed form, but it is known [16] that $\beta_0 < \frac{1}{2}\ln(\frac{b+q-1}{b-1})$ for all $q > 2$. (For $q = 2$, this value is exactly β_0 for the Ising model as quoted earlier.)

Using our techniques, we are able to prove the following:

Theorem 6.4 *The mixing time of the Glauber dynamics for the Potts model on an n-vertex b-ary tree is $O(n \log n)$ in all the following situations:*

(i) *the boundary condition is arbitrary and* $\beta < \max\left\{\beta_0, \frac{1}{2}\ln(\frac{\sqrt{b}+1}{\sqrt{b}-1})\right\}$;

(ii) *the boundary condition is constant (e.g., all sites on the boundary have spin* 1*) and β is arbitrary*;

(iii) *the boundary condition is free (i.e., the boundary spins are unconstrained) and $\beta < \beta_1$, where β_1 is the solution to the equation* $\frac{e^{2\beta_1}-1}{e^{2\beta_1}+q-1} \cdot \frac{e^{2\beta_1}-1}{e^{2\beta_1}+1} = \frac{1}{b}$.

Part (i) of this theorem shows that we get rapid mixing for arbitrary boundaries throughout the uniqueness region; also, since $\frac{1}{2}\ln(\frac{\sqrt{b}+1}{\sqrt{b}-1}) \geq \frac{1}{2}\ln(\frac{b+q-1}{b-1}) > \beta_0$ when $q \leq 2(\sqrt{b}+1)$, this result extends into the multiple phase region for many combinations of b and q. Part (ii) of the theorem is an analog of our earlier result that the Ising model with $(+)$-boundaries has optimal mixing time at all temperatures. Part (iii) is of interest for two reasons. First, since $\beta_1 > \beta_0$ always, it exhibits a natural boundary condition under which rapid mixing holds beyond the uniqueness region (but not for arbitrary β) *for all* combinations of b and q. Second, because of an intimate connection between the free boundary case and so-called "reconstruction problems" on trees [34] (in which the edges are noisy channels and the goal is to reconstruct a value transmitted from the root), we obtain an alternative proof of the best known value of the noise parameter under which reconstruction is impossible [35]. Indeed, a slight strengthening of part (iii) allows us to marginally improve on this threshold. Again, we spell out the details in [32].

[‖]A recent sequence of papers [11, 33, 17] have reduced the required number of colors further for general graphs, under the assumption that the maximum degree is $\Omega(\log n)$; the current state of the art requires $q \geq (1+\epsilon)(b+1)$ for arbitrarily small $\epsilon > 0$ [18]. However, these results do not apply in our setting where the degree $b+1$ is fixed.

7 Proofs omitted from the main text

7.1 Proofs for Section 2.1

Proof of Eqn. (3):

$$\begin{aligned}\mu_A^\eta[\mathrm{Var}_B(f)] + \mathrm{Var}_A^\eta[\mu_B(f)] &= \mu_A^\eta[\mu_B(f^2) - \mu_B(f)^2] + \mu_A^\eta[\mu_B(f)^2] - [\mu_A^\eta\mu_B(f)]^2 \\ &= \mu_A^\eta[\mu_B(f^2)] - [\mu_A^\eta\mu_B(f)]^2 \\ &= \mu_A^\eta(f^2) - \mu_A^\eta(f)^2 = \mathrm{Var}_A^\eta(f). \qquad \square\end{aligned}$$

Proof of Eqn. (5): First, for any function g and boundary configuration η we have

$$\mathrm{Var}_A^\eta[\mu_{B\setminus A}(g)] \le \mu_{B\setminus A}^\eta[\mathrm{Var}_A(g)], \tag{31}$$

as we now explain. Recall that by assumption $(\partial A) \cap B = \emptyset$, so there are no edges connecting the two disjoint subsets A and $B \setminus A$. Therefore, the distribution $\mu_{B\setminus A}$ does not depend on the configuration in A, and μ_A does not depend on the configuration in $B \setminus A$. Thus, if we fix the configuration outside A to be η, then $\mu_{B\setminus A}(g)$ (a function only of the spins in A) can be written as

$$\mu_{B\setminus A}(g) = \sum_\sigma \mu_{B\setminus A}^\eta(\sigma) g_\sigma,$$

where g_σ is g with the spins on $B \setminus A$ fixed to σ. Note that this is a convex combination of functions g_σ. Therefore, we may use the fact that variance w.r.t. a fixed measure is a convex functional, together with the fact that the measure μ_A does not depend on the configuration on $B \setminus A$, to deduce

$$\mathrm{Var}_A^\eta[\mu_{B\setminus A}(g)] \le \sum_\sigma \mu_{B\setminus A}^\eta(\sigma)\mathrm{Var}_A^\eta(g_\sigma) = \mu_{B\setminus A}^\eta[\mathrm{Var}_A(g)],$$

thus verifying (31).

Finally, equation (5) follows from (31) with $g = \mu_{B\cap A}(f)$ by writing

$$\mu[\mathrm{Var}_A(\mu_B(f))] = \mu[\mathrm{Var}_A(\mu_{B\setminus A}\mu_{B\cap A}(f))] \le \mu\mu_{B\setminus A}[\mathrm{Var}_A(\mu_{B\cap A}(f))] = \mu[\mathrm{Var}_A(\mu_{B\cap A}(f))]. \qquad \square$$

7.2 Proof of Lemma 3.5

The lemma in fact holds in a more general setting, where in place of $\widetilde{T_x}$ and $B_{x,\ell}$ we think of two arbitrary subsets A, B such that $A \cup B = T_x$. Also, in this proof we write $\nu = \mu_{T_x}^\eta$ and Var and Ent for variance and entropy with respect to ν. For part (i) we will show that if for any function g that does not depend on B we have $\mathrm{Var}[\nu_A(g)] \le \varepsilon \cdot \mathrm{Var}(g)$, then for any function f,

$$\mathrm{Var}[\nu_A(f)] \le \frac{2(1-\varepsilon)}{1-2\varepsilon} \cdot \nu[\mathrm{Var}_B(f)] + \frac{2\varepsilon}{1-2\varepsilon} \cdot \nu[\mathrm{Var}_A(f)].$$

Notice that by the convexity of variance we have $\mathrm{Var}(g_1 + g_2) \le 2[\mathrm{Var}(g_1) + \mathrm{Var}(g_2)]$ for any two functions g_1, g_2. We therefore write

$$\begin{aligned}\mathrm{Var}[\nu_A(f)] &= \mathrm{Var}[\nu_A(f) - \nu_A(\nu_B(f)) + \nu_A(\nu_B(f)] \\ &\le 2\mathrm{Var}[\nu_A(f - \nu_B(f))] + 2\mathrm{Var}[\nu_A(\nu_B(f))] \\ &\le 2\mathrm{Var}[f - \nu_B(f)] + 2\varepsilon\mathrm{Var}[\nu_B(f)] \\ &= 2\nu[\mathrm{Var}_B(f)] + 2\varepsilon(\mathrm{Var}[\nu_A(f)] + \nu[\mathrm{Var}_A(f)] - \nu[\mathrm{Var}_B(f)]),\end{aligned}$$

where we used the facts that $\mathrm{Var}[f - \nu_B(f)] = \nu[\mathrm{Var}_B(f)]$ and that $\mathrm{Var}[\nu_A(f)] + \nu[\mathrm{Var}_A(f)] = \mathrm{Var}[\nu_B(f)] + \nu[\mathrm{Var}_B(f)] = \mathrm{Var}(f)$ as in (3). We therefore conclude that $\mathrm{Var}[\nu_A(f)] \leq \frac{2(1-\varepsilon)}{1-2\varepsilon} \cdot \nu[\mathrm{Var}_B(f)] + \frac{2\varepsilon}{1-2\varepsilon} \cdot \nu[\mathrm{Var}_A(f)]$, as required.

We proceed to part (ii). Here we have to show that if for any non-negative function g that does not depend B we have $\mathrm{Ent}[\nu_A(g)] \leq \varepsilon \cdot \mathrm{Ent}(g)$, then for any non-negative function f,

$$\mathrm{Ent}[\nu_A(f)] \leq \frac{1}{1-\varepsilon'} \cdot \nu[\mathrm{Ent}_B(f)] + \frac{\varepsilon'}{1-\varepsilon'} \cdot \nu[\mathrm{Ent}_A(f)], \tag{32}$$

where $\varepsilon' = \sqrt{\varepsilon}/p$ and p stands for the minimum non-zero probability of any configuration in $T_x \setminus A$. We will in fact show that

$$\mathrm{Ent}(f) \leq \frac{1}{1-\varepsilon'}(\nu[\mathrm{Ent}_A(f)] + \nu[\mathrm{Ent}_B(f)]), \tag{33}$$

which implies (32) since $\mathrm{Ent}[\nu_A(f)] = \mathrm{Ent}(f) - \nu[\mathrm{Ent}_A(f)]$.

Before we go on with the proof, let us review some properties of entropy. First, by definition, $\mathrm{Ent}(f) = \nu(f \log \frac{f}{\nu(f)})$ and $\nu[\mathrm{Ent}_A(f)] = \nu(f \log \frac{f}{\nu_A(f)})$. Also, by the variational characterization of entropy we have $\nu_A(f \log \frac{g}{\nu_A(g)}) \leq \mathrm{Ent}_A(f)$ for all non-negative functions f and g.

We can now proceed with the proof of (33) by writing

$$\begin{aligned}
\mathrm{Ent}(f) &= \nu\left[f \log \frac{f}{\nu_B(f)}\right] + \nu\left[f \log \frac{\nu_B(f)}{\nu_A(\nu_B(f))}\right] + \nu\left[\frac{f \log \nu_A(\nu_B(f))}{\nu(f)}\right] \\
&\leq \nu\left[f \log \frac{f}{\nu_B(f)}\right] + \nu\left[f \log \frac{f}{\nu_A(f)}\right] + \nu\left[f \log \frac{\nu_A(\nu_B(f))}{\nu(f)}\right] \\
&= \nu[\mathrm{Ent}_B(f)] + \nu[\mathrm{Ent}_A(f)] + \nu\left[\nu_A(f) \log \frac{\nu_A(\nu_B(f))}{\nu(f)}\right] .
\end{aligned}$$

Therefore, (33) will follow once we show that $\nu\left[\nu_A(f) \log \frac{\nu_A(\nu_B(f))}{\nu(f)}\right] \leq \varepsilon' \,\mathrm{Ent}(f)$. We use the following claim in order to get this bound.

Claim 7.1 *Let μ be a probability measure over a space Ω where the probability of any $\sigma \in \Omega$ is either zero or at least p. Then for any two non-negative functions f and g over Ω we have*

$$\mu\left[f \log \frac{g}{\mu g}\right] \leq \frac{1}{p}\sqrt{\frac{\mu(f)}{\mu(g)} \cdot \mathrm{Ent}(f) \cdot \mathrm{Ent}(g)},$$

where Ent *is taken w.r.t. to μ.*

Assuming Claim 7.1, we conclude that

$$\nu\left[\nu_A(f) \log \frac{\nu_A(\nu_B(f))}{\nu(f)}\right] \leq \frac{1}{p}\sqrt{\mathrm{Ent}[\nu_A(f)] \cdot \mathrm{Ent}[\nu_A(\nu_B(f))]} \;\leq$$

$$\frac{1}{p}\sqrt{\varepsilon \cdot \mathrm{Ent}[\nu_A(f)] \cdot \mathrm{Ent}[\nu_B(f)]} \;\leq\; \frac{1}{p}\sqrt{\varepsilon}\,\mathrm{Ent}(f),$$

completing the proof of Lemma 3.5. We note that, since neither $\nu_A(f)$ nor $\nu_A(\nu_B(f))$ depends on A, the effective probability space in the above derivation is the marginal over $T_x \setminus A$, so indeed p can be taken as the minimum marginal probability of configurations restricted to $T_x \setminus A$.

It remains to prove claim 7.1. Consider two arbitrary non-negative functions f and g. Let χ be the indicator function of the event that $g \geq \mu(g)$. Clearly, $\chi \log \frac{g}{\mu(g)} \geq 0$ while $(1-\chi) \log \frac{g}{\mu(g)} \leq 0$. Also, since $\mu\left[\log \frac{g}{\mu(g)}\right] \leq \log \mu\left[\frac{g}{\mu(g)}\right] = 0$ then $\mu\left[(1-\chi) \log \frac{g}{\mu(g)}\right] \leq -\mu\left[\chi \log \frac{g}{\mu(g)}\right]$. Letting $f_{\max}$

and $f_{\min}$ be the maximum and minimum values of f respectively over configurations with non-zero probability, we get:

$$\begin{aligned}
\mu\left[f\log\frac{g}{\mu(g)}\right] &= \mu\left[\chi f\log\frac{g}{\mu(g)}\right]+\mu\left[(1-\chi)f\log\frac{g}{\nu(g)}\right]\\
&\le f_{\max}\cdot\mu\left[\chi\log\frac{g}{\mu(g)}\right]+f_{\min}\cdot\mu\left[(1-\chi)\log\frac{g}{\mu(g)}\right]\\
&\le (f_{\max}-f_{\min})\cdot\mu\left[\chi\log\frac{g}{\mu(g)}\right]\\
&\le \frac{1}{p}\cdot\|f-\mu(f)\|_1\cdot\mu\left[\chi\left(\frac{g}{\mu(g)}-1\right)\right]\\
&= \frac{1}{2p\cdot\mu(g)}\cdot\|f-\mu(f)\|_1\cdot\|g-\mu(g)\|_1\\
&\le \frac{1}{p}\sqrt{\frac{\mu(f)}{\mu(g)}\cdot\mathrm{Ent}(f)\cdot\mathrm{Ent}(g)},
\end{aligned}$$

where we wrote $\|\cdot\|_1$ for the ℓ_1 norm with respect to μ and used the fact that $\|f-\mu(f)\|_1^2 \le 2\mu(f)\,\mathrm{Ent}(f)$ for any non-negative function f (see, e.g., [38]). The proof of Claim 7.1 is now complete. □

7.3 Proof of Lemma 5.4

We split our analysis of $\nu(f_1 f_2)^2$ into three cases:

(a) $\mathrm{Ent}_\nu(f_2)\ge\frac{1}{\delta}$;

(b) $\delta<\mathrm{Ent}_\nu(f_2)<\frac{1}{\delta}$;

(c) $\mathrm{Ent}_\nu(f_2)\le\delta$.

Case (a). We simply bound

$$\nu(f_1f_2)^2\le\|f_1\|_\infty^2\nu(f_2)^2\le 1\le\delta\ \mathrm{Ent}_\nu(f_2)\,.$$

Case (b). We use the *entropy inequality* (see, e.g., [2]), which states that for any $t>0$,

$$\nu(f_1f_2)\le\frac{1}{t}\log\nu(e^{tf_1})+\frac{1}{t}\,\mathrm{Ent}_\nu(f_2)\,. \tag{34}$$

We choose the free parameter t in (34) equal to $\sqrt{\mathrm{Ent}_\nu(f_2)/\delta}$. Notice that, by construction, $1<t<\delta^{-1}$. Using the assumption $\nu(|f_1|>\delta)\le e^{-2/\delta}$ together with $\|f_1\|_\infty\le 1$, we get

$$\begin{aligned}
\nu(f_1f_2)^2 &\le \Big[\frac{1}{t}\log\big(e^{t\delta}+e^{t-2/\delta}\big)+\sqrt{\delta\,\mathrm{Ent}_\nu(f_2)}\,\Big]^2\\
&\le\Big[c_1\,\delta+\sqrt{\delta\,\mathrm{Ent}_\nu(f_2)}\,\Big]^2\le c_2\,\delta\,\mathrm{Ent}_\nu(f_2)
\end{aligned}$$

for suitable numerical constants c_1,c_2.

Case (c). Again we use the entropy inequality with $t=\sqrt{\mathrm{Ent}_\nu(f_2)/\delta}\le 1$, but we now simply bound the Laplace transform $\nu(e^{tf_1})$ by a Taylor expansion (in t) up to second order:

$$\begin{aligned}
\frac{1}{t}\log\nu(e^{tf_1}) &\le \frac{1}{t}\log\Big(1+e\frac{t^2}{2}\nu(f_1^2)\Big)\le e\frac{t}{2}\big[\,\delta^2+e^{-2/\delta}\,\big]\\
&=\frac{1}{2}e\big[\,\delta^2+e^{-2/\delta}\,\big]\sqrt{\mathrm{Ent}_\nu(f_2)/\delta},
\end{aligned}$$

which by (34) implies

$$\nu(f_1 f_2)^2 \le \Big[\frac{e}{2\sqrt{\delta}} \big(\delta^2 + e^{-2/\delta} \big) + \sqrt{\delta} \Big]^2 \operatorname{Ent}_\nu(f_2) \le c_3 \, \delta \operatorname{Ent}_\nu(f_2)$$

for another numerical constant c_3. □

Acknowledgments

F. Martinelli would like to thank the Miller Institute, the Dept. of Statistics and the Dept. of EECS of the University of California at Berkeley for financial support and warm hospitality. We also wish to thank E. Mossel and Y. Peres for very interesting discussions about reconstruction on trees and related topics.

References

[1] D. ALDOUS, "Random walks on finite groups and rapidly mixing Markov chains," *Séminaire de Probabilites XVII*, Lecture Notes in Mathematics **986**, pp. 243–297, Springer, Berlin, 1983.

[2] C. ANÉ, S. BLACHÈRE, D. CHAFAÏ, P. FOUGÈRES, I. GENTIL, F. MALRIEU, C. ROBERTO and G. SCHEFFER, "Sur les inégalités de Sobolev logarithmiques," *Société Mathématique de France*, 2000.

[3] R.J. BAXTER, *Exactly solved models in statistical mechanics*, Academic Press, London, 1982.

[4] N. BERGER, C. KENYON, E. MOSSEL and Y. PERES, "Glauber dynamics on trees and hyperbolic graphs," preprint, 2003. For a short version see C. KENYON, E. MOSSEL and Y. PERES, "Glauber dynamics on trees and hyperbolic graphs," *Proc. 42nd IEEE Symposium on Foundations of Computer Science* (2001), pp. 568–578.

[5] L. BERTINI, N. CANCRINI and F. CESI, "The spectral gap for a Glauber-type dynamics in a continuous gas," *Ann. Inst. H. Poincaré Probab. Statist.* **38** (2002), pp. 91–108.

[6] P. BLEHER, J. RUIZ, R.H. SCHONMANN, S. SHLOSMAN and V. ZAGREBNOV, "Rigidity of the critical phases on a Cayley tree," *Moscow Mathematical Journal* **1** (2001), pp. 345–363.

[7] P. BLEHER, J. RUIZ and V. ZAGREBNOV, "On the purity of the limiting Gibbs state for the Ising model on the Bethe lattice," *Journal of Statistical Physics* **79** (1995), pp. 473–482.

[8] T. BODINEAU and F. MARTINELLI, "Some new results on the kinetic Ising model in a pure phase," *Journal of Statistical Physics* **109** (1), 2002.

[9] G. BRIGHTWELL and P. WINKLER, "Random colorings of a Cayley tree," *Contemporary combinatorics*, Bolyai Society Mathematical Studies **10**, János Bolyai Math. Soc., Budapest, 2002, pp. 247–276.

[10] F. CESI, "Quasi-factorization of the entropy and logarithmic Sobolev inequalities for Gibbs random fields," *Probability Theory and Related Fields* **120** (2001), pp. 569–584.

[11] M. DYER and A. FRIEZE, "Randomly colouring graphs with lower bounds on girth and maximum degree," *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science*, 2001, pp. 579–587.

[12] W. EVANS, C. KENYON, Y. PERES and L.J. SCHULMAN, "Broadcasting on trees and the Ising model," *Annals of Applied Probability* **10** (2000), pp. 410–433.

[13] D. FISHER and D. HUSE, "Dynamics of droplet fluctuations in pure and random Ising systems," *Physics Review B* **35** (13), 1987.

[14] A. FRIEZE and R. KANNAN, “Log-Sobolev inequalities and sampling from log-concave distributions,” *Annals of Applied Probability* **9** (1999), pp. 14–26.

[15] H.-O. GEORGII, *Gibbs measures and phase transitions*, de Gruyter Studies in Mathematics **9**, Walter de Gruyter & Co., Berlin, 1988.

[16] O. HÄGGSTRÖM, “The random-cluster model on a homogeneous tree,” *Probability Theory and Related Fields* **104** (1996), pp. 231–253.

[17] T.P. HAYES, “Randomly coloring graphs with girth five,” *Proceedings of the 35th Annual ACM Symposium on Theory of Computing*, 2003, pp. 269–278.

[18] T.P. HAYES and E. VIGODA, “A non-Markovian coupling for randomly sampling colorings,” to appear in *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, 2003.

[19] D. IOFFE, “A note on the extremality of the disordered state for the Ising model on the Bethe lattice,” *Letters in Mathematical Physics* **37** (1996), pp. 137–143.

[20] D. IOFFE, “Extremality of the disordered state for the Ising model on general trees,” *Progress in Probability* **40** (1996), pp. 3–14.

[21] M. JERRUM and J.-B. SON, “Spectral gap and log-Sobolev constant for balanced matroids,” *Proceedings of the 43rd IEEE Symposium on Foundations of Computer Science*, 2002, pp. 721–729.

[22] M. JERRUM, J.-B. SON, P. TETALI and E. VIGODA, “Elementary bounds on Poincaré and log-Sobolev constants for decomposable Markov chains,” preprint, 2003.

[23] J. JONASSON “Uniqueness of uniform random colorings of regular trees,” *Statistics & Probability Letters* **57** (2002), pp. 243–248.

[24] J. JONASSON and J.E. STEIF, “Amenability and phase transition in the Ising model,” *Journal of Theoretical Probability* **12** (1999), pp. 549–559.

[25] F.P. KELLY, “Stochastic models of computer communication systems,” *Journal of the Royal Statistical Society* B **47** (1985), pp. 379–395.

[26] M. LUBY and E. VIGODA, “Fast convergence of the Glauber dynamics for sampling independent sets,” *Random Structures & Algorithms* **15** (1999), pp. 229–241.

[27] R. LYONS, “Phase transitions on non amenable graphs,” *J. Math. Phys.* **41** (2000), pp. 1099–1127.

[28] F. MARTINELLI, “Lectures on Glauber dynamics for discrete spin models,” *Lectures on Probability Theory and Statistics (Saint-Flour, 1997)*, Lecture notes in Mathematics **1717**, pp. 93–191, Springer, Berlin, 1998.

[29] F. MARTINELLI and E. OLIVIERI, “Approach to equilibrium of Glauber dynamics in the one phase region I: The attractive case,” *Communications in Mathematical Physics* **161** (1994), pp. 447–486.

[30] F. MARTINELLI, E. OLIVIERI and R. SCHONMANN, “For 2-D lattice spin systems weak mixing implies strong mixing,” *Comm. Math. Phys.* **165** (1994), pp. 33–47.

[31] F. MARTINELLI, A. SINCLAIR and D. WEITZ, “Glauber dynamics on trees: Boundary conditions and mixing time,” preprint, 2003, submitted to *Communications of Mathematical Physics*.

[32] F. MARTINELLI, A. SINCLAIR and D. WEITZ, “Fast mixing for independent sets, colorings and other models on trees,” preprint, 2003.

[33] M. MOLLOY, “The Glauber dynamics on colorings of a graph with high girth and maximum degree,” *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, 2002, pp. 91–98.

[34] E. MOSSEL, “Survey: Information flow on trees,” Preprint, October 2002, to appear in DIMACS volume *Graphs, Morphisms and Statistical Physics*.

[35] E. MOSSEL and Y. PERES, “Information flow on trees,” *Annals of Applied Probability*, 2003, to appear.

[36] Y. PERES and P. WINKLER, personal communication.

[37] C.J. PRESTON, *Gibbs states on countable sets*, Cambridge Tracts in Mathematics **68**, Cambridge University Press, London, 1974.

[38] L. SALOFF-COSTE, “Lectures on finite Markov chains,” *Lectures on probability theory and statistics (Saint-Flour, 1996)*, Lecture notes in Mathematics **1665**, pp. 301–413, Springer, Berlin, 1997.

[39] R.H. SCHONMANN and N.I. TANAKA, “Lack of monotonicity in ferromagnetic Ising model phase diagrams,” *Annals of Applied Probability* **8** (1998), pp. 234–245.

[40] B. SIMON, *The statistical mechanics of lattice gases, Vol. I*, Princeton Series in Physics, Princeton University Press, Princeton, NJ, 1993.

[41] F. SPITZER, “Markov random fields on an infinite tree,” *Annals of Probability* **3** (1975), pp. 387–398.

[42] D.W. STROOCK and B. ZEGARLINSKI, “The logarithmic Sobolev inequality for discrete spin systems on a lattice,” *Communications in Mathematical Physics* **149** (1992), pp. 175–194.

[43] E. VIGODA, “Improved bounds for sampling colorings,” *Journal of Mathematical Physics* **41** (2000), pp. 1555–1569.

[44] E. VIGODA, “A note on the Glauber dynamics for sampling independent sets,” Electronic Journal of Combinatorics, Volume 8(1), 2001.