Fast mixing for independent sets, colorings and other models on trees†

Fabio Martinelli‡ Alistair Sinclair§ Dror Weitz ¶

November 29, 2004

Abstract

We study the mixing time of the Glauber dynamics for general spin systems on bounded-degree trees, including the Ising model, the hard-core model (independent sets) and the antiferromagnetic Potts model at zero temperature (colorings). We generalize a framework, developed in our recent paper [20] in the context of the Ising model, for establishing mixing time $O(n \log n)$, which ties this property closely to phase transitions in the underlying model. We use this framework to obtain rapid mixing results for several models over a significantly wider range of parameter values than previously known, including situations in which the mixing time is strongly dependent on the boundary condition. We also discuss applications of our framework to reconstruction problems on trees.

---

† A preliminary version of this paper appeared in *Proceedings of the 15th ACM-SIAM Symposium on Discrete Algorithms*, January 2004.

‡ 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.

¶ School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540, U.S.A. Email: dror@ias.edu. Supported by a grant from the state of New Jersey and NSF grants DMS-0111298 and CCR-0324906. Most of this work was done while this author was a graduate student at the Computer Science Division, University of California, Berkeley, supported in part by NSF Grant CCR-0121555.

1 Introduction

1.1 Spin systems on trees

Spin systems capture a wide range of probabilistic models studied in statistical physics, applied probability, artificial intelligence and elsewhere. A (*nearest neighbor*) *spin system* on a 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) $x \in V$ a spin value $\sigma_x \in S$. The probability of finding the system in configuration σ is determined by the *Gibbs distribution*

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

Thus the pair potential specifies the likelihood of seeing a given pair of spins at adjacent sites, while the singleton potential specifies the likelihood of seeing a given spin. 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 $\Omega \subseteq S^V$ the set of all *valid* spin configurations, i.e., those for which $\mu(\sigma) > 0$. We will give several concrete examples in a moment.

Usually G is a finite portion of some regular lattice, such as $\mathbb{Z}^d$. In this paper we concentrate on what is known in statistical physics as the *Bethe lattice* $\mathbb{T}^b$, i.e., G is a complete rooted tree in which each interior vertex has $b \geq 2$ children.[1] Spin systems on trees are not only a useful simplification of their more classical counterparts on $\mathbb{Z}^d$, but have recently attracted a lot of attention in their own right as the canonical example of models on "non-amenable" graphs (i.e., those whose boundary is of comparable size to their volume) — see, e.g., [1, 2, 4, 10, 17].

A *boundary condition* corresponds to fixing the spins at the leaves of G to some specified values (e.g., all are set to the same value s). This allows us to formalize the central concept of a *phase transition*. If we let the size of the tree G grow to infinity, the Gibbs distribution tends to a limit known as the (*infinite-volume*) *Gibbs measure*. This limit may or may not depend on the boundary condition: i.e., there may be either a unique Gibbs measure, or multiple Gibbs measures ("phases") corresponding to different boundary conditions. Informally, the existence of multiple Gibbs measures corresponds to the fact that the spin configuration on the leaves can have a non-zero influence on the spin at the root even as the depth of the tree tends to infinity. A *phase transition* occurs when an infinitesimal change in the potentials leads to a switch from a unique Gibbs measure to multiple Gibbs measures. See, e.g., [9] for more background.

We now illustrate the above ideas with some concrete examples. The following four spin systems are among the most widely studied in the literature, and will serve as the motivating examples in this paper:

The (ferromagnetic) Ising model
There are two spin values $S = \{\pm 1\}$, and the potentials are $U(s_1, s_2) = -\beta s_1 s_2$, $W(s) = -\beta h s$, where β is the inverse temperature and h is the external field. Thus $\Omega = \{\pm 1\}^V$. The Gibbs distribution μ assigns higher weight to configurations in which many neighboring spins are aligned with one another. This effect increases with β, so that at high temperatures (low β) the spins behave almost independently, while at low temperatures (high β) large connected regions of equal spins tend to form. There is also a global tendency for spins to align with the sign of the external field.

[1]Strictly speaking, in the Bethe lattice all vertices (including the root) have degree $b + 1$; for convenience we assume that the root has degree b. This difference is purely technical and does not affect our results.

In fact, as is well known [25], when $h = 0$ a phase transition occurs at the critical value $\beta = \beta_0 = \frac{1}{2}\ln(\frac{b+1}{b-1})$; in other words, in the "high temperature" region $\beta \leq \beta_0$ there is a unique Gibbs measure independent of the boundary condition, but as soon as $\beta > \beta_0$ we get (at least) two different Gibbs measures corresponding to the $(+)$- and $(-)$-boundary conditions respectively; i.e., if the leaves are $(+)$ then the root has probability bounded away from $\frac{1}{2}$ of being $(+)$.[2]

The hard-core model (independent sets)
This has been used in statistical physics as a model of lattice gases [9], and also in other areas such as the modeling of communication networks [15]. Again there are just two spins $S = \{0, 1\}$, and we refer to a site as *occupied* if it has spin value 1 and *unoccupied* otherwise. The potentials are $U(1, 1) = \infty$, $U(1, 0) = U(0, 0) = 0$, $W(1) = L$ and $W(0) = 0$, 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 $\lambda = \exp(-L) > 0$, which controls the density of occupation, is referred to as the "activity."

The hard-core model also undergoes a phase transition at a critical activity $\lambda = \lambda_0 = \frac{b^b}{(b-1)^{b+1}}$ (see, e.g., [26, 15]). For $\lambda \leq \lambda_0$ there is a unique Gibbs measure independent of the boundary condition, while for $\lambda > \lambda_0$ there are (at least) two distinct phases, corresponding to the "odd" and "even" boundary conditions respectively. The even (odd) boundary condition is obtained by making the leaves of the tree all occupied if the depth is even (odd), and all unoccupied otherwise. For $\lambda > \lambda_0$, the probability of occupation of the root in the infinite-volume Gibbs measure differs for odd and even boundary conditions.

The (ferromagnetic) Potts model
This model was introduced by Potts [24] as a generalization of the Ising model to more than two spin values; see [31] for a survey. Here $S = \{1, 2, \ldots, q\}$ and the potentials are $U(s_1, s_2) = -2\beta\delta_{s_1,s_2}$, $W(s) = 0$, where β is the inverse temperature. Thus 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. Note that the Ising model (with no external field) is the special case $q = 2$. There are no hard constraints, so $\Omega = S^V$.

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 [10] 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.)

The antiferromagnetic Potts model (colorings)
In this model $S = \{1, 2, \ldots, q\}$, and the potentials are $U(s_1, s_2) = 2\beta\delta_{s_1,s_2}$, $W(s) = 0$. This is analogous to the Ising and Potts models 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; in the rest of the paper, we shall always confine our attention to this case of the model. 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. This model has been widely studied not only in statistical physics, but also in computer science because of its connection to graph coloring. See, e.g., [4] for an informative account.

[2]For notational convenience in the sequel, we will write $(+)$ and $(-)$ in place of $+1$ and -1 respectively.

For colorings on the b-ary tree it is well known that, when $q \leq b+1$, 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., [4]). Recently it has been proved that, as soon as $q \geq b+2$, the Gibbs measure is unique [14].

Remark: In this paper our discussion of the antiferromagnetic Potts model is always of the zero temperature case, and for this model we omit the temperature parameter from the rest of the discussion. $\square$

1.2 Glauber dynamics

Whereas classical statistical physics focused on *static* questions about the infinite volume Gibbs measure (such as the existence of a phase transition), the emphasis in recent years has shifted towards the study of the *Glauber dynamics*, a local Markov chain on the set of spin configurations Ω of a finite graph G. For definiteness, we describe the "heat-bath" version of Glauber dynamics: at each step, pick a site x u.a.r., and replace the spin at x by a random spin drawn from the conditional distribution given the spins of its neighbors. It is easy to check that this dynamics converges to μ as its stationary distribution, providing both an algorithm for sampling from the Gibbs distribution and a plausible model for the actual evolution of the physical system. The key question in the study of the dynamics is to determine the *mixing time*, i.e., the number of steps until the distribution is close to μ, starting from an arbitrary configuration.

Recent developments in statistical physics have revealed a remarkable connection between the mixing time and phase transitions in the special case of the Ising model: for the Ising model on an n-vertex square in $\mathbb{Z}^2$ (which also has a phase transition at a critical β_0), when $\beta < \beta_0$ the mixing time with arbitrary boundary conditions is $O(n \log n)$ (which is optimal), but as soon as $\beta > \beta_0$ there are boundary conditions for which the mixing time jumps to $\exp(\Omega(\sqrt{n}\,))$ [19]. This conforms with the intuition that the existence of multiple phases creates a "bottleneck" which dramatically slows down mixing. For the Ising model on trees the situation is even more interesting (see [2, 20]): on an n-vertex b-ary tree, the mixing time remains $O(n \log n)$ for all boundaries,[3] not only for $\beta \leq \beta_0$ but in fact for all $\beta < \beta_1$, where $\beta_1 = \frac{1}{2}\ln\left(\frac{\sqrt{b}+1}{\sqrt{b}-1}\right) > \beta_0$. As soon as $\beta > \beta_1$, the mixing time for certain boundaries becomes $n^{1+\Omega(1)}$, and the exponent is unbounded as $\beta \to \infty$. Thus there is again a sharp transition in mixing behavior, but it occurs *inside* the multiple-phase region.[4]

Finally, in the low temperature region $\beta \geq \beta_1$, the mixing time is heavily influenced by the boundary condition: recently, we proved that with the $(+)$ (or, symmetrically, $(-)$) boundary condition, the mixing time remains $O(n \log n)$ throughout this region (and hence for *all* values of β) [20]. This formalizes the intuition that the boundary condition breaks the symmetry between the two phases at low temperatures, thus eliminating the bottleneck and enabling rapid mixing.

The above discussion of the Ising model highlights two central issues in the study of the Glauber dynamics for general models. Firstly, for which range of parameter values is the mixing time "optimal" (i.e., $O(n \log n)$) for *arbitrary* boundary conditions? In particular, does this range extend throughout the region of uniqueness of the Gibbs measure, and even beyond (as it does for the Ising model)?[5] And secondly, for parameter values outside this range, are there natural boundary conditions for which the mixing time is still optimal?

[3] Throughout, a mixing time of $O(n \log n)$ hides constants that depend only on the potentials and on the degree b.

[4] This second critical value β_1 has other interpretations in terms of extremality of the Gibbs measure and the threshold for noisy data transmission on the tree [8].

[5] Far into the uniqueness region — e.g., for the Ising model at very high temperatures — it is easy to see using the Dobrushin Uniqueness Condition that the mixing time is $O(n \log n)$ (see, e.g., [29]).

In this paper we present a unified framework for answering such questions for general spin systems on trees. The framework is adapted and extended from our earlier paper [20] on the Ising model, where we bound the log-Sobolev constant (and hence the mixing time) in terms of two simple quantities derived from the Gibbs measure, leading to a clean criterion for $O(n \log n)$ mixing time.[6] These two quantities, which we call κ and γ, measure the rate of "disagreement percolation" down and up the tree respectively. Using this framework, we are able to substantially extend the range of parameter values for which $O(n \log n)$ mixing time is known for the models described above, and to explicitly relate this to properties of the Gibbs measure. In the case of specific boundaries, these are apparently the first results of this kind for any of these models. Beyond providing new results for particular models, our goal is also to demonstrate the wide applicability of our framework.

1.3 Our results

In this subsection we state our results and explain how they relate to previous work.

The Ising model
Our results for this model were outlined above, and formally read as follows:

Theorem 1.1 *For the Ising model on the n-vertex b-ary tree, the mixing time of the Glauber dynamics is $O(n \log n)$ in both of the following situations:*

(i) *the boundary condition is arbitrary, and either $\beta < \beta_1$ or $|h| > h_c(\beta)$.*

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

This theorem was proved in our recent paper [20]. However, as a convenient starting point for our other results, we recapitulate the relevant calculations in Section 4 of the present paper.

The hard-core model (independent sets)
The Glauber dynamics for the hard-core model on trees is known to have mixing time polynomial in n at all activities $\lambda > 0$ with arbitrary boundaries [13]. Moreover, a rather general result of Luby and Vigoda [16, 28] ensures a mixing time of $O(n \log n)$ when $\lambda < \frac{2}{b-1}$, with arbitrary boundaries. This latter result actually holds for any graph G of maximum degree $b + 1$.

In this paper, we prove results for the hard-core model that mirror those stated above for the Ising model. First, we show that the mixing time is $O(n \log n)$ for all activities $\lambda \le \lambda_0$ (and indeed beyond), with *arbitrary* boundary conditions. Second, for the *even* (or odd) boundary condition, we get the same result for *all* activities λ:

Theorem 1.2 *For the hard-core model on the n-vertex b-ary tree, the mixing time of the Glauber dynamics is $O(n \log n)$ in both of the following situations:*

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

(ii) *the boundary condition is even (or odd), and $\lambda > 0$ is arbitrary.*

---

[6] A recent paper of Jerrum *et al.* [13] provides alternative tools based on decomposition ideas for bounding the log-Sobolev constant; those tools work in much more general settings, but give weaker bounds than ours for the scenarios discussed in this paper. In particular, it seems unlikely that those methods are sensitive enough to isolate the regime where the mixing time is $O(n \log n)$.

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 $\lambda \leq \lambda_0$, and indeed, by analogy with the Ising model, also in some intermediate region beyond this. Note that the quantity $(\frac{b+1}{b-1})^2 - 1$ exceeds λ_0 for all $b \geq 2$, so we do indeed get $O(n \log n)$ mixing time throughout the low-activity region and beyond.[7] To the best of our knowledge this is the first such result. (Note that the result of [16, 28] mentioned earlier establishes $O(n \log n)$ mixing time only 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.) Moreover, for $b \geq 12$ the quantity $\frac{1}{\sqrt{b}-1}$ exceeds $(\frac{b+1}{b-1})^2 - 1$, so the former determines the range of λ in part (i) for large b; and since this quantity grows as $\Theta(\frac{1}{\sqrt{b}})$ compared to the $\Theta(\frac{1}{b})$ growth of λ_0, our region extends significantly beyond λ_0 for large b. We also mention that our analysis in this region has consequences for the infinite volume Gibbs measure itself, implying that, in the range of λ specified in part (i), any Gibbs measure that is the limit of finite Gibbs distributions for some fixed boundary configuration is *extremal*, again a new result. (The extremality result follows from the fact that our criterion for $O(n \log n)$ mixing time based on κ and γ also implies exponential decay of point-to-set correlations in the Gibbs distribution. See our earlier paper [20] for details.) For results on extremality with specific boundary conditions, see [5, 18] and Section 9 of the present paper.

General two-spin systems

Since our results for the Ising and hard-core models have a lot in common, it is natural to ask whether these results are instances of a more general phenomenon for systems on trees. Indeed they are: the fact that the mixing time is $O(n \log n)$ both uniformly in the boundary condition throughout the uniqueness regime, and for a constant boundary condition (e.g., all-$(+)$ or all even) for all values of the parameters of the model, holds for *any* spin system whose spin space S consists of two spin values.

Theorem 1.3 *For any two-spin system (given by potentials U and W) the following holds:*

(i) *if the system admits a unique Gibbs measure then the mixing time of the Glauber dynamics is $O(n \log n)$, uniformly in the boundary condition;*

(ii) *for the two constant boundary conditions, the mixing time is $O(n \log n)$.*

The (ferromagnetic) Potts model

Little is known about the Glauber dynamics for the Potts model on trees, beyond the facts that the mixing time is $O(n \log n)$ for arbitrary boundaries at very high temperatures (by the Dobrushin Uniqueness Condition), and is $\Omega(n^{1+\epsilon})$ for some boundaries at very low temperatures (combining results in [2, 23]). In this paper, we prove:

Theorem 1.4 *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 of 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*;

[7] For example, for $b = 2$ the result holds for all $\lambda < 8$, while $\lambda_0 = 4$; and for $b = 3$ it holds for $\lambda < 3$, while $\lambda_0 \approx 1.69$.

(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 $O(n \log n)$ mixing time 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 mixing time of the Ising model with $(+)$-boundaries is $O(n \log n)$ 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 the mixing time is $O(n \log n)$ 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 (see below), we obtain an alternative proof of the best known value of the noise parameter under which reconstruction is impossible [23]. As we observe later, a slight strengthening of part (iii) marginally improves on this threshold.

The antiferromagnetic Potts model (colorings)
The sharpest result known for the Glauber dynamics on colorings is due to Vigoda [27], who shows that for arbitrary boundary conditions the mixing time is $O(n \log n)$ provided $q > \frac{11}{6}(b+1)$. This result actually holds not only for trees but for *any* n-vertex graph G of maximum degree $b+1$. For graphs of large maximum degree and girth at least 6, this range was recently improved [7] to $q > \max\{1.489(b+1), q_0\}$, where q_0 is an absolute constant.[8] In this paper, we extend this rapid mixing result throughout the uniqueness region, except for the "critical" value $q = b+2$.

Theorem 1.5 *The mixing time of the Glauber dynamics for colorings on the* n*-vertex* b*-ary tree is* $O(n \log n)$ *for arbitrary boundary conditions and* $q \geq b+3$.

Note that this result is optimal: when $q = b+2$, it is not too hard to construct boundary conditions under which the Glauber dynamics is not connected. (See Section 7 for a discussion.)

Reconstruction problems on trees
Our final sequence of results is concerned not with the dynamics but with a quite different topic known as "reconstruction problems." Here the tree is viewed as a noisy communication network in which a value is transmitted from the root to all other nodes along the edges, with a certain probability of error in the value transmitted on each edge. These error probabilities (or "channel parameters") are the same for all edges, and errors on different edges are independent. The question is: for what values of the error probabilities can the value sent by the root be reconstructed from the values at a level far below? Reconstruction problems on trees have been widely studied in both genetics and communication theory; see, e.g., [22] for a survey.

The connection with our main results is the following. Firstly, via a standard transformation it is straightforward to relate the reconstruction problem to an associated spin system, where the potentials are determined by the error probabilities on the edges. Reconstruction is then possible only if there is significant correlation between the value at the root and the values at the leaves in the Gibbs measure for this spin system (with free boundary). (More technically, impossibility of reconstruction corresponds precisely to the free Gibbs measure being *extremal*.) Our criterion for $O(n \log n)$ mixing time of the Glauber dynamics, based on the quantities κ and γ mentioned earlier, in fact also implies a decay of correlations between the values at the root and leaves of a tree as

[8] A recent sequence of papers [6, 21, 11] 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 $\epsilon > 0$ [12], but these results do not apply in our setting where the degree $b+1$ is fixed.

the depth of the tree tends to infinity. Thus, we can use exactly the same technology in order to obtain ranges of values of the error parameters for which reconstruction is impossible. We do this in Section 9 for several popular error models, including general binary channels and symmetric multi-channels, and thus obtain simpler proofs of known thresholds for reconstruction. Moreover, for the symmetric multi-channel (for which the associated spin system is the Potts model), we are able to obtain a slight improvement over known bounds. For further background on reconstruction problems and precise statements of our results, the reader is referred to Section 9.

The remainder of the paper is organized as follows. After some basic definitions in Section 2, in Section 3 we extend the analytic framework from our previous paper to general spin systems on trees, defining the quantities κ and γ and relating them to the mixing time. Then in Sections 5, 6, 7 and 8 we specialize the analysis to the hard-core, general two-spin, colorings and Potts models respectively. Finally, we discuss the application of our results to reconstruction problems in Section 9.

2 Preliminaries

For $b \geq 2$, let $\mathbb{T}^b$ denote the infinite rooted b-ary tree (in which every vertex has b children). We will be concerned with complete finite subtrees T that are initial portions of $\mathbb{T}^b$, i.e., share the same root; 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 $(T \cup \partial T) \setminus A$).

Consider a spin system on T specified by spin values S, pair potential U and singleton potential W as in the Introduction. Let $\tau \in S^{\mathbb{T}^b}$ be a spin configuration on the infinite tree $\mathbb{T}^b$. We denote by Ω_T^τ the set of configurations $\sigma \in S^{T \cup \partial T}$ that agree with τ on ∂T; i.e., τ specifies a *boundary condition* on T. The spin at x is denoted σ_x. For any $\eta \in \Omega_T^\tau$ and any subset $A \subseteq T$, the Gibbs distribution on A conditional on the configuration outside A being η is denoted μ_A^η and is defined as follows: if $\sigma \in \Omega$ agrees with η outside A then

$$\mu_A^\eta(\sigma) \propto \exp\Big[-\big(\sum_{xy \in E(A \cup \partial A)} U(\sigma_x, \sigma_y) + \sum_{x \in A} W(\sigma_x)\big)\Big]; \tag{1}$$

otherwise, $\mu_A^\eta(\sigma) = 0$. 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 will assume that μ_A^η is well defined (i.e., that the expression in (1) is positive for at least one σ) for every A, τ and η (even when $\mu_T^\tau(\eta) = 0$); we call the spin system *permissive* in this case. (Note that this is an issue only for systems with hard constraints.) All the examples in this paper are clearly permissive. We usually abbreviate Ω_T^τ and μ_T^τ to Ω and μ respectively. When there are hard constraints (i.e., $U(s_1, s_2) = \infty$ for some s_1, s_2) we remove invalid configurations (i.e., those for which $\mu(\sigma) = 0$) from Ω.

The (*heat-bath*) *Glauber dynamics* is the following Markov chain on $\Omega = \Omega_T^\tau$. In configuration $\eta \in \Omega$, transitions are made as follows:

(i) pick a vertex $x \in T$ u.a.r., and a spin value s chosen from the distribution of the spin at x conditional on the spins of its neighbors (i.e., s has the distribution of the spin at x in $\mu_{\{x\}}^\eta$);

(ii) go to configuration $\eta^{x,s}$ obtained from η by setting the spin at x to s.

We always assume that the dynamics is *connected* on any finite region with any boundary condition; i.e., for any pair of configurations with non-zero probabilities in the Gibbs distribution, there is a finite sequence of transitions (with non-zero probabilities) from one configuration to the other. (Notice again that this is an issue only for systems with hard constraints. Also, among the specific systems we consider, the only one for which connectedness may not hold is the colorings model for certain values of b and q.)

Assuming connectedness, 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*:

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

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 for algebraic convenience only.)

When we say that the mixing time of the Glauber dynamics is $O(n\log n)$ for some boundary condition τ, we mean that for all finite T, the mixing time for μ_T^τ is $\le cn\log n$ for a constant c that depends only on b and the potentials.

3 A criterion for $O(n\log n)$ mixing time

In [20] we developed a criterion for $O(n\log n)$ mixing time of the Glauber dynamics on trees, and used it to analyze the Ising model both for arbitrary boundary conditions and in the important special case of $(+)$-boundaries. This criterion generalizes immediately to arbitrary spin systems, as we describe in this section along with some useful extensions.

The key ingredients are two quantities, which we call κ and γ, that bound the rate of percolation of disagreements down and up the tree respectively. Both are properties of the collection of Gibbs distributions $\{\mu_T^\tau\}$, where the boundary condition τ is fixed and T ranges over all finite complete subtrees of $\mathbb{T}^b$. To define κ and γ we need a little notation. For a vertex $x \in T$, write T_x for the (maximal) subtree rooted at x. When x is not the root of T, let $\mu_{T_x}^s$ denote the Gibbs distribution in which the parent of x has its spin fixed to s and the configuration on the bottom boundary of T_x is specified by τ (the global boundary condition on T).[9] For two distributions μ_1,μ_2 on Ω, $\|\mu_1-\mu_2\|_x$ denotes the variation distance between the projections of μ_1 and μ_2 onto the spin at x, i.e., $\|\mu_1-\mu_2\|_x = \frac{1}{2}\sum_{s\in S}|\mu_1(\sigma_x=s)-\mu_2(\sigma_x=s)|$. Recall that $\eta^{x,s}$ is the configuration η with the spin at x set to s.

Definition 3.1 *For a collection of Gibbs distributions $\{\mu_T^\tau\}$ as above, define $\kappa \equiv \kappa(\{\mu_T^\tau\})$ and $\gamma \equiv \gamma(\{\mu_T^\tau\})$ by*

(i) $\kappa = \sup_T \max_{z,s,s'} \|\mu_{T_z}^s - \mu_{T_z}^{s'}\|_z$;

(ii) $\gamma = \sup_T \max \|\mu_A^{\eta^{y,s}} - \mu_A^{\eta^{y,s'}}\|_z$, *where the maximum is taken over all subsets $A \subset T$, all configurations η, all sites $y \in \partial A$, all neighbors $z \in A$ of y, and all spins $s,s' \in S$.*

Remark: 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

[9] We do not specify the configuration in the rest of $T \setminus T_x$ as it has no influence on the distribution inside T_x once the spin at the parent of x is fixed.

the worst-case probability of disagreement for an *arbitrary* subset A and arbitrary boundary configuration around A, and hence depends only on the potentials of the system 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.

The intuition for these definitions comes from the following claim, which relates κ and γ to the rate of disagreement percolation in the tree. For any T and site $x \in T$, write $\widetilde{T_x}$ for $T_x \setminus \{x\}$, the subtree T_x excluding its root, and $\mu^s_{\widetilde{T_x}}$ for the Gibbs distribution when the spin at x is fixed to s. Also, for $\ell \leq \mathrm{height}(x) + 1$ write $B_{x,\ell}$ for the subtree (or "block") of height $\ell - 1$ rooted at x (i.e., $B_{x,\ell}$ has ℓ levels). For two configurations $\sigma, \sigma' \in \Omega$, let $|\sigma - \sigma'|_{x,\ell}$ denote the number of sites ℓ levels below x (i.e., on the bottom boundary of $B_{x,\ell}$) at which σ and σ' differ. Note that $|\sigma - \sigma'| \leq b^\ell$.

Claim 3.2 *For every $x \in T$ and all $\ell \leq \mathrm{height}(x) + 1$ the following hold*:

(i) *For all s, s', there is a coupling $\nu = \nu^{s,s'}$ of $\mu^s_{\widetilde{T_x}}$ and $\mu^{s'}_{\widetilde{T_x}}$ for which $\mathrm{E}_\nu|\sigma - \sigma'|_{x,\ell} \leq (\kappa b)^\ell$.*

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

The proof of this claim follows from a standard recursive coupling along paths in the tree: see [20, Claim 4.4]. Part (i) shows that κ bounds the probability of a disagreement percolating *down* the tree: i.e., when we fix a disagreement at x and recursively couple the distributions on the children of x, the expected proportion of disagreements after ℓ levels is at most κ^ℓ. Similarly, from part (ii) we see that γ bounds the probability of a disagreement percolating *up* the tree: i.e., when we fix a single disagreement at level ℓ below x, the probability of this disagreement reaching x is at most γ^ℓ.

We now state a theorem that will be our main analytical tool. The theorem gives a sufficient condition for $O(n \log n)$ mixing time in terms of the quantities κ and γ. (Recall that we always assume that the Glauber dynamics is connected for all finite regions and all boundary conditions.)

Theorem 3.3 *Consider an arbitrary (permissive) spin system and a boundary condition τ (a configuration on $\mathbb{T}^b$). If $\kappa \equiv \kappa(\{\mu^\tau_T\})$ and $\gamma \equiv \gamma(\{\mu^\tau_T\})$ satisfy $\max\{\gamma\kappa b, \gamma\} < 1$ then the mixing time of the associated Glauber dynamics is $O(n \log n)$.*

Proof: The proof follows by combining Theorems 3.4 and 5.1 of [20][10], which together imply that, under the above conditions on κ and γ, the *logarithmic Sobolev constant* of the dynamics is $\Omega(\frac{1}{n})$. By standard facts relating the log-Sobolev constant to the mixing time, this implies that the mixing time is $O(n \log n)$. $\square$

Theorem 3.3 tells us that, to prove $O(n \log n)$ mixing time, it is enough to estimate the quantities κ and γ for the spin system and boundary condition in question. As we shall see, this can be done using calculations specific to the situation at hand. In [20] we carried out these calculations for the Ising model with various boundary conditions; our goal in this paper is to perform analogous calculations for some other important models, thus demonstrating the utility of the approach.

For some of these other models, we will require two minor but useful generalizations of the above framework, which we now describe. Both generalizations stem from the observation that the role of the definitions of κ and γ is to obtain the bounds on disagreement percolation stated in Claim 3.2. In fact, in Theorem 3.3 we can replace κ and γ by any two values κ' and γ' for which the upper bounds in parts (i) and (ii) of Claim 3.2 are $O((\kappa' b)^\ell)$ and $O(\gamma'^\ell)$ respectively (where the

[10]These theorems are stated for the special case of the Ising model. However, it is easily seen that their proofs make no use of the specific form of the Ising potentials, and thus apply to arbitrary permissive spin systems on trees. See also the PhD thesis of the third author [30], where this framework is spelled out for general permissive spin systems.

$O(\,\cdot\,)$ hides constants independent of ℓ). The arguments leading to Theorem 3.3 are easily seen to hold in this slightly looser setting.

Our first generalization (which will be particularly useful for "non-attractive" systems, including systems with hard constraints) is to consider two levels of the tree at a time, rather than a single level as in Definition 3.1. Accordingly, define

$$\kappa_2 = \sup_T \max_{\substack{z,w \prec z,\\ s_1,s_1',s_2,s_2'}} \sqrt{\|\mu_{T_z}^{s_1} - \mu_{T_z}^{s_1'}\|_z \cdot \|\mu_{T_w}^{s_2} - \mu_{T_w}^{s_2'}\|_w}\ , \tag{3}$$

where $w \prec z$ denotes "w is a child of z." In fact, we may restrict the maximization to sites z of even (or odd) height. Similarly, define

$$\gamma_2 = \sup_T \max \sqrt{\|\mu_A^{\eta^{y,s_1}} - \mu_A^{\eta^{y,s_1'}}\|_z \cdot \|\mu_{A\setminus\{z\}}^{\eta^{z,s_2}} - \mu_{A\setminus\{z\}}^{\eta^{z,s_2'}}\|_w}\ , \tag{4}$$

where the maximum is over all subsets $A \subset T$, all configurations η, all neighbors $z \in A$ of y, all neighbors $w \in A$ of z, and all spins $s_1, s_1', s_2, s_2' \in S$. Note that clearly $\kappa_2 \leq \kappa$ and $\gamma_2 \leq \gamma$. Now it is easy to see that the upper bound in Claim 3.2(i) for the probability of disagreement percolating down the tree can be replaced by $(\kappa_2)^{2(\ell/2-1)}b^\ell = O((\kappa_2 b)^\ell)$. Similarly, we can replace γ^ℓ in the upper bound in Claim 3.2(ii) by $O((\gamma_2)^\ell)$. We therefore get the following generalization of Theorem 3.3:

Theorem 3.3′ *In the setting of Theorem 3.3, if κ_2 and γ_2 satisfy $\max\{\gamma_2\kappa_2 b, \gamma_2\} < 1$ then the mixing time of the associated Glauber dynamics is $O(n \log n)$.*

Our second generalization exploits the fact that, when deriving the bound on upward percolation in part (ii) of Claim 3.2, it is enough to control the probability of a disagreement percolating upwards one level from y to z only when z is sufficiently far from the boundary and the root of $B_{x,\ell}$. Let $\widehat{\gamma}$ be defined in the same way as γ, but with the maximization restricted to sets A that include the full subtree of depth d rooted at z under the orientation in which y is the parent of z; here d is an implicit parameter whose value may change from model to model, but will in each case be a constant independent of the size of T. Then the probability of disagreement percolating one level upwards to z, where z is at distance at least d from the boundary and root of $B_{x,\ell}$, is bounded by $\widehat{\gamma}$. Thus it is easy to modify the proof of Claim 3.2 so that the factor γ^ℓ in part (ii) is replaced by $\widehat{\gamma}^{\ell-2d} = O(\widehat{\gamma}^\ell)$. (We refer to [30] for a more detailed explanation of $\widehat{\gamma}$ and the bound it gives on the probability of a disagreement percolating up the tree.) Similarly, we define $\widehat{\kappa}$ as before but with the maximization restricted to z that are at distance at least d from the boundary of T. Whenever we use $\widehat{\gamma}$, we will also use $\widehat{\kappa}$ with the same value of d so that we still have $\widehat{\kappa} \leq \widehat{\gamma}$. This leads to our second generalization of Theorem 3.3:

Theorem 3.3″ *In the setting of Theorem 3.3, if $\widehat{\kappa}$ and $\widehat{\gamma}$ satisfy $\max\{\widehat{\gamma}\,\widehat{\kappa} b, \widehat{\gamma}\} < 1$ then the mixing time of the associated Glauber dynamics is $O(n \log n)$.*

4 The Ising model

In this section we will prove that the mixing time of the Glauber dynamics for the Ising model is $O(n \log n)$ in all the situations covered by Theorem 1.1. Although we already proved this in [20], we repeat the proof here, in slightly modified form, in order to illustrate ideas that will be used in the proofs for other models in subsequent sections.

In light of Theorem 3.3, in order to show $O(n\log n)$ mixing time it is enough to bound the quantities κ and γ such that $\max\{\gamma\kappa b, \gamma\} < 1$, where κ and γ can also be replaced by their variants as explained in Section 3. Thus, Theorem 1.1 will follow from the bounds on κ and γ given in Theorem 4.1 below. Recall that $\gamma \equiv \gamma(\{\mu_T^\tau\})$ depends only on the potentials of the system, while $\kappa \equiv \kappa(\{\mu_T^\tau\})$ may also depend on the boundary condition τ.

Theorem 4.1 *Consider the Ising model at inverse temperature β and external field h. Then*:

(i) *for all (β, h), $\gamma \le \frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}}$;*

(ii) *if (β, h) are such that the Gibbs measure is unique (i.e., $\beta \le \beta_0$ or $|h| > h_c(\beta)$) then for every $\varepsilon > 0$ there exists a large enough d such that $\widehat{\gamma} \le \frac{1}{b} + \varepsilon$, where d is the implicit constant in $\widehat{\gamma}$;*

(iii) *for τ the all-$(+)$ boundary configuration, if (β, h) are such that the Gibbs measure is not unique (i.e., $\beta > \beta_0$ and $|h| \le h_c(\beta)$) then $\kappa \equiv \kappa(\{\mu_T^\tau\}) \le \frac{1}{b}$.*

Before proving Theorem 4.1, we first explain how to deduce Theorem 1.1 from it. We start with part (i) of Theorem 1.1, which considers arbitrary boundary conditions and high or intermediate temperature or large external field. Recall that $\kappa \le \gamma$ for all boundary conditions. Thus, $\max\{\gamma\kappa b, 1\} < 1$ for all boundary conditions provided that $\gamma < \frac{1}{\sqrt{b}}$. However, using part (i) of Theorem 4.1, we see that this is the case for all $\beta < \frac{1}{2}\frac{\ln(\sqrt{b}+1)}{\ln(\sqrt{b}-1)}$, i.e., for all $\beta < \beta_1$, completing the proof for high and intermediate temperatures. The result for super-critical external field $|h| > h_c(\beta)$ follows immediately from part (ii) of Theorem 4.1 by applying Theorem 3.3$''$, once we recall that $\widehat{\kappa} \le \widehat{\gamma}$ for all boundary conditions. We go on to part (ii) of Theorem 1.1, where the boundary condition is all-$(+)$. Notice that the regime in which the Gibbs measure is unique is covered (for arbitrary boundary conditions) by part (i) of Theorem 1.1. For the regime in which the Gibbs measure is not unique, part (ii) of Theorem 1.1 follows immediately from part (iii) of Theorem 4.1, together with the fact that $\gamma < 1$ for all $\beta < \infty$, which is apparent from part (i) of the same theorem.

Proof of Theorem 4.1: To bound κ and γ, we need to bound a quantity of the form $\|\mu_A^{\eta^{y,+}} - \mu_A^{\eta^{y,-}}\|_z$, 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: to this end, we let $\mu_A^{\eta^{y,*}}$ 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 4.2 *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*

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

where $R = \frac{\mu_A^{\eta^{y,*}}(\sigma_z = -)}{\mu_A^{\eta^{y,*}}(\sigma_z = +)}$ is the magnetization at 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 between 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 abbreviate $\mu_A^{\eta^{y,+}}$, $\mu_A^{\eta^{y,-}}$ and $\mu_A^{\eta^{y,*}}$ to μ_A^+, μ_A^- and μ_A^* respectively. Thus $\|\mu_A^{\eta^{y,+}} - \mu_A^{\eta^{y,-}}\|_z = |\mu_A^+(\sigma_z = +) - \mu_A^-(\sigma_z = +)|$, 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 $\mu_A^-(\sigma_z = +) = \frac{1}{R^- + 1}$. $\square$

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 we cannot rule out the possibility that $R = 1$ when the subset A and the boundary configuration η are arbitrary, and this completes the proof of part (i) of Theorem 4.1. However, for $\widehat{\gamma}$ and κ we only have to consider restricted scenarios (either because A includes a deep enough full subtree in the case of $\widehat{\gamma}$, or because the boundary condition is specific in the case of κ), and as we shall see below, we get better bounds for these quantities by calculating the magnetization R in the relevant scenarios.

Before giving the details of this calculation, we make the following two remarks regarding $\widehat{\gamma}$. First recall that, in order to bound $\widehat{\gamma}$, we need to consider an arbitrary subset A that includes a deep enough subtree B rooted at z and an arbitrary boundary condition outside A. Now, notice that the Gibbs distribution on A is a convex combination of μ_B^σ as σ varies. Thus, if we establish that the magnetization R at z for the subtree B with an arbitrary boundary condition is at least a_1 and at most a_2, then this immediately implies the same for the subset A. The second point we wish to make is that the Ising model is *monotone*, i.e., by changing spins on the boundary from $(+)$ to $(-)$ the magnetization $R = \frac{\Pr(\sigma_z=-)}{\Pr(\sigma_z=+)}$ can only increase. Therefore, in order to establish upper and lower bounds on the magnetization for arbitrary boundary conditions, it is enough to bound the magnetization for the all-$(+)$ and all-$(-)$ configurations respectively. We thus concentrate on calculating the magnetization at the root of full subtrees when the boundary condition at the bottom of the subtree is either all-$(+)$ or all-$(-)$. Notice also that a full subtree with the all-$(+)$ boundary is exactly what we need to consider in order to bound κ in part (iii) of Theorem 4.1.

4.1 A recursive calculation of the magnetization at the root of full subtrees

Fix a boundary configuration τ, and for a site z with parent y let $R_{z,\ell} = \frac{p(\sigma_z=-)}{p(\sigma_z=+)}$, where $p(\cdot) = \mu_{B_{z,\ell}}^{\tau^{y,*}}(\cdot)$. (If z is the root of $\mathbb{T}^b$ then $p(\cdot) = \mu_{B_{z,\ell}}^{\tau}(\cdot)$.)

We now describe a recursive calculation of the magnetization $R_{z,\ell}$, the details of which (up to change of variables) can be found in [1] or [3]. Recall that $x \prec z$ denotes that x is a child of z. A simple direct calculation gives that $R_{z,\ell} = e^{-2\beta h}\prod_{x\prec z} F(R_{x,\ell-1})$, where $F(a) \equiv F_\beta(a) = \frac{a+e^{-2\beta}}{e^{-2\beta}a+1}$. In particular, if τ is the all-$(+)$ configuration (i.e., $R_{x,0} = 0$ for all x) then $R_{z,1} = e^{-2\beta h}[F(0)]^b$. We let $F(\infty) \equiv e^{2\beta}$ and notice that indeed, if τ is the all-$(-)$ configuration (i.e., $R_{x,0} = \infty$ for all x), then $R_{z,1} = e^{-2\beta h}[F(\infty)]^b$. From here onwards we limit our attention to the all-$(+)$ and all-$(-)$ boundary conditions. We thus define

$$J(a) \equiv J_{\beta,h}(a) = e^{-2\beta h}[F(a)]^b \tag{5}$$

and observe that $R_{z,\ell}$ equals $J^{(\ell)}(0)$ and $J^{(\ell)}(\infty)$ for τ all-$(+)$ and all-$(-)$ respectively, where $J^{(\ell)}$ stands for the ℓ-fold composition of J.

Now that we have expressed the magnetization in terms of the function J, our next step is relating J to the function K_β in Proposition 4.2, which expresses the total variation distance in terms of the magnetization. To begin, let us describe some properties of J that we shall use (refer to Fig. 1). First, 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 $[0, \infty)$. In fact, whether J has one or more fixed points corresponds exactly to whether the Gibbs measure (for the same values of β and h) is unique or not. This is because the Gibbs measure is unique if and only if the magnetization at the root of the tree of depth ℓ converges with ℓ to the same value conditioned on the all-$(+)$ and all-$(-)$ boundary configurations respectively. We denote by a_0 the least fixed point of J. Then, since $J(0) > 0$, we have $J'(a_0) \le 1$, where $J'(a) \equiv \frac{\partial J(a)}{\partial a}$ is the derivative of J. In particular, when the Gibbs measure is unique, the derivative at the unique fixed point a_0 is ≤ 1. We also observe that J has a single point of inflection a_*, i.e., the derivative J' is monotonically increasing on $[0, a_*)$ and decreasing on $[a_*, \infty)$ for some $a_* \in \mathbb{R}^+$. (This follows from the fact that the equation $J''(a) = 0$ has a unique solution, as can be verified by straightforward calculus). Therefore, J can have at most three fixed points, and furthermore, if it has two or three fixed points, then necessarily $a_* \ge a_0$, i.e., $J'(a) \le J'(a_0) \le 1$ for $a \in [0, a_0]$.

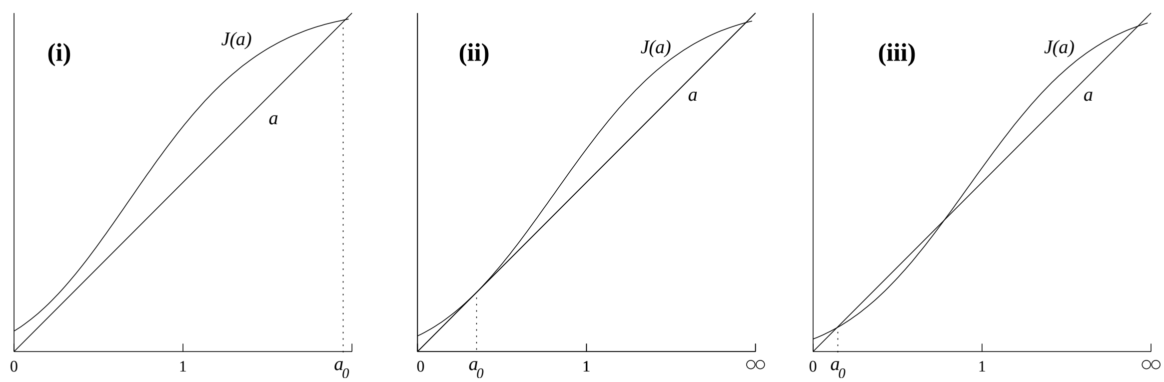

Figure 1: Curve of the function $J(a)$, used in the proof of Theorem 4.1, for $\beta > \beta_0$ and various values of the external field h. (i) $h < -h_c$; (ii) $h = -h_c(\beta)$; (iii) $h_c(\beta) > h > -h_c(\beta)$. The point a_0 is the smallest fixed point of J.

The relevance of the derivative J' and the fixed point a_0 to the discussion here is made clear by the following lemma.

Lemma 4.3 *For every $a \in \mathbb{R}^+$, $K_\beta(a) = \frac{1}{b} \cdot \frac{a}{J(a)} \cdot J'(a)$.*

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

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

We are now ready to complete the proof of the bounds on $\widehat{\gamma}$ and κ given in parts (ii) and (iii) of Theorem 4.1.

4.2 Bounding $\widehat{\gamma}$ when the Gibbs measure is unique

When the Gibbs measure is unique, the magnetization $R_{z,\ell}$ converges with ℓ to the unique fixed point a_0 of J, for which $J'(a_0) \leq 1$, and thus $K_\beta(a_0) \leq \frac{1}{b}$ by Lemma 4.3. We now observe that since a_0 is the unique fixed point, for every $\varepsilon' > 0$ there exists a large enough d such that $R_{z,d} \geq a_0 - \varepsilon'$ for the all-$(+)$ boundary condition, and $R_{z,d} \leq a_0 + \varepsilon'$ for the all-$(-)$ boundary condition. As explained above, this means that for any subset A that includes the full subtree of depth d rooted at z, and with arbitrary boundary condition outside A, the relevant magnetization $R_z \in [a_0 - \varepsilon', a_0 + \varepsilon']$. (From here onwards we write R_z for the magnetization R at z as defined in Proposition 4.2, where the subset A, the boundary condition η, and the neighbor y of z are clear from the context.) Now, since $K_\beta(a)$ is continuous in a, we deduce that that $K_\beta(R_z) \leq \frac{1}{b} + \varepsilon$ for some ε that depends on ε'. In particular, when the Gibbs measure is unique, for every $\varepsilon > 0$ there exists a large enough d such that $\widehat{\gamma} \leq \frac{1}{b} + \varepsilon$. This concludes the proof of part (ii) of Theorem 4.1.

4.3 Bounding κ for the $(+)$-boundary condition when the Gibbs measure is not unique

We now assume that τ (the global boundary configuration) is all-$(+)$ and consider (β, h) such that the Gibbs measure is not unique, i.e., $\beta > \beta_0$ and $|h| \leq h_c(\beta)$. As we shall see below, the property of this regime that we use here is that J has at least two fixed points, and therefore $J'(a) \leq 1$ for $a \in [0, a_0]$.

To calculate κ, we need to bound the variation distance $\|\mu^+_{T_z} - \mu^-_{T_z}\|_z$, which by Proposition 4.2 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. Now, since τ is all-$(+)$, then $R_z = J^{(\ell)}(0)$, where ℓ is the distance of z from the bottom boundary of T. We thus have $\kappa = \sup_T \max_{z \in T} K_\beta(R_z) = \sup_{\ell \geq 1} K_\beta(J^{(\ell)}(0))$.

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., $J^{(\ell)}(0) \leq a_0$ for all ℓ. Now by applying Lemma 4.3, since $J'(a) \leq 1$ for $a \in [0, a_0]$ in the non-uniqueness regime, and since $J(a) \geq a$ for the same a, $K_\beta(J^{(\ell)}(0)) = \frac{1}{b} \cdot \frac{J^{(\ell)}(0)}{J(J^{(\ell)}(0))} \cdot J'(J^{(\ell)}(0)) \leq \frac{1}{b}$ for all ℓ. This completes the proof of part (iii) of Theorem 4.1. □

Remark: We note that in fact $\kappa \leq \frac{1}{b}$ for $\beta \leq \beta_0$ (and arbitrary h) as well. This follows from the fact that $\gamma \leq \frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}} \leq \frac{1}{b}$ in this regime. We also note that $\widehat{\kappa} \leq \frac{1}{b} + \varepsilon$ throughout the uniqueness regime, as is apparent from part (i) of the above theorem. Indeed, the only obstacle to proving $\kappa \leq \frac{1}{b}$ with the all-$(+)$ boundary for all (β, h) is that for $\beta < \beta_0$ and $h < -h_c(\beta)$, the derivative $J'(J^{(\ell)}(0)) > 1$ for some ℓ (see Figure 1). Notice, however, that this derivative converges with ℓ to a value not larger than 1. (This was used for proving $\widehat{\gamma} \leq \frac{1}{b} + \varepsilon$ in this regime.)

5 The hard-core model (independent sets)

In this section, we will prove that the mixing time of the Glauber dynamics for sampling independent sets is $O(n \log n)$ in all the scenarios covered by Theorem 1.2. Again, we appeal to our general framework in Theorem 3.3 and its variants to deduce Theorem 1.2 from:

Theorem 5.1 *For the hard-core model with activity parameter* λ:

(i) $\gamma \leq \frac{\lambda}{1+\lambda}$;

(ii) $\gamma_2 \le \sqrt{\frac{\sqrt{1+\lambda}-1}{\sqrt{1+\lambda}+1}}$;

(iii) *for τ the 0-boundary condition (i.e., all sites are unoccupied), if the Gibbs measure is not unique (i.e., $\lambda > \lambda_0$) then $\kappa_2 \equiv \kappa_2(\{\mu_T^\tau\}) \le \frac{1}{b}$.*

Recall that $\kappa \le \gamma$, so from part (i) of this theorem we conclude that $\gamma\kappa b < 1$ when $(\frac{\lambda}{1+\lambda})^2 < \frac{1}{b}$, i.e., when $\lambda < \frac{1}{\sqrt{b}-1}$. Similarly, since $\kappa_2 \le \gamma_2$, part (ii) implies the same result when $\frac{\sqrt{1+\lambda}-1}{\sqrt{1+\lambda}+1} < \frac{1}{b}$, i.e., when $\lambda < (\frac{b+1}{b-1})^2 - 1$. Part (i) of Theorem 1.2 now follows using Theorem 3.3$''$. This also dispenses with part (ii) of Theorem 1.2 in the uniqueness regime $\lambda \le \lambda_0$ (recall that $(\frac{b+1}{b-1})^2 - 1 > \lambda_0$ for all b). Part (ii) in the non-uniqueness regime follows immediately from part (iii) of Theorem 5.1, using Theorem 3.3$'$ and the fact that $\gamma_2 < 1$. (Note that analyzing the 0-boundary for all depths of T handles both odd and even boundary conditions.)

Proof of Theorem 5.1: The proof uses similar ideas to those used in the proof of Theorem 4.1 for the Ising model. We start with a closer look at the variation distance we need to bound in order to compute κ and γ, i.e., $\|\mu_A^{\eta^{y,1}} - \mu_A^{\eta^{y,0}}\|_z$, for some η, A, $z \in A$, and neighbor $y \in \partial A$ of z. Now, from the definition of the hard-core model, in the first distribution the site z is unoccupied with certainty, and hence the variation distance between the two distributions at z is exactly the probability that z is occupied in the *second* distribution (where y is unoccupied, or equivalently, where the edge connecting y and z is removed). Let p_z stand for this last probability. Formally,

$$\|\mu_A^{\eta^{y,1}} - \mu_A^{\eta^{y,0}}\|_z = \mu_A^{\eta^{y,0}}(\sigma_z = 1) \equiv p_z. \tag{6}$$

Our main goal in the rest of this proof is to bound the probability of occupation p_z, either for all A and all boundary configurations η (in the case of γ) or for full subtrees A with the global boundary condition $\eta = \tau$ (in the case of κ).

We start with the easy observation that, for any subset A, any boundary configuration η and any site $z \in A$, $\mu_A^\eta(\sigma_z = 1) \le \frac{\lambda}{1+\lambda}$, simply because the r.h.s. is the probability of z being occupied when all its neighbors are unoccupied, and if one of its neighbors is occupied than z is unoccupied with certainty. Using (6), we deduce that $\gamma \le \frac{\lambda}{1+\lambda}$, which immediately gives part (i) of Theorem 5.1. Part (ii) follows from a similar (though rather more involved) calculation bounding the product of adjacent occupation probabilities, and part (iii) from a recursive calculation of the occupation probabilities for the 0-boundary similar to the magnetization arguments for the Ising model given in the previous section. We deal with each of these in a separate subsection below.

5.1 Bounding γ_2

Recall from the definition (4) that

$$\gamma_2^2 = \sup_T \max \|\mu_A^{\eta^{y,s_1}} - \mu_A^{\eta^{y,s_1'}}\|_z \cdot \|\mu_{A\setminus\{z\}}^{\eta^{z,s_2}} - \mu_{A\setminus\{z\}}^{\eta^{z,s_2'}}\|_w = \sup_T \max p_z p_w, \tag{7}$$

where p_z is the occupation probability as defined earlier (for a given subset A and neighbor y of $z \in A$) and p_w is the analogous quantity for the subset $A \setminus \{z\}$ and neighbor z of $w \in A \setminus \{z\}$. Of course, we could simply use the above bound on p_z and p_w separately to deduce that $\gamma_2 \le \frac{\lambda}{1+\lambda}$, but this gives no new information. We can do better by exploiting the fact that z and w cannot both be occupied to obtain a tighter bound on the product $p_z p_w$.

In what follows we abbreviate $\mu_A^{\eta^{y,0}}$ to μ, so that $p_z = \mu(\sigma_z = 1)$ and $p_w = \mu(\sigma_w = 1|\sigma_z = 0)$. Now we have

$$p_z = \mu(\sigma_z = 1|\sigma_w = 0)\mu(\sigma_w = 0) \leq \frac{\lambda}{1+\lambda}\alpha, \tag{8}$$

where we have written α to denote $\mu(\sigma_w = 0)$. On the other hand, using Bayes' rule we have

$$\begin{aligned} p_w = \mu(\sigma_w = 1|\sigma_z = 0) &= \frac{\mu(\sigma_z = 0|\sigma_w = 1)\mu(\sigma_w = 1)}{\mu(\sigma_z = 0)} \\ &= \frac{\mu(\sigma_w = 1)}{1 - p_z} \\ &\leq \frac{1-\alpha}{1 - \frac{\lambda}{1+\lambda}\alpha}, \end{aligned} \tag{9}$$

where in the last step we used (8). Substituting (8) and (9) into (7) we obtain

$$\gamma_2^2 \leq \sup_\alpha \frac{\lambda\alpha(1-\alpha)}{1+\lambda(1-\alpha)}.$$

Now for each λ the function $f_\lambda(\alpha) \equiv \frac{\lambda\alpha(1-\alpha)}{1+\lambda(1-\alpha)}$ is a concave function of α, and it is easy to check that it achieves its maximum at $\alpha = \alpha_\lambda = \frac{1+\lambda-\sqrt{1+\lambda}}{\lambda}$, where it is equal to

$$f_\lambda(\alpha_\lambda) = \frac{\sqrt{1+\lambda}-1}{\sqrt{1+\lambda}+1}.$$

This yields the bound on γ_2 claimed in part (ii) of Theorem 5.1.

Remark: For the hard-core model, we have proved rapid mixing of the Glauber dynamics with arbitrary boundary condition for $\lambda < \hat{\lambda}$, where $\hat{\lambda} > \lambda_0$, the critical value for uniqueness. However, unlike the analogous result for the Ising model (Theorem 1.1(i)) this value $\hat{\lambda}$ is not tight. Indeed, we note that the above argument can clearly be tightened in several places (at least for some values of b): e.g., the sup over α can be taken only over $\alpha \geq (1+\lambda)^{-1}$, which leads to a small improvement in the final bound. It remains an intriguing open question to establish a sharp threshold for rapid mixing with arbitrary boundary condition in the hard-core model.

5.2 Bounding κ for the all-0 boundary when the Gibbs measure is not unique

We begin with a recursive calculation of the occupation probabilities p_z in the situation where the set A is the full subtree rooted at z. A similar calculation to the one we describe below can be found in, e.g., [15]. As in the Ising model, we consider subtrees $B_{z,d}$ (disconnected from the parent y of z, or equivalently, with y unoccupied) and with a fixed boundary configuration τ. Let $R_{z,d} = \frac{p_z}{1-p_z}$ stand for the ratio of probabilities that the site z is occupied and unoccupied respectively. A simple calculation verifies that $R_{z,d} = \lambda\prod_{w\prec z}\left(\frac{1}{1+R_{w,d-1}}\right)$. For any given value of λ, we thus define the function

$$J(a) = \lambda\left(\frac{1}{1+a}\right)^b \tag{10}$$

and observe that, when the boundary condition is all-0 (respectively all-1), then $R_{z,d} = J^{(d)}(0)$ (respectively $J^{(d)}(\infty)$). Notice, however, that unlike the case of the Ising model, here J is monotonically decreasing. Furthermore, since $J(0) = \lambda > 0$, J has a *unique* fixed point for every λ; we denote this fixed point by $a_0 = a_0(\lambda)$. We also note that the derivatives of J alternate signs (the odd derivatives being negative). Now uniqueness of the Gibbs measure is equivalent to the fixed point a_0 being *attractive*, i.e., the derivative $J'(a_0) \geq -1$; indeed, when $\lambda = \lambda_0$ is critical the derivative at $a_0 = a_0(\lambda_0)$ is exactly -1. The equivalence between uniqueness of the Gibbs measure and the attractiveness of a_0 is better understood by considering the function $J_2(a) \equiv J^{(2)}(a) \equiv J(J(a))$, which corresponds to jumping two levels at a time. The main observation is that, since J is monotonically decreasing, J_2 must be monotonically increasing, and thus plays a similar role to that of J in the Ising model.

Let us now further describe some properties of the function J_2 (see Fig. 2), which can be verified using simple calculus:

1. J_2 is continuous and increasing on $[0, \infty)$, with $J_2(0) = \lambda/(1+\lambda)^b$ and $\sup_a J_2(a) = \lambda$.

2. a_0 (the unique fixed point of J) is a fixed point of J_2.

3. If the Gibbs measure is unique (i.e., $\lambda \leq \lambda_0$) then a_0 is the unique fixed point of J_2. If there are multiple Gibbs measures (i.e., $\lambda > \lambda_0$) then J_2 has three fixed points $a_1 < a_0 < a_2$, where $J(a_1) = a_2$ and $J(a_2) = a_1$.

4. The derivative $J_2'(a) \equiv \frac{\partial J_2(a)}{\partial a}$ is continuous. If a_0 is the unique fixed point of J_2 (the Gibbs measure is unique) then $J_2'(a_0) \leq 1$. If there are three fixed points then $J_2'(a_0) > 1$, and $J_2'(a) \leq 1$ for $a \in [0, a_1]$.

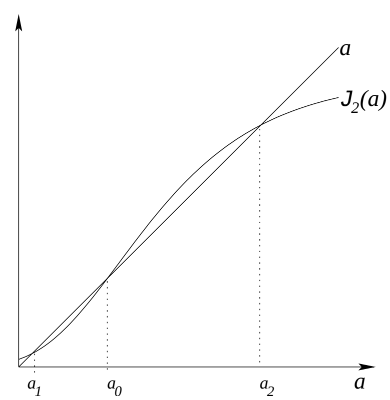

Figure 2: Curve of the function $J_2(a)$, used in the proof of Theorem 5.1, for $\lambda > \lambda_0$. The points a_1, a_0, a_2 are the fixed points of J_2 in increasing order.

It is now easy to see that, since $J_2'(a_0) = J'(a_0)^2$, then indeed the Gibbs measure is unique if and only if $J'(a_0) \geq -1$.

Before we go on, we wish to further clarify the connection between the uniqueness of the Gibbs measure and the uniqueness of the fixed point of J_2. First, notice that for odd-depth (respectively, even-depth) trees, the probability of occupation at the root is monotonically decreasing (respectively, increasing) in the boundary configuration. In particular, for both even and odd depths, the all-0 and all-1 boundaries are the extreme boundary configurations, i.e., for all boundary conditions the probability of occupation at the root is in the range delimited by the probabilities under the all-0 and all-1 boundary conditions. Thus, when J_2 has a unique fixed point, the probability p_z of occupation at the root of even-depth trees converges with the depth of the tree to the same value for all boundary configurations. This also means that p_z converges to the above value in odd-depth trees (uniformly in the boundary condition), because the limit for even-depth trees with the all-0 boundary is the same as the limit for odd-depth trees with the all-1 boundary, and vice versa.

With the above recursive calculation of p_z at hand, we can now complete our argument for bounding κ in the case that the boundary condition τ is set to all-0 and $\lambda > \lambda_0$, i.e, the Gibbs measure is not unique. Recall from the definition (3) that

$$\kappa_2^2 = \sup_T \max_{w \prec z} \|\mu^1_{T_z} - \mu^0_{T_z}\|_z \cdot \|\mu^1_{T_w} - \mu^0_{T_w}\|_w = \sup_T \max_{w \prec z} p_z p_w,$$

where p_z is the occupation probability of z in the full subtree rooted at z with boundary condition τ at the leaves and the parent of z unoccupied (and p_w is defined similarly). Recall also that it is enough to consider sites z whose height is odd. Now, since τ is the all-0 configuration, then $\frac{p_z}{1-p_z} \equiv R_z = J^{(\ell)}(0)$, where ℓ is the distance of z from the bottom boundary of T. Therefore, by letting $K(R) = \frac{R}{1+R}$ (i.e., K translates the ratio $R = \frac{p}{1-p}$ to p), we see that $\kappa_2^2 = \sup_{\ell \geq 1} K[J(J_2^{(\ell)}(0))] \cdot K[J_2^{(\ell)}(0)]$. We will use the properties of J_2 in order to show that the last expression is bounded by $\frac{1}{b^2}$. Notice that, since we are in the regime of non-uniqueness of the Gibbs measure, the sequence $J^{(\ell)}(0)$ does not converge with ℓ to a_0, but oscillates around it, i.e., $J^{(2\ell)}(0) = J_2^{(\ell)}(0)$ converges to a_1 while $J^{(2\ell+1)}(0) = J(J_2^{(\ell)})$ converges to a_2.

A straightforward calculation shows that the derivative $J'(a) = -b \cdot \frac{J(a)}{1+a}$. We thus have

$$\begin{aligned} J_2'(a) &= J'(J(a)) \cdot J'(a) \\ &= b\left(\frac{J(J(a))}{1+J(a)}\right) \cdot b\left(\frac{J(a)}{1+a}\right) \\ &= b^2 \cdot \frac{J_2(a)}{a} \cdot \frac{J(a)}{1+J(a)} \cdot \frac{a}{1+a} \\ &= b^2 \cdot \frac{J_2(a)}{a} \cdot K(J(a)) \cdot K(a). \end{aligned}$$

Applying this equality with $a = J_2^{(\ell)}(0)$, and recalling that $J_2^{(\ell)}(0)$ converges to a_1 from below and that $J_2(a) \geq a$ and $J_2'(a) < 1$ for all $a \in [0, a_1]$, we conclude that for all ℓ,

$$K[J(J_2^{(\ell)}(0))] \cdot K[J_2^{(\ell)}(0)] \;=\; \frac{1}{b^2} \cdot \frac{J_2^{(\ell)}(0)}{J_2(J_2^{(\ell)}(0))} \cdot J_2'(J_2^{(\ell)}(0))) \;\leq\; \frac{1}{b^2},$$

as required. This completes the proof of part (iii) of Theorem 5.1. $\square$

Remark: It is interesting to note that we can use this type of analysis (specifically, the fact that the derivative $J'(a_0) \geq -1$) to show that, throughout the uniqueness regime, $\widehat{\gamma} \leq \frac{1}{b} + \epsilon$ (as we did in Theorem 4.1(ii) for the Ising model). This provides an alternative proof of $O(n \log n)$ mixing time with arbitrary boundary condition throughout the uniqueness regime $\lambda \leq \lambda_0$ (though of course this is already covered by part (ii) of Theorem 5.1, which holds over a wider range of λ).

6 General two-spin systems

It is not too difficult to see that most of our analysis of the hard-core model was based on the same high-level ideas as the analysis of the Ising model. Indeed, these ideas are part of a general theory that holds for any spin system for which the spin space S consists of two values. In this section we develop this general theory in order to show that, for general two-spin systems, the mixing time of the Glauber dynamics is $O(n \log n)$ in the scenarios covered by Theorem 1.3.

In order to simplify our discussion of general two-spin systems we use the following notation. First, w.l.o.g. we assume that $S = \{-,+\}$. Furthermore, we can assume that the value of the pair potential $U(-,+) < \infty$ since otherwise the system is not permissive (specifically, it is a trivial system with only two legal configurations: all-$(+)$ and all-$(-)$ respectively). Now, since the Gibbs distribution remains unaffected by adding a uniform constant value to the potential, we can assume w.l.o.g. that $U(-,+) = 0$, and hence that the pair potential is specified by the two values $U(-,-)$ and $U(+,+)$. We let $\lambda_- = \exp(-U(-,-))$ and $\lambda_+ = \exp(-U(+,+))$. In a similar manner, we can assume w.l.o.g. that the self potential $W(+) = 0$, and let $\lambda = \exp(-W(-))$. Notice that the Gibbs distribution assigns to configuration σ a probability proportional to $\lambda^{\#\{-\}} \cdot \lambda_-^{\#\{-,-\}} \cdot \lambda_+^{\#\{+,+\}}$, where $\#\{-\}, \#\{-,-\}$ and $\#\{+,+\}$ stand for the number of sites whose spin is $(-)$, edges whose spins are $\{-,-\}$ and edges whose spins are $\{+,+\}$ respectively in σ. From here onwards we assume that a two-spin system is specified by the three parameters $(\lambda, \lambda_-, \lambda_+)$. For example, the Ising model with parameters (β, h) is given by $\lambda = e^{-2\beta h}$ and $\lambda_- = \lambda_+ = e^{2\beta}$. The hard-core model with activity λ is given by λ, $\lambda_- = 0$ and $\lambda_+ = 1$, where we have identified the spins 0 and 1 of the hard-core model with $(+)$ and $(-)$ respectively. Before stating our results for two-spins systems, we note that the case $\lambda_- = \lambda_+ = 0$ corresponds to a trivial non-permissive system that has only two legal configurations (the "odd" and "even" configurations respectively, in which the spin values alternate down the levels of the tree). We thus assume w.l.o.g. that in any given system $\lambda_+ > 0$.

As in the cases of the Ising and hard-core models, Theorem 1.3 follows from bounds on κ and γ:

Theorem 6.1 *For any two-spin system $(\lambda, \lambda_-, \lambda_+)$ on the regular b-ary tree:*

(i) *if the Gibbs measure is unique then, for every $\varepsilon > 0$, there exists a large enough d such that $\widehat{\gamma} \le \frac{1}{b} + \varepsilon$;*

(ii) *for τ the all-$(+)$ configuration, if the Gibbs measure is not unique then $\kappa_2 \equiv \kappa_2(\{\mu_T^\tau\}) \le \frac{1}{b}$.*

We observe that $\gamma < 1$ for every two-spin system (because we are assuming $\lambda_+ > 0$, which implies that for any boundary condition, the spin at any given site is $(+)$ with positive probability), and hence Theorem 1.3 follows from Theorem 6.1 by applying either Theorem 3.3$''$ in case the Gibbs measure is unique, or Theorem 3.3$'$ for the $(+)$-boundary in case the Gibbs measure is not unique.

Proof of Theorem 6.1: The first step in the proof is relating the total variation distance at z between two distributions that differ in a single boundary site y to the "magnetization" at z. The following is a generalization of Proposition 4.2:

Proposition 6.2 *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*

$$\|\mu_A^{\eta^{y,+}} - \mu_A^{\eta^{y,-}}\|_z = |K(R)|,$$

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

$$K(a) = \frac{\lambda_+}{a + \lambda_+} - \frac{1}{a\lambda_- + 1}.$$

Proof: First, as was already explained in the proof of Proposition 4.2, we may assume w.l.o.g. that the edge between y and z is the only one connecting y to A. As in the previous proof, we abbreviate $\mu_A^{\eta^{y,+}}$, $\mu_A^{\eta^{y,-}}$ and $\mu_A^{\eta^{y,*}}$ to μ_A^+, μ_A^- and μ_A^* respectively. Thus $\|\mu_A^{\eta^{y,+}} - \mu_A^{\eta^{y,-}}\|_z = |\mu_A^+(\sigma_z = +) - \mu_A^-(\sigma_z = +)|$, 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 $\frac{\mu_A^-(\sigma_z = -)}{\mu_A^-(\sigma_z = +)}$. Again, since the only influence of y on A is through z, we have $R^+ = \frac{R}{\lambda_+}$ and $R^- = R\lambda_-$. The proposition now follows once we notice that, by definition of R^+ and R^-, $\mu_A^+(\sigma_z = +) = \frac{1}{R^+ + 1}$ and $\mu_A^-(\sigma_z = +) = \frac{1}{R^- + 1}$. $\square$

The next step is generalizing the recursive calculation of the magnetization at the root of full subtrees. As we did for the Ising model, fix a boundary configuration τ, and for a site z with parent y let $R_{z,\ell} = \frac{p(\sigma_z = -)}{p(\sigma_z = +)}$, where $p(\cdot) = \mu_{B_{z,\ell}}^{\tau^{y,*}}(\cdot)$. (If z is the root of $\mathbb{T}^b$ then $p(\cdot) = \mu_{B_{z,\ell}}^{\tau}(\cdot)$.) A direct calculation (similar to that for the Ising model) gives that $R_{z,\ell} = \lambda \prod_{x \prec z} F(R_{x,\ell-1})$, where $F(a) = \frac{\lambda_- a + 1}{a + \lambda_+}$. In particular, if τ is the all-$(+)$ configuration (i.e., $R_{x,0} = 0$ for all x) then $R_{z,1} = \lambda[F(0)]^b$. Again, we let $F(\infty) \equiv \lambda_-$ and notice that indeed, if τ is the all-$(-)$ configuration (i.e., $R_{x,0} = \infty$ for all x), then $R_{z,1} = \lambda[F(\infty)]^b$. As before, this motivates us to define

$$J(a) = \lambda[F(a)]^b, \tag{11}$$

where we notice that $R_{z,\ell}$ equals $J^{(\ell)}(0)$ and $J^{(\ell)}(\infty)$ for τ all-$(+)$ and all-$(-)$ respectively.

We now observe that the relationship established for the Ising and hard-core models between K and the derivative of J holds for general two-spin systems. In particular, a straightforward calculation verifies that $K(a) = aF'(a)/F(a)$ for all a, and therefore

$$J'(a) \;=\; bJ(a)\frac{F'(a)}{F(a)} \;=\; b \cdot \frac{J(a)}{a} \cdot K(a). \tag{12}$$

The proof of Theorem 6.1 will be concluded once we notice that the function J here has the same properties and relationships with the uniqueness of the Gibbs measure as those mentioned earlier for the Ising and hard-core models. To this end, we separate the discussion into two classes of systems, where we think of the Ising and hard-core models as typical representatives of each class. We say that a system is *ferromagnetic* (respectively, *antiferromagnetic*) if $\lambda_- \cdot \lambda_+ > 1$ (respectively, if $\lambda_- \cdot \lambda_+ < 1$). Notice that in a ferromagnetic system neighboring spins are positively correlated, i.e., the spin at site z is more likely to be $(+)$ conditioned on its neighbor being $(+)$ than conditioned on its neighbor being $(-)$. In an antiferromagnetic system, the opposite effect takes place. Indeed, if $\lambda_- \cdot \lambda_+ = 1$ then the spin at z is independent of the spins of its neighbors. Notice that in the latter case the Gibbs distribution is a product distribution, and trivially $\kappa = \gamma = 0$ for all boundary conditions.

We now describe the properties of the function J, first for ferromagnetic and then for antiferromagnetic systems.

6.1 Ferromagnetic systems

For a ferromagnetic system, the function J has all the properties we described earlier in the Ising model case. Specifically:

1. J is continuous and increasing on $[0, \infty)$, with $J(0) = \lambda(1/\lambda_+)^b > 0$ and $\sup_a J(a) = \lambda(\lambda_-)^b < \infty$.
2. J has at least one fixed point in $[0, \infty)$. The fixed point is unique if and only if the system admits a unique Gibbs measure.
3. $J'(a_0) \le 1$, where a_0 denotes the least fixed point of J. In particular, if the Gibbs measure is unique then the derivative at the unique fixed point is $J'(a_0) \le 1$.

4. If a_0 is not the unique fixed point then $J'(a) \leq 1$ for $a \in [0, a_0]$. (This follows from the fact that J has a single point of inflection a_*, i.e., the derivative J' is monotonically increasing on $[0, a_*)$ and decreasing on $[a_*, \infty)$ for some $a_* \in \mathbb{R}^+$, which also means that J has at most three fixed points.)

We note that the relationship with the uniqueness of the Gibbs measure follows from the fact that the all-$(-)$ and all-$(+)$ configurations are the *minimal* and *maximal* boundary conditions respectively, as in the case of the Ising model, i.e., for any boundary condition the probability that the spin at z is $(+)$ is bounded below and above by the same probability under the all-$(-)$ and all-$(+)$ boundary conditions respectively.

Now, part (i) of Theorem 6.1 follows by the same argument used for the Ising model in the uniqueness regime (see the proof of Theorem 4.1) since the variation distance for the fixed-point magnetization $|K(a_0)| = \frac{1}{b}|J'(a_0)| \leq \frac{1}{b}$. For part (ii) we repeat the observation that, for τ the all-$(+)$ configuration, $\kappa \equiv \kappa(\{\mu_T^\tau\}) = \sup_{\ell \geq 1} K[J^{(\ell)}(0)]$. Since $|K(a)| = \frac{1}{b}\frac{J(a)}{a}|J'(a)| \leq \frac{1}{b}$ for $a \in [0, a_0]$, this implies $\kappa \leq \frac{1}{b}$. Notice that for ferromagnetic systems we get the stronger result that κ is bounded rather than just κ_2. (Clearly $\kappa_2 \leq \kappa$ always.) This completes the proof of Theorem 6.1 for ferromagnetic systems.

6.2 Antiferromagnetic systems

For an antiferromagnetic system, the function J has all the properties described earlier for the hard-core model, and we again introduce the function $J_2(a) \equiv J(J(a))$. Specifically, we have:

1. J is continuous and decreasing on $[0, \infty)$, with $0 < J(0) = \lambda(1/\lambda_+)^b < \infty$ and $J(\infty) = \lambda(\lambda_-)^b \geq 0$.

2. J has a unique fixed point a_0.

3. The derivatives of J alternate signs (the odd derivatives are negative).

4. J_2 is continuous and increasing on $[0, \infty)$, with $J_2(0) = J(J(0)) > J(\infty) \geq 0$ and $\sup_a J_2(a) \leq J(0) < \infty$.

5. a_0 (the unique fixed point of J) is a fixed point of J_2.

6. If the system admits a unique Gibbs measure then a_0 is the unique fixed point of J_2. If there are multiple Gibbs measures then J_2 has three fixed points $a_1 < a_0 < a_2$, where $J(a_1) = a_2$ and $J(a_2) = a_1$.

7. The derivative $J_2'(a)$ is continuous. If a_0 is the unique fixed point of J_2 (the Gibbs measure is unique) then $J_2'(a_0) \leq 1$. If there are three fixed points then $J_2'(a_0) > 1$, and $J_2'(a) < 1$ for $a \in [0, a_1]$. Consequently, the system admits a unique Gibbs measure if and only if $|J'(a_0)| = \sqrt{J_2'(a_0)} \leq 1$.

Again, the connection with the uniqueness of the Gibbs measure stems from the fact that the all-$(+)$ and all-$(-)$ are extreme configurations (though, as in the hard-core model, the direction of the monotonicity depends on the parity of the depth of the tree).

We also observe that

$$\begin{aligned} J_2'(a) &= J'(J(a)) \cdot J'(a) \\ &= b \cdot \frac{J(J(a))}{J(a)} \cdot K(J(a)) \cdot b \cdot \frac{J(a)}{a} \cdot K(a) \\ &= b^2 \cdot \frac{J_2(a)}{a} \cdot K(J(a))K(a). \end{aligned}$$

Now, part (i) of Theorem 6.1 (the uniqueness case) follows by the same argument used for the ferromagnetic case, since $|K(a_0)| = \frac{1}{b}|J'(a_0)| \le \frac{1}{b}$ when the Gibbs measure is unique. Part (ii) of Theorem 6.1 follows from the same arguments used for the hard-core model in the non-uniqueness regime, once we notice that, as in the hard-core model, for τ the all-$(+)$ configuration, $\kappa \equiv \kappa(\{\mu_T^\tau\}) = \sup_{\ell\ge 1} |K[J(J_2^{(\ell)}(0))]| \cdot |K[J_2^{(\ell)}(0)]|$. This completes the proof of Theorem 6.1 for antiferromagnetic systems. $\square$

7 The antiferromagnetic Potts model (colorings)

In this section we will prove Theorem 1.5: that the mixing time of the Glauber dynamics for sampling proper colorings is $O(n \log n)$ for all boundary conditions provided that the number of colors $q \ge b+3$, i.e., whenever the infinite-volume Gibbs measure is unique, except for the critical value $q = b+2$. In order to again make use of the machinery from Section 3 involving κ and γ, we prove the following:

Theorem 7.1 *For the colorings model with q colors, if the infinite-volume Gibbs measure is unique then for every $\varepsilon > 0$ we have $\widehat{\gamma} \le \frac{1}{q-1} + \varepsilon$ (for a suitable choice $d = d(\varepsilon)$ of the implicit constant in $\widehat{\gamma}$).*

Since in [14] it was shown that the Gibbs measure is unique for all $q \ge b+2$, we conclude that for these values of q, $\widehat{\gamma} \le \frac{1}{q-1} + \varepsilon < \frac{1}{b}$. Theorem 1.5 now follows from Theorem 3.3$''$ as usual.

Remarks:

- Notice that when $q = b+2$, even though $\widehat{\gamma} < 1/b$ (and thus $\widehat{\gamma}\widehat{\kappa}b < 1$), we cannot deduce $O(n \log n)$ mixing time because, for this value of q, the Glauber dynamics is not connected for some boundary conditions. (Specifically, this can occur for a subtree with appropriate boundary colors on the parent of the root as well as on the leaves.) Recall that our general framework for $O(n \log n)$ mixing time, presented in Theorem 3.3 and its variants, requires that the dynamics be connected for all sub-regions and all boundary conditions; if this condition does not hold, then we cannot deduce a mixing time of $O(n \log n)$ for any boundary condition (even one for which the dynamics is in fact connected). See [30] for more on the relevance of this condition to our framework.
- If we consider a slightly modified heat-bath dynamics, where at each step the configuration of a random *edge* (rather than the spin of a random single site) is updated, then the fact that $\widehat{\gamma} < \frac{1}{b}$ even for the critical value $q = b+2$ implies $O(n \log n)$ mixing time uniformly in the boundary condition, since the edge dynamics remains connected for $q = b+2$. See [30] for details.
- The fact that $\widehat{\gamma} < \frac{1}{b}$ means not only that the influence of *any* boundary configuration on the spin at the root decays with the distance of the boundary from the root (as is already implied by the fact that the Gibbs measure is unique), but that it decays *exponentially fast*. This fact is of independent interest, and to the best of our understanding was not obvious from the proof of uniqueness given in [14].

Proof of Theorem 7.1: The idea of the proof is the following. Consider a subset A, a site $y \in \partial A$ and $z \in A$, where z is a neighbor of y. Fix a disagreement at y, i.e., set two different spins s_1 and s_2 at y. We wish to bound $\|\mu_A^{\eta^{y,s_1}} - \mu_A^{\eta^{y,s_2}}\|_z$. Suppose now that, without the condition at y, the color of z is equally likely to be any of the q possible colors (as is the case when the boundary around A is "free"). Thus, the distribution of the color of z in $\mu_A^{\eta^{y,s_1}}$ is uniform over the $q-1$ colors other than s_1, and in $\mu_A^{\eta^{y,s_2}}$ it is uniform over the colors other than s_2. We then get that $\|\mu_A^{\eta^{y,s_1}} - \mu_A^{\eta^{y,s_2}}\|_z = \frac{1}{q-1}$, because we can couple the two distributions such that spin s_2 in $\mu_A^{\eta^{y,s_1}}$ is coupled with spin s_1 in $\mu_A^{\eta^{y,s_2}}$ and the two spins at z agree otherwise. (It is easy to see that this is the optimal coupling.) In our proof, we use the hypothesis that the Gibbs measure is unique in order to approximate the Gibbs distribution over A under any boundary configuration by the free boundary case, and hence get that the variation distance is arbitrarily close (as a function of the implicit parameter d) to $\frac{1}{q-1}$.

Let us proceed with the formal proof. Recall that in order to bound $\widehat{\gamma}$, we need to consider a subset A that includes the full subtree of depth d rooted at z, and bound the variation distance $\max_{s_1,s_2} \|\mu_A^{\eta^{y,s_1}} - \mu_A^{\eta^{y,s_2}}\|_z$ for an arbitrary boundary configuration η, where y is the (unique) neighbor of z in ∂A.

Now it is easy to see that, for the colorings model,

$$\|\mu_A^{\eta^{y,s_1}} - \mu_A^{\eta^{y,s_2}}\|_z = \max\{\mu_A^{\eta^{y,s_1}}(\sigma_z = s_2), \mu_A^{\eta^{y,s_2}}(\sigma_z = s_1)\}.$$

This identity follows from an argument similar to that used above for the free boundary case, as we now explain. Observe that, for every color s that differs from both s_1 and s_2, we have $\mu_A^{\eta^{y,s_1}}(\sigma_z = s) = \frac{1-\mu_A^{\eta^{y,s_1}}(\sigma_z=s_2)}{1-\mu_A^{\eta^{y,s_2}}(\sigma_z=s_1)} \cdot \mu_A^{\eta^{y,s_2}}(\sigma_z = s)$. Hence, if $\mu_A^{\eta^{y,s_1}}(\sigma_z = s_2) \geq \mu_A^{\eta^{y,s_2}}(\sigma_z = s_1)$, then $\mu_A^{\eta^{y,s_1}}(\sigma_z = s) \leq \mu_A^{\eta^{y,s_2}}(\sigma_z = s)$ for every $s \neq s_2$, and so the event $\mathcal{E} = \{\sigma_z = s_2\}$ maximizes the expression $|\mu_A^{\eta^{y,s_1}}(\mathcal{E}) - \mu_A^{\eta^{y,s_2}}(\mathcal{E})|$ over all events $\mathcal{E}$ that only depend on σ_z.

As in the previous models we analyzed, it is now convenient to consider the distribution induced by removing the edge from z to y (i.e., with a "free" condition at y). Recall that this distribution is denoted $\mu_A^{\eta^{y,*}}$. Let $p_z(s) = \mu_A^{\eta^{y,*}}(\sigma_z = s)$, and notice that for the colorings model $\mu_A^{\eta^{y,s_1}}(\sigma_z = s_2) = \frac{p_z(s_2)}{1-p_z(s_1)}$ simply because $\mu_A^{\eta^{y,s_1}}(\,\cdot\,) = \mu_A^{\eta^{y,*}}(\cdot \,|\, \sigma_z \neq s_1)$. Thus, $\max_{s_1,s_2} \|\mu_A^{\eta^{y,s_1}} - \mu_A^{\eta^{y,s_2}}\|_z = \max_{s_1,s_2} \mu_A^{\eta^{y,s_1}}(\sigma_z = s_2) = \max_{s_1,s_2} \frac{p_z(s_2)}{1-p_z(s_1)}$.

To obtain the claimed bound on $\widehat{\gamma}$ we have to show that, for all sets A as above, and all boundary configurations η, $\max_{s_1,s_2} \frac{p_z(s_2)}{1-p_z(s_1)} \leq \frac{1}{q-1} + \varepsilon$. It is at this point that we use the assumption that the Gibbs measure is unique. This means that, if d (the depth of the full subtree contained in A) is large enough, the distribution $p_z(\,\cdot\,)$ is arbitrarily close to the uniform distribution, regardless of the boundary configuration η. Thus, for every $\varepsilon' > 0$, there exists a (large enough) constant d such that $p_z(s) \leq \frac{1+\varepsilon'}{q}$ for all colors s. Hence, $\max_{s_1,s_2} \frac{p_z(s_2)}{1-p_z(s_1)} \leq \frac{1+\varepsilon'}{q-1-\varepsilon'} \leq \frac{1}{q-1} + \varepsilon$ for some ε that goes to zero as $d \to \infty$, as required. $\square$

8 The ferromagnetic Potts model

In this section we will prove that, for the Potts model, the mixing time is $O(n \log n)$ in all the situations described in Theorem 1.4. As in the discussion of other models in previous sections, we use the machinery from Section 3. The following theorem sets out the relevant properties of κ and γ.

Theorem 8.1 *For the Potts model with q colors at inverse temperature β the following hold*:

(i) $\gamma \le \frac{e^{2\beta}-1}{e^{2\beta}+1}$;

(ii) *if the Gibbs measure is unique (i.e., $\beta < \beta_0$) then $\widehat{\gamma} \le \frac{e^{2\beta}-1}{e^{2\beta}+q-1} + \varepsilon < \frac{1}{b}$ (for a small enough ε that depends on the choice of d, the implicit constant in $\widehat{\gamma}$)*;

(iii) *if the Gibbs measure is not unique and the boundary condition τ is constant then $\kappa \equiv \kappa(\{\mu_T^\tau\}) \le \frac{1}{b}$*;

(iv) *if the boundary condition τ is free then $\kappa = \frac{e^{2\beta}-1}{e^{2\beta}+q-1}$.*

Part (i) of Theorem 1.4 follows from parts (i) and (ii) of Theorem 8.1 and the fact that, for any boundary condition, $\kappa \le \gamma$. Part (ii) of Theorem 1.4 follows from parts (ii) and (iii) of Theorem 8.1 and the fact that $\gamma < 1$ (as is apparent from part (i) of the same theorem). Finally, part (iii) of Theorem 1.4 follows from parts (i) and (iv) of Theorem 8.1.

Proof of Theorem 8.1: Much as we did for the previously discussed models, the first step we take in order to bound κ and γ is expressing the influence of a boundary spin as a function of the distribution of its neighboring interior spin, when the boundary spin is free. Generalizing Proposition 4.2 from the Ising model to the Potts model gives:

Proposition 8.2 *For any subset $A \subseteq T$, any boundary configuration η, any pair of spins (s_1, s_2), any site $y \in \partial A$ and any neighbor $z \in A$ of y, we have*

$$\|\mu_A^{\eta^{y,s_1}} - \mu_A^{\eta^{y,s_2}}\|_z = K(p(s_1), p(s_2)),$$

where $p(s) = \mu_A^{\eta^{y,*}}(\sigma_z = s)$ and the function K is defined by

$$K(p_1, p_2) = \max\left\{\frac{e^{2\beta}p_1}{(e^{2\beta}-1)p_1+1} - \frac{p_1}{(e^{2\beta}-1)p_2+1}, \frac{e^{2\beta}p_2}{(e^{2\beta}-1)p_2+1} - \frac{p_2}{(e^{2\beta}-1)p_1+1}\right\}, \quad (13)$$

and we notice that $K(p_1, p_2)$ is the first term in the maximum if and only if $p_1 \ge p_2$.

Proof: Let $p^s(s') = \mu_A^{\eta^{y,s}}(\sigma_z = s')$. Then $p^s(s) = \frac{e^{2\beta}p(s)}{e^{2\beta}p(s)+1-p(s)} = \frac{e^{2\beta}p(s)}{(e^{2\beta}-1)p(s)+1}$, and for $s' \ne s$, $p^s(s') = \frac{p(s')}{e^{2\beta}p(s)+1-p(s)} = \frac{p(s')}{(e^{2\beta}-1)p(s)+1}$, by definition of the Potts model. Now, the proposition follows once we notice that $\|\mu_A^{\eta^{y,s_1}} - \mu_A^{\eta^{y,s_2}}\|_z = \max\{p^{s_1}(s_1) - p^{s_2}(s_1), p^{s_2}(s_2) - p^{s_1}(s_2)\}$. The reason for this equality is that if $p(s_1) \ge p(s_2)$ then $p^{s_1}(s) \le p^{s_2}(s)$ for all $s \ne s_1$ (and $p^{s_1}(s_1) \ge p^{s_2}(s_1)$), as a simple calculation verifies. $\square$

We can now easily dispense with parts (ii) and (iv) of Theorem 8.1. For part (iv) we simply observe that, for τ the free boundary condition, the distribution $p(\cdot)$ at the root of the tree is uniform, i.e., $p(s) = \frac{1}{q}$ for all s, and therefore $\kappa = K(\frac{1}{q}, \frac{1}{q}) = \frac{e^{2\beta}-1}{e^{2\beta}+q-1}$, as required. For part (ii), observe that when the Gibbs measure is unique the distribution of the spin at the root of the tree converges as the depth increases to the uniform distribution, uniformly in the boundary condition. Thus, if A includes the full subtree of depth d rooted at z then, for every s, $p(s)$ converges with d to $\frac{1}{q}$, i.e., $\widehat{\gamma} \le K(\frac{1}{q} + \varepsilon', \frac{1}{q} + \varepsilon') = K(\frac{1}{q}, \frac{1}{q}) + \varepsilon \le \frac{e^{2\beta}-1}{e^{2\beta}+q-1} + \varepsilon$ for some ε' and ε that go to zero as $d \to \infty$, as required. (Notice that the Gibbs measure is unique in the regime $\beta < \beta_0$, for which $\frac{e^{2\beta}-1}{e^{2\beta}+q-1} < \frac{1}{b}$.)

We go on to prove part (i). Here we have to consider an arbitrary boundary configuration, and we cannot assume the distribution $p(\cdot)$ is close to uniform because the bound should also apply when the Gibbs distribution is not unique. Thus, the approach we take is to simply calculate the maximum of $K(p_1,p_2)$ over all possible distributions. As a first step, let $K_{\max}(a) = \max_{p_1,p_2}\{K(p_1,p_2) : p_1+p_2=a\}$ denote the maximum restricted to distributions in which the sum of the two probabilities is a. We now observe that $K_{\max}(a)$ is strictly increasing in a. This is a consequence of the following two facts. First, $K(p,p)$ is strictly increasing in p. Second, for $p_1>p_2$, $K(p_1,p_3) > K(p_1,p_2)$ for every $p_3 \in (p_2,p_1]$. We therefore conclude that $\max_{p_1,p_2} K(p_1,p_2) = K_{\max}(1) = \max_p\left\{\frac{e^{2\beta}p}{(e^{2\beta}-1)p+1} - \frac{p}{(e^{2\beta}-1)(1-p)+1}\right\}$. It is now easy to verify that the expression in the maximization on the r.h.s. coincides with K_β defined for the Ising model in Proposition 4.2, under the change of variables p to $R=\frac{1-p}{p}$, and that this expression is maximized for $p=\frac{1}{2}$. We therefore conclude that $\gamma \le \max_{p_1,p_2} K(p_1,p_2) = K(\frac{1}{2},\frac{1}{2}) = \frac{e^{2\beta}-1}{e^{2\beta}+1}$.

We now go on to prove part (iii) of Theorem 8.1 (the last remaining part). Here the boundary is constant. W.l.o.g. we assume it is all 1. In order to bound κ we need to consider the distribution $p(\cdot)$ of the spin at the root of maximal subtrees with boundary condition all 1 at the bottom. Notice that, by symmetry, $p(s)$ is uniform in $s\neq 1$. Therefore, this distribution is completely specified by $p(1)$ since for $s\neq 1$, $p(s)=\frac{1-p(1)}{q-1}$. As we shall see below, the fact that the distribution at the root is one-parameter allows for an analysis that is similar to that carried out for two-spin systems in Section 6.

To start, notice that by Proposition 8.2, $\|\mu^1_{T_z}-\mu^s_{T_z}\| = K(p(1),\frac{1-p_z(1)}{q-1})$, where we recall that $\mu^s_{T_z}=\mu^{\eta^{y,s}}_{T_z}$, and where $p_z(s)=\mu^{\eta^{y,*}}_{T_z}(\sigma_z=s)$. Similarly, for s_1,s_2 both different from 1, $\|\mu^{s_1}_T - \mu^{s_2}_T\| = K(\frac{1-p_z(1)}{q-1},\frac{1-p_z(1)}{q-1})$. We now observe that, since the system is ferromagnetic, $p_z(1)\ge\frac{1}{q}$ for all z. An explicit calculation reveals that for every $p\ge\frac{1}{q}$, $K(p,\frac{1-p}{q-1}) \ge K(\frac{1-p}{q-1},\frac{1-p}{q-1})$. Thus in order to bound κ, it is enough to bound $K(p_z(1),\frac{1-p_z(1)}{q-1})$ for every z. It is now convenient to consider the ratio $R=\frac{1-p(1)}{p(1)}$ and define

$$K(R) \;\equiv\; K\left(p(1),\frac{1-p(1)}{q-1}\right) \;=\; \frac{e^{2\beta}}{e^{2\beta}+R} - \frac{1}{(\frac{e^{2\beta}+q-2}{q-1})R+1}\,. \tag{14}$$

Observe that $\kappa=\sup_T\sup_z K(R_z)$. Thus, we need to show that this supremum is at most $\frac{1}{b}$. We now use the fact that the distribution at the root is one-parameter once again, this time to recursively calculate R_z, as we did earlier for two-spin systems. In particular, we notice that $R_z = (q-1)\prod_{w\prec z}F(R_w)$, where $F(a)=((\frac{e^{2\beta}+q-2}{q-1})a+1)/(a+e^{2\beta})$. As in the two-spin case, we let $J(a)=(q-1)[F(a)]^b$ and observe that $R_z=J^{(\ell)}(0)$, where ℓ is the height of z. Notice that the functions $K(a)$ and $J(a)$ as defined here correspond exactly to the same two functions, defined for the two-spin system with parameters $(\lambda,\lambda_-,\lambda_+)=(q-1,\,\frac{e^{2\beta}+q-2}{q-1},\,e^{2\beta})$. In other words, if we translate the spin 1 of the Potts model to $(+)$ and any non-1 spin to $(-)$, the Potts model with all-1 boundary corresponds exactly to the above two-spin system with all-$(+)$ boundary. To get some intuition for this translation, notice that $\lambda=q-1$ arises from the fact that a $(-)$-spin represents $q-1$ different spins (uniformly weighted) of the Potts model. For the same reason, $\lambda_-=\frac{e^{2\beta}+q-2}{q-1}$ arises from the fact that, given a non-1 spin, the interaction with another non-1 spin is composed of a fraction $1/(q-1)$ times $e^{2\beta}$ (when the neighboring site is of the same Potts spin), and a fraction $(q-2)/(q-1)$ times 1 (for the interaction with the other $q-2$ spins).

Given this correspondence, it is now clear that κ for the all-1 boundary in the Potts model (the supremum over $K(R_z)$ given above) is exactly κ for the all-$(+)$ boundary in the two-spin system.

Thus, we can conclude the proof of Theorem 8.1 (iii) by using Theorem 6.1 (ii) for the two-spin system, once we show that non-uniqueness of the Gibbs measure for the Potts model implies non-uniqueness of the Gibbs measure in the two-spin system.

Now, if the Gibbs measure is not unique for Potts then, conditioned on the all-1 boundary configuration, the probability that the spin at the root is 1 is greater than $1/q$ for arbitrarily large trees (since the model is ferromagnetic). This implies that, in the two-spin system, the probability of $(+)$ is $> 1/q$, which immediately implies non-uniqueness in this system because $a_* = q-1$ is always a fixed point of the function J defined above. (In fact, it is not too difficult to see that the reverse implication holds as well, i.e., that the Gibbs measure is unique for the Potts model with parameters (q, β) if and only if it is unique for the corresponding two-spin system.)

We conclude that, for the Potts model in the regime of non-uniqueness of the Gibbs measure and for a constant boundary configuration, $\kappa \leq \frac{1}{b}$. This completes the proof of Theorem 8.1 part (iii). $\square$

9 Reconstruction problems

In this section we discuss reconstruction problems on trees, and show how the techniques we developed earlier can be applied to establish new results in this context. In the models considered here, information is transmitted from the root to all other nodes along the edges of the tree, with a certain probability of error in the value transmitted on each edge (a branching random walk). These error probabilities are the same for all edges, and errors on different edges are independent. Models of this kind arise in genetics, describing the propagation of a genetic property from an ancestor to its descendants, as well as in communication theory, where the process represents a tree network in which each edge is a noisy channel. Usually, the question of interest is the following: for the given error probabilities, can the value at the root be reconstructed from the values at a level far below? In other words, is there significant correlation between the value at the root and the values at the leaves?

This is where our earlier framework comes into play. In our development of the criterion for rapid mixing based on κ, γ (given in detail in our previous paper [20] and summarized in Section 3 of the present paper), we showed as an intermediate step that, under the same criterion, the correlation between the spin at the root and the configuration at the leaves goes to zero as the depth of the tree tends to infinity. In this section we will adapt this theory to the reconstruction setting, and thus establish ranges of values of the error probabilities for which reconstruction is impossible, using the same κ, γ criterion. Notice that here, in contrast to previous sections, we are not concerned with dynamical questions (i.e., convergence to equilibrium), but only with properties of the equilibrium state.

The rest of this section is organized as follows: we first define the information flow model precisely, then rewrite κ and γ in terms of this model, and finally apply the criterion to specific choices of channels.

9.1 Definitions

We consider networks on the infinite tree $\mathbb{T}^b$. A *q-state channel* is given by a stochastic matrix $\mathbf{P} = \{p_{ij} : i, j \in S\}$, where $S = \{1, \ldots, q\}$ is the alphabet of possible symbols. We always assume that $\mathbf{P}$ defines an ergodic (irreducible and aperiodic) Markov chain, and therefore has a unique stationary distribution $\pi = (\pi_1, \ldots, \pi_q)$, with $\pi_i > 0$ for all i. A configuration $\sigma \in S^{\mathbb{T}^b}$ is then generated according to the following process. First, the symbol at the root is chosen according to π. Thereafter, the configuration is generated recursively: if a site $x \in \mathbb{T}^b$ was assigned the symbol $i \in S$ then each of its children is independently assigned a symbol j, where for each $j \in S$ the probability of assigning j is p_{ij}. We denote by ν the probability measure on $S^{\mathbb{T}^b}$ resulting from this process.

Let us now discuss some properties of the measure ν. First, since the symbol at the root is chosen from the stationary distribution π, for every $x \in \mathbb{T}^b$ the marginal of ν on σ_x is also π. Second, by definition, the projection of ν on S^T, where T is any initial portion of $\mathbb{T}^b$, is exactly the same as the distribution resulting from executing the same process on T. Third, since the process is Markovian, it is not hard to check that ν is a Gibbs measure: i.e., for any finite subset $A \subset \mathbb{T}^b$ and boundary configuration η, letting $\nu_A^\eta \equiv \nu(\cdot \mid \eta_{\mathbb{T}^b \setminus A})$, the marginal of ν_A^η on S^A depends only on $\eta_{\partial A}$. In fact, ν_A^η can be constructed equivalently as follows. Without loss of generality assume that A is connected, and let x be the node in $A \cup \partial A$ that is closest to the root of $\mathbb{T}^b$ (i.e., x is the root of $A \cup \partial A$). Run the process described above on $A \cup \partial A$, starting at x. Then ν_A^η is the resulting distribution conditioned on the event that the configuration of ∂A is given by η.[11]

We can also relate ν to an (infinite-volume) Gibbs measure associated with a spin system as defined in Section 2. Let $\mathbf{P} = \{p_{ij}\}$ be a channel and π its stationary distribution. Define the edge potential $U(i, j) = -\ln(p_{ij}/\pi_j)$ and the site potential $W(i) = -\ln \pi_i$. These edge potentials are symmetric, in accordance with our definition of spin systems in Section 2, provided the channel is *reversible*, i.e., it satisfies the detailed balance conditions $\pi_i p_{ij} = \pi_j p_{ji}$ for all $i, j \in S$. For this reason we will limit our attention in this paper to reversible channels (and all the examples we discuss will have this property).[12] Now it is not hard to check (see, e.g., [9, Chapter 12]) that ν is a Gibbs measure for the spin system with the above potentials. As an example, if $\mathbf{P}$ is the *binary symmetric channel*, in which $q = 2$ and $p_{ii} = p \geq \frac{1}{2}$ for $i = 1, 2$, then ν is the free Gibbs measure of the Ising model with inverse temperature $\beta = \frac{1}{2}\ln(\frac{p}{1-p})$ and no external field ($h = 0$).[13]

Before we continue, we note that reversible channels have another property which makes the above correspondence with spin systems of interest. First, notice that reversibility implies that the distribution ν_A^η can be generated as described above, except that now the node x at which we start the process can be any node in $A \cup \partial A$ (with the edges of $A \cup \partial A$ reoriented so that x becomes the root). Consider now the information flow process on the *Bethe lattice* rather than on the tree $\mathbb{T}^b$. Recall that the Bethe lattice differs from $\mathbb{T}^b$ only in that the root has $b + 1$ children, rather than b; i.e., it is regular of degree $b + 1$. We will denote the Bethe lattice $\widehat{\mathbb{T}}^b$. Since in our definition of the information flow process the value at the root is chosen from the stationary distribution of the channel, for reversible channels on $\widehat{\mathbb{T}}^b$ the resulting measure ν is *translation-invariant*, and in particular, is a translation-invariant Gibbs measure of the corresponding spin system. Translation-invariant Gibbs measures are of special interest in statistical physics, and this is another motivation for considering the information flow models described in this section: they give convenient representations of these special measures.

Now that the information flow process is defined, we can formulate the reconstruction problem

[11]In our original formulation ν_A^η is defined over the infinite configuration space, while here it is defined only over $S^{A \cup \partial A}$. However, this is a trivial point since in ν_A^η the configuration outside A is deterministically set to η.

[12]Our framework also applies to non-reversible channels, which give rise to asymmetric edge potentials. In turn, one can make sense of spin systems with asymmetric potentials by viewing the tree as a directed graph with edges oriented away from the root.

[13]The reader may notice that applying the above formula for translating channels to potentials to the binary symmetric channel yields potentials that differ somewhat from those in the definition of the Ising model in Section 1. However, it will be seen that the difference is the addition of a constant to the edge potential, and a (different) constant to the site potential. It is easy to check that such shifts merely multiply the weights of all configurations in the Gibbs distribution by a uniform factor, and hence do not change the spin system.

precisely. Naturally, reconstructing the value at the root from values deep down the tree is possible only if the value at the root is correlated with those from which we try to reconstruct it. Several equivalent formulations of this correlation exist in the literature. We give here the simplest formulation (for others see, e.g., [22, 18]). For a symbol $s \in S$, let ν^s be defined as ν conditioned on the event that the symbol at the root is s. For two probability measures μ_1, μ_2, let $\|\mu_1 - \mu_2\|_\ell$ denote the total variation distance between the projections of μ_1 and μ_2 on the configuration space of the ℓth level of the tree.

Definition 9.1 *We say that reconstruction is impossible for the channel $\mathbf{P}$ on $\mathbb{T}^b$ if for every $s_1, s_2 \in S$, $\|\nu^{s_1} - \nu^{s_2}\|_\ell$ tends to zero as $\ell \to \infty$.*

We mention in passing that, if we view ν as the Gibbs measure of an associated spin system, the fact that reconstruction is impossible for ν is equivalent to saying that ν is an *extremal* Gibbs measure of the spin system. (See, e.g., [9] for a discussion of extremal Gibbs measures.)

9.2 A criterion for non-reconstructibility

In our previous paper [20] we developed a new criterion for establishing decay of correlations for Gibbs measures on trees, based on the two quantities κ, γ. (This criterion also implies rapid mixing of the Glauber dynamics since, by an independent argument we gave in the same paper, decay of correlations implies rapid mixing. The framework of Section 3 in the present paper is based on this further implication; here, however, we are interested only in the decay of correlations part). Since the measure ν is also a Gibbs measure, it is relatively straightforward to see that the same criterion can be used to establish decay of correlations for ν, i.e., that reconstruction is impossible. In order to use this criterion, we first redefine κ, γ in terms of the information flow model.

For a symbol $s \in S$ and a node $z \in \mathbb{T}^b$ other than the root, let $\nu^s_{T_z}$ be the distribution ν conditioned on the event that the symbol at the parent of z is s. Rewriting the definitions of κ and γ (Definition 3.1), we have:

Definition 9.2 *Let $\mathbf{P}$ be a channel on the $\mathbb{T}^b$, and ν its associated Gibbs measure as derived above. Then the quantities $\kappa \equiv \kappa(\nu)$ and $\gamma \equiv \gamma(\nu)$ are defined by*

(i) $\kappa = \sup_{z \in \mathbb{T}^b} \max_{s,s'} \|\nu^s_{T_z} - \nu^{s'}_{T_z}\|_z$;

(ii) $\gamma = \sup_{A \subset \mathbb{T}^b} \max \|\nu^{\eta^{y,s}}_A - \nu^{\eta^{y,s'}}_A\|_z$, *where the maximum is taken over all boundary conditions η,*[14] *all sites $y \in \partial A$, all neighbors $z \in A$ of y, and all spins $s, s' \in S$.*

Notice from the definition of ν that κ takes the particularly simple form $\kappa = \max_{i,j \in S} \frac{1}{2} \sum_k |p_{ik} - p_{jk}|$; in the case of a binary channel ($q = 2$), this becomes $|p_{11} - p_{21}| = |p_{12} - p_{22}|$. For γ we do not have such a clean general formula, and will need to bound it separately for specific models later on.

We can now state our criterion for non-reconstructibility, which is analogous to our criterion for rapid mixing from Section 3.

Theorem 9.3 *Consider an arbitrary (ergodic and permissive) channel $\mathbf{P}$ on $\mathbb{T}^b$. If $\kappa \equiv \kappa(\nu)$ and $\gamma \equiv \gamma(\nu)$ satisfy $\gamma\kappa b < 1$ then reconstruction is impossible for $\mathbf{P}$ on $\mathbb{T}^b$.*

[14]Notice that ν^η_A may not be well-defined for some η if $p_{ij} = 0$ for some (i, j). For this reason we require that the channel be *permissive*, i.e., that the spin-system corresponding to it is permissive. It is easy to see that in this case ν^η_A is well-defined for every η if A is a connected subset.

Proof: We first note that Claim 3.2 (after making the appropriate notational modifications) holds in the present context as well since it only uses the fact that, in the measure under consideration, once a value is fixed at a site x the configuration on a subtree T_z, where z is a child of x, is independent of the configuration on the rest of the tree. It is easily seen that ν satisfies this property. We now refer to Theorem 4.3 of [20], which is based on a claim analogous to Claim 3.2 and thus applies to any measure for which Claim 3.2 holds.[15] This theorem states that ν satisfies the so-called "variance mixing" condition $\mathrm{VM}(\ell, \epsilon(\ell))$ with $\epsilon(\ell) \to 0$ as $\ell \to \infty$. (VM is defined in [20, Definition 3.1].) But by standard arguments this easily implies that $\|\nu^{s_1} - \nu^{s_2}\|_\ell$ goes to zero as $\ell \to \infty$ for all $s_1, s_2 \in S$, as required. $\square$

Remarks:

- The reader will notice that, in contrast to our earlier Theorem 3.3, the criterion in Theorem 9.3 involves only the product $\gamma\kappa b$ and not $\max\{\gamma\kappa b, \gamma\}$; in other words, we no longer require that $\gamma < 1$. This is because we are aiming here only for decay of correlations as expressed by the variance mixing condition, whereas to ensure $O(n \log n)$ mixing time we need a stronger "entropy mixing" condition. The reader interested in the details of this issue may compare Theorems 4.3 and 5.1 of [20].
- We have stated Theorem 9.3 for the tree $\mathbb{T}^b$. It can be verified that the same criterion applies for the Bethe lattice $\widehat{\mathbb{T}}^b$ (where κ and γ are defined w.r.t. the associated Gibbs measure ν); this follows by translating the machinery of [20] from $\mathbb{T}^b$ to the (very similar) $\widehat{\mathbb{T}}^b$. In fact, notice that since κ depends only on the transition probabilities of the channel, it has the same value on both $\mathbb{T}^b$ and $\widehat{\mathbb{T}}^b$. Moreover, although γ may take different values in the two settings, the only possible source of this difference are distributions ν_A^η, where A includes the root (of either $\mathbb{T}^b$ or $\widehat{\mathbb{T}}^b$). But for reversible channels on $\widehat{\mathbb{T}}^b$, translation invariance means that in the definition of γ we can w.l.o.g. take the maximum over subsets A that do not include the root. This implies that the value of γ on $\widehat{\mathbb{T}}^b$ is at most that on $\mathbb{T}^b$, and hence if the criterion in Theorem 9.3 holds for a channel $\mathbf{P}$ on $\mathbb{T}^b$ it also holds for $\mathbf{P}$ on $\widehat{\mathbb{T}}^b$. The same remark applies to Theorems 9.3′ and 9.3″ below.

As in the case of rapid mixing, we can also relax the above criterion slightly (following the same line of reasoning as in Section 3 of the present paper). Define $\kappa_2, \gamma_2, \widehat{\kappa}$ and $\widehat{\gamma}$ in an analogous way to Section 3.

Theorem 9.3′ *In the setting of Theorem 9.3, if κ_2 and γ_2 satisfy $\gamma_2\kappa_2 b < 1$ then reconstruction is impossible.*

Theorem 9.3″ *In the setting of Theorem 9.3, if $\widehat{\kappa}$ and $\widehat{\gamma}$ satisfy $\widehat{\gamma}\widehat{\kappa}b < 1$ then reconstruction is impossible.*

9.3 Applications

We now proceed to apply our criterion for various popular choices of channels. In most cases our results mirror the best bounds in the literature, and in some specific cases we even manage to push the bounds a little further. However, in all cases the main novelty of our argument is its simplicity: we need only calculate κ and γ and apply the criterion in Theorem 9.3 or its variants.

General binary channel
Here we give bounds that apply to any binary ($q = 2$) channel $\mathbf{P} = \{p_{11}, p_{12}, p_{21}, p_{22}\}$. In this generality, the best known bound is due to Martin [18]. We now show how essentially the same bound follows as an immediate consequence of our machinery.

[15] Again, strictly speaking this theorem is stated for the special case of the Ising model. However, it is easily seen that the theorem extends to an arbitrary permissive spin system on trees. We refer to the PhD thesis of the third author [30, Theorem 5.13] for the general formulation.

Theorem 9.4 *Let $\mathbf{P}$ be a binary channel. Reconstruction for $\mathbf{P}$ on $\mathbb{T}^b$ is impossible provided that*

$$\left(\sqrt{p_{11}p_{22}} - \sqrt{p_{12}p_{21}}\right)^2 < \frac{1}{b}\,. \tag{15}$$

Remarks:

- In light of the second remark after the proof of Theorem 9.3, the same non-reconstructibility bound holds for the Bethe lattice $\widehat{\mathbb{T}}^b$.
- The bound in [18] is slightly stronger than ours since there the range of non-reconstructibility includes the case of equality in (15). On the other hand, our bound has implications for the asymptotic independence between the symbol at the root and the configuration on a level far below that are stronger than just non-reconstructibility (see [20] for details). This is also the reason our bound does not cover the case of equality in (15).
- It is known (see, e.g., [22]) that reconstruction *is* possible when $(p_{11}-p_{21})^2 > \frac{1}{b}$. Thus, for *symmetric*[16] binary channels the bound in Theorem 9.4 is tight. However, for asymmetric channels there is a gap, and there are specific channels well within this gap for which reconstruction is known to be possible and others for which it is known to be impossible.

In order to prove Theorem 9.4 we use a simple calculation that is summarized in the following lemma.

Lemma 9.5 *For a binary channel $\mathbf{P}$ on $\mathbb{T}^b$,*

$$\gamma \;\equiv\; \gamma(\nu) \;\le\; \frac{|\sqrt{p_{11}p_{22}} - \sqrt{p_{12}p_{21}}|}{\sqrt{p_{11}p_{22}} + \sqrt{p_{12}p_{21}}}. \tag{16}$$

Theorem 9.4 can now be derived using our criterion for non-reconstructibility in Theorem 9.3, together with the fact that for binary channels $\kappa = |p_{11} - p_{21}| = |p_{12} - p_{22}| = |p_{11}p_{22} - p_{12}p_{21}|$.

Proof of Lemma 9.5: First, notice that since we always assume the channel is ergodic we need only consider the case in which at most one of the transition probabilities $\{p_{11}, p_{12}, p_{21}, p_{22}\}$ is 0. Moreover, if exactly one of these probabilities is 0 then the expression on the r.h.s. of (16) equals 1, which is a trivial bound on γ. Therefore, from here onwards we assume w.l.o.g. that all four transition probabilities are strictly positive. Consider now an arbitrary finite subset A, an arbitrary boundary configuration η and two neighboring sites y, z, with $y \in \partial A$ and $z \in A$. We have to show that $\|\nu_A^{\eta^{y,1}} - \nu_A^{\eta^{y,2}}\|_z$ is bounded above by the r.h.s. of (16). As mentioned before, the distributions $\nu_A^{\eta^{y,1}}$ and $\nu_A^{\eta^{y,2}}$ can be generated by running the information flow process on $A \cup \partial A$ starting from y (where y is regarded as the root) and conditioning on the events that the configuration on ∂A is $\eta^{y,1}$ and $\eta^{y,2}$ respectively. Indeed, for the rest of this proof the probability space under consideration will be the one arising from running the information flow process on $A \cup \partial A$ starting from y. We write σ for the resulting random configuration and let $\partial^* A = (\partial A)\setminus y$. Since the channel is binary, it is enough to bound $|\Pr(\sigma_z = 1 \,|\, \sigma_y = 1, \sigma_{\partial^* A} = \eta) - \Pr(\sigma_z = 1 \,|\, \sigma_y = 2, \sigma_{\partial^* A} = \eta)|$. Using Bayes' rule, we have

$$\Pr(\sigma_z = 1 \,|\, \sigma_y = 1, \sigma_{\partial^* A} = \eta) \;=\; \frac{\Pr(\sigma_z = 1 \,|\, \sigma_y = 1)\Pr(\sigma_{\partial^* A} = \eta \,|\, \sigma_z = 1, \sigma_y = 1)}{\Pr(\sigma_{\partial^* A} = \eta \,|\, \sigma_y = 1)} \;=$$

$$\frac{p_{11} \cdot \Pr(\sigma_{\partial^* A} = \eta \,|\, \sigma_z = 1, \sigma_y = 1)}{p_{11} \cdot \Pr(\sigma_{\partial^* A} = \eta \,|\, \sigma_z = 1, \sigma_y = 1) \;+\; p_{12} \cdot \Pr(\sigma_{\partial^* A} = \eta \,|\, \sigma_z = 2, \sigma_y = 1)},$$

[16] A channel is symmetric iff the associated matrix $\mathbf{P}$ is symmetric.

with an analogous expression for $\Pr(\sigma_z = 1 \,|\, \sigma_y = 2, \sigma_{\partial^* A} = \eta)$. We now notice that

$$\frac{\Pr(\sigma_{\partial^* A} = \eta \,|\, \sigma_z = 2, \sigma_y = 1)}{\Pr(\sigma_{\partial^* A} = \eta \,|\, \sigma_z = 1, \sigma_y = 1)} \;=\; \frac{\Pr(\sigma_{\partial^* A} = \eta \,|\, \sigma_z = 2, \sigma_y = 2)}{\Pr(\sigma_{\partial^* A} = \eta \,|\, \sigma_z = 1, \sigma_y = 2)} \equiv R.$$

This is because, conditioned on a fixed value at y, the configurations on the subtrees rooted at the children of y are independent of each other and since, conditioned on a fixed value at z, the configuration on the subtree rooted at z is independent of y. We conclude that

$$\|\nu_A^{\eta^{y,1}} - \nu_A^{\eta^{y,2}}\|_z \;=\; \left|\frac{p_{11}}{p_{11} + p_{12}R} - \frac{p_{21}}{p_{21} + p_{22}R}\right|. \tag{17}$$

Now straightforward calculus verifies that, as a function of R, the maximum value of the r.h.s. of (17) is achieved at $R = \sqrt{\frac{p_{11}p_{21}}{p_{12}p_{22}}}$ and that this maximum value is $|\frac{\sqrt{p_{11}p_{22}} - \sqrt{p_{12}p_{21}}}{\sqrt{p_{11}p_{22}} + \sqrt{p_{12}p_{21}}}|$, thus completing the proof of the lemma. $\square$

Binary channel with one-sided deterministic error

We go on to consider a binary channel with one-sided deterministic error, i.e., $p_{22} = 0$. We write the probability of error for the first symbol as $p_{12} = \frac{w}{1+w}$, where w is a positive real number. On the Bethe lattice $\widehat{\mathbb{T}}^b$, it is not hard to check (using the recipe given in Section 9.1) that the measure ν arising from this channel corresponds to the unique translation-invariant Gibbs measure of the hard-core (independent sets) spin system with activity parameter $\lambda = w(1+w)^b$.[17] This measure has been widely discussed in the literature (see, e.g., [18, 5]); a further reason for discussing it here is that, for this specific binary channel, we will be able to improve on the general bound for binary channels given above. To get a better intuition for the measure ν, we note that it is the analog of the free-boundary measure for the Ising model with no external field. For a pictorial illustration, we refer the reader back to Figure 2: whereas the fixed points a_1, a_2 correspond to the limiting probabilities of occupation at the root under the "odd" and "even" boundary conditions respectively, the middle fixed point a_0 corresponds to the measure ν we consider here.

The best known bound for this channel is given in [18], where it is shown that reconstruction is impossible whenever $\frac{w}{1+w}\left(\frac{\ln(1+\lambda)}{\ln(1+w)} - 1\right) < 1$. The importance of this bound is that the range of non-reconstructibility in terms of λ does not vanish as $b \to \infty$. Specifically, reconstruction is impossible for $\lambda \le e - 1$ *independently* of b. We will now give another bound for non-reconstructibility using our general technique. While our bound does vanish as $b \to \infty$ when expressed in terms of λ (and hence in this sense is weaker than that above), it is actually slightly better for small values of b (see Table 1). Furthermore, it illustrates how the analysis we carried out in the spin systems context (for the hard-core model) can be easily translated to achieve interesting results in the reconstruction context.

Theorem 9.6 *For the one-sided deterministic error binary channel with parameter w on $\widehat{\mathbb{T}}^b$, reconstruction is impossible whenever*

$$\frac{w}{1+w}\sqrt{\frac{\sqrt{1+\lambda} - 1}{\sqrt{1+\lambda} + 1}} < \frac{1}{b}\,.$$

The reason the bound here is better than that for general binary channels in Theorem 9.4 is that, for this particular channel, we are able to get a better bound on γ than the one in Lemma 9.5

[17] Strictly speaking, the above recipe does not yield the potentials that define the hard-core model. However, it easy to verify that on a regular graph such as $\widehat{\mathbb{T}}^b$ the spin systems arising from the two sets of potentials are equivalent.

b	Martin [18]	Theorem 9.6
2	11.9	14.6
3	6.51	6.86
4	4.91	4.67

Table 1: Non-reconstruction thresholds (in terms of λ) for the binary channel with one-sided deterministic error and small values of b. The thresholds given by Martin [18] are compared with the thresholds from Theorem 9.6 of the present paper. (Reconstruction is impossible for all values of λ below the given threshold.) Whereas our bound is better for $b \in \{2, 3\}$, Martin's bound is better for all larger values of b. Furthermore, Martin's bound is above $e - 1$ for all values of b while ours vanishes as $b \to \infty$.

(which for this channel takes the trivial value of 1). To do so we simply translate the bound we obtained for the hard-core model in Section 5. (Actually we use the bound on γ_2, which is even better). Specifically, we use the fact that ν can equivalently be described as a Gibbs measure arising from the hard-core spin system with activity parameter $\lambda = w(1+w)^b$ (and thus the conditional probabilities in finite subsets are the same as for this spin system). Translating the bound on γ_2 from part (ii) of Theorem 5.1 (and noting that the same calculation works for the Bethe lattice $\widehat{\mathbb{T}}^b$) yields:

Lemma 9.7 *For the one-sided deterministic error binary channel on $\widehat{\mathbb{T}}^b$,*

$$\gamma_2 \leq \sqrt{\frac{\sqrt{1+\lambda}-1}{\sqrt{1+\lambda}+1}}.$$

Theorem 9.6 now follows immediately from our criterion for non-reconstructibility given in Theorem 9.3′ and the fact that for binary channels $\kappa = |p_{12} - p_{22}| = \frac{w}{1+w}$.

Symmetric multi-channel

We now go on to discuss multi-channels, i.e., channels in which the number of symbols q may be larger than two. Here we limit the discussion to symmetric channels, i.e., for some $\delta \in [0, \frac{1}{q-1}]$, $p_{ij} = \delta$ for $i \neq j$ and $p_{ii} = 1 - (q-1)\delta$ for all $i \in S = \{1, \ldots, q\}$. We note that the measure ν arising from this multi-channel corresponds to the free measure of the Potts model with inverse temperature $\beta = \frac{1}{2}\ln(\frac{1}{\delta} - (q-1))$. (Notice that for $\delta \leq \frac{1}{q}$ the correspondence is with the ferromagnetic Potts model, and for $\delta > \frac{1}{q}$ the correspondence is with the antiferromagnetic Potts model, i.e., $\beta < 0$.) The best known bound for this channel is due to Mossel and Peres [23] and states that reconstruction is impossible whenever $\frac{(1-q\delta)^2}{1-(q-2)\delta} \leq \frac{1}{b}$. While this bound is tight for the binary case, there is still a gap for $q \geq 3$. (See [22] for a survey of known reconstruction thresholds.) We give a simple proof of the above bound using our techniques, and marginally improve on it for $q \geq 3$.

Theorem 9.8 *Reconstruction is impossible for the symmetric multi-channel on $\mathbb{T}^b$ provided that*

$$(1-\epsilon)\,\frac{(1-q\delta)^2}{1-(q-2)\delta} < \frac{1}{b}\,,$$

where $\epsilon = \epsilon(b, q, \delta) \geq 0$ with equality if and only if $q = 2$ or $\delta = \frac{1}{q-1}$. [The exact definition of $\epsilon(b, q, \delta)$ is rather involved and appears in the proof of Lemma 9.9 below.]

As for the previous channels we have considered, Theorem 9.8 follows from our criterion for non-reconstructibility in Theorem 9.3 and bounds on κ, γ for this channel. Recall that $\kappa = \max_{i,j\in S} \frac{1}{2}\sum_k |p_{ik} - p_{jk}| = |1 - q\delta|$. As was the case for the previous channels, the more interesting ingredient is the bound on γ:

Lemma 9.9 *For the symmetric multi-channel on $\mathbb{T}^b$,*

$$\gamma \leq (1-\epsilon)\frac{|1-q\delta|}{1-(q-2)\delta},$$

where $\epsilon = \epsilon(b, q, \delta) \geq 0$ is as claimed in Theorem 9.8.

Theorem 9.8 follows immediately from this lemma as in our previous examples.

Proof of Lemma 9.9: First notice that, by translating the bound on γ for the Potts model given in Theorem 8.1(i) to the present setting, we immediately obtain $\gamma \leq \frac{|1-q\delta|}{1-(q-2)\delta}$. (Although the bound in Theorem 8.1(i) is given only for the case $\beta \geq 0$, i.e., $\epsilon \leq \frac{1}{q}$, it is easy to see that if we replace the bound with its absolute value, the proof — with minor modifications to take account of the change of sign — remains valid for $\beta < 0$ as well.) In order to obtain the factor $1 - \epsilon$ improvement we need to delve into the details of the proof of Theorem 8.1(i). Recall the functions $K, K_{\max}$ and the probabilities p_1, p_2 defined in that proof. The idea for obtaining the improvement is that, if $q \geq 3$ and there are no deterministic errors (i.e., $\delta < \frac{1}{q-1}$), then regardless of the boundary condition $a = p_1 + p_2 \leq a_{\max} < 1$, where $a_{\max}$ depends on (b, q, ϵ). Recalling the arguments in the original proof, this means that $\gamma \leq K_{\max}(a_{\max}) < K_{\max}(1) = \frac{|1-q\delta|}{1-(q-2)\delta}$, where we have used the fact that $K_{\max}(a)$ is strictly increasing in a. To see that indeed $a_{\max} < 1$ for $q \geq 3$ when there are no deterministic errors, we simply observe that every symbol appears with positive probability at the root, conditioned on any boundary condition. For example, using Bayes' rule, $\frac{\delta^b}{q}$ is a trivial lower bound on this conditional probability for the ferromagnetic case ($\delta \leq \frac{1}{q}$), while $\frac{[1-(q-1)\delta]^b}{q}$ is a lower bound for the antiferromagnetic case ($\delta > \frac{1}{q}$). □

References

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

[2] N. Berger, C. Kenyon, E. Mossel and Y. Peres, "Glauber dynamics on trees and hyperbolic graphs," preprint, 2003. Preliminary version: C. Kenyon, E. Mossel and Y. Peres, "Glauber dynamics on trees and hyperbolic graphs," *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science*, 2001, pp. 568–578.

[3] 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.

[4] G.R. Brightwell and P. Winkler, "Random colorings of a Cayley tree," *Contemporary combinatorics*, Bolyai Soc. Math. Studies **10**, János Bolyai Math. Soc., Budapest, 2002, pp. 247–276.

[5] G.R. Brightwell and P. Winkler, "A second threshold for the hard-core model on a Bethe lattice," *Random Structures & Algorithms* **24** (2004), pp. 303–314.

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

[7] M. Dyer, A. Frieze, T.P. Hayes and E. Vigoda, "Randomly colouring constant degree graphs," *Proceedings of the 45th IEEE Symposium on Foundations of Computer Science*, 2004, to appear.

[8] 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.

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

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

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

[12] T.P. HAYES and E. VIGODA, "A non-Markovian coupling for randomly sampling colorings," *Proceedings of the 44th IEEE Symposium on Foundations of Computer Science*, 2003, pp. 618–627.

[13] M. JERRUM, J.-B. SON, P. TETALI and E. VIGODA, "Elementary bounds on Poincaré and log-Sobolev constants for decomposable Markov chains," INI Report NI03006-CMP. To appear in *Annals of Applied Probability*, 2004.

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

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

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

[17] R. LYONS, Phase transitions on non amenable graphs, *Journal of Mathematical Physics* **41** (2000), pp. 1099–1127.

[18] J.B. MARTIN, "Reconstruction thresholds on regular trees," in *Proceedings of Discrete Random Walks, DRW'03*, C. Banderier and C. Krattenthaler eds., Discrete Mathematics and Theoretical Computer Science Proceedings Vol. AC, 2003, pp. 191–204.

[19] 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.

[20] F. MARTINELLI, A. SINCLAIR and D. WEITZ, "The Ising model on trees: Boundary conditions and mixing time," Tech. Report UCB//CSD-03-1256, Dept. of EECS, UC Berkeley, July 2003. Available from http://sunsite.Berkeley.EDU/NCSTRL/ Slightly modified version under the title "Glauber dynamics on trees: boundary conditions and mixing time" appeared in *Communications in Mathematical Physics* **250** (2004), pp. 301–334.

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

[22] E. MOSSEL, "Survey: Information flow on trees," in *Graphs, Morphisms and Statistical Physics*, J. Nešetřil and P. Winkler eds., DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 2004, pp. 155–170.

[23] E. MOSSEL and Y. PERES, "Information flow on trees," *Annals of Applied Probability* **13** (2003), pp. 817–844.

[24] R.B. POTTS, "Some generalized order-disorder transformations," *Proceedings of the Cambridge Philosophical Society* **48** (1952), pp. 106–109.

[25] C.J. PRESTON, *Gibbs states on countable sets*, Cambridge University Press, London, 1974.

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

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

[28] E. VIGODA, "A note on the Glauber dynamics for sampling independent sets," *Electronic Journal of Combinatorics* **8(1)** (2001).

[29] D. WEITZ "Combinatorial criteria for uniqueness of the Gibbs measure," preprint, 2004. To appear in *Random Structures & Algorithms*.

[30] D. WEITZ "Mixing in time and space for discrete spin systems," Ph.D. dissertation, University of California at Berkeley, May 2004.

[31] F.Y. WU, "The Potts model," *Reviews of Modern Physics* **54** (1982), pp. 235–268.