

# Mixing Time of Critical Ising Model on Trees is Polynomial in the Height

Jian Ding<sup>1</sup>, Eyal Lubetzky<sup>2</sup>, Yuval Peres<sup>2</sup>

<sup>1</sup> Department of Statistics, UC Berkeley, Berkeley, CA 94720,  
USA. E-mail: jding@stat.berkeley.edu

<sup>2</sup> Microsoft Research, One Microsoft Way, Redmond, WA 98052-6399,  
USA. E-mail: eyal@microsoft.com; peres@microsoft.com

Received: 6 February 2009 / Accepted: 25 September 2009

Published online: 8 January 2010 – © The Author(s) 2010. This article is published with open access at Springerlink.com

**Abstract:** In the heat-bath Glauber dynamics for the Ising model on the lattice, physicists believe that the spectral gap of the continuous-time chain exhibits the following behavior. For some critical inverse-temperature  $\beta_c$ , the inverse-gap is  $O(1)$  for  $\beta < \beta_c$ , polynomial in the surface area for  $\beta = \beta_c$  and exponential in it for  $\beta > \beta_c$ . This has been proved for  $\mathbb{Z}^2$  except at criticality. So far, the only underlying geometry where the critical behavior has been confirmed is the complete graph. Recently, the dynamics for the Ising model on a regular tree, also known as the Bethe lattice, has been intensively studied. The facts that the inverse-gap is bounded for  $\beta < \beta_c$  and exponential for  $\beta > \beta_c$  were established, where  $\beta_c$  is the critical spin-glass parameter, and the tree-height  $h$  plays the role of the surface area.

In this work, we complete the picture for the inverse-gap of the Ising model on the  $b$ -ary tree, by showing that it is indeed polynomial in  $h$  at criticality. The degree of our polynomial bound does not depend on  $b$ , and furthermore, this result holds under any boundary condition. We also obtain analogous bounds for the mixing-time of the chain. In addition, we study the near critical behavior, and show that for  $\beta > \beta_c$ , the inverse-gap and mixing-time are both  $\exp[\Theta((\beta - \beta_c)h)]$ .

## 1. Introduction

The *Ising Model* on a finite graph  $G = (V, E)$  with inverse-temperature  $\beta \geq 0$  and no external magnetic field is defined as follows. Its set of possible *configurations* is  $\Omega = \{\pm 1\}^V$ , where each  $\sigma \in \Omega$  assigns positive or negative *spins* to the vertices of  $G$ . In the free boundary case, the probability that the system is at a given configuration  $\sigma$  is given by the *Gibbs distribution*

$$\mu_G(\sigma) = \frac{1}{Z(\beta)} \exp \left( \beta \sum_{xy \in E} \sigma(x)\sigma(y) \right),$$

where  $Z(\beta)$  is the partition function. In the presence of a boundary condition  $\tau \in \{\pm 1\}^{\partial V}$  (that fixes the spins of some subset  $\partial V \subset V$  of the sites), we let  $\mu_G^\tau(\sigma)$  denote the Gibbs measure conditioned on  $\sigma$  agreeing with  $\tau$  on  $\partial V$ .

The heat-bath Glauber dynamics for the Ising model on  $G$  is the Markov chain with the following transition rule: At each step, a vertex is chosen uniformly at random, and its spin is updated according to  $\mu_G^\tau$  conditioned on the spins of all the other vertices. It is easy to verify that this chain is reversible with respect to the Gibbs distribution  $\mu_G^\tau$ . The continuous-time version of the dynamics associates each site with an independent Poisson clock of unit rate, determining the update times of this site as above (note that the continuous dynamics is  $|V|$  times faster than the discrete dynamics).

The *spectral-gap* of a reversible discrete-time chain, denoted by  $\text{gap}$ , is  $1 - \lambda$ , where  $\lambda$  is the largest nontrivial eigenvalue of its transition kernel. The spectral-gap of the continuous-time process is defined analogously via the spectrum of its generator, and in the special case of Glauber dynamics for  $\mu_G^\tau$ , this gap is precisely  $|V|$  times the discrete-time gap. This parameter governs the rate of convergence to equilibrium in  $L^2(\mu_G^\tau)$ .

In the classical Ising model, the underlying geometry is the  $d$ -dimensional lattice, and there is a critical inverse-temperature  $\beta_c$  where the static Gibbs measure exhibits a phase transition with respect to long-range correlations between spins. While the main focus of the physics community is on critical behavior (see the 20 volumes of [14]), so far, most of the rigorous mathematical analysis was confined to the non-critical regimes.

Supported by many experiments and studies in the theory of dynamical critical phenomena, physicists believe that the spectral-gap of the continuous-time dynamics on lattices has the following *critical slowing down* behavior (e.g., [16, 20, 25, 38]): At high temperatures ( $\beta < \beta_c$ ) the inverse-gap is  $O(1)$ , at the critical  $\beta_c$  it is polynomial in the surface-area and at low temperatures it is exponential in it. This is known for  $\mathbb{Z}^2$  except at the critical  $\beta_c$ , and establishing the order of the gap at criticality seems extremely challenging. In fact, the only underlying geometry, where the critical behavior of the spectral-gap has been fully established, is the complete graph (see [9]).

The important case of the Ising model on a regular tree, known as the Bethe lattice, has been intensively studied (e.g., [3–7, 15, 17, 18, 23, 27, 31]). On this canonical example of a *non-amenable* graph (one whose boundary is proportional to its volume), the model exhibits a rich behavior. For example, it has two distinct critical inverse-temperatures: one for uniqueness of the Gibbs state, and another for the purity of the free-boundary state. The latter,  $\beta_c$ , coincides with the critical spin-glass parameter.

As we later describe, previous results on the Ising model on a regular tree imply that the correct parameter to play the role of the surface-area is the tree-height  $h$ : It was shown that the inverse-gap is  $O(1)$  for  $\beta < \beta_c$  and exponential in  $h$  for  $\beta > \beta_c$ , yet its critical behavior remained unknown.

In this work, we complete the picture of the spectral-gap of the dynamics for the critical Ising model on a regular tree, by establishing that it is indeed polynomial in  $h$ . Furthermore, this holds under any boundary condition, and an analogous result is obtained for the  $L^1$  (total-variation) mixing time, denoted by  $t_{\text{MIX}}$  (formally defined in Subsect. 2.1).

**Theorem 1.** Fix  $b \geq 2$  and let  $\beta_c = \text{arctanh}(1/\sqrt{b})$  denote the critical inverse-temperature for the Ising model on the  $b$ -ary tree of height  $h$ . Then there exists some constant  $c > 0$  independent of  $b$ , so that the following holds: For any boundary condition  $\tau$ , the continuous-time Glauber dynamics for the above critical Ising model satisfies  $\text{gap}^{-1} \leq t_{\text{MIX}} = O(h^c)$ .

One of the main obstacles in proving the above result is the arbitrary boundary condition, due to which the spin system loses its symmetry (and the task of analyzing the dynamics becomes considerably more involved). Note that, although boundary conditions are believed to only accelerate the mixing of the dynamics, even tracking the effect of the (symmetric) all-plus boundary on lattices for  $\beta > \beta_c$  is a formidable open problem (see [26]).

In light of the above theorem and the known fact that the inverse-gap is exponential in  $h$  at low temperatures ( $\beta > \beta_c$  fixed), it is natural to ask how the transition between these two phases occurs, and in particular, what the critical exponent of  $\beta - \beta_c$  is. This is answered by the following theorem, which establishes that  $\log(\text{gap}^{-1}) \asymp (\beta - \beta_c)h + \log h$  for small  $\beta - \beta_c$ . Moreover, this result also holds for  $\beta = \beta_c + o(1)$ , and thus pinpoints the transition to a polynomial inverse-gap at  $\beta - \beta_c \asymp \frac{\log h}{h}$ .

**Theorem 2.** *For some  $\varepsilon_0 > 0$ , any  $b \geq 2$  fixed and all  $\beta_c < \beta < \beta_c + \varepsilon_0$ , where  $\beta_c = \text{arctanh}(1/\sqrt{b})$  is the critical spin-glass parameter, the following holds: The continuous-time Glauber dynamics for the Ising model on a  $b$ -ary tree with inverse-temperature  $\beta$  and free boundary satisfies*

$$\begin{aligned} \text{gap}^{-1} &= h^{\Theta(1)} && \text{if } \beta = \beta_c + O\left(\frac{\log h}{h}\right), \\ \text{gap}^{-1} &= \exp[\Theta((\beta - \beta_c)h)] && \text{otherwise.} \end{aligned} \tag{1.1}$$

Furthermore, both upper bounds hold under any boundary condition  $\tau$ , and (1.1) remains valid if  $\text{gap}^{-1}$  is replaced by  $t_{\text{MIX}}$ .

In the above theorem and in what follows, the notation  $f = \Theta(g)$  stands for  $f = O(g)$  and  $g = O(f)$ .

Finally, our results include new lower bounds on the critical inverse-gap and the total-variation mixing-time (see Theorem 3). The lower bound on  $\text{gap}^{-1}$  refutes a conjecture of [3], according to which the continuous-time inverse-gap is linear in  $h$ . Our lower bound on  $t_{\text{MIX}}$  is of independent interest: Although in our setting the ratio between  $t_{\text{MIX}}$  and  $\text{gap}^{-1}$  is at most poly-logarithmic in  $n$ , the number of sites, we were able to provide a lower bound of order  $\log n$  on this ratio without resorting to eigenfunction analysis.

*1.1. Background.* The thoroughly studied question of whether the free boundary state is pure (or extremal) in the Ising model on the Bethe lattice can be formulated as follows: Does the effect that a typical boundary has on the spin at the root vanish as the size of the tree tends to infinity? It is well-known that one can sample a configuration for the tree according to the Gibbs distribution with free boundary by propagating spins along the tree (from a site to its children) with an appropriate bias (see Subsect. 2.2 for details). Hence, the above question is equivalent to asking whether the spin at the root can be reconstructed from its leaves, and as such has applications in Information Theory and Phylogeny (see [15] for further details).

In sufficiently high temperatures, there is a unique Gibbs state for the Ising model on a  $b$ -ary tree ( $b \geq 2$ ), hence in particular the free boundary state is pure. The phase-transition with respect to the uniqueness of the Gibbs distribution occurs at the inverse-temperature  $\beta_u = \text{arctanh}(1/b)$ , as established in 1974 by Preston [34].

In [6], the authors studied the critical spin-glass model on the Bethe lattice (see also [5, 7]), i.e., the Ising model with a boundary of i.i.d. uniform spins. Following that work, it was finally shown in [4] that the phase-transition in the free-boundary extremality has

the same critical inverse-temperature as in the spin-glass model,  $\beta_c = \operatorname{arctanh}(1/\sqrt{b})$ . That is, the free-boundary state is pure iff  $\beta \leq \beta_c$ . This was later reproved in [17, 18].

The inverse-gap of the Glauber dynamics for the Ising model on a graph  $G$  was related in [3] to the *cut-width* of the graph,  $\xi(G)$ , defined as follows: It is the minimum integer  $m$ , such that for some labeling of the vertices  $\{v_1, \dots, v_n\}$  and any  $k \in [n]$ , there are at most  $m$  edges between  $\{v_1, \dots, v_k\}$  and  $\{v_{k+1}, \dots, v_n\}$ . The authors of [3] proved that for any bounded degree graph  $G$ , the continuous-time gap satisfies  $\mathfrak{gap}^{-1} = \exp[O(\xi(G)\beta)]$ .

Recalling the aforementioned picture of the phase-transition of the gap, this supports the claim that the cut-width is the correct extension of the surface-area to general graphs. One can easily verify that for  $\mathbb{Z}_L^d$  (the  $d$ -dimensional box of side-length  $L$ ) the cut-width has the same order as the surface-area  $L^{d-1}$ , while for a regular tree of height  $h$  it is of order  $h$ .

Indeed, for the Ising model on a  $b$ -ary tree with  $h$  levels and free boundary, it was shown in [3] that the inverse-gap is  $O(1)$  for all  $\beta < \beta_c$ , whereas for  $\beta > \beta_c$  it satisfies  $\log \mathfrak{gap}^{-1} \asymp h$  (with constants that depend on  $b$  and  $\beta$ ). The behavior of the gap at criticality was left as an open problem: it is proved in [3] that the critical  $\mathfrak{gap}^{-1}$  is at least linear in  $h$  and conjectured that this is tight. A weaker conjecture of [3] states that  $\mathfrak{gap}^{-1} = \exp(o(h))$ .

Further results on the dynamics were obtained in [27], showing that the log-Sobolev constant  $\alpha_s$  (defined in Sect. 2) is uniformly bounded away from zero for  $\beta < \beta_c$  in the free-boundary case, as well as for any  $\beta$  under the all-plus boundary condition. While this implies that  $\mathfrak{gap}^{-1} = O(1)$  in these regimes, it sheds no new light on the behavior of the parameters  $\mathfrak{gap}$ ,  $\alpha_s$  in our setting of the critical Ising model on a regular tree with free-boundary.

*1.2. The critical inverse-gap and mixing-time.* Theorems 1,2, stated above, establish that on a regular tree of height  $h$ , the critical and near-critical continuous-time  $\mathfrak{gap}^{-1}$  and  $t_{\text{MIX}}$  are polynomial in  $h$ . In particular, this confirms the conjecture of [3] that the critical inverse-gap is  $\exp(o(h))$ .

Moreover, our upper bounds hold for any boundary condition, while matching the behavior of the free-boundary case: Indeed, in this case the critical inverse-gap is polynomial in  $h$  (as [3] showed it is at least linear), and for  $\beta - \beta_c > 0$  small we do have that  $\log(\mathfrak{gap}^{-1}) \asymp (\beta - \beta_c)h$ . For comparison, recall that under the all-plus boundary condition, [27] showed that  $\mathfrak{gap}^{-1} = O(1)$  at all temperatures.

We next address the conjecture of [3] that the critical inverse-gap is in fact linear in  $h$ . The proof that the critical  $\mathfrak{gap}^{-1}$  has order at least  $h$  uses the same argument that gives a tight lower bound at high temperatures: Applying the Dirichlet form (see Subsect. 2.4) to the sum of spins at the leaves as a test-function. Hence, the idea behind the above conjecture is that the sum of spins at the boundary (that can be thought of as the magnetization) approximates the second eigenfunction also for  $\beta = \beta_c$ .

The following theorem refutes this conjecture. In addition, this theorem provides a nontrivial lower bound on  $t_{\text{MIX}}$  that separates it from  $\mathfrak{gap}^{-1}$  (thus far, our bounds in Theorems 1,2 applied to both parameters as one).

**Theorem 3.** *Fix  $b \geq 2$  and let  $\beta_c = \operatorname{arctanh}(1/\sqrt{b})$  be the critical inverse-temperature for the Ising model on the  $b$ -ary tree with  $n$  vertices. Then the corresponding continuous-time Glauber dynamics with free boundary condition satisfies:*

$$\mathfrak{gap}^{-1} \geq c_1 \log^2 n, \quad (1.2)$$

$$t_{\text{MIX}} \geq c_2 \log^3 n, \tag{1.3}$$

for some  $c_1, c_2 > 0$ . Furthermore,  $t_{\text{MIX}} \geq c \text{gap}^{-1} \log n$  for some  $c > 0$ .

Indeed, the above theorem implies that  $\text{gap}^{-1}$  has order at least  $h^2$  and  $t_{\text{MIX}}$  has order at-least  $h^3$ , where  $h$  is again the height of the tree. It is a well known fact that the log-Sobolev constant satisfies  $\alpha_s \leq \frac{1}{2} \text{gap}$ , hence in our setting we have  $\alpha_s = O(h^{-2}) = o(1)$ .

We note that by related results on the log-Sobolev constant, it follows that in the Ising model on a regular tree, for any temperature and with any boundary condition we have  $t_{\text{MIX}} = O(\text{gap}^{-1} \log^2 n)$ . In light of this, establishing a lower bound of order  $\log n$  on the ratio between  $t_{\text{MIX}}$  and  $\text{gap}^{-1}$  is quite delicate (e.g., proving such a bound usually involves constructing a distinguishing statistic via a suitable eigenfunction (Wilson’s method [39])).

*1.3. Techniques and proof ideas.* To prove the main theorem, our general approach is a recursive analysis of the spectral-gap via an appropriate block-dynamics (roughly put, multiple sites comprising a block are updated simultaneously in each step of this dynamics; see Subsect. 2.5 for a formal definition). This provides an estimate of the spectral-gap of the single-site dynamics in terms of those of the individual blocks and the block-dynamics chain itself (see [25]). However, as opposed to most applications of the block-dynamics method, where the blocks are of relatively small size, in our setting we must partition a tree of height  $h$  to subtrees of height linear in  $h$ . This imposes arbitrary boundary conditions on the individual blocks, and highly complicates the analysis of the block-dynamics chain.

In order to estimate the gap of the block-dynamics chain, we apply the method of Decomposition of Markov chains (for details on this method see Subsect. 2.6). Combining this method with a few other ideas (such as establishing contraction and controlling the external field in certain chains), the proof of Theorem 1 is reduced into the following spatial-mixing/reconstruction type problem. Consider the procedure, where we assign the spins of the boundary given the value at the root of the tree, then reconstruct the root from the values at the boundary. The key quantity required in our setting is the difference in the expected outcome of the root, comparing the cases where its initial spin was either positive or negative.

This quantity was studied by [31] in the free-boundary case, where it was related to capacity-type parameters of the tree (see [15] for a related result corresponding to the high temperature regime). Unfortunately, in our case we have an arbitrary boundary condition, imposed by the block-dynamics. This eliminates the symmetry of the system, which was a crucial part of the arguments of [31]. The most delicate step in the proof of Theorem 1 is the extension of these results of [31] to any boundary condition. This is achieved by carefully tracking down the effect of the boundary on the expected reconstruction result in each site, combined with correlation inequalities and an analytical study of the corresponding log-likelihood-ratio function.

The lower bound on the critical inverse-gap reflects the change in the structure of the dominant eigenfunctions between high and low temperatures. At high temperatures, the sum of spins on the boundary gives the correct order of the gap. At low temperatures, a useful lower bound on  $\text{gap}^{-1}$  was shown in [3] via the recursive-majority function (intuitively, this reflects the behavior at the root: Although this spin may occasionally flip its value, at low temperature it constantly tends to revert to its biased state). Our

results show that at criticality, a lower bound improving upon that of [3] is obtained by essentially merging the above two functions into a weighted sum of spins, where the weight of a spin is determined by its tree level.

To establish a lower bound on  $t_{\text{MIX}}$  of order  $\text{gap}^{-1}h$ , we consider a certain *speed-up* version of the dynamics: a block-dynamics, whose blocks are a mixture of singletons and large subtrees. The key ingredient here is the Censoring Inequality of Peres and Winkler [33], that shows that this dynamics indeed mixes as fast as the usual (single-site) one. We then consider a series of modified versions of this dynamics, and study their mixing with respect to the total-variation and Hellinger distances. In the end, we arrive at a product chain, whose components are each the single-site dynamics on a subtree of height linear in  $h$ . This latter chain provides the required lower bound on  $t_{\text{MIX}}$ .

*1.4. Organization.* The rest of this paper is organized as follows. Sect. 2 contains several preliminary facts and definitions. In Sect. 3 we prove a spatial-mixing type result on the critical and near-critical Ising model on a tree with an arbitrary boundary condition. This then serves as one of the key ingredients in the proof of the main result, Theorem 1, which appears in Section 4. In Sect. 5 we prove Theorem 3, providing the lower bounds for the critical inverse-gap and mixing-time. Section 6 contains the proof of Theorem 2, addressing the near-critical behavior of  $\text{gap}^{-1}$  and  $t_{\text{MIX}}$ . The final section, Sect. 7, is devoted to concluding remarks and open problems.

## 2. Preliminaries

*2.1. Total-variation mixing.* Let  $(X_t)$  be an aperiodic irreducible Markov chain on a finite state space  $\Omega$ , with stationary distribution  $\pi$ . For any two distributions  $\phi, \psi$  on  $\Omega$ , the *total-variation distance* of  $\phi$  and  $\psi$  is defined as

$$\|\phi - \psi\|_{\text{TV}} \triangleq \sup_{A \subset \Omega} |\phi(A) - \psi(A)| = \frac{1}{2} \sum_{x \in \Omega} |\phi(x) - \psi(x)|.$$

The (worst-case) total-variation *mixing-time* of  $(X_t)$ , denoted by  $t_{\text{MIX}}(\varepsilon)$  for  $0 < \varepsilon < 1$ , is defined to be

$$t_{\text{MIX}}(\varepsilon) \triangleq \min \left\{ t : \max_{x \in \Omega} \|\mathbb{P}_x(X_t \in \cdot) - \pi\|_{\text{TV}} \leq \varepsilon \right\},$$

where  $\mathbb{P}_x$  denotes the probability given that  $X_0 = x$ . As it is easy and well known (cf., e.g., [1, Chap. 4]) that the spectral-gap of  $(X_t)$  satisfies  $\text{gap}^{-1} \leq t_{\text{MIX}}(1/e)$ , it will be convenient to use the abbreviation

$$t_{\text{MIX}} \triangleq t_{\text{MIX}}(1/e).$$

Analogously, for a continuous-time chain on  $\Omega$  with heat-kernel  $H_t$ , we define  $t_{\text{MIX}}$  as the minimum  $t$  such that  $\max_{x \in \Omega} \|H_t(x, \cdot) - \pi\|_{\text{TV}} \leq 1/e$ .

*2.2. The Ising model on trees.* When the underlying geometry of the Ising model is a tree with free boundary condition, the Gibbs measure has a natural constructive representation. This appears in the following well known claim (see, e.g., [15] for more details).

**Claim 2.1.** Consider the Ising model on a tree  $T$  rooted at  $\rho$  with free boundary condition and at the inverse-temperature  $\beta$ . For all  $e \in E(T)$ , let  $\eta_e \in \{\pm 1\}$  be i.i.d. random variables with  $\mathbb{P}(\eta_e = 1) = (1 + \tanh \beta)/2$ . Furthermore, let  $\sigma(\rho)$  be a uniform spin, independent of  $\{\eta_e\}$ , and for  $v \neq \rho$ ,

$$\sigma(v) = \sigma(\rho) \prod_{e \in \mathcal{P}(\rho, v)} \eta_e, \quad \text{where } \mathcal{P}(\rho, v) \text{ is the simple path from } \rho \text{ to } v.$$

Then the distribution of the resulting  $\sigma$  is the corresponding Gibbs measure.

In light of the above claim, one is able to sample a configuration according to Gibbs distribution on a tree with free boundary condition using the following simple scheme: Assign a uniform spin at the root  $\rho$ , then scan the tree from top to bottom, successively assigning each site with a spin according to the value at its parent. More precisely, a vertex is assigned the same spin as its parent with probability  $(1 + \tanh \beta)/2$ , and the opposite one otherwise. Equivalently, a vertex inherits the spin of its parent with probability  $\tanh \beta$ , and otherwise it receives an independent uniform spin. Finally, for the conditional Gibbs distribution given a plus spin at the root  $\rho$ , we assign  $\rho$  a plus spin rather than a uniform spin, and carry on as above.

However, notice that the above *does not* hold for the Ising model in the presence of a boundary condition, which may impose a different external influence on different sites.

**2.3.  $L^2$ -capacity.** The authors of [31] studied certain spatial mixing properties of the Ising model on trees (with free or all-plus boundary conditions), and related them to the  $L^p$ -capacity of the underlying tree. In Sect. 3, we extend some of the results of [31] to the (highly asymmetric) case of a tree with an arbitrary boundary condition, and relate a certain “decay of correlation” property to the  $L^2$ -capacity of the tree, defined as follows.

Let  $T$  be a tree rooted at  $\rho$ , denote its leaves by  $\partial T$ , and throughout the paper, write  $(u, v) \in E(T)$  for the directed edge between a vertex  $u$  and its child  $v$ . We further define  $\text{dist}(u, v)$  as the length (in edges) of the simple path connecting  $u$  and  $v$  in  $T$ .

For each  $e \in E(T)$ , assign the *resistance*  $R_e \geq 0$  to the edge  $e$ . We say that a non-negative function  $f : E(T) \rightarrow \mathbb{R}$  is a *flow* on  $T$  if the following holds for all  $(u, v) \in E(T)$  with  $v \notin \partial T$ :

$$f(u, v) = \sum_{(v, w) \in E(T)} f(v, w),$$

that is, the incoming flow equals the outgoing flow on each internal vertex  $v$  in  $T$ . For any flow  $f$ , define its *strength*  $|f|$  and *voltage*  $V(f)$  by

$$|f| \triangleq \sum_{(\rho, v) \in E(T)} f(\rho, v), \quad V(f) \triangleq \sup \left\{ \sum_{e \in \mathcal{P}(\rho, w)} f(e)R_e : w \in \partial T \right\},$$

where  $\mathcal{P}(\rho, w)$  denotes the simple path from  $\rho$  to  $w$ . Given these definitions, we now define the  $L^2$ -capacity  $\text{cap}_2(T)$  to be

$$\text{cap}_2(T) \triangleq \sup\{|f| : f \text{ is a flow with } V(f) \leq 1\}.$$

For results on the  $L^2$ -capacity of general networks (and more generally,  $L^p$ -capacities, where the expression  $f(e)R_e$  in the above definition of  $V(f)$  is replaced by its  $(p - 1)$



power), as part of the Discrete Nonlinear Potential Theory, cf., e.g., [28, 36, 37] and the references therein.

For our proofs, we will use the well-known fact that the  $L^2$ -capacity of the tree  $T$  is precisely the *effective conductance* between the root  $\rho$  and the leaves  $\partial T$ , denoted by  $C_{\text{eff}}(\rho \leftrightarrow \partial T)$ . See, e.g., [24] for further information on electrical networks.

**2.4. Spectral gap and log-Sobolev constant.** Our bound on the mixing time of Glauber dynamics for the Ising model on trees will be derived from a recursive analysis of the spectral gap of this chain. This analysis uses spatial-mixing type results (and their relation to the above mentioned  $L^2$  capacity) as a building block. We next describe how the mixing-time can be bounded via the spectral-gap in our setting.

The spectral gap and log-Sobolev constant of a reversible Markov chain with stationary distribution  $\pi$  are given by the following Dirichlet forms (see, e.g., [1, Chap. 3,8]):

$$\text{gap} = \inf_f \frac{\mathcal{E}(f)}{\text{Var}_\pi(f)}, \quad \alpha_s = \inf_f \frac{\mathcal{E}(f)}{\text{Ent}(f)}, \tag{2.1}$$

where

$$\mathcal{E}(f) = \langle (I - P)f, f \rangle_\pi = \frac{1}{2} \sum_{x, y \in \Omega} [f(x) - f(y)]^2 \pi(x)P(x, y), \tag{2.2}$$

$$\text{Ent}_\pi(f) = \mathbb{E}_\pi \left( f^2 \log(f^2 / \mathbb{E}_\pi f^2) \right). \tag{2.3}$$

By bounding the log-Sobolev constant, one may obtain remarkably sharp upper bounds on the  $L^2$  mixing-time: cf., e.g., [10–13, 35]. The following result of Diaconis and Saloff-Coste [12, Th. 3.7] (see also [35, Cor. 2.2.7]) demonstrates this powerful method; its next formulation for discrete-time appears in [1, Chap. 8]. As we are interested in total-variation mixing, we write this bound in terms of  $t_{\text{MIX}}$ , though it in fact holds also for the (larger)  $L^2$  mixing-time.

**Theorem 2.2.** ([12, 35], reformulated) *For any reversible finite Markov chain with stationary distribution  $\pi$ ,*

$$t_{\text{MIX}}(1/e) \leq \frac{1}{4} \alpha_s^{-1} (\log \log(1/\pi^*) + 4),$$

where  $\pi^* = \min_x \pi(x)$ .

We can then apply a result of [27], which provides a useful bound on  $\alpha_s$  in terms of  $\text{gap}$  in our setting, and obtain an upper bound on the mixing-time.

**Theorem 2.3.** ([27, Th. 5.7]) *There exists some  $c > 0$  such that the Ising model on the  $b$ -ary tree with  $n$  vertices satisfies  $\alpha_s \geq c \cdot \text{gap} / \log n$ .*

Note that the proof of the last theorem holds for any  $\beta$  and under any boundary condition. Combining Theorems 2.2 and 2.3, and noticing that  $\pi_* \geq 2^{-n} \exp(-2\beta n)$  (as there are  $2^n$  configurations, and the ratio between the maximum and minimum probability of a configuration is at most  $\exp(2\beta n)$ ), we obtain the following corollary:



**Corollary 2.4.** *The Glauber dynamics for the Ising model on a  $b$ -ary tree with  $n$  vertices satisfies  $t_{\text{MIX}} = O(\alpha_s^{-1} \log n) = O(\text{gap}_{\mathbb{P}}^{-1} \log^2 n)$  for any  $\beta$  and any boundary condition.*

The above corollary reduces the task of obtaining an upper bound for the mixing-time into establishing a suitable lower bound on the spectral gap. This will be achieved using a block dynamics analysis.

*2.5. From single site dynamics to block dynamics.* Consider a cover of  $V$  by a collection of subsets  $\{B_1, \dots, B_k\}$ , which we will refer to as “blocks”. The *block dynamics* corresponding to  $B_1, \dots, B_k$  is the Markov chain, where at each step a uniformly chosen block is updated according to the stationary distribution given the rest of the system. That is, the entire set of spins of the chosen block is updated simultaneously, whereas all other spins remain unchanged. One can verify that the block dynamics is reversible with respect to the Gibbs distribution  $\mu_n$ .

Recall that, given a subset of the sites  $U \subset V$ , a boundary condition  $\eta$  imposed on  $U$  is the restriction of the sites  $U^c = V \setminus U$ , so that all agree with  $\eta$  throughout the dynamics, i.e., only sites in  $U$  are considered for updates. It will sometimes be useful to consider  $\eta \in \Omega$  (rather than a configuration of the sites  $U^c$ ), in which case only its restriction to  $U^c$  is accounted for.

The following theorem shows the useful connection between the single-site dynamics and the block dynamics. This theorem appears in [25] in a more general setting, and following is its reformulation for the special case of Glauber dynamics for the Ising model on a finite graph with an arbitrary boundary condition. Though the original theorem is stated for the continuous-time dynamics, its proof naturally extends to the discrete-time case; we provide its details for completeness.

**Proposition 2.5.** ([25, Prop. 3.4], restated) *Consider the discrete time Glauber dynamics on a  $b$ -ary tree with boundary condition  $\eta$ . Let  $\text{gap}_U^\eta$  be the spectral-gap of the single-site dynamics on a subset  $U \subset V$  of the sites, and  $\text{gap}_B^\eta$  be the spectral-gap of the block dynamics corresponding to  $B_1, \dots, B_k$ , an arbitrary cover of a vertex set  $W \subset V$ . The following holds:*

$$\text{gap}_W^\eta \geq \frac{k}{|W|} \text{gap}_B^\eta \inf_i \inf_\varphi |B_i| \text{gap}_{B_i}^\varphi \left( \sup_{x \in W} \#\{i : B_i \ni x\} \right)^{-1}.$$

*Proof.* Let  $P$  denote the transition kernel of the above Glauber dynamics. Defining

$$g \triangleq \inf_i \inf_\varphi |B_i| \text{gap}_{B_i}^\varphi,$$

the Dirichlet form (2.1) gives that, for any function  $f$ ,

$$\text{Var}_{B_i}^\varphi(f) \leq \frac{\mathcal{E}_{B_i}^\varphi(f)}{\text{gap}_{B_i}^\varphi} \leq \frac{|B_i|}{g} \mathcal{E}_{B_i}^\varphi(f).$$

Combining this with definition (2.2) of  $\mathcal{E}(\cdot)$ ,

$$\mathcal{E}_B^\eta(f) = \frac{1}{k} \sum_{\varphi \in \Omega} \mu_W^\eta(\varphi) \sum_i \text{Var}_{B_i}^\varphi(f) \leq \frac{1}{kg} \sum_{\varphi \in \Omega} \mu_W^\eta(\varphi) \sum_i |B_i| \mathcal{E}_{B_i}^\varphi(f).$$

On the other hand, definition (2.2) again implies that

$$\begin{aligned}
& \sum_{\varphi \in \Omega} \mu_W^\eta(\varphi) \sum_i |B_i| \mathcal{E}_{B_i}^\varphi(f) \\
&= \sum_{\varphi \in \Omega} \mu_W^\eta(\varphi) \frac{1}{2} \sum_{\sigma \in \Omega} \sum_i \sum_{x \in B_i} \mu_{B_i}^\varphi(\sigma) |B_i| P_{B_i}^\varphi(\sigma, \sigma^x) [f(\sigma) - f(\sigma^x)]^2 \\
&\leq \frac{1}{2} \sup_{x \in W} \#\{i : B_i \ni x\} \sum_{\sigma \in \Omega} \mu_W^\eta(\sigma) \sum_{x \in W} |W| P_W^\sigma(\sigma, \sigma^x) [f(\sigma) - f(\sigma^x)]^2 \\
&= |W| \sup_{x \in W} \#\{i : B_i \ni x\} \mathcal{E}_W^\eta(f),
\end{aligned}$$

where  $\sigma^x$  is the configuration obtained from  $\sigma$  by flipping the spin at  $x$ , and we used the fact that

$$|B_i| P_{B_i}^\sigma(\sigma, \sigma^x) = |W| P_W^\sigma(\sigma, \sigma^x) \quad \text{for any } i \in [k] \text{ and } x \in B_i.$$

Altogether, we obtain that

$$\mathcal{E}_B^\eta(f) \leq \frac{|W|}{kg} \sup_{x \in W} \#\{i : B_i \ni x\} \mathcal{E}_W^\eta(f).$$

Recalling that the single-site dynamics and the block-dynamics have the same stationary measure,

$$\frac{\mathcal{E}_B^\eta(f)}{\text{Var}_W^\eta(f)} = \frac{\mathcal{E}_B^\eta(f)}{\text{Var}_B^\eta(f)} \geq \text{gap}_B^\eta$$

(where we again applied inequality (2.1)), thus

$$\frac{\mathcal{E}_W^\eta(f)}{\text{Var}_W^\eta(f)} \geq \frac{k}{|W|} g \left( \sup_{x \in W} \#\{i : B_i \ni x\} \right)^{-1} \text{gap}_B^\eta.$$

The proof is now completed by choosing  $f$  to be the eigenfunction that corresponds to the second eigenvalue of  $P_W^\eta$  (achieving  $\text{gap}_W^\eta$ ), with a final application of (2.1).  $\square$

The above proposition can be applied, as part of the spectral gap analysis, to reduce the size of the base graph (though with an arbitrary boundary condition), provided that one can estimate the gap of the corresponding block dynamics chain.

**2.6. Decomposition of Markov chains.** In order to bound the spectral gap of the block dynamics, we require a result of [19], which analyzes the spectral gap of a Markov chain via its decomposition into a *projection* chain and a *restriction* chain.

Consider an ergodic Markov chain on a finite state space  $\Omega$  with transition kernel  $P : \Omega \times \Omega \rightarrow [0, 1]$  and stationary distribution  $\pi : \Omega \rightarrow [0, 1]$ . We assume that the Markov chain is *time-reversible*, that is to say, it satisfies the detailed balance condition

$$\pi(x)P(x, y) = \pi(y)P(y, x) \text{ for all } x, y \in \Omega.$$

Let  $\Omega = \Omega_0 \cup \dots \cup \Omega_{m-1}$  be a decomposition of the state space into  $m$  disjoint sets. Writing  $[m] \triangleq \{0, \dots, m-1\}$ , we define  $\bar{\pi} : [m] \rightarrow [0, 1]$  as

$$\bar{\pi}(i) \triangleq \pi(\Omega_i) = \sum_{x \in \Omega_i} \pi(x)$$

and define  $\bar{P} : [m] \times [m] \rightarrow [0, 1]$  to be

$$\bar{P}(i, j) \triangleq \frac{1}{\bar{\pi}(i)} \sum_{x \in \Omega_i, y \in \Omega_j} \pi(x) P(x, y).$$

The Markov chain on the state space  $[m]$  whose transition kernel is  $\bar{P}$  is called the *projection chain*, induced by the partition  $\Omega_0, \dots, \Omega_{m-1}$ . It is easy to verify that, as the original Markov chain is reversible with respect to  $\pi$ , the projection chain is reversible with respect to the stationary distribution  $\bar{\pi}$ .

In addition, each  $\Omega_i$  induces a *restriction chain*, whose transition kernel  $P_i : \Omega_i \times \Omega_i \rightarrow [0, 1]$  is given by

$$P_i(x, y) = \begin{cases} P(x, y), & \text{if } x \neq y, \\ 1 - \sum_{z \in \Omega_i \setminus \{x\}} P(x, z), & \text{if } x = y. \end{cases}$$

Again, the restriction chain inherits its reversibility from the original chain, and has a stationary measure  $\pi_i$ , which is simply  $\pi$  restricted to  $\Omega_i$ :

$$\pi_i(x) \triangleq \pi(x) / \bar{\pi}(i) \text{ for all } x \in \Omega_i.$$

In most applications, the projection chain and the different restriction chains are all irreducible, and thus the various stationary distributions  $\bar{\pi}$  and  $\pi_0, \dots, \pi_{m-1}$  are all unique.

The following result provides a lower bound on the spectral gap of the original Markov chain given its above described decomposition:

**Theorem 2.6** ([19, Th. 1]). *Let  $P$  be the transition kernel of a finite reversible Markov chain, and let  $\text{gap}$  denote its spectral gap. Consider the decomposition of the chain into a projection chain and  $m$  restriction chains, and denote their corresponding spectral gaps by  $\text{gap}$  and  $\text{gap}_0, \dots, \text{gap}_{m-1}$ . Define*

$$\text{gap}_{\min} \triangleq \min_{i \in [m]} \text{gap}_i, \quad \gamma \triangleq \max_{i \in [m]} \max_{x \in \Omega_i} \sum_{y \in \Omega \setminus \Omega_i} P(x, y).$$

Then  $\text{gap}$ , the spectral gap of the original Markov chain, satisfies:

$$\text{gap} \geq \frac{\text{gap}}{3} \wedge \frac{\text{gap} \cdot \text{gap}_{\min}}{3\gamma + \text{gap}}.$$

The main part of Sect. 4 will be devoted to the analysis of the projection chain, in an effort to bound the spectral gap of our block dynamics via the above theorem.

### 3. Spatial Mixing of Ising Model on Trees

In this section, we will establish a spatial-mixing type result for the Ising model on a general (not necessarily regular) finite tree under an arbitrary boundary condition. This result (namely, Proposition 3.1) will later serve as the main ingredient in the proof of Theorem 1 (see Sect. 4). Throughout this section, let  $\beta > 0$  be an arbitrary inverse-temperature and  $\theta = \tanh \beta$ .

We begin with a few notations. Let  $T$  be a tree rooted at  $\rho$  with a boundary condition  $\tau \in \{\pm 1\}^{\partial T}$  on its leaves, and  $\mu^\tau$  be the corresponding Gibbs measure.

For any  $v \in T$ , denote by  $T_v$  the subtree of  $T$  containing  $v$  and its all descendants. In addition, for any  $B \subset A \subset T$  and  $\sigma \in \{\pm 1\}^A$ , denote by  $\sigma_B$  the restriction of  $\sigma$  to the sites of  $B$ . We then write  $\mu_v^\tau$  for the Gibbs measure on the subtree  $T_v$  given the boundary  $\tau_{\partial T_v}$ .

Consider  $\hat{T} \subset T \setminus \partial T$ , a subtree of  $T$  that contains the root  $\rho$ , and write  $\hat{T}_v = T_v \cap \hat{T}$ . Similar to the above definitions for  $T$ , we denote by  $\hat{\mu}^\xi$  the Gibbs measure on  $\hat{T}$  given the boundary condition  $\xi \in \{\pm 1\}^{\partial \hat{T}}$ , and let  $\hat{\mu}_v^\xi$  be the Gibbs measure on  $\hat{T}_v$  given the boundary  $\xi_{\partial \hat{T}_v}$ .

The following two measures are the conditional distributions of  $\mu_v^\tau$  on the boundary of the subtree  $\hat{T}_v$  given the spin at its root  $v$ :

$$\begin{aligned} Q_v^+(\xi) &\triangleq \mu_v^\tau \left( \sigma_{\partial \hat{T}_v} = \xi_{\partial \hat{T}_v} \mid \sigma(v) = 1 \right) && \text{for } \xi \in \{\pm 1\}^{\partial \hat{T}}, \\ Q_v^-(\xi) &\triangleq \mu_v^\tau \left( \sigma_{\partial \hat{T}_v} = \xi_{\partial \hat{T}_v} \mid \sigma(v) = -1 \right) && \text{for } \xi \in \{\pm 1\}^{\partial \hat{T}}. \end{aligned}$$

We can now state the main result of this section, which addresses the problem of reconstructing the spin at the root of the tree from its boundary.

**Proposition 3.1.** *Let  $\hat{T}$  be as above, let  $0 < \theta \leq \frac{3}{4}$  and define*

$$\Delta \triangleq \int \hat{\mu}^\xi(\sigma(\rho) = 1) dQ_\rho^+(\xi) - \int \hat{\mu}^\xi(\sigma(\rho) = 1) dQ_\rho^-(\xi).$$

*Then there exists an absolute constant  $\kappa > \frac{1}{100}$  such that*

$$\Delta \leq \frac{\text{cap}_2(\hat{T})}{\kappa(1 - \theta)},$$

*where the resistances are assigned to be  $R_{(u,v)} = \theta^{-2 \text{dist}(\rho,v)}$ . Furthermore, this also holds for any external field  $h \in \mathbb{R}$  on the root  $\rho$ .*

To prove the above theorem, we consider the notion of the log-likelihood ratio at a vertex  $v$  with respect to  $\hat{T}_v$  given the boundary  $\xi_{\partial \hat{T}_v}$ :

$$x_v^\xi = \log \left( \frac{\hat{\mu}_v^\xi(\sigma(v) = +1)}{\hat{\mu}_v^\xi(\sigma(v) = -1)} \right), \tag{3.1}$$

as well as the following quantity, analogous to  $\Delta$  (defined in Proposition 3.1):

$$m_v \triangleq \int x_v^\xi dQ_v^+ - \int x_v^\xi dQ_v^-. \tag{3.2}$$

As we will later explain,  $m_v \geq 0$  for any  $v \in T$ , and we seek an upper bound on this quantity. One of the main results of [31] was such an estimate for the case of *free* boundary condition, yet in our setting we have an *arbitrary* boundary condition (adding a considerable amount of difficulties to the analysis). The following theorem extends the upper bound on  $m_\rho$  to any boundary; to avoid confusion, we formulate this bound in terms of the same absolute constant  $\kappa$  given in Proposition 3.1.

**Theorem 3.2.** *Let  $\hat{T}$  and  $m_\rho$  be as above, and let  $0 < \theta \leq \frac{3}{4}$ . Then there exists an absolute constant  $\kappa > \frac{1}{100}$  such that*

$$m_\rho \leq \frac{\text{cap}_2(\hat{T})}{\kappa(1 - \theta)/4},$$

where the resistances are assigned to be  $R_{(u,v)} = \theta^{-2 \text{dist}(\rho,v)}$ .

*Proof of Theorem 3.2.* As mentioned above, the novelty (and also the main challenge) in the result stated in Theorem 3.2 is the presence of the arbitrary boundary condition  $\tau$ , which eliminates most of the symmetry that one has in the free boundary case. Note that this symmetry was a crucial ingredient in the proof of [31] for the free boundary case (namely, in that case  $Q_v^+$  and  $Q_v^-$  are naturally symmetric).

In order to tackle this obstacle, we need to track down the precise influence of the boundary condition  $\tau$  on each vertex  $v \in T$ . We then incorporate this information in the recursive analysis that appeared (in a slightly different form) in [23]. This enables us to relate the recursion relation of the  $m_v$ -s to that of the  $L^2$ -capacity.

The following quantity captures the above mentioned influence of  $\tau$  on a given vertex  $v \in T$ :

$$x_v^* = \log \left( \frac{\mu_v^\tau(\sigma(v) = 1)}{\mu_v^\tau(\sigma(v) = -1)} \right). \tag{3.3}$$

Notice that  $x_v^*$  has a similar form to  $x_v^\xi$  (defined in (3.1)), and is essentially the log-likelihood ratio at  $v$  induced by the boundary condition  $\tau$ . The quantity  $x_v^\xi$  is then the log-likelihood ratio that in addition considers the extra constraints imposed by  $\xi$ . Also note that a free boundary condition corresponds to the case where  $x_v^* = 0$  for all  $v \in T$ .

To witness the effect of  $x_v^*$ , consider the probabilities of propagating a spin from a parent  $v$  to its child  $w$ , formally defined by

$$p_{v,w}^\tau(1, 1) \triangleq \mu_v^\tau(\sigma(w) = 1 \mid \sigma(v) = 1),$$

$$p_{v,w}^\tau(1, -1) \triangleq \mu_v^\tau(\sigma(w) = -1 \mid \sigma(v) = 1);$$

we define  $p_{v,w}^\tau(-1, 1)$  and  $p_{v,w}^\tau(-1, -1)$  analogously. The next simple lemma shows the relation between  $x_v^*$  and these probabilities.

**Lemma 3.3.** *The following holds for any  $(v, w) \in T$ :*

$$p_{v,w}^\tau(1, 1) - p_{v,w}^\tau(-1, 1) = D_w^* \theta,$$

where  $D_w^* \triangleq (\cosh \beta)^2 / (\cosh^2 \beta + \cosh^2(x_w^*/2) - 1)$ .

*Proof.* Recalling definition (3.3) of  $x_v^*$ , we can translate the boundary condition  $\tau$  into an external field  $x_w^*/2$  on the vertex  $w$  when studying the distribution of its spin. Hence,

$$\begin{aligned} p_{v,w}^\tau(1, 1) - p_{v,w}^\tau(-1, 1) &= \frac{e^{\beta+x_w^*/2}}{e^{\beta+x_w^*/2} + e^{-\beta-x_w^*/2}} - \frac{e^{-\beta+x_w^*/2}}{e^{-\beta+x_w^*/2} + e^{\beta-x_w^*/2}} \\ &= \frac{e^{2\beta} - e^{-2\beta}}{e^{x_w^*} + e^{-x_w^*} + e^{2\beta} + e^{-2\beta}} \\ &= \frac{\cosh^2 \beta}{\cosh^2 \beta + \cosh^2(x_w^*/2) - 1} \tanh \beta, \end{aligned}$$

as required.  $\square$

*Remark 3.4.* In the free boundary case, we have  $p_{v,w}(1, 1) - p_{v,w}(-1, 1) = \theta$ . For a boundary condition  $\tau$ , the coefficient  $0 < D_w^* \leq 1$  represents the contribution of this boundary to the propagator probability.

We now turn our attention to  $m_v$ . As mentioned before, the fact that  $m_v \geq 0$  follows from its definition (3.2). Indeed, the monotonicity of the Ising model implies that the measure  $Q_v^+$  stochastically dominates the measure  $Q_v^-$  (with respect to the natural partial order on the configurations of  $\partial\hat{T}_v$ ). For instance, it is easy to see this by propagating 1 and  $-1$  spins from the root to the bottom, and applying a monotone coupling on these two processes. Finally,  $x_v^\xi$  is monotone increasing in  $\xi$  (again by the monotonicity of the Ising model), thus  $m_v \geq 0$ .

The first step in establishing the recursion relation of  $m_v$  (that would lead to the desired upper bound) would be to relate  $m_v$  to some quantities associated with its children, as stated next.

**Lemma 3.5.** *For any  $v \in \hat{T} \setminus \partial\hat{T}$ , we have that*

$$m_v = \theta \sum_{w:(v,w) \in \hat{T}} D_w^* \left( \int f(x_w^\xi) dQ_w^+(\xi) - \int f(x_w^\xi) dQ_w^-(\xi) \right),$$

where

$$f(x) = \log \left( \frac{\cosh(x/2) + \theta \sinh(x/2)}{\cosh(x/2) - \theta \sinh(x/2)} \right). \tag{3.4}$$

*Proof.* We need the following well-known lemma, that appeared in [23] in a slightly different form; see also [2] and [31, Lemma 4.1].

**Lemma 3.6** ([23,31] (reformulated)). *Let  $f$  be as in (3.4). For all  $v \in \hat{T} \setminus \partial\hat{T}$  and  $\xi \in \{\pm 1\}^{\partial\hat{T}}$ , the following holds:  $x_v^\xi = \sum_{w:(v,w) \in \hat{T}} f(x_w^\xi)$ .*

According to this lemma, we obtain that

$$m_v = \sum_{w:(v,w) \in \hat{T}} \left( \int f(x_w^\xi) dQ_v^+(\xi) - \int f(x_w^\xi) dQ_v^-(\xi) \right). \tag{3.5}$$

Noting that  $x_w^\xi$  is actually a function of  $\xi_{\partial\hat{T}_w}$ , we get that

$$\int f(x_w^\xi) dQ_v^+(\xi) = \int f(x_w^\xi) d(p_{v,w}^\tau(1, 1)Q_w^+(\xi) + p_{v,w}^\tau(1, -1)Q_w^-(\xi)),$$

and similarly, we have

$$\int f(x_w^\xi) dQ_v^-(\xi) = \int f(x_w^\xi) d(p_{v,w}^\tau(-1, 1)Q_w^+(\xi) + p_{v,w}^\tau(-1, -1)Q_w^-(\xi)).$$

Combining these two equalities, we deduce that

$$\begin{aligned} & \int f(x_w^\xi) dQ_v^+(\xi) - \int f(x_w^\xi) dQ_v^-(\xi) \\ &= (p_{v,w}^\tau(1, 1) - p_{v,w}^\tau(-1, 1)) \left( \int f(x_w^\xi) dQ_w^+(\xi) - \int f(x_w^\xi) dQ_w^-(\xi) \right) \\ &= \theta D_w^* \left( \int f(x_w^\xi) dQ_w^+(\xi) - \int f(x_w^\xi) dQ_w^-(\xi) \right), \end{aligned} \tag{3.6}$$

where in the last inequality we applied Lemma 3.3. Plugging (3.6) into (3.5) now completes the proof of the lemma.  $\square$

Observe that in the free boundary case,  $Q_v^+(\xi) = Q_v^(-\xi)$  for any  $\xi$ . Unfortunately, the presence of the boundary  $\tau$  breaks this symmetry, causing the distributions  $Q_v^+$  and  $Q_v^-$  to become skewed. Nevertheless, we can still relate these two distributions through the help of  $x_v^*$ , as formulated by the following lemma.

**Lemma 3.7.** *For  $v \in T$ , let  $Q_v$  be the following distribution over  $\{\pm 1\}^{\partial \hat{T}}$ :*

$$Q_v(\xi) = Q_v^\tau(\xi) \stackrel{\Delta}{=} \mu_v^\tau \left( \sigma_{\partial \hat{T}_v} = \xi_{\partial \hat{T}_v} \right).$$

*Then for all  $\xi \in \{\pm 1\}^{\partial \hat{T}}$ , we have*

$$Q_v^+(\xi) - Q_v^-(\xi) = C_v^* \left( \tanh \frac{x_v^\xi}{2} - \tanh \frac{x_v^*}{2} \right) Q_v(\xi),$$

where  $C_v^* = 2 \cosh^2(x_v^*/2)$ .

*Proof.* It is clear from the definitions of  $Q_v^+$ ,  $Q_v^-$  and  $Q_v$  that

$$\begin{aligned} Q_v^+(\xi) &= \frac{Q_v(\xi) \mu_v^\tau(\sigma(v) = 1 \mid \xi)}{\mu_v^\tau(\sigma(v) = 1)} = \frac{1 + \tanh(x_v^\xi/2)}{1 + \tanh(x_v^*/2)} Q_v(\xi), \\ Q_v^-(\xi) &= \frac{Q_v(\xi) \mu_v^\tau(\sigma(v) = -1 \mid \xi)}{\mu_v^\tau(\sigma(v) = -1)} = \frac{1 - \tanh(x_v^\xi/2)}{1 - \tanh(x_v^*/2)} Q_v(\xi). \end{aligned}$$

Hence, a straightforward calculation gives that

$$\begin{aligned} Q_v^+(\xi) - Q_v^-(\xi) &= \frac{2 \left( \tanh(x_v^\xi/2) - \tanh(x_v^*/2) \right)}{(1 + \tanh(x_v^*/2))(1 - \tanh(x_v^*/2))} Q_v(\xi) \\ &= 2 \cosh^2 \left( \frac{x_v^*}{2} \right) \left( \tanh \frac{x_v^\xi}{2} - \tanh \frac{x_v^*}{2} \right) Q_v(\xi), \end{aligned}$$

as required.  $\square$



The following technical lemma will allow us to combine Lemmas 3.5, 3.7 and obtain an upper bound on  $m_v$  in terms of  $\{m_w : (v, w) \in \hat{T}\}$ . Note that the constant  $\kappa$  mentioned next is in fact the absolute constant  $\kappa$  in the statement of Theorem 3.2.

**Lemma 3.8.** *Let  $f$  be defined as in (3.4) for some  $0 < \theta \leq \frac{3}{4}$ . Then*

$$|f(x) - f(y)| \leq 2f(|x - y|/2) \text{ for any } x, y \in \mathbb{R}, \tag{3.7}$$

and there exists a universal constant  $\kappa > \frac{1}{100}$  such that for any two constants  $C_1, C_2 \geq 1$  with  $C_2 \geq 1 + (\frac{1}{2}C_1 - 1)(1 - \theta^2)$  and any  $\delta > 0$  we have

$$f(\delta) (1 + 4\kappa(1 - \theta)C_1\delta \tanh(\delta/2)) \leq C_2\theta\delta. \tag{3.8}$$

*Proof.* We first show (3.7). Put  $\delta = |x - y|$ , and define

$$h(\delta) = \sup_t |f(t + \delta) - f(t)|.$$

We claim that

$$h(\delta) = f(\delta/2) - f(-\delta/2) = 2f(\delta/2). \tag{3.9}$$

The second equality follows from the fact that  $f(x)$  is an odd function. To establish the first equality, a straightforward calculation gives that

$$f'(x) = \frac{\theta}{1 + (1 - \theta^2) \sinh^2(x/2)},$$

and it follows that  $f'(x)$  is an even non-negative function which is decreasing in  $x \geq 0$ . The following simple claim therefore immediately implies (3.9):

**Claim 3.9.** *Let  $g(t) \geq 0$  be an even function that is decreasing in  $t \geq 0$ . Then  $G(t) = \int_0^t g(x)dx$  has  $G(t + \delta) - G(t) \leq 2G(\delta/2)$  for any  $t$  and  $\delta > 0$ .*

*Proof.* Fix  $\delta > 0$  and define  $F(t)$  as follows:

$$F(t) = G(t + \delta) - G(t).$$

We therefore have  $F'(t) = g(t + \delta) - g(t)$ . Noticing that

$$\begin{cases} |t + \delta| \geq |t| & \text{if } t \geq -\frac{\delta}{2} \\ |t + \delta| \leq |t| & \text{otherwise} \end{cases},$$

the assumption on  $g$  now gives that  $F'(t) \leq 0$  while  $t \geq -\frac{\delta}{2}$  and otherwise  $F'(t) \geq 0$ . Altogether, we deduce that

$$F(t) \leq F(-\frac{\delta}{2}) = G(\frac{\delta}{2}) - G(-\frac{\delta}{2}) = 2G(\frac{\delta}{2}),$$

as required.  $\square$

It remains to show that (3.8) holds for some  $\kappa = \kappa(\theta_0) > 0$ . Clearly, it suffices to establish this statement for  $C_2 = \left[1 + \left(\frac{1}{2}C_1 - 1\right)(1 - \theta^2)\right] \vee 1$  and any  $C_1 \geq 1$ . Rearranging (3.8), we are interested in a lower bound on

$$\inf_{\theta \leq \theta_0, t > 0, C_1 \geq 1} \frac{\left[ \left(1 + \left(\frac{C_1}{2} - 1\right)(1 - \theta^2)\right) \vee 1 \right] \theta t - f(t)}{4C_1(1 - \theta)tf(t) \tanh\left(\frac{t}{2}\right)}. \tag{3.10}$$

First, consider the case  $1 \leq C_1 < 2$ . We then have  $C_2 = 1$ , and the expression being minimized in (3.10) takes the form:

$$\frac{\theta t - f(t)}{4C_1(1 - \theta)tf(t) \tanh\left(\frac{t}{2}\right)} > \frac{\theta t - f(t)}{8(1 - \theta)tf(t) \tanh\left(\frac{t}{2}\right)} \triangleq \Gamma(t, \theta),$$

where the inequality is by our assumption that  $C_1 < 2$ . We therefore have that  $\inf_{\theta \leq \theta_0, t > 0} \Gamma(t, \theta)$  minimizes (3.10) for  $C_1 < 2$ , and will next show that this is also the case for  $C_1 \geq 2$  under a certain condition. Indeed, letting

$$g(t, \theta, C_1) \triangleq \frac{\left[1 + \left(\frac{C_1}{2} - 1\right)(1 - \theta^2)\right] \theta t - f(t)}{4C_1(1 - \theta)tf(t) \tanh(t/2)},$$

it is easy to verify that the following holds:

$$\frac{\partial g}{\partial C_1} = \frac{f(t) - \theta^3 t}{4C_1^2(1 - \theta)tf(t) \tanh\left(\frac{t}{2}\right)},$$

hence  $g$  is increasing in  $C_1$  for every  $\theta, t$  such that  $f(t) > \theta^3 t$ . Therefore,

$$g(t, \theta, C_1) \geq g(t, \theta, 2) = \Gamma(t, \theta) \text{ for all } t, \theta \text{ such that } f(t) > \theta^3 t.$$

Before analyzing  $\Gamma(t, \theta)$ , we will treat the values of  $\theta, t$  such that  $f(t) \leq \theta^3 t$ . Assume that case, and notice that the numerator of  $g$  then satisfies

$$\begin{aligned} & \left[1 + \left(\frac{C_1}{2} - 1\right)(1 - \theta^2)\right] \theta t - f(t) \\ & \geq \left[1 + \left(\frac{C_1}{2} - 1\right)(1 - \theta^2) - \theta^2\right] \theta t = \theta(1 - \theta^2)t \frac{C_1}{2}, \end{aligned}$$

and thereby the dependency on  $C_1$  vanishes:

$$g(t, \theta, C_1) \geq \frac{\theta(1 - \theta^2)t/2}{4(1 - \theta)tf(t) \tanh(t/2)} = \frac{\theta(1 + \theta)}{8f(t) \tanh(t/2)}.$$

Since both  $\tanh(t/2)$  and  $f(t)$  are monotone increasing in  $t$  and are bounded from above by 1 and  $\log\left(\frac{1+\theta}{1-\theta}\right)$  respectively, we get

$$g(t, \theta, C_1) \geq \frac{\theta(1 + \theta)}{8 \log\left(\frac{1+\theta}{1-\theta}\right)} \geq \frac{\theta(1 + \theta)}{8 \frac{2\theta}{1-\theta}} = \frac{1 - \theta^2}{16} > \frac{1}{40}, \tag{3.11}$$

where the second inequality is by the fact that  $\log(1 + x) \leq x$  for any  $x > 0$ , and the last inequality follows by the assumption  $\theta \leq \frac{3}{4}$ .

It thus remains to establish a uniform lower bound on  $\Gamma(t, \theta)$ . In what follows, our choice of constants was a compromise between simplicity and the quality of the lower bound, and we note that one can easily choose constants that are slightly more optimal.

Assume first that  $\theta \geq \theta_0 \geq 0$  for some  $\theta_0$  to be defined later. Notice that

$$\tilde{f}(t, \theta) \triangleq \frac{1}{\theta} f(t, \theta) = 2 \sum_{i=0}^{\infty} \frac{\tanh^{2i+1}(t/2)}{2i+1} \theta^{2i},$$

and so  $\tilde{f}(t, \theta)$  is strictly increasing in  $\theta$  for any  $t > 0$ . Since

$$\begin{aligned} \Gamma(t, \theta) &= \frac{\theta t - f(t)}{8(1-\theta)t f(t) \tanh(\frac{t}{2})} \\ &\geq \frac{\theta t - f(t)}{8(1-\theta_0)t f(t) \tanh(\frac{t}{2})} = \frac{t - \tilde{f}(t)}{8(1-\theta_0)t \tilde{f}(t) \tanh(\frac{t}{2})}, \end{aligned}$$

we have that  $\Gamma$  is monotone decreasing in  $\theta$  for any such  $t$ , and therefore  $\Gamma(t, \theta) \geq \frac{1}{8(1-\theta_0)} \tilde{\Gamma}(t)$ , where  $\tilde{\Gamma}$  is defined as follows:

$$\tilde{\Gamma}(t) \triangleq \frac{\theta t - f(t, \theta)}{t f(t, \theta) \tanh(\frac{t}{2})} \quad \text{at } \theta = \frac{3}{4}. \tag{3.12}$$

Recall that the Taylor expansion of  $f(t, \theta)$  around 0 is  $\theta t - \frac{\theta(1-\theta^2)}{12} t^3 + O(t^5)$ . It is easy to verify that for  $\theta = \frac{3}{4}$  this function satisfies

$$f(t, \theta) \leq \theta t - \frac{(\theta t)^3}{20} \quad \text{for } \theta = \frac{3}{4} \text{ and any } 0 < t \leq 3.$$

Adding the fact that  $\tanh(x) \leq x$  for all  $x \geq 0$ , we immediately obtain that

$$\tilde{\Gamma} \geq \frac{\theta^3 t^3}{20t(\theta t)(t/2)} = \frac{\theta^2}{10} > \frac{1}{20} \quad \text{for all } 0 < t \leq 3.$$

On the other hand, for  $t \geq 3$  we can use the uniform upper bounds of 1 and  $\log(\frac{1+\theta}{1-\theta})$  for  $\tanh(t/2)$  and  $f(t)$  respectively, and gain that

$$\tilde{\Gamma} \geq \frac{\theta t - \log(\frac{1+\theta}{1-\theta})}{t \log(\frac{1+\theta}{1-\theta})} = \frac{\theta}{\log(\frac{1+\theta}{1-\theta})} - \frac{1}{t} \geq \frac{1}{20} \quad \text{for all } t \geq 3.$$

Altogether, as  $\Gamma \geq \frac{1}{8(1-\theta_0)} \tilde{\Gamma}$ , we can conclude that  $\Gamma \geq [160(1-\theta_0)]^{-1}$ .

Note that the trivial choice of  $\theta_0 = 0$  already provides a uniform lower bound of  $\frac{1}{160}$  for  $\Gamma$  (and hence also for  $\kappa$ , as the lower bound in (3.11) is only larger). However, this bound can be improved by choosing another  $\theta_0$  and treating the case  $0 < \theta \leq \theta_0$  separately. To demonstrate this, take for instance  $\theta_0 = \frac{1}{2}$ . Since the above analysis gave that  $\tilde{\Gamma} \geq \frac{1}{20}$  whenever  $\theta \leq \frac{3}{4}$ , it follows that

$$\Gamma \geq \frac{1}{160(1-\theta_0)} = \frac{1}{80} \quad \text{for all } \frac{1}{2} \leq \theta \leq \frac{3}{4}.$$

For  $\theta \leq \theta_0$ , we essentially repeat this analysis of  $\tilde{\Gamma}$ , only this time the respective value of  $\theta$  (that is, the maximum value it can attain) is  $\frac{1}{2}$ . One can thus verify that in that case,

$$f(t, \theta) \leq \theta t - \frac{(\theta t)^3}{6} \quad \text{for } \theta = \frac{1}{2} \text{ and any } 0 < t \leq 2.7,$$

and the above argument then shows that

$$\tilde{\Gamma} \geq \frac{\theta^2}{3} = \frac{1}{12} \quad \text{for all } 0 < t \leq 2.7.$$

On the other hand,

$$\tilde{\Gamma} \geq \frac{\theta}{\log\left(\frac{1+\theta}{1-\theta}\right)} - \frac{1}{t} \geq \frac{1}{12} \quad \text{for all } t \geq 2.7,$$

thus for  $\theta = \frac{1}{2}$  we have  $\tilde{\Gamma} \geq \frac{1}{12}$  for all  $t > 0$ . This converts into the lower bound  $\Gamma \geq \frac{1}{96}$ , thus completing the proof with a final value of  $\kappa = \frac{1}{96}$ .  $\square$

*Remark 3.10.* Note that the only places where we used the fact that  $\theta \leq \frac{3}{4}$  are the lower bound on  $g(t, \theta, C_1)$  in (3.11) and the analysis of  $\tilde{\Gamma}$ , as defined in (3.12). In both cases, we actually only need to have  $\theta \leq \theta_1$  for some constant  $\theta_1 < 1$ , whose precise value might affect the final value of  $\kappa$ .

Using the above lemma, we are now ready to obtain the final ingredient required for the proof of the recursion relation of  $m_v$ , as incorporated in Lemma 3.11. This lemma provides a recursive bound on a quantity that resembles  $m_v$ , where instead of integrating over  $x_v^\xi$ , we integrate over  $f(x_v^\xi)$ .

**Lemma 3.11.** *Let  $f$  and  $D_v^*$  be as in (3.4) and Lemma 3.3 respectively. There exists a universal constant  $\kappa > \frac{1}{100}$  so that for  $K = \frac{1}{4}(1 - \theta)\kappa$  we have*

$$\int f(x_v^\xi) dQ_v^+(\xi) - \int f(x_v^\xi) dQ_v^-(\xi) \leq \frac{\theta m_v}{D_v^*(1 + K m_v)}.$$

*Proof.* Clearly,

$$\begin{aligned} & \int f(x_v^\xi) dQ_v^+(\xi) - \int f(x_v^\xi) dQ_v^-(\xi) \\ &= \int (f(x_v^\xi) - f(x_v^*)) dQ_v^+(\xi) - \int (f(x_v^\xi) - f(x_v^*)) dQ_v^-(\xi). \end{aligned}$$

Applying Lemma 3.7, we then obtain that

$$\begin{aligned} & \int f(x_v^\xi) dQ_v^+(\xi) - \int f(x_v^\xi) dQ_v^-(\xi) \\ &= C_v^* \int (f(x_v^\xi) - f(x_v^*)) \left( \tanh \frac{x_v^\xi}{2} - \tanh \frac{x_v^*}{2} \right) dQ_v(\xi), \end{aligned}$$

and similarly,

$$m_v = C_v^* \int (x_v^\xi - x_v^*) \left( \tanh \frac{x_v^\xi}{2} - \tanh \frac{x_v^*}{2} \right) dQ_v(\xi).$$

Let

$$\begin{aligned} F(x) &= (f(x) - f(x_v^*)) (\tanh(x/2) - \tanh(x_v^*/2)), \\ G(x) &= (x - x_v^*) (\tanh(x/2) - \tanh(x_v^*/2)), \end{aligned}$$

and define  $\Lambda$  to be the probability measure on  $\mathbb{R}$  as:

$$\Lambda(x) \triangleq Q_v(\{\xi : x_v^\xi = x\}).$$

According to this definition, we have

$$\int F(x_v^\xi) dQ_v(\xi) = \int F(x) d\Lambda, \quad \text{and} \quad \int G(x_v^\xi) dQ_v(\xi) = \int G(x) d\Lambda,$$

and thus, by the above arguments,

$$\begin{aligned} &\int f(x_v^\xi) dQ_v^+(\xi) - \int f(x_v^\xi) dQ_v^-(\xi) = C_v^* \int F(x) d\Lambda, \\ m_v &= C_v^* \int G(x) d\Lambda. \end{aligned} \tag{3.13}$$

Furthermore, notice that by (3.7) and the fact that  $f$  is odd and increasing for  $x \geq 0$ ,

$$F(x) \leq 2f\left(\frac{x - x_v^*}{2}\right) \left( \tanh \frac{x}{2} - \tanh \frac{x_v^*}{2} \right),$$

and so

$$\begin{aligned} &\int f(x_v^\xi) dQ_v^+(\xi) - \int f(x_v^\xi) dQ_v^-(\xi) \\ &\leq 2C_v^* \int f\left(\frac{x - x_v^*}{2}\right) \left( \tanh \frac{x}{2} - \tanh \frac{x_v^*}{2} \right) d\Lambda. \end{aligned} \tag{3.14}$$

In our next argument, we will estimate  $\int G(x) d\Lambda$  and  $\int G(x) d\Lambda$  according to the behavior of  $F$  and  $G$  about  $x_v^*$ . Assume that  $x_v^* \geq 0$ , and note that, although the case of  $x_v^* \leq 0$  can be treated similarly, we claim that this assumption does not lose generality. Indeed, if  $x_v^* < 0$ , one can consider the boundary condition of  $\tau' = -\tau$ , which would give the following by symmetry:

$$x_v^{*'} = -x_v^*, \quad X'_v(-\xi) = -x_v^\xi(\xi), \quad Q'_v(-\xi) = Q_v(\xi).$$

Therefore, as  $f(\cdot)$  and  $\tanh(\cdot)$  are both odd functions, we have that  $\int F(x) d\Lambda$  and  $\int G(x) d\Lambda$  will not change under the modified boundary condition, and yet  $x_v^{*' } \geq 0$  as required.

Define

$$I^- \triangleq (-\infty, x_v^*] \quad , \quad I^+ \triangleq [x_v^*, \infty).$$

First, consider the case where for either  $I = I^+$  or  $I = I^-$  we have

$$\begin{cases} \int_I F(x)d\Lambda \geq \frac{1}{2} \int F(x)d\Lambda, \\ \int_I G(x)d\Lambda \geq \frac{1}{2} \int F(x)d\Lambda. \end{cases} \quad (3.15)$$

In this case, the following holds:

$$\begin{aligned} \left( \int F(x)d\Lambda \right) \left( \int G(x)d\Lambda \right) &\leq 4 \left( \int_I F(x)d\Lambda \right) \left( \int_I G(x)d\Lambda \right) \\ &\leq 4 \int_I F(x)G(x)d\Lambda \leq 4 \int F(x)G(x)d\Lambda, \end{aligned}$$

where in the second line we applied the FKG-inequality, using the fact that both  $F$  and  $G$  are decreasing in  $I^-$  and increasing in  $I^+$ . The last inequality followed from the fact that  $F$  and  $G$  are always non-negative. Note that

$$\int F(x)G(x)d\Lambda = \int (f(x) - f(x_v^*)) (x - x_v^*) \left( \tanh \frac{x}{2} - \tanh \frac{x_v^*}{2} \right)^2 d\Lambda,$$

and recall that Claim 3.9 applied onto  $\tanh(x)$  (which indeed has an even non-negative derivative  $\cosh^{-2}(x)$  that is decreasing in  $x \geq 0$ ) gives

$$\tanh \frac{x}{2} - \tanh \frac{y}{2} \leq 2 \tanh \left( \frac{x-y}{4} \right) \text{ for any } x > y.$$

Noticing that each of the factors comprising  $F(x)G(x)$  has the same sign as that of  $(x - x_v^*)$ , and combining this with (3.7), it thus follows that

$$\begin{aligned} &\left( \int F(x)d\Lambda \right) \left( \int G(x)d\Lambda \right) \\ &\leq 16 \int f \left( \frac{x - x_v^*}{2} \right) (x - x_v^*) \left( \tanh \frac{x}{2} - \tanh \frac{x_v^*}{2} \right) \tanh \left( \frac{x - x_v^*}{4} \right) d\Lambda. \end{aligned} \quad (3.16)$$

Second, consider the case where for  $I^+$  and  $I^-$  as above, we have

$$\begin{cases} \int_{I^+} F(x)d\Lambda \geq \frac{1}{2} \int F(x)d\Lambda, \\ \int_{I^-} G(x)d\Lambda \geq \frac{1}{2} \int G(x)d\Lambda. \end{cases} \quad (3.17)$$

The following definitions of  $\tilde{F}$  and  $\tilde{G}$  thus capture a significant contribution of  $F$  and  $G$  to  $\int Fd\Lambda$  and  $\int Gd\Lambda$  respectively:

$$\begin{cases} \tilde{F}(s) \triangleq F(x_v^* + s) \\ \tilde{G}(s) \triangleq G(x_v^* - s) \end{cases} \text{ for any } s \geq 0. \quad (3.18)$$

By further defining the probability measure  $\tilde{\Lambda}$  on  $[0, \infty)$  to be

$$\tilde{\Lambda}(s) \triangleq \Lambda(x_v^* - s) \mathbf{1}_{\{s \neq 0\}} + \Lambda(x_v^* + s) \text{ for any } s \geq 0, \quad (3.19)$$

we obtain that

$$\begin{aligned} \int F(x)d\Lambda &\leq 2 \int_{I^+} F(x)d\Lambda \leq 2 \int_0^\infty \tilde{F}(x)d\tilde{\Lambda}, \\ \int G(x)d\Lambda &\leq 2 \int_{I^-} G(x)d\Lambda \leq 2 \int_0^\infty \tilde{G}(x)d\tilde{\Lambda}. \end{aligned}$$

With both  $\tilde{F}$  and  $\tilde{G}$  being monotone increasing on  $[0, \infty)$ , applying the FKG-inequality with respect to  $\tilde{\Lambda}$  now gives

$$\begin{aligned} \left(\int F(x)d\Lambda\right) \left(\int G(x)d\Lambda\right) &\leq 4 \int_0^\infty \tilde{F}(x)\tilde{G}(x)d\tilde{\Lambda} \\ &= 4 \int_0^\infty (f(x_v^* + s) - f(x_v^*)) \left(\tanh \frac{x_v^* + s}{2} - \tanh \frac{x_v^*}{2}\right) \\ &\quad \cdot (-s) \left(\tanh \frac{x_v^* - s}{2} - \tanh \frac{x_v^*}{2}\right) d\tilde{\Lambda}. \end{aligned}$$

Returning to the measure  $\Lambda$ , the last expression takes the form

$$\begin{aligned} &4 \int_{I^+} (f(x) - f(x_v^*)) \left(\tanh \frac{x}{2} - \tanh \frac{x_v^*}{2}\right) \\ &\quad \cdot (x - x_v^*) \left(\tanh \frac{x_v^*}{2} - \tanh \frac{2x_v^* - x}{2}\right) d\Lambda \\ &+ 4 \int_{I^-} (f(2x_v^* - x) - f(x_v^*)) \left(\tanh \frac{2x_v^* - x}{2} - \tanh \frac{x_v^*}{2}\right) \\ &\quad \cdot (x - x_v^*) \left(\tanh \frac{x}{2} - \tanh \frac{x_v^*}{2}\right) d\Lambda. \end{aligned}$$

We now apply (3.7) and Claim 3.9 (while leaving the term  $(\tanh \frac{x}{2} - \tanh \frac{x_v^*}{2})$  unchanged in both integrals) to obtain that

$$\begin{aligned} &\left(\int F(x)d\Lambda\right) \left(\int G(x)d\Lambda\right) \\ &\leq 16 \int f\left(\frac{x - x_v^*}{2}\right) \left(\tanh \frac{x}{2} - \tanh \frac{x_v^*}{2}\right) (x - x_v^*) \tanh\left(\frac{x - x_v^*}{4}\right) d\Lambda. \end{aligned}$$

That is, we have obtained the same bound as in (3.16).

It remains to deal with the third case where for  $I^+$  and  $I^-$  as above,

$$\begin{cases} \int_{I^-} F(x)d\Lambda \geq \frac{1}{2} \int F(x)d\Lambda, \\ \int_{I^+} G(x)d\Lambda \geq \frac{1}{2} \int G(x)d\Lambda. \end{cases} \tag{3.20}$$

In this case, we modify the definition (3.18) of  $\tilde{F}$  and  $\tilde{G}$  appropriately:

$$\begin{cases} \tilde{F}(s) \triangleq F(x_v^* + s) \\ \tilde{G}(s) \triangleq G(x_v^* - s) \end{cases} \text{ for any } s \geq 0,$$



and let  $\tilde{\Lambda}$  remain the same, as given in (3.19). It then follows that

$$\begin{aligned} \int F(x)d\Lambda &\leq 2 \int_{I^-} F(x)d\Lambda \leq 2 \int_0^\infty \tilde{F}(x)d\tilde{\Lambda}, \\ \int G(x)d\Lambda &\leq 2 \int_{I^+} G(x)d\Lambda \leq 2 \int_0^\infty \tilde{G}(x)d\tilde{\Lambda}, \end{aligned}$$

with  $\tilde{F}$  and  $\tilde{G}$  monotone increasing on  $[0, \infty)$ . By the FKG-inequality,

$$\begin{aligned} \left(\int F(x)d\Lambda\right) \left(\int G(x)d\Lambda\right) &\leq 4 \int_0^\infty \tilde{F}(x)\tilde{G}(x)d\tilde{\Lambda} \\ &= 4 \int_0^\infty (f(x_v^* - s) - f(x_v^*)) \left(\tanh \frac{x_v^* - s}{2} - \tanh \frac{x_v^*}{2}\right) \\ &\quad \cdot s \left(\tanh \frac{x_v^* + s}{2} - \tanh \frac{x_v^*}{2}\right) d\tilde{\Lambda}. \end{aligned} \tag{3.21}$$

As before, we now switch back to  $\Lambda$  and infer from (3.7) and Claim 3.9 that

$$\begin{aligned} \left(\int F(x)d\Lambda\right) \left(\int G(x)d\Lambda\right) \\ \leq 16 \int f\left(\frac{x - x_v^*}{2}\right) \left(\tanh \frac{x}{2} - \tanh \frac{x_v^*}{2}\right) (x - x_v^*) \tanh\left(\frac{x - x_v^*}{4}\right) d\Lambda, \end{aligned}$$

that is, (3.16) holds for each of the 3 possible cases (3.15), (3.17) and (3.20).

Altogether, this implies that

$$\begin{aligned} \left(\int f(x_v^\xi)dQ_v^+(\xi) - \int f(x_v^\xi)dQ_v^-(\xi)\right) m_v \\ = (C_v^*)^2 \left(\int F(x)d\Lambda\right) \left(\int G(x)d\Lambda\right) \\ \leq 16C_v^{*2} \int f\left(\frac{x - x_v^*}{2}\right) \left(\tanh \frac{x}{2} - \tanh \frac{x_v^*}{2}\right) (x - x_v^*) \tanh\left(\frac{x - x_v^*}{4}\right) d\Lambda. \end{aligned}$$

Therefore, recalling (3.14) and choosing  $K = \frac{1}{4}(1 - \theta)\kappa$ , where  $\kappa$  is as given in Lemma 3.8, we have

$$\begin{aligned} \left(\int f(x_v^\xi)dQ_v^+(\xi) - \int f(x_v^\xi)dQ_v^-(\xi)\right) (1 + Km_v) \\ \leq 2C_v^* \int f\left(\frac{x - x_v^*}{2}\right) \left[1 + 4\kappa(1 - \theta)C_v^* \frac{x - x_v^*}{2} \tanh \frac{x - x_v^*}{4}\right] \\ \cdot \left(\tanh \frac{x}{2} - \tanh \frac{x_v^*}{2}\right) d\Lambda \\ \leq 2C_v^* \int (1/D_v^*)\theta \frac{x - x_v^*}{2} \left(\tanh \frac{x}{2} - \tanh \frac{x_v^*}{2}\right) d\Lambda = \theta \frac{C_v^*}{D_v^*} \int G(x)d\Lambda, \end{aligned}$$

where the inequality in the last line is by Lemma 3.8 for  $\delta = |x - x_v^*|/2$  (the case  $x < x_v^*$  follows once again from the fact that  $f$  is odd) and a choice of  $C_1 = C_v^* =$

$2 \cosh^2(x_v^*/2) \geq 2$  and  $C_2 = (1/D_v^*)$  (recall that, by definition,  $1/D_v^* = 1 + (\frac{1}{2}C_v^* - 1)(1 - \theta^2) \geq 1$ , satisfying the requirements of the lemma). Therefore, (3.13) now implies that

$$\int f(x_v^\xi) dQ_v^+(\xi) - \int f(x_v^\xi) dQ_v^-(\xi) \leq \frac{\theta m_v}{D_v^*(1 + \kappa m_v)},$$

as required.  $\square$

Combining Lemmas 3.5 and 3.11, we deduce that there exists a universal constant  $\kappa > 0$  such that

$$m_v \leq \sum_{w:(v,w) \in \hat{T}} \frac{\theta^2 m_w}{1 + \frac{1}{4}\kappa(1 - \theta)m_w}. \tag{3.22}$$

The proof will now follow from a theorem of [31], that links a function on the vertices of a tree  $T$  with its  $L^2$ -capacity according to certain resistances.

**Theorem 3.12** ([31, Th. 3.2] (reformulated)). *Let  $T$  be a finite tree, and suppose that there exists some  $K > 0$  and positive constants  $\{a_v : v \in T\}$  such that for every  $v \in T$  and  $x \geq 0$ ,*

$$g_v(x) \leq a_v x / (1 + Kx).$$

*Then any solution to the system  $x_v = \sum_{w:(v,w) \in T} g_w(x_w)$  satisfies*

$$x_\rho \leq \text{cap}_2(T) / K,$$

*where the resistances are given by  $R_{(u,v)} = \prod_{(x,y) \in \mathcal{P}(\rho,v)} a_y^{-1}$ , with  $\mathcal{P}(\rho, v)$  denoting the simple path between  $\rho$  and  $v$ .*

Together with inequality (3.22), the above theorem immediately gives

$$m_\rho \leq \frac{\text{cap}_2(\hat{T})}{\kappa(1 - \theta)/4},$$

completing the proof of Theorem 3.2.  $\square$

*Proof of Proposition 3.1.* In order to obtain the required result from Theorem 3.2, recall the definition of  $x_v^\xi$  for  $v \in T$ , according to which we can write

$$\hat{\mu}^\xi(\sigma(\rho) = 1) = (1 + \tanh(x_\rho^\xi/2 + h)) / 2,$$

where  $h$  is the mentioned external field at the root  $\rho$ . By monotone coupling, we can construct a probability measure  $Q_c$  on the space  $\{(\xi, \xi') : \xi \geq \xi'\}$  such that the two marginal distributions correspond to  $Q_\rho^+$  and  $Q_\rho^-$  respectively. It therefore follows that

$$\begin{aligned} \Delta &= \int \left( \hat{\mu}^\xi(\sigma(\rho) = 1) - \hat{\mu}^{\xi'}(\sigma(\rho) = 1) \right) dQ_c \\ &= \frac{1}{2} \int \left( \tanh(x_\rho^\xi/2 + h) - \tanh(x_\rho^{\xi'}/2 + h) \right) dQ_c \\ &\leq \frac{1}{2} \int \frac{x_\rho^\xi - x_\rho^{\xi'}}{2} dQ_c = \frac{1}{4} m_\rho \leq \frac{\text{cap}_2(\hat{T})}{\kappa(1 - \theta)}, \end{aligned}$$

where the last inequality follows from Theorem 3.2 using the same value of  $\kappa \geq \frac{1}{100}$ . This completes the proof.  $\square$

### 4. Upper Bound on the Inverse-Gap and Mixing Time

This section is devoted to the proof of the main theorem, Theorem 1, from which it follows that the mixing time of the continuous-time Glauber dynamics for the Ising model on a  $b$ -ary tree (with any boundary condition) is poly-logarithmic in the tree size.

Recalling the log-Sobolev results described in Sect. 2, it suffices to show an upper bound of  $O(n \log^M n)$  on inverse-gap of the discrete-time chain (equivalently, a lower bound on its  $\mathfrak{gap}$ ), which would then imply an upper bound of  $O(n \log^{M+2} n)$  for the  $L^2$  mixing-time (and hence also for the total-variation mixing-time).

The proof comprises several elements, and notably, uses a block dynamics in order to obtain the required upper bound inductively. Namely, we partition a tree on  $n$  vertices to blocks of size roughly  $n^{1-\alpha}$  each, for some small  $\alpha > 0$ , and use an induction hypothesis that treats the worst case boundary condition. The main effort is then to establish a lower bound on the spectral-gap of the block dynamics (as opposed to each of its individual blocks). This is achieved by Theorem 4.1 (stated later), whose proof hinges on the spatial-mixing result of Sect. 3, combined with the Markov chain decomposition method.

Throughout this section, let  $b \geq 2$  be fixed, denote by  $\beta_c = \operatorname{arctanh}(1/\sqrt{b})$  the critical inverse-temperature and let  $\theta = \tanh \beta_c$ .

*4.1. Block dynamics for the tree.* In what follows, we describe our choice of blocks for the above mentioned block dynamics. Let  $h$  denote the height of our  $b$ -ary tree (that is, there are  $b^h$  leaves in the tree), and define

$$\ell \triangleq \alpha h, \quad r \triangleq h - \ell, \tag{4.1}$$

where  $0 < \alpha < \frac{1}{2}$  is some (small) constant to be selected later.

For any  $v \in T$ , let  $B(v, k)$  be the subtree of height  $k - 1$  rooted at  $v$ , that is,  $B(v, k)$  consists of  $k$  levels (except when  $v$  is less than  $k$  levels away from the bottom of  $T$ ). We further let  $H_k$  denote the  $k^{\text{th}}$  level of the tree  $T$ , that according to this notation contains  $b^k$  vertices.

Next, define the set of blocks  $\mathcal{B}$  as:

$$\mathcal{B} \triangleq \{B(v, r) : v \in H_\ell \cup \{\rho\}\} \quad \text{for } \ell, r \text{ as above.} \tag{4.2}$$

That is, each block is a  $b$ -ary tree with  $r$  levels, where one of these blocks is rooted at  $\rho$ , and will be referred to as the *distinguished block*, whereas the others are rooted at the vertices of  $H_\ell$  (see Fig. 1).

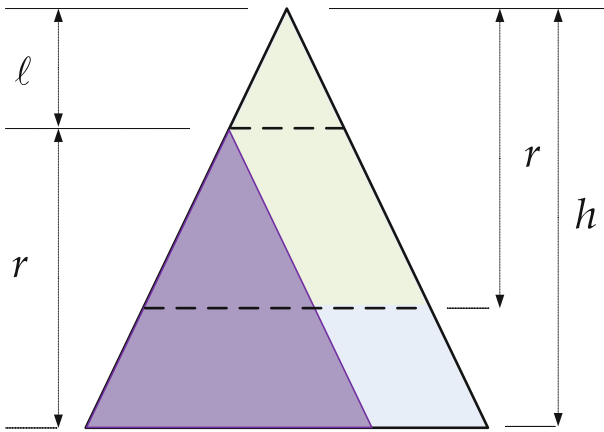
The following theorem establishes a lower bound on the spectral gap of the above-specified block dynamics (with blocks  $\mathcal{B}$ ).

**Theorem 4.1.** *Consider the Ising model on the  $b$ -ary tree at the critical inverse-temperature  $\beta_c$  and with an arbitrary boundary  $\tau$  and an arbitrary external field on the root. Let  $\mathfrak{gap}_\mathcal{B}^\tau$  be the spectral gap of the corresponding block dynamics with blocks  $\mathcal{B}$  as in (4.2). The following then holds:*

$$\mathfrak{gap}_\mathcal{B}^\tau \geq \frac{1}{4(b^\ell + 1)} \left( 1 - \frac{\alpha}{\kappa(1 - \theta)(1 - 2\alpha)} \right),$$

where  $\kappa > 0$  is the absolute constant given in Theorem 3.2.

Given the above theorem, we can now derive a proof for the main result.



**Fig. 1.** Block dynamics for the Ising model on the tree: illustration shows the distinguished block  $B(\rho, r)$  as well as a representative block of the form  $B(v, r)$  for  $v \in H_\ell$

4.2. *Proof of Theorem 1.* By definition, as  $b \geq 2$ , we have that

$$\theta = \tanh \beta_c = \frac{1}{\sqrt{b}} \leq \frac{1}{\sqrt{2}},$$

hence we can readily choose an absolute constant  $0 < \alpha < 1$  such that

$$c(\alpha) \triangleq \frac{1}{8} \left( 1 - \frac{\alpha}{\kappa(1-\theta)(1-2\alpha)} \right) > 0.$$

Let  $n_h = \sum_{j=0}^{h-1} b^j$  be the number of vertices in a  $b$ -ary tree of height  $h$  excluding its leaves, and let  $\text{gap}_h^{\tau;\xi}$  be the spectral gap of the (single-site) discrete-time Glauber dynamics for the Ising model on a  $b$ -ary tree of height  $h$  with boundary  $\tau$  and external field  $\xi$  on the root (in the special case of a free boundary condition,  $n_h$  should instead include the leaves). Define

$$g_h = n_h \min_{\tau, \xi} \text{gap}_h^{\tau;\xi}.$$

Recalling the definition of  $\mathcal{B}$  according to the above choice of  $\alpha$ , we have that each of its blocks is a tree of height  $r = (1 - \alpha)h$ , and that

$$\sup_{v \in T} \#\{B \in \mathcal{B} : v \in B\} = 2,$$

as each of the vertices in levels  $\ell, \ell + 1, \dots, r$  is covered precisely twice in  $\mathcal{B}$ , while every other vertex is covered precisely once.

Hence, by Proposition 2.5 and Theorem 4.1, it now follows that for any  $h \geq 1/\alpha$  (such that our choices of  $\ell, r$  in (4.1) are both non-zero) we have

$$g_h \geq \left( \frac{1}{4(b^\ell + 1)} \left( 1 - \frac{\alpha}{\kappa(1-\theta)(1-2\alpha)} \right) \right) g_r \cdot \frac{1}{2} = c(\alpha)g_{(1-\alpha)h}.$$

Having established the induction step, we now observe that, as  $\alpha$  is constant, clearly  $g_k \geq c'$  holds for any  $k \leq 1/\alpha$  and some fixed  $c' = c'(\alpha) > 0$ . Hence,

$$g_h \geq c' (c(\alpha))^{\log_{1-\alpha}(1/h)} = c' h^{-\log\left(\frac{1}{c(\alpha)}\right)/\log\left(\frac{1}{1-\alpha}\right)},$$

that is, there exists an absolute constant  $M$  (affected by our choice of the absolute constants  $\kappa, \alpha$ ) so that the inverse-gap of the continuous-time dynamics with an arbitrary boundary condition  $\tau$  is at most  $g_h^{-1} = O(h^M)$ , as required.  $\square$

**4.3. Proof of Theorem 4.1.** In order to obtain the desired lower bound on the spectral gap of the block dynamics, we will apply the method of decomposition of Markov chains, described in Subsect. 2.6. To this end, we will partition our configuration according to the spins of the subset

$$S \triangleq B(\rho, \ell - 1).$$

Note that  $S$  is strictly contained in the distinguished block  $B(\rho, r)$ , and does not intersect any other  $B \in \mathcal{B}$ . For  $\eta \in \{\pm 1\}^S$ , denote the set of configurations which agree with  $\eta$  by

$$\Omega_\eta \triangleq \{\sigma \in \Omega : \sigma_S = \eta\}.$$

Following the definitions in Subsect. 2.6, we can now naturally decompose the block dynamics into a projection chain  $\bar{P}$  on  $\{\pm 1\}^S$  and restriction chains  $P_\eta$  on  $\Omega_\eta$  for each  $\eta \in \{\pm 1\}^S$ . With Theorem 2.6 in mind, we now need to provide suitable lower bounds on  $\mathfrak{g}\bar{\mathfrak{a}}\mathfrak{p}^\tau$  and  $\mathfrak{g}\mathfrak{a}\mathfrak{p}_\eta^\tau$ , the respective spectral gaps of  $\bar{P}$  and  $P_\eta$  given the boundary condition  $\tau$ .

We begin with the lower bound on the restriction chain  $\mathfrak{g}\mathfrak{a}\mathfrak{p}_\eta^\tau$ , formulated in the next lemma.

**Lemma 4.2.** *For any boundary  $\tau$ , external field on the root and  $\eta \in \{\pm 1\}^S$ , the spectral gap of the restriction chain  $P_\eta$  satisfies  $\mathfrak{g}\mathfrak{a}\mathfrak{p}_\eta^\tau \geq 1/(b^\ell + 1)$ .*

*Proof.* Recall that the restriction chain  $P_\eta$  moves from  $\sigma \in \Omega_\eta$  to  $\sigma' \in \Omega_\eta$  (that is,  $\sigma$  and  $\sigma'$  both agree with  $\eta$  on  $S$ ) according to the original law of the chain, and remains at  $\sigma$  instead of moving to any  $\sigma' \notin \Omega_\eta$ . By definition of our block dynamics, this means that with probability  $b^\ell/(b^\ell + 1)$  we apply a transition kernel  $Q_1$ , that selects one of the blocks rooted at  $H_\ell$  to be updated according to its usual law (since  $S$  and all of these blocks are pairwise disjoint). On the other hand, with probability  $1/(b^\ell + 1)$ , we apply a transition kernel  $Q_2$  that updates the distinguished block, yet only allows updates that keep  $S$  unchanged (otherwise, the chain remains in place).

We next claim that the update of the distinguished block can only increase the value of  $\mathfrak{g}\mathfrak{a}\mathfrak{p}_\eta^\tau$ . To see this, consider the chain  $P'_\eta$ , in which the distinguished block is never updated; that is,  $Q_2$  described above is replaced by the identity. Clearly, since each of the vertices of  $T \setminus S$  appears in (precisely) one of the non-distinguished blocks, the stationary distribution of  $P'_\eta$  is again  $\mu_{\tau;\eta}$ , the Gibbs distribution with boundary conditions  $\eta$  and  $\tau$ . Therefore, recalling the Dirichlet form (2.2), for any  $f$  we clearly have

$$\begin{aligned} \mathcal{E}_{P'_\eta}(f) &= \frac{1}{2} \sum_{x,y \in \Omega_\eta} [f(x) - f(y)]^2 \mu^{\tau;\eta}(x) P'_\eta(x, y) \\ &\leq \frac{1}{2} \sum_{x,y \in \Omega_\eta} [f(x) - f(y)]^2 \mu^{\tau;\eta}(x) P_\eta(x, y) = \mathcal{E}_{P_\eta}(f), \end{aligned}$$

and thus, by the spectral gap bound in terms of the Dirichlet form (2.1),

$$\text{gap}(P_\eta) \geq \text{gap}(P'_\eta). \tag{4.3}$$

It remains to analyze the chain  $P'_\eta$ , which is in fact a *product chain*, and as such its eigenvalues can be directly expressed in terms of the eigenvalues of its component chains. This well known fact is stated in the following straightforward claim (cf., e.g., [1, Chap. 4] and [22, Lemma 12.11]); we include its proof for completeness.

**Claim 4.3.** *For  $j \in [d]$ , let  $P_j$  be a transition kernel on  $\Omega_j$  with eigenvalues  $\Lambda_j$ . Let  $\nu$  be a probability distribution on  $[d]$ , and define  $P'$ , the transition matrix of the product chain of the  $P_j$ -s on  $\Omega' = \Omega_1 \times \Omega_2 \times \dots \times \Omega_d$ , by*

$$P'((x_1, \dots, x_d), (y_1, \dots, y_d)) = \sum_{j=1}^d \nu(j) P_j(x_j, y_j) \prod_{i:i \neq j} \mathbf{1}_{\{x_i=y_i\}}.$$

*Then  $P'$  has eigenvalues  $\left\{ \sum_{j=1}^d \nu(j) \lambda_j : \lambda_j \in \Lambda_j \right\}$  (with multiplicities).*

*Proof.* Clearly, by induction it suffices to prove the lemma for  $d = 2$ . In this case, it is easy to verify that the transition kernel  $\tilde{P}$  can be written as

$$\tilde{P} = \nu(1)(P_1 \otimes I_{\Omega_2}) + \nu(2)(I_{\Omega_1} \otimes P_2),$$

where  $\otimes$  denotes the matrix tensor-product. Thus, by tensor arithmetic, for any  $u, v$ , eigenvectors of  $P_1, P_2$  with corresponding eigenvalues  $\lambda_1, \lambda_2$  respectively,  $(u \otimes v)$  is an eigenvector of  $\tilde{P}$  with a corresponding eigenvalue of  $\nu(1)\lambda_1 + \nu(2)\lambda_2$ , as required.  $\square$

In our setting, first notice that  $Q_1$  itself is a product chain, whose components are the  $b^\ell$  chains, uniformly selected, updating each of the non-distinguished blocks. By definition, a single block-update replaces the contents of the block with a sample according to the stationary distribution conditioned on its boundary. Therefore, each of the above mentioned component chains has a single eigenvalue of 1 whereas all its other eigenvalues are 0.

It thus follows that  $P'_\eta$  (a lazy version of  $Q_1$ ) is another product chain, giving  $Q_1$  probability  $b^\ell/(b^\ell + 1)$  and the identity chain probability  $1/(b^\ell + 1)$ . By Claim 4.3, we conclude that the possible eigenvalues of  $P'_\eta$  are precisely

$$\left\{ \frac{1}{b^\ell + 1} + \frac{1}{b^\ell + 1} \sum_{j=1}^{b^\ell} \lambda_j : \lambda_j \in \{0, 1\} \right\}.$$

In particular,  $\text{gap}(P'_\eta) = 1/(b^\ell + 1)$ , and (4.3) now completes the proof.  $\square$

It remains to provide a bound on  $\text{gap} \bar{P}^\tau$ , the spectral gap of the projection chain in the decomposition of the block dynamics according to  $S$ . This is the main part of our proof of the lower bound for the spectral gap of the block dynamics, on which the entire proof of Theorem 1 hinges. To obtain this bound, we relate the projection chain to the spatial-mixing properties of the critical Ising model on the tree under various boundary conditions, studied in Sect. 3.

**Lemma 4.4.** *For any boundary  $\tau$  and external field on the root, the spectral gap of the projection chain  $\bar{P}$  on the space  $\{\pm 1\}^S$  satisfies*

$$\text{gap}_{\bar{P}}^\tau \geq \frac{1}{b^\ell + 1} \left( 1 - \frac{\alpha}{\kappa(1-\theta)(1-2\alpha)} \right),$$

where  $\kappa > 0$  is the absolute constant given in Proposition 3.1.

We prove this lemma by establishing a certain contraction property of the projection chain  $\bar{P}$ . Recall that  $\bar{P}(\eta, \eta')$ , for  $\eta, \eta' \in \{\pm 1\}^S$ , is the probability that completing  $\eta$  into a state  $\sigma$  according to the stationary distribution (with boundary  $\eta$  and  $\tau$ ) and then applying the block dynamics transition, gives some  $\sigma'$  that agrees with  $\eta'$  on  $S$ .

Let  $S^* = H_{\ell-1}$  denote the bottom level of  $S$ , and notice that in the above definition of the transition kernel of  $\bar{P}$ , the value of the spins in  $S \setminus S^*$  do not affect the transition probabilities. Therefore, the projection of the chain  $\bar{P}$  onto  $S^*$  is itself a Markov chain, which we denote by  $\bar{P}^*$ . In fact, we claim that the eigenvalues of  $\bar{P}$  and those of  $\bar{P}^*$  are precisely the same (with the exception of additional 0-eigenvalues in  $\bar{P}$ ). To see this, first notice that the eigenfunctions of  $\bar{P}^*$  can be naturally extended into eigenfunctions of  $\bar{P}$  with the same eigenvalues (as  $\bar{P}^*$  is a projection of  $\bar{P}$ ). Furthermore, whenever  $\eta_1 \neq \eta_2 \in S$  agree on  $S^*$ , they have the same transition probabilities to any  $\eta' \in S$ , thus contributing a 0-eigenvalue to  $\bar{P}$ . It is then easy to see that all other eigenvalues of  $\bar{P}$  (beyond those that originated from  $\bar{P}^*$ ) must be 0. Altogether,

$$\text{gap}(\bar{P}^*) = \text{gap}(\bar{P}) \quad (= \text{gap}_{\bar{P}}^\tau), \tag{4.4}$$

and it remains to give a lower bound for  $\text{gap}(\bar{P}^*)$ . The next lemma shows that  $\bar{P}^*$  is contracting with respect to Hamming distance on  $\{\pm 1\}^{S^*}$ .

**Lemma 4.5.** *Let  $\bar{X}_t^*$  and  $\bar{Y}_t^*$  be instances of the chain  $\bar{P}^*$ , starting from  $\varphi$  and  $\psi$  respectively. Then there exists a coupling such that*

$$\mathbb{E}_{\varphi, \psi} \text{dist}(\bar{X}_1^*, \bar{Y}_1^*) \leq \left( \frac{b^\ell}{b^\ell + 1} + \frac{1}{b^\ell + 1} \cdot \frac{1 + (b-1)\ell}{bk(1-\theta)(r-\ell)} \right) \text{dist}(\varphi, \psi).$$

*Proof.* Clearly, if  $\varphi = \psi$  the lemma trivially holds via the identity coupling. In order to understand the setting when  $\varphi \neq \psi$ , recall the definition of the chain  $\bar{P}^*$ , which has the following two possible types of moves  $E_1$  and  $E_2$ :

- (1) With probability  $1 - \frac{1}{b^\ell + 1}$ , the block dynamics updates one of the non-distinguished blocks: denote this event by  $E_1$ . Since this operation does not affect the value of the spins in the subset  $S$  (and in particular, in  $S^*$ ), the projection chain  $\bar{P}$  remains in place in this case (and so does  $\bar{P}^*$ ).
- (2) With probability  $\frac{1}{b^\ell + 1}$ , the distinguished block is being updated: denote this event by  $E_2$ . By the discussion above, this is equivalent to the following. Let  $\eta$  denote the current state of the chain  $\bar{P}^*$ . First,  $T \setminus S$  is assigned values according to the stationary distribution with boundary  $\eta$  and  $\tau$ . Then, the distinguished block  $B(\rho, r)$  is updated given all other spins in the tree, and the resulting value of  $S$  (and hence also of  $S^*$ ) is determined by the new state of the projection chain.

By the triangle inequality, it suffices to consider the case of  $\text{dist}(\varphi, \psi) = 1$ . Suppose therefore that  $\varphi$  and  $\psi$  agree everywhere on  $S^*$  except at some vertex  $\varrho$ , and that without loss of generality,

$$\varphi(\varrho) = 1, \quad \psi(\varrho) = -1.$$



Crucially, the above mentioned procedure for the event  $E_2$  is precisely captured by the spatial-mixing properties that were studied in Sect. 3. Namely, a spin of some site  $v \in S^*$  is propagated down the tree  $T_v$  (with boundary condition  $\tau$ ), and then the new value of  $S^*$  is reconstructed from level  $r + 1$ , the external boundary of  $B(\varrho, r)$ . We construct a monotone coupling that will accomplish the required contraction property.

First, when propagating the sites  $v \in S^*$  with  $v \neq \varrho$ , we use the identity coupling (recall that  $\varphi(v) = \psi(v)$  for all  $v \neq \varrho$ ). Second, consider the process that the spin at  $\varrho$  undergoes. For  $\varphi$ , a positive spin is propagated to  $T_\varrho$  (with boundary condition  $\tau$ ) and then reconstructed from level  $r + 1$  in the tree  $T$  (which corresponds to level  $r - \ell + 1$  in the subtree  $T_\varrho$ ), with an additional boundary condition from  $T \setminus T_\varrho$  that translates into some external field. For  $\psi$ , a negative spin is propagated analogously, and notice that in its reconstruction, the exact same external field applies (as  $T \setminus T_\varrho$  was guaranteed to be the same for  $\varphi$  and  $\psi$ ).

Therefore, applying Proposition 3.1 on the tree  $T_\varrho$  with respect to the subtree  $\hat{T} = B(\varrho, r - \ell + 1)$ , we can deduce that

$$\mathbb{E}_{\varphi, \psi} (\bar{X}_1^*(\varrho) - \bar{Y}_1^*(\varrho) \mid E_2) \leq 2 \frac{\text{cap}_2(B(\varrho, r - \ell + 1))}{\kappa(1 - \theta)}, \tag{4.5}$$

where  $\kappa > \frac{1}{100}$ , and the resistances are assigned as

$$R_{(u, v)} = (\tanh \beta_c)^{-2 \text{dist}(\varrho, v)}.$$

We now turn to estimating the  $L^2$ -capacity, which is equivalent to the effective conductance between  $\varrho$  and  $\partial B(\varrho, r - \ell + 1)$ . This will follow from the well-known Nash-Williams Criterion (cf., e.g., [24]). Here and in what follows,  $R_{\text{eff}} \triangleq 1/C_{\text{eff}}$  denotes the effective resistance.

**Lemma 4.6.** (Nash-Williams Criterion [30]) *If  $\{\Pi_j\}_{j=1}^J$  is a sequence of pairwise disjoint cutsets in a network  $G$  that separate a vertex  $v$  from some set  $A$ , then*

$$R_{\text{eff}}(v \leftrightarrow A) \geq \sum_j \left( \sum_{e \in \Pi_j} \frac{1}{R_e} \right)^{-1}.$$

In our case,  $G$  is the  $b$ -ary tree  $B(\varrho, r - \ell + 1)$ , and it is natural to select its different levels as the cutsets  $\Pi_j$ . It then follows that

$$R_{\text{eff}}(\varrho \leftrightarrow \partial B(\varrho, r - \ell + 1)) \geq \sum_{k=1}^{r-\ell+1} (b^k \theta^{2k})^{-1} = r - \ell + 1, \tag{4.6}$$

where we used the fact that  $\tanh \beta_c = \theta = 1/\sqrt{b}$ . It therefore follows that

$$\text{cap}_2(B(\varrho, r - \ell + 1)) \leq \frac{1}{r - \ell},$$

which, together with (4.5), implies that

$$\mathbb{E}_{\varphi, \psi} (\bar{X}_1^*(\varrho) - \bar{Y}_1^*(\varrho) \mid E_2) \leq \frac{2}{\kappa(1 - \theta)(r - \ell)}. \tag{4.7}$$

Unfortunately, aside from controlling the probability that the spin at  $\varrho$  will coalesce in  $\varphi$  and  $\psi$ , we must also consider the probability that  $\varrho$  would remain different, and that this difference might be propagated to other vertices in  $S^*$  (as part of the update of  $B(\rho, r)$ ). Assume therefore that we updated the spin at  $\varrho$  and indeed  $\bar{X}_1^*(\varrho) \neq \bar{Y}_1^*(\varrho)$ , and next move on to updating the remaining vertices of  $S^*$ . Since our propagation processes corresponding to  $\bar{X}^*$  and  $\bar{Y}^*$  gave every vertex in  $T \setminus T_\varrho$  the same spin, it follows that each vertex  $v \in S^*$ ,  $v \neq \varrho$ , has the same external field in  $\bar{X}^*$  and  $\bar{Y}^*$ , with the exception of the effect of the spin at  $\varrho$ .

We may therefore apply the next lemma of [3], which guarantees that we can ignore this mentioned common external field when bounding the probability of propagating the difference in  $\varrho$ .

**Lemma 4.7.** ([3, Lemma 4.1]) *Let  $T$  be a finite tree and let  $v \neq w$  be vertices in  $T$ . Let  $\{J_e \geq 0 : e \in E(T)\}$  be the interactions on  $T$ , and let  $\{H(u) \in \mathbb{R} : u \in V(T)\}$  be an external field on the vertices of  $T$ . We consider the following conditional Gibbs measures:*

$$\begin{aligned} \mu^{+,H} &: \text{the Gibbs measure with external field } H \text{ conditioned on } \sigma(v) = 1, \\ \mu^{-,H} &: \text{the Gibbs measure with external field } H \text{ conditioned on } \sigma(v) = -1. \end{aligned}$$

Then  $\mu^{+,H}(\sigma(w)) - \mu^{-,H}(\sigma(w))$  achieves its maximum at  $H \equiv 0$ .

In light of the discussion above, Lemma 4.7 gives that

$$\begin{aligned} &\mathbb{E}_{\varphi, \psi} \left( \frac{1}{2} \sum_{v \in S^*} (\bar{X}_1^*(v) - \bar{Y}_1^*(v)) \mid E_2 \right) \\ &\leq \frac{1}{2} \mathbb{E}_{\varphi, \psi} (\bar{X}_1^*(\varrho) - \bar{Y}_1^*(\varrho) \mid E_2) \left( 1 + \sum_{k=1}^{\ell-1} \frac{b-1}{b} b^k \theta^{2k} \right) \\ &\leq \frac{1 + (b-1)(\ell-1)/b}{\kappa(1-\theta)(r-\ell)} = \frac{1 + (b-1)\ell}{b\kappa(1-\theta)(r-\ell)}, \end{aligned}$$

where in the first inequality we used the propagation property of the Ising model on the tree (Claim 2.1), and in the second one we used the fact that  $\theta = \tanh(\beta_c) = 1/\sqrt{b}$ , as well as the estimate in (4.7).

We conclude that there exists a monotone coupling of  $\bar{X}_t^*$  and  $\bar{Y}_t^*$  with

$$\mathbb{E}_{\varphi, \psi} (\text{dist}(\bar{X}_1^*, \bar{Y}_1^*) \mid E_2) \leq \frac{1 + (b-1)\ell}{b\kappa(1-\theta)(r-\ell)},$$

which then directly gives that

$$\mathbb{E}_{\varphi, \psi} (\text{dist}(\bar{X}_1^*, \bar{Y}_1^*)) \leq \frac{b^\ell}{b^\ell + 1} + \frac{1}{b^\ell + 1} \cdot \frac{1 + (b-1)\ell}{b\kappa(1-\theta)(r-\ell)},$$

as required.  $\square$

The above contraction property will now readily infer the required bound for the spectral gap of  $\bar{P}^*$  (and hence also for  $\mathfrak{g}\bar{\text{a}}\mathfrak{p}^\tau$ ).

*Proof of Lemma 4.4.* The following lemma of Chen [8] relates the contraction of the chain with its spectral gap:

**Lemma 4.8.** ([8]) *Let  $P$  be a transition kernel for a Markov chain on a metric space  $\Omega$ . Suppose there exists a constant  $\iota$  such that for each  $x, y \in \Omega$ , there is a coupling  $(X_1, Y_1)$  of  $P(x, \cdot)$  and  $P(y, \cdot)$  satisfying*

$$\mathbb{E}_{x,y}(\text{dist}(X_1, Y_1)) \leq \iota \text{dist}(x, y). \tag{4.8}$$

Then the spectral gap of  $P$  satisfies  $\text{gap} \geq 1 - \iota$ .

By Lemma 4.5, the requirement (4.8) is satisfied with

$$\iota = \frac{b^\ell}{b^\ell + 1} + \frac{1}{b^\ell + 1} \cdot \frac{1 + (b-1)\ell}{b\kappa(1-\theta)(r-\ell)},$$

and hence

$$\begin{aligned} \text{gap}(\bar{P}^*) &\geq 1 - \iota = \frac{1}{b^\ell + 1} \left( 1 - \frac{1 + (b-1)\ell}{b\kappa(1-\theta)(r-\ell)} \right) \\ &\geq \frac{1}{b^\ell + 1} \left( 1 - \frac{\alpha}{\kappa(1-\theta)(1-2\alpha)} \right), \end{aligned} \tag{4.9}$$

where in the last inequality we increased  $1 + (b-1)\ell$  into  $b\ell$  to simplify the final expression. This lower bound on  $\text{gap}(\bar{P}^*)$  translates via (4.4) into the desired lower bound on the spectral gap of the projection chain,  $\text{g}\bar{\text{a}}\text{p}^\tau$ .  $\square$

We are now ready to provide a lower bound on the spectral gap of the block dynamics,  $\text{gap}_\mathcal{B}^\tau$ , and thereby conclude the proof of Theorem 4.1. By applying Theorem 2.6 to our decomposition of the block dynamics chain  $P_\mathcal{B}^\tau$ ,

$$\text{gap}_\mathcal{B}^\tau \geq \frac{\text{g}\bar{\text{a}}\text{p}}{3} \wedge \frac{\text{g}\bar{\text{a}}\text{p} \cdot \text{gap}_{\min}}{3\gamma + \text{g}\bar{\text{a}}\text{p}}, \tag{4.10}$$

where

$$\text{gap}_{\min} \triangleq \min_{\eta \in \{\pm 1\}^S} \text{gap}_\eta^\tau, \quad \gamma \triangleq \max_{\eta \in \{\pm 1\}^S} \max_{x \in \Omega_\eta} \sum_{y \in \Omega \setminus \Omega_\eta} P_\mathcal{B}^\tau(x, y).$$

Lemma 4.2 gives that  $\text{gap}_{\min} \geq 1/(b^\ell + 1)$ , and clearly, as the spins in  $S$  can only change if the distinguished block is updated,  $\gamma \leq 1/(b^\ell + 1)$ . Combining these two inequalities, we obtain that

$$\frac{\text{g}\bar{\text{a}}\text{p} \cdot \text{gap}_{\min}}{3\gamma + \text{g}\bar{\text{a}}\text{p}} = \frac{\text{gap}_{\min}}{1 + 3\gamma/\text{g}\bar{\text{a}}\text{p}} \geq \frac{1}{(b^\ell + 1) + 3/\text{g}\bar{\text{a}}\text{p}} \geq \frac{1}{4} \left( \frac{1}{b^\ell + 1} \wedge \text{g}\bar{\text{a}}\text{p} \right) \tag{4.11}$$

with room to spare. Together with (4.10), this implies that

$$\text{gap}_\mathcal{B}^\tau \geq \frac{1}{4(b^\ell + 1)} \wedge \frac{1}{4} \text{g}\bar{\text{a}}\text{p},$$

and Lemma 4.4 now gives that

$$\text{gap}_\mathcal{B}^\tau \geq \frac{1}{4(b^\ell + 1)} \left( 1 - \frac{\alpha}{\kappa(1-\theta)(1-2\alpha)} \right),$$

as required. This concludes the proof of Theorem 4.1, and completes the proof of the upper bound on the mixing time.  $\square$

*Remark 4.9.* Throughout the proof of Theorem 1 we modified some of the constants (e.g., (4.9), (4.11), etc.) in order to simplify the final expressions obtained. By doing the calculations (slightly) more carefully, one can obtain an absolute constant of about 300 for the upper bound in Theorem 1.

### 5. Lower Bounds on the Mixing Time and Inverse-gap

In this section, we prove Theorem 3, which provides lower bounds on the inverse-gap and mixing time of the critical Ising model on a  $b$ -ary tree with free boundary. Throughout this section, let  $b \geq 2$  be fixed, and set  $\theta = \tanh \beta_c = \frac{1}{\sqrt{b}}$ .

*5.1. Lower bound on the inverse-gap.* The required lower bound will be obtained by an application of the Dirichlet form (2.1), using a certain weighted sum of the spins as the corresponding test function.

*Proof of Theorem 3, inequality (1.2).* Let  $T$  be a  $b$ -ary tree, rooted at  $\rho$ , with  $h$  levels (and  $n = \sum_{k=0}^h b^k$  vertices). We will show that

$$\text{gap}^{-1} \geq \frac{b-1}{6b} n h^2.$$

For simplicity, we use the abbreviation  $d(v) \triangleq \text{dist}(\rho, v)$ , and define

$$g(\sigma) \triangleq \sum_{v \in T} \theta^{d(v)} \sigma(v) \text{ for } \sigma \in \Omega.$$

By the Dirichlet form (2.1), and since  $P(\sigma, \sigma') \leq \frac{1}{n}$  for any  $\sigma, \sigma' \in \Omega$  in the discrete-time dynamics, we have that

$$\begin{aligned} \mathcal{E}(g) &= \frac{1}{2} \sum_{\sigma, \sigma'} [g(\sigma) - g(\sigma')]^2 \mu(\sigma) P(\sigma, \sigma') \\ &\leq \frac{1}{2} \max_{\sigma} \sum_{\sigma'} [g(\sigma) - g(\sigma')]^2 P(\sigma, \sigma') \leq \frac{1}{2} \sum_{k=0}^h \frac{b^k}{n} (2\theta^k)^2 \leq \frac{2(h+1)}{n}. \end{aligned}$$

On the other hand, the variance of  $g$  can be estimated as follows:

$$\begin{aligned} \text{Var}_{\mu} g &= \text{Var}_{\mu} \left( \sum_{v \in T} \theta^{d(v)} \sigma(v) \right) = \sum_{u, w \in T} \theta^{d(u)+d(w)} \text{Cov}_{\mu}(\sigma(u), \sigma(w)) \\ &= \sum_{u, v, w \in T} \theta^{d(u)+d(w)} \text{Cov}_{\mu}(\sigma(u), \sigma(w)) \mathbf{1}_{\{u \wedge w = v\}}, \end{aligned}$$

where the notation  $(u \wedge w)$  denotes their most immediate common ancestor (i.e., their common ancestor  $z$  with the largest  $d(z)$ ). Notice that for each  $v \in T$ , the number of  $u, w$  that are of distance  $i, j$  from  $v$  respectively and have  $v = u \wedge w$  is precisely  $b^i \cdot (b-1)b^{j-1}$ , since determining  $u$  immediately rules  $b^{j-1}$  candidates for  $w$ . Furthermore, by Claim 2.1 we have

$$\text{Cov}_{\mu}(\sigma(u), \sigma(w)) = \theta^{d(u)+d(w)-2d(v)},$$

and so

$$\begin{aligned} \text{Var}_\mu g &= \sum_{u,v,w \in T} \theta^{d(u)+d(w)} \theta^{d(u)+d(w)-2d(v)} \mathbf{1}_{\{u \wedge w = v\}} \\ &\geq \sum_{k=0}^h b^k \sum_{i=0}^{h-k} \sum_{j=0}^{h-k} b^i (b-1) b^{j-1} \theta^{2k+i+j} \theta^{i+j} \\ &= \frac{b-1}{b} \sum_{k=0}^h (h-k)^2 = \frac{b-1}{6b} h(h+1)(2h+1) \geq \frac{b-1}{3b} h^2(h+1). \end{aligned}$$

Altogether, we can conclude that

$$\text{gap} \leq \frac{\mathcal{E}(g)}{\text{Var}_\mu g} = \frac{6b}{b-1} \cdot \frac{1}{nh^2}, \tag{5.1}$$

as required.  $\square$

*5.2. Lower bound on the mixing-time.* In order to obtain the required lower bound on the mixing time, we consider a “speed-up” version of the dynamics, namely a custom block-dynamics comprising a mixture of singletons and large subtrees. We will show that, even for this faster version of the dynamics, the mixing time has order at least  $n \log^3 n$ .

Let  $T$  be a  $b$ -ary tree with  $h$  levels (and  $n = \sum_{k=0}^h b^k$  vertices). Consider two integers  $1 \leq \ell < r \leq h$ , to be specified later. For every  $v \in H_\ell$ , select one of its descendants in  $H_r$  arbitrarily, and denote it by  $w_v$ . Write  $W = \{w_v : v \in H_\ell\}$  as the set of all such vertices. Further define

$$B_v \triangleq (T_v \setminus T_{w_v}) \cup \{w_v\} \quad (\text{for each } v \in H_\ell).$$

The *speed-up dynamics*,  $(X_t)$ , is precisely the block-dynamics with respect to

$$\mathcal{B} = \{B_v : v \in H_\ell\} \cup \bigcup_{u \notin W} \{u\}$$

(see Fig. 2). In other words, the transition rule of the speed-up dynamics is the following:

- (i) Select a vertex  $u \in V(T)$  uniformly at random.
- (ii) If  $u \notin W$ , update this site according to the usual rule of the Glauber dynamics.
- (iii) Otherwise, update  $B_v$  given the rest of the spins, where  $v \in H_\ell$  is the unique vertex with  $u = w_v$ .

The following theorem of [33] guarantees that, starting from all-plus configuration, the speed-up Glauber dynamics indeed mixes faster than the original one. In what follows, write  $\mu \preceq \nu$  if  $\mu$  stochastically dominates  $\nu$ .

**Theorem 5.1.** ([33] and also see [32, Theorem 16.5]) *Let  $\Omega$  be the configuration space of a monotone spin-system with stationary distribution  $\pi$ . Let  $\mu$  denote the distribution on  $\Omega$  which results from successive updates at sites  $v_1, \dots, v_m$ , beginning at the top configuration. Define  $\nu$  similarly but with updates only at a subsequence  $v_{i_1}, \dots, v_{i_k}$ . Then  $\mu \preceq \nu$ , and  $\|\mu - \pi\|_{\text{TV}} \leq \|\nu - \pi\|_{\text{TV}}$ . Moreover, this also holds if the sequence  $v_1, \dots, v_m$  and the subsequence  $i_1, \dots, i_k$  are chosen at random according to any prescribed distribution.*

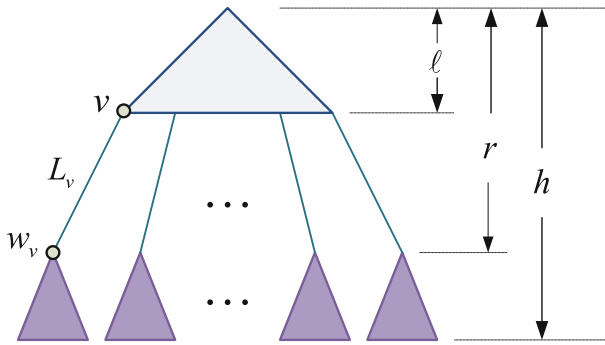


Fig. 2. Speed-up dynamics for the Ising model on the tree

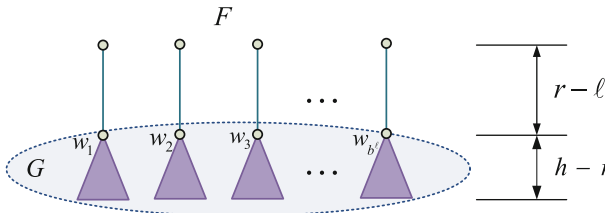


Fig. 3. Speed-up dynamics on the forest  $F$  and the sub-forest  $G$

To see that indeed the speed-up dynamics  $X_t$  is at least as fast as the usual dynamics, first note that any vertex  $u \notin W$  is updated according to the original rule of the Glauber dynamics. Second, instead of updating the block  $B_v$ , we can simulate this operation by initially updating  $w_v$  (given its neighbors), and then performing sufficiently many single-site updates in  $B_v$ . This approximates the speed-up dynamics arbitrarily well, and comprises a superset of the single-site updates of the usual dynamics. The above theorem thus completes this argument.

It remains to estimate the mixing time of the speed-up dynamics  $X_t$ . To this end, define another set of blocks as follows: for every  $v \in H_\ell$ , let  $L_v$  denote the simple path between  $v$  and  $w_v$  (inclusive), define the forest

$$F \triangleq \bigcup_{v \in H_\ell} (L_v \cup T_{w_v}),$$

and put

$$\mathcal{B}_F \triangleq \{L_v : v \in H_\ell\} \cup \bigcup_{u \in F \setminus W} \{u\}.$$

We define  $Y_t$ , the speed-up dynamics on  $F$ , to be the block-dynamics with respect to  $\mathcal{B}_F$  above (see Fig. 3). This should not be confused with running a dynamics on a subset of  $T$  with a boundary condition of the remaining vertices; rather than that,  $Y_t$  should be thought of as a dynamics on a separate graph  $F$ , which is endowed with a natural one-to-one mapping to the vertices of  $T$ . Further note that, except for the singleton blocks in  $\mathcal{B}$ , every block  $B_v \in \mathcal{B}$  in the block-dynamics  $X_t$  has a counterpart  $L_v \subset B_v$  in  $Y_t$ .

The next lemma compares the continuous-time versions of  $X_t$  and  $Y_t$  (where each block is updated at rate 1), and shows that on a certain subset of the vertices, they typically remain the same for a substantial amount of time.

**Lemma 5.2.** *Let  $(X_t)$  and  $(Y_t)$  be the continuous-time speed-up dynamics on  $T$  and  $F$  respectively, as defined above. Let  $G = \bigcup_{v \in H_\ell} T_{w_v}$  and define*

$$\tau = \inf_t \{X_t(u) \neq Y_t(u) \text{ for some } u \in V(G)\}.$$

*Then there exists a coupling of  $X_t$  and  $Y_t$  such that*

$$\mathbb{P}(\tau > t) \geq \exp(-\theta^{r-\ell} b^\ell t).$$

*Proof.* For two configurations  $\sigma \in \{\pm 1\}^T$  and  $\eta \in \{\pm 1\}^F$ , denote their Hamming distance on  $F$  by

$$\text{dist}(\sigma, \eta) = \sum_{v \in F} \mathbf{1}_{\{\sigma(v) \neq \eta(v)\}}.$$

The coupling of  $X_t$  and  $Y_t$  up to time  $\tau$  can be constructed as follows:

- (1) Whenever a singleton block  $\{u\}$  with  $u \in T \setminus F$  is being updated in  $X_t$ , the chain  $Y_t$  remains in place.
- (2) Otherwise, when a block  $B$  is updated in  $X_t$ , we update  $B \cap F$  (the unique  $B' \in \mathcal{B}_F$  with  $B' \subset B$ ) in  $Y_t$  so as to minimize  $\text{dist}(X_t, Y_t)$ .

For any  $w \in W$ , define the stopping time

$$\tau_w = \inf\{t : X_t(w) \neq Y_t(w)\},$$

and notice that in the above defined coupling we have  $\tau = \min_{w \in W} \tau_w$ , since  $W$  separates  $G \setminus W$  from  $F$ .

Let  $v \in H_\ell$  and  $w = w_v \in W$ , and suppose that block  $B_v$  is to be updated at time  $t < \tau_w$  in  $X_t$ , and hence, as defined above,  $L_v$  is to be updated in  $Y_t$ . By definition, at this time these two blocks have the same boundary except for at  $v$ , where there is a boundary condition in  $T$  (the parent of  $v$ ) and none in  $F$  (recall  $v$  is the root of one of the trees in  $F$ ).

We now wish to give an upper bound on the probability that this update will result in  $X_t(w) \neq Y_t(w)$ . By the monotonicity of the Ising model, it suffices to give an upper bound for this event in the case where  $v$  has some parent  $z$  in  $F$ , and  $X_t(z) \neq Y_t(z)$ . In this case, we can bound the probability that  $X_t(w) \neq Y_t(w)$  (in the maximal coupling) by an expression of the form

$$\frac{1}{2} \left( \mu^{+,H}(\sigma(w)) - \mu^{-,H}(\sigma(w)) \right)$$

as described in Lemma 4.7, where the external field  $H$  corresponds to the value of the spins in  $T_w \setminus \{w\}$ . Lemma 4.7 then allows us to omit the external field  $H$  at  $w$ , translating the problem into estimating the probability that a difference propagates from  $v$  to  $w$ . By Claim 2.1, we deduce that

$$\mathbb{P}(X_t(w) \neq Y_t(w)) \leq \theta^{r-\ell},$$



and therefore

$$\mathbb{P}(t < \tau_w) \geq \exp\left(-\theta^{r-\ell}t\right).$$

Using the fact  $|W| = b^\ell$ , it follows that

$$\mathbb{P}(t < \tau) = \mathbb{P}\left(t < \min_{w \in W} \tau_w\right) \geq \exp(-\theta^{r-\ell}b^\ell t),$$

as required.  $\square$

With the above estimate on the probability that  $X_t$  and  $Y_t$  are equal on the subgraph  $G$  up to a certain time-point, we can now proceed to studying the projection of  $X_t$  on  $G$  via that of  $Y_t$  (being a product chain,  $Y_t$  is much simpler to analyze).

To be precise, let  $\tilde{X}_t$  and  $\tilde{Y}_t$  denote the respective projections of  $X_t$  and  $Y_t$  onto  $G$ , which as a reminder is the union of all trees  $T_{w_v}$ . Notice that  $\tilde{Y}_t$  is precisely the continuous-time single-site Glauber dynamics on  $G$ , since the block update of  $L_v$  in  $F$  translates simply into the single-site update of  $w_v$  in  $G$ . On the other hand,  $\tilde{X}_t$  is not even necessarily a Markov chain. We next prove a lower bound on the mixing time of the Markov chain  $\tilde{Y}_t$ .

**Lemma 5.3.** *Let  $\tilde{H}_t$  be the transition kernel of  $\tilde{Y}_t$ , and let  $\mu_G$  denote its corresponding stationary measure. Let  $\mathfrak{g}_{\text{ap}}'$  denote the spectral-gap of the continuous-time single-site dynamics on a  $b$ -ary tree of height  $h - r$ . Then*

$$\|\tilde{H}_t(\underline{1}, \cdot) - \mu_G\|_{\text{TV}} > \frac{3}{5} \text{ for any } t \leq \frac{\ell \log b - 2}{2\mathfrak{g}_{\text{ap}}'},$$

where  $\underline{1}$  denotes the all-plus configuration.

*Proof.* Let  $T'$  denote a  $b$ -ary tree of height  $h - r$  and  $n'$  vertices. Let  $P'$  be the transition kernel of the corresponding discrete-time single-site Glauber dynamics on  $T'$ , let  $H'_t$  be the transition kernel of the continuous-time version of this dynamics, and let  $\mu'$  be their corresponding stationary measure.

By definition of  $G$  as a disjoint union of  $b^\ell$  copies of  $T'$ , clearly  $\tilde{Y}_t$  is a product of  $b^\ell$  copies of identical and independent component chains on  $T'$ . We can therefore reduce the analysis of  $\tilde{Y}_t$  into that of  $H'_t$ , where the second eigenvalue of its discrete-time counterpart  $P'$  plays a useful role.

The following lemma ensures that  $P'$  has an increasing eigenfunction corresponding to its second largest eigenvalue  $\lambda'$ .

**Lemma 5.4** ([29, Lemma 3]). *The second eigenvalue of the discrete-time Glauber dynamics for the Ising model has an increasing eigenfunction.*

Since the eigenspace of  $\lambda'$  has an increasing eigenfunction, it also contains a monotone eigenfunction  $f$  such that  $|f(\underline{1})| = \|f\|_\infty$ . Therefore, the transition kernel of the continuous-time chain satisfies

$$\begin{aligned} (H'_t f)(\underline{1}) &= \left( \sum_{k=0}^{\infty} e^{-tn'} \frac{(tn')^k}{k!} (P')^k f \right)(\underline{1}) \\ &= e^{-tn'} \sum_{k=0}^{\infty} \frac{(tn'\lambda')^k}{k!} f(\underline{1}) = e^{-n'(1-\lambda')t} f(\underline{1}). \end{aligned} \tag{5.2}$$

Since  $\int f d\mu' = 0$ , we have that

$$|(H'_t f)(\underline{1})| = \left| \sum_y (H'_t(\underline{1}, y) f(y) - f(y) \mu'(y)) \right| \leq 2 \|f\|_\infty \|H_t(\underline{1}, \cdot) - \mu'\|_{\text{TV}}.$$

Plugging in (5.2) and using the fact that  $|f(\underline{1})| = \|f\|_\infty$ , it follows that

$$\|H'_t(\underline{1}, \cdot) - \mu'\|_{\text{TV}} \geq \frac{1}{2} e^{-n'(1-\lambda')t}. \tag{5.3}$$

In order to relate the product chain  $\tilde{Y}_t$  to its component chain  $Y'_t$ , we will consider the Hellinger distance between certain distributions, defined next (for further details, see, e.g., [21]). First, define the *Hellinger integral* (also known as the *Hellinger affinity*) of two distribution  $\mu$  and  $\nu$  on  $\Omega$  to be

$$I_{\mathcal{H}}(\mu, \nu) \triangleq \sum_{x \in \Omega} \sqrt{\mu(x)\nu(x)}.$$

The *Hellinger distance* is now defined as

$$d_{\mathcal{H}}(\mu, \nu) \triangleq \sqrt{2 - 2I_{\mathcal{H}}(\mu, \nu)}.$$

Clearly, for any two distributions  $\mu$  and  $\nu$ ,

$$I_{\mathcal{H}}(\mu, \nu) = \sum_{x \in \Omega} \sqrt{\mu(x)\nu(x)} \geq \sum_{x \in \Omega} \mu(x) \wedge \nu(x) = 1 - \|\mu - \nu\|_{\text{TV}},$$

and so  $d_{\mathcal{H}}$  provides the following lower bound on the total variation distance:

$$\|\mu - \nu\|_{\text{TV}} \geq 1 - I_{\mathcal{H}}(\mu, \nu) = \frac{1}{2} d_{\mathcal{H}}^2(\mu, \nu). \tag{5.4}$$

Furthermore, the Hellinger distance also provides an upper bound on  $d_{\text{TV}}$ , as the next simple inequality (e.g., [15, Lemma 4.2 (i)]) shows:

$$\|\mu - \nu\|_{\text{TV}} \leq d_{\mathcal{H}}(\mu, \nu). \tag{5.5}$$

To justify this choice of a distance when working with product chains, notice that any two product measures  $\mu = \prod_{i=1}^n \mu^{(i)}$  and  $\nu = \prod_{i=1}^n \nu^{(i)}$  satisfy

$$I_{\mathcal{H}}(\mu, \nu) = \prod_{i=1}^n I_{\mathcal{H}}(\mu^{(i)}, \nu^{(i)}). \tag{5.6}$$

Next, we consider the Hellinger integral of our component chains  $H'_t$ . Indeed, combining the definition of  $d_{\mathcal{H}}$  with (5.5), we get that

$$I_{\mathcal{H}}(H'_t(\underline{1}, \cdot), \mu') \leq 1 - \frac{1}{2} \|H'_t(\underline{1}, \cdot) - \mu'\|_{\text{TV}}^2 \leq 1 - \frac{1}{8} e^{-2n'(1-\lambda')t},$$

where the last inequality is by (5.3). Therefore, applying (5.6) to the product chain  $\tilde{H}_t$  (the product of  $b^\ell$  copies of  $H'_t$ ), we can now deduce that

$$I_{\mathcal{H}}(\tilde{H}_t(\underline{1}, \cdot), \mu_G) \leq \left( 1 - \frac{1}{8} e^{-2(1-\lambda')nt'} \right)^{b^\ell}.$$

At this point, (5.4) gives that

$$\|\tilde{H}_t(\underline{1}, \cdot) - \mu_G\|_{\text{TV}} \geq 1 - \left(1 - \frac{e^{-2(1-\lambda')tn'}}{8}\right)^{b^\ell}.$$

Recall that by definition,  $\text{gap}'$  is the spectral-gap of  $H'_t$ , the continuous-time version of  $P'$ , and so  $\text{gap}' = n'(1 - \lambda')$ . Hence, if

$$t \leq \frac{\ell \log b - 2}{2\text{gap}'}$$

then

$$\|\tilde{H}_t(\underline{1}, \cdot) - \mu_G\|_{\text{TV}} \geq 1 - \exp\left(-e^2/8\right) > \frac{3}{5},$$

as required.  $\square$

The final ingredient required is the comparison between  $\mu_G$  (the Gibbs distribution on  $G$ ), and the projection of  $\mu$  (the Gibbs distribution for  $T$ ) onto the graph  $G$ . The following lemma provides an upper bound on the total-variation distance between these two measures.

**Lemma 5.5.** *Let  $\mu$  and  $\mu_G$  be the Gibbs distributions for  $T$  and  $G$  resp., and let  $\tilde{\mu}$  denote the projection of  $\mu$  onto  $G$ , that is:*

$$\tilde{\mu}(\eta) = \mu(\{\sigma \in \{\pm 1\}^T : \sigma_G = \eta\}) \quad (\text{for } \eta \in \{\pm 1\}^G).$$

Then  $\|\mu_G - \tilde{\mu}\|_{\text{TV}} \leq b^{2\ell}\theta^{2(r-\ell)}$ .

*Proof.* Recalling that  $G$  is a disjoint union of trees  $\{T_w : w \in W\}$ , clearly the configurations of these trees are independent according to  $\mu_G$ . On the other hand, with respect to  $\tilde{\mu}$ , these configurations are correlated through their first (bottom-most) common ancestor. Further notice that, by definition, the distance between  $w_i \neq w_j \in W$  in  $T$  is at least  $2(r - \ell + 1)$ , as they belong to subtrees of distinct vertices in  $H_\ell$ .

To bound the effect of the above mentioned correlation, we construct a coupling between  $\mu_G$  and  $\tilde{\mu}$  iteratively on the trees  $\{T_w : w \in W\}$ , generating the corresponding configurations  $\eta$  and  $\tilde{\eta}$ , as follows. Order  $W$  arbitrarily as  $W = \{w_1, \dots, w_{b^\ell}\}$ , and begin by coupling  $\mu_G$  and  $\tilde{\mu}$  on  $T_{w_1}$  via the identity coupling. Now, given a coupling on  $\cup_{i < k} T_{w_i}$ , we extend the coupling to  $T_{w_k}$  using a maximal coupling. Indeed, by essentially the same reasoning used for the coupling of the processes  $X_t$  and  $Y_t$  on  $G$  in Lemma 5.2, the probability that some already determined  $w_i$  (for  $i < k$ ) would affect  $w_k$  is at most  $\theta^{2(r-\ell+1)}$ . Summing these probabilities, we have that

$$\mathbb{P}\left(\eta_{T_{w_k}} \neq \tilde{\eta}_{T_{w_k}}\right) = \mathbb{P}(\eta(w_k) \neq \tilde{\eta}(w_k)) \leq (k - 1)\theta^{2(r-\ell+1)}.$$

Altogether, taking another union bound over all  $k \in [b^\ell]$ , we conclude that

$$\|\mu_G - \tilde{\mu}\|_{\text{TV}} \leq \mathbb{P}(\eta \neq \tilde{\eta}) \leq b^{2\ell}\theta^{2(r-\ell)},$$

completing the proof.  $\square$

We are now ready to prove the required lower bound on  $t_{\text{MIX}}$ .

*Proof of Theorem 3, Inequality (1.3).* As we have argued above (see Theorem 5.1 and the explanation thereafter), it suffices to establish a lower bound on the mixing time of the speed-up dynamics  $X_t$  on  $T$ . By considering the projection of this chain onto  $G$ , we have that

$$\|\mathbb{P}_1(X_t \in \cdot) - \mu\|_{\text{TV}} \geq \|\mathbb{P}_1(\tilde{X}_t \in \cdot) - \tilde{\mu}\|_{\text{TV}},$$

and recalling the definition of  $\tau$  as  $\inf_t \{(X_t)_G \neq (Y_t)_G\}$ ,

$$\begin{aligned} \|\mathbb{P}_1(\tilde{X}_t \in \cdot) - \tilde{\mu}\|_{\text{TV}} &\geq \|\mathbb{P}_1(\tilde{Y}_t \in \cdot) - \tilde{\mu}\|_{\text{TV}} - \mathbb{P}(\tau \leq t) \\ &\geq \|\mathbb{P}_1(\tilde{Y}_t \in \cdot) - \mu_G\|_{\text{TV}} - \mathbb{P}(\tau \leq t) - \|\mu_G - \tilde{\mu}\|_{\text{TV}}. \end{aligned}$$

Let  $\text{gap}$  and  $\text{gap}'$  denote the spectral-gaps of the continuous-time single-site dynamics on a  $b$ -ary tree with  $h$  levels and  $h - r$  levels respectively (and free boundary condition), and choose  $t$  such that

$$t \leq \frac{\ell \log b - 2}{2\text{gap}'}. \tag{5.7}$$

Applying Lemmas 5.2, 5.3 and 5.5, we obtain that

$$\|\mathbb{P}_1(X_t \in \cdot) - \mu\|_{\text{TV}} \geq \frac{3}{5} - (1 - \exp(-\theta^{r-\ell} b^{\ell t})) - b^{2\ell} \theta^{2(r-\ell)}. \tag{5.8}$$

Now, selecting

$$\ell = \frac{h}{5} \text{ and } r = \frac{4h}{5},$$

and recalling that  $b\theta^2 = 1$ , we have that the last two terms in (5.8) both tend to 0 as  $h \rightarrow \infty$ , and so

$$\|\mathbb{P}_1(X_t \in \cdot) - \mu\|_{\text{TV}} \geq \frac{3}{5} - o(1).$$

In particular, for a sufficiently large  $h$ , this distance is at least  $1/e$ , hence by definition the continuous-time dynamics satisfies  $t_{\text{MIX}} \geq t$ . We can now plug in our estimates for  $\text{gap}'$  to obtain the required lower bounds on  $t_{\text{MIX}}$ .

First, recall that by (5.1),

$$\text{gap}' \leq \frac{6b}{b-1} \cdot \frac{1}{(h-r)^2},$$

and so the following choice of  $t$  satisfies (5.7):

$$t \triangleq \frac{(b-1)}{12b} (h-r)^2 (\ell \log b - 2).$$

It follows that the mixing-time of the continuous-time dynamics satisfies

$$t_{\text{MIX}} \geq t \geq \left( \frac{(b-1) \log b}{1500b} + o(1) \right) h^3,$$

and the natural translation of this lower bound into the discrete-time version of the dynamics yields the lower bound in (1.3).

Second, let  $g(h)$  be the continuous-time inverse-gap of the dynamics on the  $b$ -ary tree of height  $h$  with free-boundary condition, and recall that by Theorem 1, we have that  $g$  is polynomial in  $h$ . In particular,

$$g(h) \leq Cg(h/5) \quad \text{for some fixed } C > 0 \text{ and all } h.$$

Since by definition  $(\text{gap}')^{-1} = g(h - r) = g(h/5)$  and  $\text{gap}^{-1} = g(h)$ , we can choose  $t$  to be the right-hand-side of (5.7) and obtain that for any large  $h$ ,

$$t_{\text{MIX}} \geq t \geq C' \text{gap}^{-1} h \quad \text{for some } C' > 0 \text{ fixed.}$$

Clearly, this statement also holds when both  $t_{\text{MIX}}$  and  $\text{gap}$  correspond to the discrete-time version of the dynamics, completing the proof.  $\square$

### 6. Phase Transition to Polynomial Mixing

This section contains the proof of Theorem 2, which addresses the near critical Ising model on the tree, and namely, the transition of its (continuous-time) inverse-gap and mixing-time from polynomial to exponential in the tree-height. Theorem 2 will follow directly from the next theorem:

**Theorem 6.1.** *Fix  $b \geq 2$ , let  $\varepsilon = \varepsilon(h)$  satisfy  $0 < \varepsilon < \varepsilon_0$  for a suitably small constant  $\varepsilon_0$ , and let  $\beta = \text{arctanh}(\sqrt{(1 + \varepsilon)/b})$ . The following holds for the continuous-time Glauber dynamics for the Ising model on the  $b$ -ary tree with  $h$  levels at the inverse-temperature  $\beta$ :*

- (i) *For some  $c_1 > 0$  fixed, the dynamics with free boundary satisfies*

$$\text{gap}^{-1} \geq c_1 ((1/\varepsilon) \wedge h)^2 (1 + \varepsilon)^h. \tag{6.1}$$

- (ii) *For some absolute constant  $c_2 > 0$  and any boundary condition  $\tau$ ,*

$$\text{gap}^{-1} \leq t_{\text{MIX}} \leq e^{c_2(\varepsilon h + \log h)}. \tag{6.2}$$

Throughout this section, let  $b \geq 2$  be some fixed integer, and let  $T$  be a  $b$ -ary tree with height  $h$  and  $n$  vertices. Define  $\theta = \sqrt{(1 + \varepsilon)/b}$ , where  $\varepsilon = \varepsilon(n)$  satisfies  $0 < \varepsilon \leq \varepsilon_0$  (for some suitably small constant  $\varepsilon_0 < \frac{1}{8}$  to be later specified), and as usual write  $\beta = \text{arctanh}(\theta)$ .

*Proof of Theorem 6.1.* The proof follows the same arguments of the proof of Theorems 1 and 3. Namely, the upper bound uses an inductive step using a similar block dynamics, and the decomposition of this chain to establish a bound on its gap (as in Sect. 4) via the spatial mixing properties of the Ising model on the tree (studied in Sect. 3). The lower bound will again follow from the Dirichlet form, using a testing function analogous to the one used in Sect. 5. As most of the arguments carry to the new regime of  $\beta$  in a straightforward manner, we will only specify the main adjustments one needs to make in order to extend Theorems 1 and 3 to obtain Theorem 6.1.

*Upper bound on the inverse-gap.* Let  $\frac{1}{100} < \kappa < 1$  be the universal constant that was introduced in Lemma 3.8 (and appears in Proposition 3.1 and Theorem 3.2), and define

$$\varepsilon_0 \triangleq \frac{\kappa}{20} \leq \frac{1}{8}.$$

As  $b \geq 2$  and  $\varepsilon < \varepsilon_0 \leq \frac{1}{8}$ , we have that  $\theta \leq \frac{3}{4}$ , hence Proposition 3.1 and Theorem 3.2 both hold in this supercritical setting. It therefore remains to extend the arguments in Sect. 4 (that use Proposition 3.1 as one of the ingredients in the proof of the upper bound on  $\text{gap}^{-1}$ ) to this new regime of  $\beta$ .

Begin by defining the same block dynamics as in (4.2), only with respect to the following choice of  $\ell$  and  $r$  (replacing their definition (4.1)):

$$\alpha \triangleq \varepsilon_0 = \kappa/20, \tag{6.3}$$

$$\ell \triangleq \alpha [(1/\varepsilon) \wedge h], \quad r \triangleq h - \ell. \tag{6.4}$$

Following the same notations of Sect. 4, we now need to revisit the arguments of Lemma 4.5, and extend them to the new value of  $\theta = \tanh \beta = \sqrt{(1+\varepsilon)/b}$ . This comprises the following two elements:

- (1) Bounding the  $L^2$ -capacity  $\text{cap}_2(B(\varrho, r - \ell))$ .
- (2) Estimating the probability that a difference in one spin would propagate to other spins, when coupling two instances of the chain  $\tilde{P}^*$ .

Recalling the Nash-Williams Criterion (Lemma 4.6) and its application in inequality (4.6), the effective resistance between  $\varrho$  and  $\partial B(\varrho, r - \ell)$  is at least

$$\sum_{k=1}^{r-\ell+1} (b^k \theta^{2k})^{-1} = \sum_{k=1}^{r-\ell+1} (1+\varepsilon)^{-k} = \frac{1}{\varepsilon} \left( 1 - (1+\varepsilon)^{-(r-\ell+1)} \right),$$

which implies that

$$\text{cap}_2(B(\varrho, r - \ell + 1)) \leq \frac{\varepsilon}{1 - (1+\varepsilon)^{-(r-\ell)}}. \tag{6.5}$$

Now, if  $\varepsilon \geq 1/h$ , we have

$$\begin{aligned} 1 - (1+\varepsilon)^{-(r-\ell)} &= 1 - (1+\varepsilon)^{-(h-2\alpha/\varepsilon)} \geq 1 - (1+\varepsilon)^{-(1-2\alpha)/\varepsilon} \\ &\geq 1 - e^{-(1-2\alpha)} \geq \frac{1-2\alpha}{2}, \end{aligned}$$

where the last inequality uses the fact that  $\exp(-x) \leq 1 - x + \frac{x^2}{2}$  and that  $\alpha > 0$ . Similarly, if  $\varepsilon < 1/h$  then

$$\begin{aligned} 1 - (1+\varepsilon)^{-(r-\ell)} &= 1 - (1+\varepsilon)^{-h(1-2\alpha)} \geq 1 - e^{-\varepsilon h(1-2\alpha)} \\ &\geq \varepsilon h(1-2\alpha) - \frac{(\varepsilon h(1-2\alpha))^2}{2} \geq \varepsilon h \frac{1-2\alpha}{2}, \end{aligned}$$

where in the last inequality we plugged in the fact that  $\varepsilon h < 1$ . Combining the last two equations with (6.5), we deduce that

$$\text{cap}_2(B(\varrho, r - \ell + 1)) \leq \frac{2(\varepsilon \vee (1/h))}{1 - 2\alpha}.$$

Using (4.5), it then follows that

$$\mathbb{E}_{\varphi, \psi} (\bar{X}_1^*(\varrho) - \bar{Y}_1^*(\varrho) \mid E_2) \leq \frac{2(\varepsilon \vee (1/h))}{\kappa(1-\theta)(1-2\alpha)}.$$

By repeating the next arguments of Lemma 4.5 (without any additional essential changes), we obtain that under the monotone coupling,

$$\begin{aligned} \mathbb{E}_{\varphi, \psi} (\text{dist}(\bar{X}_1^*, \bar{Y}_1^*) \mid E_2) &\leq \frac{2(\varepsilon \vee (1/h))}{\kappa(1-\theta)(1-2\alpha)} \left( 1 + \sum_{k=1}^{\ell-1} \frac{b-1}{b} b^k \theta^{2k} \right) \\ &= \frac{2(\varepsilon \vee (1/h))}{\kappa(1-\theta)(1-2\alpha)} \left( \frac{1}{b} + \frac{b-1}{b} \frac{(1+\varepsilon)^{\alpha[(1/\varepsilon) \wedge h]} - 1}{\varepsilon} \right) \\ &\leq \frac{2(\varepsilon \vee (1/h))}{\kappa(1-\theta)(1-2\alpha)} \frac{(1+\varepsilon)^{\alpha[(1/\varepsilon) \wedge h]} - 1}{\varepsilon} \leq \frac{2(\varepsilon \vee (1/h))}{\kappa(1-\theta)(1-2\alpha)} \frac{e^{\alpha[1 \wedge \varepsilon h]} - 1}{\varepsilon} \\ &\leq \frac{2(\varepsilon \vee (1/h))}{\kappa(1-\theta)(1-2\alpha)} \frac{2\alpha[1 \wedge \varepsilon h]}{\varepsilon} = \frac{4\alpha}{\kappa(1-\theta)(1-2\alpha)}, \end{aligned}$$

where in the last line we used the fact that  $e^x - 1 < 2x$  for all  $0 \leq x \leq 1$ . Again defining  $g_h = n_h \min_{\tau} \mathfrak{gap}_{\mathcal{P}_h^{\tau}}$ , we note that all the remaining arguments in Sect. 4 apply in our case without requiring any modifications, hence the following recursion holds for  $g_h$ :

$$g_h \geq c(\alpha)g_r = c(\alpha)g_{h-\alpha[(1/\varepsilon) \wedge h]}, \tag{6.6}$$

where

$$c(\alpha) \triangleq \frac{1}{8} \left( 1 - \frac{4\alpha}{\kappa(1-\theta)(1-2\alpha)} \right).$$

Recalling the definition (6.3) of  $\alpha$ , since  $\theta \leq \frac{3}{4}$  and  $\kappa < 1$  we have that

$$\frac{4\alpha}{\kappa(1-\theta)(1-2\alpha)} = \frac{2}{(1-\theta)(10-\kappa)} < \frac{8}{9},$$

and so  $c(\alpha) > 0$ . We now apply the next recursion over  $g_{h_k}$ :

$$h_0 = h, \quad h_{k+1} = \begin{cases} h_k - (\alpha/\varepsilon) & \text{if } h_k \geq (1/\varepsilon), \\ (1-\alpha)h_k & \text{if } h_k \leq (1/\varepsilon). \end{cases}$$

Notice that by our definition (6.3), we have  $\varepsilon < \varepsilon_0 = \alpha$ . With this in mind, definition (6.4) now implies that for any  $h > 1/\alpha$  we have  $\ell, r \geq 1$ . Thus, letting  $K = \min\{k : h_k \leq 1/\alpha\}$ , we can conclude from (6.6) that

$$\begin{aligned} g_{h_k} &\geq c(\alpha)g_{h_{k+1}} \text{ for all } k < K, \text{ and hence} \\ g_h &\geq (c(\alpha))^K g_{h_K}. \end{aligned}$$

By the definitions of  $h_k$  and  $K$ , clearly

$$K \leq \frac{\varepsilon}{\alpha} h + \log_{1/(1-\alpha)} (h \wedge (1/\varepsilon)) = O(\varepsilon h + \log h).$$

Since  $h_K \leq 1/\alpha$ , clearly  $g_{h_K} > c'$  for some constant  $c' = c'(\alpha) > 0$ , giving

$$g_h \geq c' (c(\alpha))^K \geq e^{-M(\varepsilon h + \log h)}$$

for some constant  $M = M(\alpha) > 0$  and any sufficiently large  $n$ . By definition of  $g_h$ , this provides an upper bound on  $\text{gap}^{-1}$ , and as  $t_{\text{MIX}} = O(\text{gap}^{-1} \log^2 n)$  (see Cor. 2.4 in Sect. 2), we obtain the upper bound on  $t_{\text{MIX}}$  that appears in (6.2).

*Lower bound on the inverse-gap.* We now turn to establishing a lower bound on the inverse-gap. Define the test function  $g$  to be the same one given in Subsect. 5.1:

$$g(\sigma) = \sum_{v \in T} \theta^{\text{dist}(\rho, v)} \sigma(v).$$

By the same calculations as in the proof of Theorem 3 (Subsect. 5.1), we have that

$$\mathcal{E}(g) \leq \frac{1}{2} \sum_{k=0}^h \frac{b^k}{n} (2\theta^k)^2 = \frac{2}{n} \sum_{k=0}^h (1+\varepsilon)^k = \frac{2}{n} \frac{(1+\varepsilon)^{h+1} - 1}{\varepsilon}, \tag{6.7}$$

whereas

$$\begin{aligned} \text{Var}_\mu(g) &= \frac{b-1}{b} \sum_{k=0}^h b^k \theta^{2k} \left( \sum_{i=0}^{h-k} b^i \theta^{2i} \right)^2 = \frac{b-1}{b} \sum_{k=0}^h (1+\varepsilon)^k \left( \sum_{i=0}^{h-k} (1+\varepsilon)^i \right)^2 \\ &= \frac{b-1}{b} \sum_{k=0}^h (1+\varepsilon)^k \left( \frac{(1+\varepsilon)^{h-k+1} - 1}{\varepsilon} \right)^2 \\ &= \frac{b-1}{b\varepsilon^2} \sum_{k=0}^h \left( (1+\varepsilon)^{2h-k+2} - 2(1+\varepsilon)^{h+1} + (1+\varepsilon)^k \right) \\ &= \frac{b-1}{b\varepsilon^3} \left( (1+\varepsilon)^{2h+3} - (2h+3)\varepsilon(1+\varepsilon)^{h+1} - 1 \right). \end{aligned} \tag{6.8}$$

When  $\varepsilon \geq 8/h$  we have

$$\begin{aligned} \frac{1}{2}(1+\varepsilon)^{h+2} - (2h+3)\varepsilon &\geq \frac{\varepsilon}{2}(h+2) + \frac{\varepsilon^2}{2} \binom{h+2}{2} - (2h+3)\varepsilon \\ &\geq (h+2)\varepsilon \left( \frac{1}{2} + \varepsilon \frac{h+1}{4} - 2 \right) \geq 4, \end{aligned}$$

and therefore in this case (6.8) gives

$$\text{Var}_\mu(g) \geq \frac{b-1}{2b} \frac{(1+\varepsilon)^{2h+3}}{\varepsilon^3}. \tag{6.9}$$

Combining (6.7) and (6.9), the Dirichlet form (2.1) now gives that

$$\text{gap} \leq \frac{4b}{b-1} \frac{\varepsilon^2}{n(1+\varepsilon)^h} \quad \text{for } \varepsilon \geq 8/h. \tag{6.10}$$



On the other hand, when  $0 \leq \varepsilon < 8/h$  we still have  $b\theta^2 \geq 1$ , and hence

$$\text{Var}_\mu(g) = \frac{b-1}{b} \sum_{k=0}^h b^k \theta^{2k} \left( \sum_{i=0}^{h-k} b^i \theta^{2i} \right)^2 \geq \frac{b-1}{b} \sum_{k=0}^h (h-k)^2 \geq \frac{b-1}{3b} h^3.$$

In addition, using the fact that the expression  $[(1 + \varepsilon)^{h+1} - 1]/\varepsilon$  in (6.7) is monotone increasing in  $\varepsilon$ , in this case we have

$$\mathcal{E}(g) \leq \frac{2}{n} \frac{(1 + (8/h))^{h+1} - 1}{8/h} \leq e^7 h/n,$$

where the last inequality holds for any  $h \geq 20$ . Altogether, the Dirichlet form (2.1) yields (for such values of  $h$ )

$$\text{gap} \leq \frac{3e^7 b}{b-1} \frac{1}{nh^2} \quad \text{for } 0 < \varepsilon \leq 8/h. \tag{6.11}$$

Combining (6.10) and (6.11), we conclude that

$$\text{gap} \leq \frac{3e^{15} b}{b-1} \left[ n(1 + \varepsilon)^h ((1/\varepsilon) \wedge h)^2 \right]^{-1},$$

where we used the fact that  $(1 + \varepsilon)^h \leq e$ . This gives the lower bound on  $\text{gap}^{-1}$  that appears in (6.1), completing the proof of Theorem 6.1.  $\square$

### 7. Concluding Remarks and Open Problems

- We have established that in the continuous-time Glauber dynamics for the critical Ising model on a regular tree with arbitrary boundary condition, both the inverse-gap and the mixing-time are polynomial in the tree-height  $h$ . This completes the picture for the phase-transition of the inverse-gap (bounded at high temperatures, polynomial at criticality and exponential at low temperatures), as conjectured by the physicists for lattices. Moreover, this provides the first proof of this phenomenon for any underlying geometry other than the complete graph.
- In addition, we studied the near-critical behavior of the inverse-gap and mixing-time. Our results yield the critical exponent of  $\beta - \beta_c$ , as well as pinpoint the threshold at which these parameters cease to be polynomial in the height.
- For further study, it would now be interesting to determine the precise power of  $h$  in the order of each the parameters  $\text{gap}^{-1}$  and  $t_{\text{MIX}}$  at the critical temperature. In the free-boundary case, our lower bounds for these parameters in Theorem 3 provide candidates for these exponents:

**Question 7.1.** Fix  $b \geq 2$  and let  $\beta_c = \text{arctanh}(1/\sqrt{b})$  be the critical inverse-temperature for the Ising model on a  $b$ -ary tree of height  $h$ . Does the corresponding continuous-time Glauber dynamics with free boundary condition satisfy  $\text{gap}^{-1} \asymp h^2$  and  $t_{\text{MIX}} \asymp h^3$ ?

- Both at critical and at near-critical temperatures, our upper bounds for the inverse-gap and mixing-time under an arbitrary boundary condition matched the behavior in the free-boundary case. This suggests that a boundary condition can only accelerate the mixing of the dynamics, and is further supported by the behavior of the model under the all-plus boundary, as established in [27]. We therefore conjecture the following monotonicity of  $\text{gap}^{-1}$  and  $t_{\text{MIX}}$  with respect to the boundary condition:

**Conjecture 7.2.** Fix  $b \geq 2$  and  $\beta > 0$ , and consider the Ising model on a  $b$ -ary tree with parameter  $\beta$ . Denote by  $\text{gap}$  and  $t_{\text{MIX}}$  the spectral-gap and mixing time for the Glauber dynamics with free boundary, and by  $\text{gap}^\tau$  and  $t_{\text{MIX}}^\tau$  those with boundary condition  $\tau$ . Then

$$\text{gap} \leq \text{gap}^\tau \quad \text{and} \quad t_{\text{MIX}} \geq t_{\text{MIX}}^\tau \quad \text{for any } \tau.$$

- A related statement was proved in [26] for two-dimensional lattices at low temperature: It was shown that, in that setting, the spectral-gap under the all-plus boundary condition is substantially larger than the spectral-gap under the free boundary condition. In light of this, it would be interesting to verify whether the monotonicity property, described in Conjecture 7.2, holds for the Ising model on an arbitrary finite graph.

*Acknowledgements.* We thank the anonymous referees for helpful comments.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

1. Aldous, D., Fill J.A.: *Reversible Markov Chains and Random Walks on Graphs*, In preparation, <http://www.stat.berkeley.edu/~aldous/RWG/book.html>
2. Baxter, R.J.: *Exactly Solved Models in Statistical Mechanics*. London: Academic Press Inc. [Harcourt Brace Jovanovich Publishers], 1989. Reprint of the 1982 original
3. Berger, N., Kenyon, C., Mossel, E., Peres, Y.: Glauber dynamics on trees and hyperbolic graphs. *Prob. Th. Rel. Fields* **131**(3), 311–340 (2005)
4. Bleher, P.M., Ruiz, J., Zagrebnov, V.A.: On the purity of the limiting Gibbs state for the Ising model on the Bethe lattice. *J. Stat. Phys.* **79**(1-2), 473–482 (1995)
5. Carlson, J.M., Chayes, J.T., Chayes, L., Sethna, J.P., Thouless, D.J.: Bethe lattice spin glass: the effects of a ferromagnetic bias and external fields. I. Bifurcation analysis. *J. Stat. Phys.* **61**(5-6), 987–1067 (1990)
6. Carlson, J.M., Chayes, J.T., Chayes, L., Sethna, J.P., Thouless, D.J.: Critical Behavior of the Bethe Lattice Spin Glass. *Europhys. Lett.* **106**(5), 355–360 (1988)
7. Carlson, J.M., Chayes, J.T., Sethna, J.P., Thouless, D.J.: Bethe lattice spin glass: the effects of a ferromagnetic bias and external fields. II. Magnetized spin-glass phase and the de Almeida-Thouless line. *J. Stat. Phys.* **61**(5-6), 1069–1084 (1990)
8. Chen, M.-F.: *Trilogy of couplings and general formulas for lower bound of spectral gap*. *Probability towards 2000* (New York, 1995), *Lecture Notes in Statist.*, Vol. **128**, New York: Springer, 1998, pp. 123–136
9. Ding, J., Lubetzky, E., Peres, Y.: The mixing time evolution of Glauber dynamics for the mean-field Ising model. *Commun. Math. Phys.* **289**(2), 725–764 (2009)
10. Diaconis, P., Saloff-Coste, L.: Comparison techniques for random walk on finite groups. *Ann. Probab.* **21**(4), 2131–2156 (1993)
11. Diaconis, P., Saloff-Coste, L.: Comparison theorems for reversible Markov chains. *Ann. Appl. Probab.* **3**(3), 696–730 (1993)
12. Diaconis, P., Saloff-Coste, L.: Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.* **6**(3), 695–750 (1996)
13. Diaconis, P., Saloff-Coste, L.: Nash inequalities for finite Markov chains. *J. Theor. Prob.* **9**(2), 459–510 (1996)

14. Domb, C., Lebowitz, J.L. (eds.): *Phase Transitions and Critical Phenomena*. Vol. **20**, San Diego, CA: Academic Press, 2001
15. Evans, W., Kenyon, C., Peres, Y., Schulman, L.J.: Broadcasting on trees and the Ising model. *Ann. Appl. Probab.* **10**(2), 410–433 (2000)
16. Hohenberg, P.C., Halperin, B.I.: Theory of dynamic critical phenomena. *Rev. Mod. Phys.* **49**(3), 435–479 (1977)
17. Ioffe, D.: *Extremality of the disordered state for the Ising model on general trees*. *Trees (Versailles, 1995)*, *Progr. Probab.*, Vol. **40**, Basel: Birkhäuser, 1996, pp. 3–14 (English, with English and French summaries)
18. Ioffe, D.: On the extremality of the disordered state for the Ising model on the Bethe lattice. *Lett. Math. Phys.* **37**(2), 137–143 (1996)
19. Jerrum, M., Son, J.-B., Tetali, P., Vigoda, E.: Elementary bounds on Poincaré and log-Sobolev constants for decomposable Markov chains. *Ann. Appl. Probab.* **14**(4), 1741–1765 (2004)
20. Lauritsen, K.B., Fogedby, H.C.: Critical exponents from power spectra. *J. Stat. Phys.* **72**(1), 189–205 (1993)
21. Le Cam, L.: *Notes on Asymptotic Methods in Statistical Decision Theory*. Centre de Recherches Mathématiques, Montreal, Que.: Université de Montréal, 1974
22. Levin, D.A., Peres, Y., Wilmer, E.: *Markov Chains and Mixing Times*. Amer. Math. Soc., 2008
23. Lyons, R.: The Ising model and percolation on trees and tree-like graphs. *Commun. Math. Phys.* **125**(2), 337–353 (1989)
24. Lyons, R., Peres, Y.: *Probability on Trees and Networks*. Cambridge University Press, 2008, In preparation. Current version is available at <http://mypage.iu.edu/~rdlyons/prbtree/book.pdf>
25. Martinelli, F.: Lectures on Glauber dynamics for discrete spin models. In: *Lectures on Probability Theory and Statistics (Saint-Flour, 1997)*, *Lecture Notes in Math.*, Vol. **1717**, Berlin: Springer, 1999, 93–191
26. Martinelli, F.: On the two-dimensional dynamical Ising model in the phase coexistence region. *J. Stat. Phys.* **76**(5-6), 1179–1246 (1994)
27. Martinelli, F., Sinclair, A., Weitz, D.: Glauber dynamics on trees: boundary conditions and mixing time. *Commun. Math. Phys.* **250**(2), 301–334 (2004)
28. Murakami, A., Yamasaki, M.: Nonlinear potentials on an infinite network. *Mem. Fac. Sci. Shimane Univ.* **26**, 15–28 (1992)
29. Nacu, Ş.: Glauber Dynamics on the Cycle is Monotone. *Prob. Th. Re. Fields* **127**, 177–185 (2003)
30. Nash-Williams, C.St.J.A.: Random walk and electric currents in networks. *Proc. Cambridge Philos. Soc.* **55**, 181–194 (1959)
31. Pemantle, R., Peres, Y.: *The critical Ising model on trees, concave recursions and nonlinear capacity*, *Ann. Probab.*, to appear, [http://arxiv.org/abs/math/0503137v2\[math.PR\]](http://arxiv.org/abs/math/0503137v2[math.PR]), 2006
32. Peres, Y.: *Lectures on “Mixing for Markov Chains and Spin Systems” (University of British Columbia, August 2005)*. Summary available at <http://www.stat.berkeley.edu/~peres/ubc.pdf>
33. Peres, Y., Winkler, P.: Can extra updates delay mixing? In preparation
34. Preston, C.J.: *Gibbs States on Countable Sets*. Cambridge Tracts in Mathematics, No. **68**, London: Cambridge University Press, 1974
35. Saloff-Coste, L.: Lectures on finite Markov chains. In: *Lectures on Probability Theory and Statistics (Saint-Flour, 1996)*, *Lecture Notes in Math.*, Vol. **1665**, Berlin: Springer, 1997, pp. 301–413
36. Soardi, P.M.: Morphisms and currents in infinite nonlinear resistive networks. *Potential Anal.* **2**(4), 315–347 (1993)
37. Soardi, P.M.: *Potential Theory on Infinite Networks*. Lecture Notes in Mathematics, Vol. **1590**, Berlin: Springer-Verlag, 1994
38. Wang, F.-G., Hu, C.-K.: Universality in dynamic critical phenomena. *Phys. Rev. E* **56**(2), 2310–2313 (1997)
39. Wilson, D.B.: Mixing times of Lozenge tiling and card shuffling Markov chains. *Ann. Appl. Probab.* **14**(1), 274–325 (2004)

Communicated by F. Toninelli