Combinatorial Criteria for Uniqueness of Gibbs Measures

Dror Weitz [†]
School of Mathematics,
Institute for Advanced Study,
Princeton, NJ 08540, U.S.A.
dror@ias.edu

April 12, 2005

Abstract

We generalize previously known conditions for uniqueness of the Gibbs measure in statistical physics models by presenting conditions of any finite size for models on any underlying graph. We give two dual conditions, one requiring that the total influence *on* a site is small, and the other that the total influence *of* a site is small. Our proofs are combinatorial in nature and use tools from the analysis of discrete Markov chains, in particular the path coupling method. The implications of our conditions for the mixing time of natural Markov chains associated with the models are discussed as well. We also present some examples of models for which the conditions hold.

[†]Supported by a grant from the State of New Jersey and by NSF grants DMS-0111298 and CCR-0324906. Most of this work was done while the author was studying at the Computer Science Division, University of California, Berkeley, supported in part by NSF grants CCR-9820951 and CCR-0121555, and by DARPA cooperative agreement F30602-00-2-060. Part of this work was done while the author was an intern at Microsoft Research.

1 Introduction

An important aspect of the study of spin systems in statistical mechanics is investigating the properties of the system when in macroscopic equilibrium. A *spin system* is composed of sites, which are the vertices of some infinite, locally finite graph (e.g., the d-dimensional Cartesian lattice $\mathbb{Z}^d$). A configuration of the system is an assignment of one of a (finite) set of spins to each site. The sites interact locally, according to the specification of the system, such that different combinations of spins on neighboring sites have different relative likelihoods. This interaction gives rise to a well defined probability distribution over configurations of any finite subset (volume) of the sites, conditional on a fixed configuration of the sites outside this subset. A *macroscopic equilibrium*, or a *Gibbs measure*, is a probability measure over the configurations of all the sites that is compatible with the conditional distributions on all finite volumes.

It is well known that, for any system of local interactions, at least one Gibbs measure always exists. However, a given system may admit multiple Gibbs measures, and one of the central issues in statistical physics is determining whether a spin system admits a unique or multiple Gibbs measures. The motivation behind this classification is locating the boundary between systems that admit a unique Gibbs measure and those admitting multiple ones. Finding this boundary is important because it identifies the points at which different systems undergo a *phase transition* in their macroscopic behavior from a unique possible equilibrium to multiple ones, a phenomenon that has additional physical manifestations. For example, the uniqueness of the Gibbs measure is equivalent to asymptotic independence between the configuration of a finite volume and the "boundary" configuration outside a large ball around this volume, and thus the phase transition points described above correspond to the emergence of long-range correlations (i.e., order) in the system. This also explains why discrete mathematicians and probabilists are interested in this subject: the question of uniqueness can be viewed combinatorially as comparing two finite distributions (conditioned on two different boundary configurations), and asking whether or not their difference goes to zero as the boundary ball recedes to infinity.

It is often the case that the distributions described above do not have succinct representations, so that analyzing the asymptotics directly is impossible. Thus, it is important to give *finite* conditions which imply uniqueness of the Gibbs measure. By "finite conditions" we mean conditions that depend only on distributions over configurations in volumes of at most some constant size, and hence can be verified by direct calculation. Dobrushin [4] was the first to give such a condition, which has become widely known as the "Dobrushin Uniqueness Condition". This condition considers only the distributions at single sites. Later, Dobrushin and Shlosman [5] gave a more general condition which may depend on larger volumes (but still of finite size). However, unlike the original Dobrushin condition, their condition is applicable only when the underlying graph of sites is an integer lattice $\mathbb{Z}^d$. Additional versions of the Dobrushin-Shlosman condition were given by others (e.g., Stroock and Zegarlinski [20]), but still only in the context of $\mathbb{Z}^d$.

In this paper, we generalize the above conditions by considering *both* larger volumes *and* any underlying graph. Naturally, all such conditions require that the influence spins at different sites have on each other is "small" in an appropriate sense. However, although they do not mention this explicitly, some of the conditions in the literature require that the total influence *on* a site is small while others require that the total influence *of* a site is small. We make a clear distinction between these two cases, giving two dual conditions, both of them in the generality described above.

Our proofs are combinatorial in nature and involve a dynamical analysis similar to that carried out in the analysis of Markov chains. We make heavy use of couplings, especially the *path coupling* method [2]. As we discuss in the text, the connection with Markov chains is part of a more general framework of connections between decay of correlations in the Gibbs measure (spatial mixing) and the mixing time of a corresponding Markov chain (temporal mixing) [20, 14, 3, 13, 8, 1].

We apply our conditions to prove uniqueness of the Gibbs measure for various models. Although

the models we discuss are already known to admit a unique Gibbs measure by other methods, for most of them, our results extend the range of parameters for which uniqueness is established using "finite size" conditions of the Dobrushin type. In addition, our applications illustrate how our two conditions may be used in different scenarios and clarify the differences between them.

The organization of the paper is as follows. In Section 2 we give definitions and necessary background on spin systems, Gibbs measures and coupling analysis, and state our results precisely. Section 3 contains the proofs of these theorems. In Section 4 we give a few extensions of our results and discuss the implications of our conditions in the Markov chain setting. Finally, in Section 5 we apply our conditions to various models, thus (re)proving that they admit a unique Gibbs measure.

2 Definitions, preliminaries and statements of results

2.1 Spin systems and Gibbs measures

Let $G = (V, E)$ be a countably infinite undirected graph that is locally finite (i.e., of bounded degree). Let $\mathcal{S}$ be a finite alphabet referred to as the *spin space*. A *configuration* is then an element $\sigma \in \Omega := \mathcal{S}^V$, or an assignment of spins to V.

We use the following terminology and notation. Elements of V are called *sites*. Subsets of V are called *regions*, and denoted by upper-case Greek letters. If Λ is a region, then $\Lambda^c := V \setminus \Lambda$ and $\partial\Lambda := \{x \in \Lambda^c \mid \exists y \in \Lambda \text{ s.t. } \{x, y\} \in E\}$ is the outer boundary of Λ. For a configuration σ we write σ_x for the spin at site x under σ, and similarly, σ_Λ for the configuration on Λ. When we write "$\sigma = \tau$ on Λ" we mean that $\sigma_\Lambda = \tau_\Lambda$. Similarly, "$\sigma = \tau$ off Λ" means that $\sigma_{\Lambda^c} = \tau_{\Lambda^c}$.

We consider spin systems with nearest neighbor interactions: each edge $\{x, y\} \in E$ is associated with a symmetric pair potential $U_{\{x,y\}} : \mathcal{S} \times \mathcal{S} \to \mathbb{R} \cup \{\infty\}$, and each vertex $x \in V$ is associated with a self potential $U_x : \mathcal{S} \to \mathbb{R} \cup \{\infty\}$. Then, for a finite region Λ, the Hamiltonian $H_\Lambda : \Omega \to \mathbb{R} \cup \{\infty\}$ is defined as

$$H_\Lambda(\sigma) := \sum_{\{x,y\} \in E \,:\, \{x,y\} \cap \Lambda \neq \emptyset} U_{\{x,y\}}(\sigma_x, \sigma_y) + \sum_{x \in \Lambda} U_x(\sigma_x).$$

The value this Hamiltonian assigns can be considered as the contribution to the energy of σ coming from Λ. Let τ specify a boundary condition. The *finite region Gibbs distribution* on Λ conditioned on τ is defined as:

$$\gamma_\Lambda^\tau(\sigma) := \begin{cases} \frac{1}{Z_\Lambda^\tau} \exp(-H_\Lambda(\sigma)) & \text{if } \sigma = \tau \text{ off } \Lambda \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

where Z_Λ^τ is the appropriate normalizing factor. Notice that by definition, the distribution on the configurations of Λ depends only on $\tau_{\partial\Lambda}$. In order to guarantee that γ_Λ^τ is well defined, it is in fact defined only for *feasible* boundary conditions. A configuration τ is said to be feasible if and only if $U_{x,y}(\tau_x, \tau_y)$ and $U_x(\tau_x)$ are finite for every edge $\{x, y\} \in E$ and every site $x \in V$[†]. Naturally, we require that the potentials give rise to at least one feasible configuration. Notice that for feasible τ, γ_Λ^τ is indeed well defined, i.e., there is always at least one configuration σ to which the distribution assigns positive probability (τ itself is such a configuration). Notice also that only feasible configurations σ may be in the support of the distribution.

Example 1 Probably the best known spin system is the *Ising* model. In this case, the spin space is $\mathcal{S} = \{-1, +1\}$, while $U_{\{x,y\}}(s_1, s_2) = -\beta \cdot s_1 \cdot s_2$ and $U_x(s) = -\beta \cdot h \cdot s$, where $\beta \in \mathbb{R}$ is the inverse temperature and $h \in \mathbb{R}$ is the external field. Thus, the energy of a configuration is linear in the number of edges with disagreeing spins, as well as the number of spins with sign opposite to that of h.

[†]Infeasible configurations exist only in systems with *hard constraints*, where some potentials may take infinite values. See Example 2

Example 2 Another famous example is the hard-core model (independent sets). In this model the spin space is $\mathcal{S} = \{0, 1\}$, and the potentials are $U_{\{x,y\}}(1, 1) = \infty$ and $U_{\{x,y\}} = 0$ otherwise, and $U_x(s) = -s \cdot \ln \lambda$, where λ is the activity parameter. The interpretation here is that a spin of 1 stands for an occupied site so a configuration specifies a subset of occupied sites. The infinite energy the edge potential assigns to a pair of occupied sites means that there is a hard constraint forbidding two neighboring sites from both being occupied. Thus, in this model, a configuration τ is feasible if and only if it specifies an independent subset of V. Furthermore, the finite Gibbs distributions are over independent sets σ, with the probability of σ being proportional to $\lambda^{|\sigma|}$, where $|\sigma|$ is the size (i.e., the number of occupied sites) of the independent set σ.

It is not difficult to verify that any finite region Gibbs distribution satisfies what are called the "DLR compatibility conditions", namely, for every Λ, any feasible τ and σ that agree off Λ, and every $\Delta \subseteq \Lambda$,

$$\gamma_\Lambda^\tau(\,\cdot\,|\sigma_{\Delta^c}) = \gamma_\Delta^\sigma. \tag{2}$$

An immediate consequence is that γ_Λ^τ is *stationary* under γ_Δ. We illustrate what stationarity means with the following two-step process (over configurations on Λ). In the first step, a configuration σ is chosen according to γ_Λ^τ. In the second step, a configuration is chosen according to γ_Δ^σ. Stationarity means that the resulting distribution of the two-step process is the same as if we only execute the first step, namely, choosing from γ_Λ^τ.

The collection of all the finite region Gibbs distributions γ_Λ^τ as Λ and τ vary is referred to as the *specification* γ. Clearly, γ is completely determined by the set of pair and singleton potentials, but in the sequel it will often be more convenient to consider the specification γ, rather than the potentials, as representing the spin system. The notion of DLR compatibility motivates the definition of probability measures on the infinite space that are compatible with a specification γ.

Definition 2.1 *A probability measure μ over the subset of feasible configurations is called a* Gibbs measure *for the specification γ if, for every finite region Λ and μ-almost every configuration σ,*

$$\mu(\,\cdot\,|\sigma_{\Lambda^c}) = \gamma_\Lambda^\sigma.$$

The physical intuition for a Gibbs measure is that it describes a macroscopic equilibrium, i.e., all parts of the system are in equilibrium with their boundaries.

It is well known that for any specification γ derived as above, a Gibbs measure always exists. However, several Gibbs measures (or "phases") for a given specification may coexist (see, e.g., [9] or [10] for details and more on Gibbs measures). As explained in the Introduction, a central goal is to classify a given specification as admitting either a unique Gibbs measure or multiple ones. Usually, a description of a spin system includes a macroscopic parameter (such as temperature in the Ising model) and the aim is to classify the range of parameter values into two regimes, one where the Gibbs measure is unique, and the other where there are multiple Gibbs measures. For example, the Ising model on the square integer lattice $\mathbb{Z}^2$ with no external field admits a unique Gibbs measure when the temperature is above a known critical value $T_c = 1/\beta_c$, and two distinct Gibbs measures when the temperature is below T_c. One of these Gibbs measures is the limit of γ_Λ^τ as Λ goes to $\mathbb{Z}^2$, where the boundary configuration τ is the all-$(+)$ configuration. The other Gibbs measure is the same limit where τ is the all-$(-)$ configuration.

Since the Gibbs measure is unique if and only if the limit of finite volume Gibbs distributions is unique, the notion of uniqueness can be interpreted as an asymptotic independence between the configuration of a finite region and a distant boundary configuration. In order to write the above in a formal way, we introduce the following notation. Let μ_1 and μ_2 be two probability measures on Ω, and Λ be a finite region. Then

$$\|\mu_1 - \mu_2\|_\Lambda \;:=\; \max_{A \subseteq \mathcal{S}^\Lambda} |\mu_1(A) - \mu_2(A)|, \tag{3}$$

i.e., $\|\mu_1 - \mu_2\|_\Lambda$ is the *total variation distance* between the projections of μ_1 and μ_2 on $\mathcal{S}^\Lambda$. The Gibbs measure for the specification γ is unique if and only if the following condition holds.

Proposition 2.2 *A specification γ admits a unique Gibbs measure if and only if for every finite region Λ there exists an infinite sequence of regions $\Lambda \subset \Psi_1 \subset \Psi_2 \subset \ldots \subset \Psi_m \subset \ldots$ such that every $x \in V$ is in Ψ_m for some m and*

$$\sup_{\tau,\sigma} \|\gamma^\tau_{\Psi_m} - \gamma^\sigma_{\Psi_m}\|_\Lambda \underset{m\to\infty}{\longrightarrow} 0. \tag{4}$$

Proof: The proposition is standard and we prove the forward direction (the one relevant for our purposes) for completeness; the reverse direction follows from the fact that for any boundary condition τ, the limit of $\gamma^\tau_{\Psi_m}$ as $m \to \infty$ (at least along appropriate subsequences) is a Gibbs measure (see, e.g., Chapter 2 in [10] or Chapter 4 in [9] for details). Before we continue with the proof, we mention that two Gibbs measures μ_1, μ_2 are considered the same if for every finite region Λ, the projections of μ_1, μ_2 on $\mathcal{S}^\Lambda$ are the same. (Again, see [10] or [9] for details and justification). We go on with the proof and fix an arbitrary finite region Λ. If μ is a Gibbs measure then by definition, for every $\Psi \supseteq \Lambda$, the projection of μ on $\mathcal{S}^\Lambda$ is a convex combination of the projections of γ^σ_Ψ on $\mathcal{S}^\Lambda$ as σ varies. Thus, if μ_1 and μ_2 are two Gibbs measures then for every m, $\|\mu_1 - \mu_2\|_\Lambda \leq \sup_{\tau,\sigma} \|\gamma^\tau_{\Psi_m} - \gamma^\sigma_{\Psi_m}\|_\Lambda$. Now, if (4) holds then by taking $m \to \infty$ we conclude that the projections of μ_1 and μ_2 on $\mathcal{S}^\Lambda$ are the same and hence that $\mu_1 = \mu_2$. □

Clearly, it is not possible to mechanically verify the condition in Proposition 2.2 since it involves inspecting infinitely many distributions in regions of arbitrary size. Thus, it is important to give general tools that allow one, by performing finite calculations, to classify a given specification as admitting either a unique or multiple Gibbs measures. Examples of such tools are the Dobrushin [4] and Dobrushin-Shlosman [5] conditions. These are conditions on the finite Gibbs distributions over an appropriate set of bounded diameter regions which ensure that the Gibbs measure is unique. These conditions have turned out to be very useful, since they can be verified by direct calculation for a number of models in appropriate parameter ranges, thus implying uniqueness of the Gibbs measure in a rather straightforward way. In this paper, we generalize the Dobrushin and Dobrushin-Shlosman conditions, extending their applicability as well as providing new insights into the theory underlying them.

2.2 Dynamics

The conditions we give (and their proofs) employ notions and tools used in the construction and analysis of local Markov chains that are designed to sample from the Gibbs distribution. In Markov chains of this type a step is a random update of a finite size region, i.e., in each step of the chain the configuration in some finite region (block) is replaced by a configuration chosen from some distribution, where this distribution is determined by the current configuration. These distributions are constructed in a way that guarantees convergence of the chain to the Gibbs distribution. In this subsection we set the notation for and define local updates in precise terms that are useful for our discussion.

Let $\{\Theta_i\}_{i=1,2,\ldots}$ be a collection of finite regions (blocks) that cover G finitely many times, i.e., each site is included in at least one and at most finitely many Θ_i. In addition, each Θ_i is assigned a positive weight w_i. We refer to an element of the collection as a *block* (rather than a region) in order to distinguish blocks from other regions mentioned in our discussion. In a local update rule, the collection of blocks $\{\Theta_i\}$ specifies the blocks whose configuration may be updated. The relevance of the weights w_i is that the updated block is chosen at random from some finite subset of the collection, and the probability of choosing Θ_i is proportional to w_i.

Once a weighted collection of blocks is given, the second ingredient needed in order to complete the specification of an update rule is the collection of distributions that govern the result of an

update. As mentioned above, the distribution over resulting configurations depends on the current configuration. Thus, we need to specify a collection of distributions $\kappa = \{\kappa_i^\tau\}$, indexed by the current configuration τ and the index i of the block to be updated. Naturally, these distributions have to be consistent with the Gibbs measure:

Definition 2.3 *We say that κ is a* local update rule *for the specification γ w.r.t. the collection of blocks $\{\Theta_i\}$ if $\kappa = \{\kappa_i^\tau\}$ is a collection of probability distributions such that:*

1. *for every configuration τ and every i, κ_i^τ is a probability distribution on the configurations that agree with τ off Θ_i;*
2. *for every feasible τ and i, $\gamma^\tau_{\Theta_i}$ is stationary under κ_i, where the notion of stationarity was explained following equation (2);*
3. *the projections of κ_i^τ and κ_i^σ on $\mathcal{S}^{\Theta_i}$ are the same whenever τ and σ agree on $\Theta_i \cup \partial\Theta_i$, i.e., the distribution κ_i^τ (on the configurations in Θ_i) depends only on $\tau_{\Theta_i \cup \partial\Theta_i}$.*

Property 1 guarantees that only Θ_i is updated under κ_i^τ while the rest of the configuration remains unchanged. Property 2 ensures that the update rule is consistent with the Gibbs distribution[‡]. Property 3 is a locality requirement in the sense that the result of an update of Θ_i depends only on the current configuration in Θ_i and its immediate neighborhood; this is a natural requirement since the Gibbs distribution in Θ_i is also local, i.e., depends only on $\partial\Theta_i$.

We wish to emphasize the following facts regarding local update rules. First, unlike γ, we require that κ_i^τ is defined even for infeasible τ. However, the stationarity requirement does not apply to infeasible configurations, and thus, unless τ agrees with some feasible configuration on $\Theta_i \cup \partial\Theta_i$, the specification γ imposes no restriction on the distribution κ_i^τ. Second, the fact that $\gamma^\tau_{\Theta_i}$ is stationary under κ_i for every feasible τ implies by (2) that, for any $\Lambda \supseteq \Theta_i$ and every feasible τ, γ^τ_Λ is stationary under κ_i, i.e., any Gibbs distribution in any region that includes Θ_i is unaffected by an update of Θ_i. Third, unlike $\gamma^\tau_{\Theta_i}$, κ_i^τ may depend on the configuration inside Θ_i (as well as the configuration on $\partial\Theta_i$). Nevertheless, a natural choice for κ_i^τ is simply $\gamma^\tau_{\Theta_i}$ (known as the "heat-bath" update), but other possible and reasonable choices exist. As an example of other possible local update rules, consider a "Metropolis" update where κ_i^τ is the distribution resulting from the following process. First, update the configuration in Θ_i by choosing it u.a.r. from $\mathcal{S}^{\Theta_i}$; suppose the resulting configuration is σ. Then, output σ ("accept") with probability $\min\left\{\frac{\exp(-H_{\Theta_i}(\sigma))}{\exp(-H_{\Theta_i}(\tau))}, 1\right\}$ and otherwise output τ ("reject"). There are other examples of more sophisticated update rules which are specific to certain models (e.g., the update rule for proper colorings described in [21], or the one for independent sets [7] which we discuss in Section 5).

We now go on with our definitions. Since (by property 1 of Definition 2.3) the spin of a given site may change only when updating a block that includes this site, we will often need to refer to the subset of such blocks. For a site x, let $B(x) := \{i \mid x \in \Theta_i\}$. Similarly, for a region Λ, $B(\Lambda) := \{i \mid \Lambda \cap \Theta_i \neq \emptyset\} = \bigcup_{x\in\Lambda} B(x)$. Finite subsets of block indices arise throughout our discussion, and for such a subset S we write $w_S := \sum_{i\in S} w_i$ for its aggregated weight.

A common tool for analyzing Markov chains that use a local update rule is to couple the updates of Θ_i starting from two different configurations. A *coupling* of two distributions μ_1 and μ_2 is any joint distribution whose marginals are μ_1 and μ_2. For any two configurations η and ξ that differ in exactly one site, let $K_i(\eta,\xi)$ be a coupling of κ_i^η and κ_i^ξ. (These atomic couplings determine a coupling $K(\eta,\xi)$ for arbitrary pairs of configurations η,ξ that differ in more than one site, using the path coupling construction explained in Section 3.2). If η and ξ agree on $\Theta_i \cup \partial\Theta_i$, $K_i(\eta,\xi)$ is

[‡]Notice that we only require that the Gibbs distribution is stationary w.r.t. the update and not that it is the unique stationary distribution, as is the case when constructing a Markov chain for sampling from the Gibbs distribution. See Section 4.2 for a discussion on the Markov chains setting.

always defined as the coupling where the two configurations agree on Θ_i with probability 1. We call the collection $\{K_i\}$, denoted K, a *coupled update rule* for γ. From here onwards, when we refer to a coupled update rule K, we assume it implicitly specifies the collection of blocks $\{\Theta_i\}$, their weights $\{w_i\}$ and the local update rule κ according to which K is defined.

Our aim is to give conditions on K that imply uniqueness of the Gibbs measure for the specification γ. Namely, our theorems will be of the form: "If there exists a coupled update rule K for the specification γ such that K satisfies certain conditions, then there is a unique Gibbs measure that is consistent with γ." The conditions on K will require that under a coupled update, the average "distance" between the two coupled configurations is small. Our notion of distance is specified per site. Let $\rho = \{\rho_x\}_{x\in G}$ be a collection of metrics on the spin space $\mathcal{S}$. We write $\rho_x(\eta,\xi)$ for $\rho_x(\eta_x,\xi_x)$, and abuse this notation when considering a coupling Q by writing $\rho_x(Q)$ for the average distance (w.r.t. the joint distribution Q) between the two coupled configurations. Our notion of distance is extended to regions by summing over single sites, i.e., we let $\rho_\Lambda(\eta,\xi) := \sum_{x\in\Lambda}\rho_x(\eta,\xi)$. To illustrate the above notion of distance, we note that in the applications given in Section 5 the metrics we use are of the form $\rho_x = u_x \cdot \rho_\delta$, where $u_x \in \mathbb{R}^+$ is a weight associated with the site x and $\rho_\delta(s_1,s_2) = 1$ if $s_1 \neq s_2$ (and naturally, $\rho_\delta(s_1,s_2) = 0$ if $s_1 = s_2$). In this case, $\rho_x(Q)$ is just u_x times the probability that the spins at x differ under the coupling Q, and $\rho_\Lambda(Q)$ is the average weighted *Hamming* distance between the two coupled configurations in Λ.

Our theorems below consider collections of metrics with the following two natural properties. The first property states that the distance at any single site is bounded by a uniform constant: we say that a collection of metrics $\{\rho_x\}$ is *bounded* if $\sup_{x\in V}\max_{s_1,s_2\in\mathcal{S}}\rho_x(s_1,s_2)$ is finite, where we recall that V is the set of vertices of G. The second, stronger property states that the total distance in arbitrarily large regions is bounded by a uniform constant: we say that a collection of metrics $\rho = \{\rho_x\}$ is *summable* if $\sum_{x\in V}\max_{s_1,s_2\in\mathcal{S}}\rho_x(s_1,s_2)$ is finite.

2.3 Results

Once a coupled update rule K and a collection of metrics $\rho = \{\rho_x\}$ are fixed, we are in a position to define the influence of a site y on another site x (w.r.t. K and ρ) in an analogous way to the definition of the "matrix of dependencies" in Dobrushin's condition [4].

Definition 2.4 *For a given coupled update rule K and collection of metrics ρ, define the* influence of site y on site x*, denoted $I_{x\leftarrow y}$, as the smallest constant for which, for all pairs of configurations (η,ξ) s.t. $\eta = \xi$ off y,*

$$\sum_{i\in B(x)} w_i\rho_x(K_i(\eta,\xi)) \;\le\; \rho_y(\eta,\xi)I_{x\leftarrow y}.$$

The motivation for the above definition is that $I_{x\leftarrow y}\,/\,w_{B(x)}$ is an upper bound on the average distance between the coupled spins at x (relative to the initial distance between the spins at y) at the end of the following procedure: starting from two configurations that may differ only at y, choose a block $\Theta_i \in B(x)$ with probability $w_i\,/\,w_{B(x)}$ and perform a coupled update of Θ_i. Note that $I_{x\leftarrow y} = 0$ if $y \notin \bigcup_{i\in B(x)}(\Theta_i \cup \partial\Theta_i)$ (i.e., only sites in or adjacent to blocks containing x may have non-zero influence on x). We write $I_{x\leftarrow} := \sum_y I_{x\leftarrow y}$ for the sum of influences of all sites on the site x (and notice by the previous remark that this sum is finite). Our first theorem states that, if the normalized total influence on every site w.r.t. a bounded collection of metrics is less than 1, then the Gibbs measure is unique.

Theorem 2.5 *If a specification γ admits a coupled update rule K together with a bounded collection of metrics ρ for which $\sup_x\left\{I_{x\leftarrow}\,/\,w_{B(x)}\right\} < 1$, then the Gibbs measure for γ is unique.*

We note that our requirement that the metric collection be bounded is necessary. In our discussion of applications in Section 5 we give an example of a specification that admits multiple Gibbs

measures but for which there exists a coupled update rule and an *unbounded* metric collection that satisfy the condition in Theorem 2.5.

Remark: Previously known conditions involving the total influence on a site are the single site Dobrushin condition [4] and the condition referred to as $DSU(Y)$ by Stroock and Zegarlinski [20]. Both conditions only consider the case in which $\rho_x = \rho_\delta$ for all x, where ρ_δ was defined at the end of Section 2.2. In addition, the Dobrushin condition only considers the case in which each Θ_i is a single site. The condition of Stroock and Zegarlinski, while considering blocks of larger size as we do, only considers the special case where the underlying graph G is an integer lattice $\mathbb{Z}^d$. Thus, our Theorem 2.5 is a generalization of both Dobrushin's condition and the Stroock and Zegarlinski one.

In our second theorem we consider a natural *dual* condition, namely, that the total influence *of* every site is small. Following the line established in the previous condition, we write $I_{\leftarrow y}$ for the total influence of site y. Although it might seem natural to define $I_{\leftarrow y}$ as $\sum_x I_{x \leftarrow y}$, the appropriate definition turns out to be a slightly more relaxed one obtained by changing the order of quantification over pairs of configurations:

Definition 2.6 *For a given coupled update rule K and collection of metrics ρ, define the* total influence of site y, *denoted $I_{\leftarrow y}$, as the smallest constant for which, for all pairs of configurations (η, ξ) s.t. $\eta = \xi$ off y,*

$$\sum_i w_i \rho_{\Theta_i}(K_i(\eta, \xi)) \;\leq\; \rho_y(\eta, \xi) I_{\leftarrow y}.$$

Again, there are only finitely many non-zero terms in the sum since there are only finitely many blocks Θ_i which are affected by y. The relevance of this definition comes from the fact that $I_{\leftarrow y}$ is related to the average total distance resulting from an update of a block randomly chosen from those affected by y, when starting from two configurations that differ only at y. (The exact relationship between $I_{\leftarrow y}$ and this distance is rather involved; the detailed bound is given in Section 3). To see the connection to the previous definition of influence, notice that $I_{\leftarrow y} \leq \sum_x I_{x \leftarrow y}$. In fact, the only difference between these two expressions is that in $\sum_x I_{x \leftarrow y}$ the quantification over pairs of configurations is taken separately for each x, while in the definition of $I_{\leftarrow y}$ the quantification is taken once, before summing over x (the summation over x comes from the expansion of ρ_{Θ_i}).

Compared to the condition in Theorem 2.5, our condition for uniqueness based on the influence *of* a site places a stronger restriction on the metric collection we are allowed to use by requiring that it be summable.

Theorem 2.7 *If a specification γ admits a coupled update rule K together with a summable collection of metrics ρ that satisfy $\sup_y w_{B(y)} < \infty$, $\inf_y w_{B(y)} > 0$ and $\sup_y \left\{ I_{\leftarrow y} \,/\, w_{B(y)} \right\} < 1$, then the Gibbs measure for γ is unique.*

Again, the requirement that the metric collection be summable is necessary as is illustrated in Section 5, where we also show that the condition $\sup_y w_{B(y)} < \infty$ is necessary. It is not clear whether the requirement that $\inf_y w_{B(y)} > 0$ is necessary or just an artifact of our proof.

Remark: A previously known condition involving the total influence of a site was given by Dobrushin and Shlosman [5]. However, they only considered the case where the underlying graph G is an integer lattice $\mathbb{Z}^d$ and the collection of blocks $\{\Theta_i\}$ is the set of all translations of some fixed subset Θ. In addition, in their condition there is freedom to specify only one metric ρ, so that $\rho_x = \rho$ for all $x \in \mathbb{Z}^d$. Notice that this means, in our language, that the resulting collection of metrics is not summable, which at first sight seems not to fit the framework of Theorem 2.7. However, Theorem 2.7 can still be seen as a generalization of the Dobrushin-Shlosman condition as we now explain. Suppose, as in the Dobrushin-Shlosman setting, that there exists a coupled update rule K and a single metric ρ for which the condition in Theorem 2.7 holds with $\rho_x = \rho$ for all $x \in \mathbb{Z}^d$, and that the diameter of the blocks Θ_i used by K is bounded by some constant r. We can then construct a slightly modified collection of metrics by letting $\rho'_x = (1+\epsilon)^{-|x|}\rho$, where $|x|$ stands for the distance of the site x from the origin of $\mathbb{Z}^d$ and $\epsilon > 0$ is a small enough constant. Since the volume of a ball around the origin of $\mathbb{Z}^d$ grows subexponentially with the ball's radius, ρ' is clearly summable for any $\epsilon > 0$.

On the other hand, it is not too difficult to see that if the condition in Theorem 2.7 holds w.r.t. ρ, and if ϵ is small enough, then the condition also holds with ρ replaced by ρ'. The reason for this is that the influence of a site can increase by a factor of at most $(1+\epsilon)^r$ when replacing ρ by ρ'. In fact, in their proof Dobrushin and Shlosman use a similar construction to the above. Furthermore, the fact that in their condition the metric is the same for all sites restricts their condition to models on $\mathbb{Z}^d$ (or, more precisely, to models on graphs of sub-exponential growth). By allowing different metrics for different sites (but requiring that the collection is summable) we are able to handle arbitrary graphs with no restriction on their geometry.

Remark: At this point it is also worth mentioning that in the literature, the Dobrushin-Shlosman condition is sometimes referred to as a direct extension of the single-site Dobrushin condition although in fact the two conditions are dual in nature. The reason for this misconception is that the Dobrushin-Shlosman condition was only stated for translation invariant update rules (for ease of notation), allowing the authors to write it in terms of the total influence *on* a site (or on a block) even though the property they used in the proof is that the total influence *of* a site is small (inequalities 2.24 and 2.26 in the proof of Lemma 2.2 in [5]). To clarify this point further, notice that for specifications on $\mathbb{Z}^d$, when the coupled update rule is translation invariant and the metrics ρ_x are uniform in x, then the matrix of dependencies is translation invariant as well, i.e., $I_{x\leftarrow y}$ depends only on $x-y$ (the difference between the two d-dimensional vectors x and y). Therefore, $\sum_x I_{x\leftarrow y} = \sum_y I_{x\leftarrow y}$ and thus $\sup_y I_{\leftarrow y} \le \sup_y \sum_x I_{x\leftarrow y} = \sup_x \sum_y I_{x\leftarrow y} = \sup_x I_{x\leftarrow}$. In other words, in this setting, if the condition involving the total influence on a site holds (Theorem 2.5) then so does the condition involving the total influence of a site (Theorem 2.7).

Remark: As we will discuss in more detail in section 4.2, conditions like the ones in Theorems 2.5 and 2.7 also imply that the Markov chain corresponding to the update rule κ has essentially optimal mixing time. This is in fact part of a more general framework of connections between optimal mixing time of the Markov chain and uniqueness of the Gibbs measure. We also note that in the context of Markov chains the duality between influence *on* and influence *of* a site was already mentioned in [2], where it was referred to as a duality between conditions on the rows and on the columns of the dependency matrix respectively.

3 Proofs

3.1 Framework

Our theorems state that under certain conditions the Gibbs measure for a given specification γ is unique. Thus, following Proposition 2.2, we will show that if the hypothesis of the theorems is true then for every finite region Λ we can find an infinite sequence of finite regions $\{\Lambda_m\}_{m=0,1,2,\ldots}$ such that $\Lambda = \Lambda_0 \subseteq \Lambda_1 \subseteq \ldots \subseteq \Lambda_m \subseteq \ldots$ and for every two (boundary) configurations σ and τ, $\|\gamma^\sigma_{\Lambda_m} - \gamma^\tau_{\Lambda_m}\|_\Lambda \to 0$ as $m \to \infty$.

The construction of the sequence $\{\Lambda_m\}$ depends on the collection of blocks $\{\Theta_i\}$ used by the coupled update rule given in the hypothesis of the theorems. For a subset of block indices S, let $\Phi(S) := \bigcup_{i\in S}(\Theta_i \cup \partial\Theta_i)$ stand for the region of sites that may influence the result of an update of a block from S. Then the sequence $\{\Lambda_m\}$ is defined recursively as $\Lambda_0 = \Lambda$ and $\Lambda_{m+1} = \Phi(B(\Lambda_m))$. The important property of this sequence is that, if $x \in \Lambda_m$, then all the sites that have non-zero influence on x (via a coupled update) are included in Λ_{m+1}. Notice also that, since every site is included in at least one block Θ_i, then $\Psi \subseteq \Phi(B(\Psi))$ and therefore $\Lambda_m \subseteq \Lambda_{m+1}$. It is also easy to see that the sequence $\{\Lambda_m\}$ consumes V, i.e., that every site $x \in V$ is in some Λ_m.

The proofs of both our theorems will take the following form. For an arbitrary finite region Λ and arbitrary boundary configurations σ and τ, using the given coupled update rule we will construct a coupling Q_m of $\gamma^\sigma_{\Lambda_m}$ and $\gamma^\tau_{\Lambda_m}$ such that $\rho_\Lambda(Q_m)$ is exponentially small in m, and in particular, vanishes as m increases. This will conclude the proofs since, if η and ξ stand for the two coupled configurations under Q_m, then

$$\|\gamma^\sigma_{\Lambda_m} - \gamma^\tau_{\Lambda_m}\|_\Lambda \;\le\; \Pr{}_{Q_m}(\eta_\Lambda \neq \xi_\Lambda) \;\le\; \frac{\rho_\Lambda(Q_m)}{\min_{\eta_\Lambda\neq\xi_\Lambda} \rho_\Lambda(\eta_\Lambda,\xi_\Lambda)}\,, \tag{5}$$

and $\min_{\eta_\Lambda\neq\xi_\Lambda} \rho_\Lambda(\eta_\Lambda,\xi_\Lambda) > 0$ because ρ_Λ is a metric on $\mathcal{S}^\Lambda$.

3.2 Path coupling

When a coupled update rule K is given then $K_i(\eta, \xi)$ is specified only for pairs (η, ξ) that differ in a single site. Based on these atomic couplings, in this subsection we extend this definition to coupled updates for arbitrary pairs of starting configurations. Before doing so, we set notation for an update of a random block. Let S be a finite set of natural numbers indexing blocks. We write $\kappa_S^\eta := (\sum_{i \in S} w_i \kappa_i^\eta) \,/\, w_S$ for the distribution resulting from updating a random block from S starting from configuration η, where the probability of updating Θ_i for $i \in S$ is proportional to w_i. Similarly, we write $K_S(\eta, \xi) := (\sum_{i \in S} w_i K_i(\eta, \xi)) \,/\, w_S$ for a coupled update of a random block from the set S starting from configurations η and ξ that differ at a single site. Notice that $K_S(\eta, \xi)$ is indeed a coupling of κ_S^η and κ_S^ξ.

We now extend the definition of K_S to arbitrary pairs of starting configurations. We first consider pairs (η, ξ) that agree on $\Phi(S)$ (but may differ in arbitrarily many sites elsewhere). Notice that such pairs induce the same distribution on configurations of $\Phi(S)$ when updating a random block from S, and thus we define $K_S(\eta, \xi)$ as the trivial coupling where the two resulting configurations agree on $\Phi(S)$ with probability 1. For general η and ξ, K_S is defined using a *path coupling*. Path couplings (in a more general setting) were introduced in [2] where they were used to upper bound the mixing time of certain Markov chains, although similar ideas were already used in the proofs of the uniqueness conditions in [4] and [5].

The idea of a path coupling is to interpolate over differences at single sites, thus reducing the definition of the coupling for general starting pairs (η, ξ) to those that differ at exactly one site. Although in the literature the interpolation is usually taken only over the sites at which η and ξ differ, here, in order to ease notation, we interpolate over all sites in $\Phi(S)$. Let $z_1, z_2, \ldots, z_n$ be an enumeration of the sites in $\Phi(S)$, where $n = |\Phi(S)|$. Given η and ξ, we then construct a sequence of configurations $\eta^{(0)}, \eta^{(1)}, \ldots, \eta^{(n)}$ such that $\eta^{(0)} = \eta$, and for $1 \leq j \leq n$, $\eta_x^{(j)} = \eta_x^{(j-1)}$ for all $x \neq z_j$ while $\eta_{z_j}^{(j)} = \xi_{z_j}$. Observe that for every $1 \leq j \leq n$, $\eta^{(j)}$ agrees with ξ on $\{z_1, \ldots, z_j\}$ and with η on $\{z_{j+1}, \ldots, z_n\}$. In particular, $\eta^{(n)}$ agrees with ξ on $\Phi(S)$. Furthermore, $\eta^{(j-1)}$ and $\eta^{(j)}$ may only disagree at z_j, and $\eta^{(j)} = \eta^{(j-1)}$ if and only if η and ξ assign the same spin to z_j.

Using the above notation, observe that the couplings $K_S(\eta^{(j-1)}, \eta^{(j)})$ are already defined for all $1 \leq j \leq n$, as is the (trivial) coupling $K_S(\eta^{(n)}, \xi)$. We go on to construct the coupling $K_S(\eta, \xi)$. Recall that $K_S(\eta, \xi)$ should be a coupling of κ_S^η and κ_S^ξ, i.e, a coupling of the update of a random block Θ_i, where $i \in S$, starting from η and ξ respectively. To construct this coupling, first choose a configuration $\sigma^{(0)}$ from κ_S^η. Then, choose a configuration $\sigma^{(1)}$ from $\kappa_S^{\eta^{(1)}}$ according to the coupling $K_S(\eta, \eta^{(1)})$ conditioned on $\sigma^{(0)}$ being the first configuration in the pair. It is easy to verify that the unconditional distribution of $\sigma^{(1)}$ is indeed $\kappa_S^{\eta^{(1)}}$. Continuing inductively, in step j, choose a configuration $\sigma^{(j)}$ from $\kappa_S^{\eta^{(j)}}$ according to the coupling $K_S(\eta^{(j-1)}, \eta_j)$ conditioned on $\sigma^{(j-1)}$. Finally, choose a configuration $\sigma^{(n+1)}$ from κ_S^ξ according to the trivial coupling $K_S(\eta^{(n)}, \xi)$ conditioned on $\sigma^{(n)}$. (The last coupling changes the configuration outside $\Phi(S)$ from η to ξ). Notice that the joint distribution of $\sigma^{(0)}, \sigma^{(1)}, \ldots, \sigma^{(n+1)}$ is a simultaneous coupling of the distributions $\kappa_S^\eta, \kappa_S^{\eta^{(1)}}, \ldots, \kappa_S^{\eta^{(n)}}, \kappa_S^\xi$. We define $K_S(\eta, \xi)$ as the joint distribution of $\sigma^{(0)}$ and $\sigma^{(n+1)}$, which is indeed a coupling of κ_S^η and κ_S^ξ.

The coupling $K_S(\eta, \xi)$ defined above has the following important property, which can be verified using the triangle inequality for metrics together with the fact that in the above construction the joint distribution of $\sigma^{(j-1)}$ and $\sigma^{(j)}$ is $K_S(\eta^{(j-1)}, \eta^{(j)})$, by definition. For every region $\Delta \subseteq \Phi(S)$,

$$\rho_\Delta(K_S(\eta, \xi)) \;\leq\; \sum_{j=1}^{n} \rho_\Delta(K_S(\eta^{(j-1)}, \eta^{(j)})). \tag{6}$$

Now that a coupled update is defined for any two starting configurations, we can define an

operator on couplings which, for a given coupling Q, specifies the the result of a coupled update when starting from two configurations chosen from Q.

Definition 3.1 *Let Q be a coupling of two probability distributions μ_1 and μ_2 on Ω. Define*

$$F_S(Q) \;:=\; Q \cdot K_S \;=\; \sum_{\eta,\xi} Q(\eta,\xi) K_S(\eta,\xi),$$

where $Q(\eta,\xi)$ is the measure of the pair (η,ξ) under the joint distribution Q. Equivalently, viewing Q as a probability distribution on $\Omega\times\Omega$ and K_S as a Markov kernel on $\Omega\times\Omega$, $F_S(Q)$ stands for the distribution resulting from taking one step in the Markov chain defined by K_S when the starting state is chosen according to Q.

Remark: Even though the space of pairs of configurations is infinite, we used a finite sum notation in Definition 3.1 since in what follows Q will always be a finite distribution, i.e., the support of Q will be a finite subset of pairs of configurations.

Notice that if K is a coupled update rule for γ, and if Q is a coupling of γ^σ_Ψ and γ^τ_Ψ for some $\Psi \supseteq \bigcup_{i\in S}\Theta_i$ and two arbitrary (boundary) configurations σ and τ, then $F_S(Q)$ is a coupling of these two distributions as well. This is because both distributions are stationary under an update of Θ_i for any $i\in S$.

As a final piece of notation, F_S^t stands for t applications of F_S and is the analogue of performing t coupled steps in a Markov chain.

3.3 Influence on a site

In this subsection we give the proof of Theorem 2.5, namely, that when the influence *on* every site is small, the Gibbs measure is unique. Theorem 2.5 is an immediate consequence of the following theorem.

Theorem 3.2 *Let γ be a specification, K a coupled update rule for γ and $\rho=\{\rho_x\}$ a collection of metrics. For any $\delta>0$, let $\alpha=\delta+\sup_x\left\{I_{x\leftarrow}\,/\,w_{B(x)}\right\}$, where $I_{x\leftarrow}$ is defined w.r.t. K and ρ. Then, for every finite region Λ, every positive integer m and any two boundary configurations σ and τ, there is a coupling Q_m of $\gamma^\sigma_{\Lambda_m}$ and $\gamma^\tau_{\Lambda_m}$ s.t. $\rho_\Lambda(Q_m)\le c|\Lambda|\alpha^m$, where $c=\max_{x\in\Lambda_m}\max_{s_1,s_2\in\mathcal{S}}\rho_x(s_1,s_2)$ and the definition of Λ_m is as in Section 3.1.*

Notice that if $\sup_x\left\{I_{x\leftarrow}\,/\,w_{B(x)}\right\}<1$ as in the hypothesis of Theorem 2.5 then there exists $\delta>0$ such that $\alpha=\delta+\sup_x I_{x\leftarrow}\,/\,w_{B(x)}<1$. Furthermore, for a bounded collection of metrics (as in the hypothesis of Theorem 2.5), $c=\max_{x\in\Lambda_m}\max_{s_1,s_2\in\mathcal{S}}\rho_x(s_1,s_2)$ is bounded by a constant independent of m. Thus, Theorem 2.5 follows from Theorem 3.2 as explained at the end of Section 3.1.

The proof of Theorem 3.2 is based on the following lemma, which for an update of a random block gives an upper bound on the average distance at a site x as a function of the initial distances in the neighborhood of x.

Lemma 3.3 *Fix a coupled update rule K and a collection of metrics ρ. Let Q be any coupling, x any site and S any finite subset of block indices such that $B(x)\subseteq S$. Then*

$$\rho_x(F_S(Q)) \;\le\; \left(1-\frac{w_{B(x)}}{w_S}\right)\rho_x(Q) \;+\; \frac{I_{x\leftarrow}}{w_S}\sup_{y\in\Phi(B(x))}\rho_y(Q). \tag{7}$$

Proof: The idea here is that the first term on the r.h.s. of (7) represents the contribution to the distance at x when the updated block is not in $B(x)$ (in which case the two spins at x remain unchanged as does the distance at x) while the second term represents the contribution to the distance when the updated block is one from $B(x)$, in which case the distance can be bounded by

the total influence on x times the maximum distance of a site that may influence x, as explained below. We proceed with the formal proof. By definition,

$$\rho_x(F_S(Q)) \;=\; \rho_x\left(\sum_{\eta,\xi} Q(\eta,\xi)\, K_S(\eta,\xi)\right) \;=\; \sum_{\eta,\xi} Q(\eta,\xi)\, \rho_x(K_S(\eta,\xi)).$$

We now recall the notation used in the construction of the path coupling in Section 3.2, i.e., let $z_1,\ldots,z_n$ enumerate the sites of $\Phi(S)$, where $n=|\Phi(S)|$, and for given η and ξ let $\eta=\eta^{(0)},\ldots,\eta^{(n)}$ be the corresponding sequence of configurations. Then, using (6),

$$\begin{aligned}
\rho_x(F_S(Q)) &\le \sum_{\eta,\xi} Q(\eta,\xi) \sum_{j=1}^{n} \rho_x(K_S(\eta^{(j-1)},\eta^{(j)})) \\
&= \frac{1}{w_S}\sum_{\eta,\xi} Q(\eta,\xi) \sum_{j=1}^{n}\left[\sum_{i\in S\setminus B(x)} w_i\rho_x(K_i(\eta^{(j-1)},\eta^{(j)})) + \sum_{i\in B(x)} w_i\rho_x(K_i(\eta^{(j-1)},\eta^{(j)}))\right] \\
&= \frac{1}{w_S}\sum_{\eta,\xi} Q(\eta,\xi)\left[\rho_x(\eta,\xi)\sum_{i\in S\setminus B(x)} w_i + \sum_{j=1}^{n}\sum_{i\in B(x)} w_i\rho_x(K_i(\eta^{(j-1)},\eta^{(j)}))\right] \\
&= \left(1-\frac{w_{B(x)}}{w_S}\right)\rho_x(Q) + \frac{1}{w_S}\sum_{j=1}^{n}\sum_{\eta,\xi} Q(\eta,\xi)\sum_{i\in B(x)} w_i\rho_x(K_i(\eta^{(j-1)},\eta^{(j)})),
\end{aligned}$$

where we made use of the facts that for $i\notin B(x)$, $\rho_x(K_i(\eta^{(j-1)},\eta^{(j)}) = \rho_x(\eta^{(j-1)},\eta^{(j)})$ and that $\sum_{j=1}^{n}\rho_x(\eta^{(j-1)},\eta^{(j)}) = \rho_x(\eta,\xi)$. What remains to be shown is that

$$\sum_{j=1}^{n}\sum_{\eta,\xi} Q(\eta,\xi)\sum_{i\in B(x)} w_i\rho_x(K_i(\eta^{(j-1)},\eta^{(j)})) \;\le\; I_{x\leftarrow}\sup_{y\in\Phi(B(x))}\rho_y(Q). \tag{8}$$

Notice, however, that since $\eta^{(j-1)}$ and $\eta^{(j)}$ may differ only at z_j then $\sum_{i\in B(x)} w_i\rho_x(K_i(\eta^{(j-1)},\eta^{(j)})) \le \rho_{z_j}(\eta^{(j-1)},\eta^{(j)})I_{x\leftarrow z_j}$. Thus, the l.h.s. of (8) is bounded by

$$\begin{aligned}
\sum_{j=1}^{n} I_{x\leftarrow z_j}\sum_{\eta,\xi} Q(\eta,\xi)\rho_{z_j}(\eta^{(j-1)},\eta^{(j)}) &= \sum_{j=1}^{n} I_{x\leftarrow z_j}\sum_{\eta,\xi} Q(\eta,\xi)\rho_{z_j}(\eta,\xi) \\
&= \sum_{j=1}^{n} I_{x\leftarrow z_j}\rho_{z_j}(Q) \\
&\le \sup_{y\in\Phi(B(x))}\{\rho_y(Q)\}\sum_{y} I_{x\leftarrow y} \\
&= I_{x\leftarrow}\sup_{y\in\Phi(B(x))}\rho_y(Q),
\end{aligned}$$

where we used the fact that $I_{x\leftarrow y}=0$ for $y\notin\Phi(B(x))$. □

Lemma 3.3 is useful since it uses only first order information about Q in order to bound $\rho_x(F_S(Q))$, i.e., we only need to know bounds on the average distances at single sites regardless of how these distances depend on each other under Q. In the proof of Theorem 3.2 below, we use Lemma 3.3 iteratively to improve the bounds on single site distances.

Proof of Theorem 3.2: For the δ given in the theorem, let $t_m = \lceil \left(w_{B(\Lambda_{m-1})} / \min_{x\in\Lambda_{m-1}} \left\{w_{B(x)}\right\}\right) \ln \frac{1}{\delta} \rceil$. We will show that for any coupling Q, every $0 \le k \le m$, $x \in \Lambda_{m-k}$, and $t \ge kt_m$,

$$\rho_x(F^t_{B(\Lambda_{m-1})}(Q)) \;\le\; c\alpha^k, \tag{9}$$

where c and α are as defined in Theorem 3.2. The theorem follows from (9) as explained next. Take any coupling Q of $\gamma^\sigma_{\Lambda_m}$ and $\gamma^\tau_{\Lambda_m}$ (for example, the product coupling). Then, for every t, $F^t_{B(\Lambda_{m-1})}(Q)$ is also a coupling of $\gamma^\sigma_{\Lambda_m}$ and $\gamma^\tau_{\Lambda_m}$ because the update rule is consistent with γ and all the blocks that might be updated in the process are included in $\Phi(B(\Lambda_{m-1})) = \Lambda_m$ by definition. Thus, by setting $t = mt_m$ we get a coupling Q_m for which $\rho_x(Q_m) \le c\alpha^m$ for every $x \in \Lambda_0 = \Lambda$. Hence, $\rho_\Lambda(Q_m) \le c|\Lambda|\alpha^m$, as required.

We go on to prove (9). Notice that the bound in (9) improves as time increases but only when the distance of x from the boundary increases as well, i.e., we only have to consider sites in Λ_{m-k}. The idea of the proof is that once we have established a bound for sites in Λ_{m-k}, we can improve this bound for a site $x \in \Lambda_{m-k-1}$ by updating a random block from the ones that cover x, since all the sites that influence x are in Λ_{m-k}. The chosen time parameter ensures that we will indeed update a block from those that cover x with high probability.

The formal proof proceeds by induction on k. The base case ($k = 0$) is clear since $\rho_x(Q) \le \max_{s_1,s_2\in\mathcal{S}} \rho_x(s_1, s_2) \le c$ for every $x \in \Lambda_m$ by definition of c. We assume (9) for k and *show* for $k+1$. Fix an arbitrary $x \in \Lambda_{m-k-1}$. We have to show that for every $t \ge (k+1)t_m$, $\rho_x(F^t_{B(\Lambda_{m-1})}(Q)) \le c\alpha^{k+1}$. Notice that $y \in \Lambda_{m-k}$ for every $y \in \Phi(B(x))$ and hence we can use the induction hypothesis together with Lemma 3.3 to get that, for every $t > kt_m$,

$$\rho_x(F^t_{B(\Lambda_{m-1})}(Q)) \;\le\; \left(1 - \frac{w_{B(x)}}{w_{B(\Lambda_{m-1})}}\right) \rho_x(F^{t-1}_{B(\Lambda_{m-1})}(Q)) + \frac{I_{x\leftarrow}}{w_{B(\Lambda_{m-1})}} c\alpha^k.$$

Therefore,

$$\rho_x(F^t_{B(\Lambda_{m-1})}(Q)) - \frac{I_{x\leftarrow}}{w_{B(x)}} c\alpha^k \;\le\; \left(1 - \frac{w_{B(x)}}{w_{B(\Lambda_{m-1})}}\right) \left[\rho_x(F^{t-1}_{B(\Lambda_{m-1})}(Q)) - \frac{I_{x\leftarrow}}{w_{B(x)}} c\alpha^k\right]$$

and hence, since by the induction hypothesis $\rho_x(F^{kt_m}_{B(\Lambda_{m-1})}(Q)) \le c\alpha^k$, then for all $t \ge kt_m$,

$$\rho_x(F^t_{B(\Lambda_{m-1})}(Q)) \;\le\; \frac{I_{x\leftarrow}}{w_{B(x)}} c\alpha^k + \left(1 - \frac{w_{B(x)}}{w_{B(\Lambda_{m-1})}}\right)^{t-kt_m} c\alpha^k.$$

In particular, for all $t \ge (k+1)t_m$,

$$\rho_x(F^t_{B(\Lambda_{m-1})}(Q)) \;\le\; \frac{I_{x\leftarrow}}{w_{B(x)}} c\alpha^k + \delta c\alpha^k \;\le\; c\alpha^{k+1}.$$

This concludes the proof of (9) and thus completes the proof of Theorem 3.2. □

3.4 Influence of a site

In this section we prove Theorem 2.7, namely, that when the influence *of* every site is small, the Gibbs measure is unique. In contrast to the proof in the previous section, where we used the bound on the influence *on* a site to show that the distance at *every* site decreases exponentially as we recede from the boundary, here we will use the bound on the influence *of* a site to show that the *total* distance decreases exponentially. Theorem 2.5 is an immediate consequence of the following theorem.

Theorem 3.4 *Let γ be a specification, K a coupled update rule for γ and $\rho = \{\rho_x\}$ a collection of metrics. For any $\delta > 0$, let $\alpha = \delta + \sup_y \{I_{\leftarrow y}\} \,/\, \left(\sup_y \{I_{\leftarrow y}\} + \inf_y \left\{w_{B(y)} - I_{\leftarrow y}\right\}\right)$, where $I_{\leftarrow y}$ is defined w.r.t. K and ρ. Then, for every region Λ, any positive integer m and any two boundary configurations σ and τ, there is a coupling Q of $\gamma^{\sigma}_{\Lambda_{m+1}}$ and $\gamma^{\tau}_{\Lambda_{m+1}}$ s.t. $\rho_\Lambda(Q) \leq c\alpha^m$, where $c = \max_{\eta,\xi} \rho_{\Lambda_m}(\eta,\xi)$ and Λ_m is as defined in Section 3.1.*

Notice that when ρ is summable then the combination of the conditions $\sup_y w_{B(y)} < \infty$, $\inf_y w_{B(y)} > 0$ and $\sup_y \left\{I_{\leftarrow y} \,/\, w_{B(y)}\right\} < 1$ in the hypothesis of Theorem 2.7 is equivalent to the condition $\sup_y \{I_{\leftarrow y}\} \,/\, \left(\sup_y \{I_{\leftarrow y}\} + \inf_y \left\{w_{B(y)} - I_{\leftarrow y}\right\}\right) < 1$. Therefore, for K and ρ as in Theorem 2.7, there exists $\delta > 0$ for which $\alpha < 1$, where α is as defined in Theorem 3.4. Furthermore, the summability of the collection of metrics in Theorem 2.7 implies that $c = \max_{\eta,\xi} \rho_{\Lambda_m}(\eta,\xi)$ is bounded by a constant independent of m. Thus, Theorem 2.7 follows from Theorem 3.4 as explained at the end of Section 3.1.

The proof of Theorem 3.4 is based on the following lemma, which is similar in spirit to Lemma 3.3, but rather than bounding the average distance at a single site, here we bound the average total distance in a region Δ as a function of the initial average total distance in the neighborhood of Δ, when updating of a random block.

Lemma 3.5 *Fix a coupled update rule K and a collection of metrics ρ. Let Q be any coupling, Δ any region and S any finite subset of block indices such that $B(\Delta) \subseteq S$. Let $\mathrm{MAX} = \max_{y \in \Phi(B(\Delta))} \{I_{\leftarrow y}\}$ and $\mathrm{MIN} = \min_{y \in \Phi(B(\Delta))} \left\{w_{B(y)} - I_{\leftarrow y}\right\}$. Then*

$$\rho_\Delta(F_S(Q)) \;\leq\; \left(1 - \frac{\mathrm{MAX} + \mathrm{MIN}}{w_S}\right) \rho_\Delta(Q) \;+\; \frac{\mathrm{MAX}}{w_S} \rho_{\Phi(B(\Delta))}(Q). \tag{10}$$

Proof: We start by using the path coupling bound (6) to get

$$\rho_\Delta(F_S(Q)) \;\leq\; \sum_{\eta,\xi} Q(\eta,\xi) \sum_{j=1}^{n} \rho_\Delta(K_S(\eta^{(j-1)}, \eta^{(j)})), \tag{11}$$

where $n = |\Phi(S)|$ and the sequence of configurations $\eta^{(j)}$ is as defined in the construction of the path coupling. In turn, we can bound $\rho_\Delta(K_S(\eta^{(j-1)}, \eta^{(j)}))$ depending on the location of z_j (the only site at which $\eta^{(j-1)}$ and $\eta^{(j)}$ may differ) as follows:

$$\rho_\Delta(K_S(\eta^{(j-1)}, \eta^{(j)})) \;\leq\; \rho_{z_j}(\eta^{(j-1)}, \eta^{(j)}) \times \begin{cases} I_{\leftarrow z_j} \,/\, w_S + 1 - w_{B(z_j)} \,/\, w_S & z_j \in \Delta; \\ I_{\leftarrow z_j} \,/\, w_S & z_j \in \Phi(B(\Delta)) \setminus \Delta; \\ 0 & z_j \notin \Phi(B(\Delta)). \end{cases} \tag{12}$$

Notice that (12) follows from the fact that $K_S = \left(\sum_{i \in S} w_i K_i\right) / w_S$, the definition of $I_{\leftarrow z_j}$ and the following four observations. First, when $z_j \in \Theta_i$, $\rho_\Delta(K_i(\eta^{(j-1)}, \eta^{(j)})) \leq \rho_{\Theta_i}(K_i(\eta^{(j-1)}, \eta^{(j)}))$ because all the sites outside Θ_i remain unchanged and thus the coupled spins of all sites outside Θ_i agree with certainty in the coupling $K_i(\eta^{(j-1)}, \eta^{(j)})$. Second, when $z_j \in \Delta \setminus \Theta_i$ then in addition to the distance at Θ_i, there may be positive distance at z_j, which is not accounted for by the distance in Θ_i but which needs to be accounted for as part of the distance in Δ. Thus, in this case, $\rho_\Delta(K_i(\eta^{(j-1)}, \eta^{(j)})) \leq \rho_{z_j}(\eta^{(j-1)}, \eta^{(j)})) + \rho_{\Theta_i}(K_i(\eta^{(j-1)}, \eta^{(j)}))$, where we used the fact that the distance at z_j remains unchanged by the update of Θ_i. Third, when $z_j \notin \Delta$, $\rho_\Delta(K_i(\eta^{(j-1)}, \eta^{(j)})) \leq \rho_{\Theta_i}(K_i(\eta^{(j-1)}, \eta^{(j)}))$ regardless of whether $z_j \in \Theta_i$ or not because there is no need to count the distance at z_j. Fourth, if $z_j \notin \Phi(B(\Delta))$ then z_j cannot influence the resulting configuration in Δ, i.e., $\rho_\Delta(K_i(\eta^{(j-1)}, \eta^{(j)})) = 0$ for all i. This is because the only updates that may incur a non-zero distance at Δ are of blocks for which $\Theta_i \cap \Delta \neq \emptyset$, but then $\eta^{(j-1)}$ and $\eta^{(j)}$ agree on $\Theta_i \cup \partial\Theta_i$ since $z_j \notin \Phi(B(\Delta))$ so the distance in Θ_i remains zero.

Now, by plugging the bounds in (12) into the r.h.s. of (11), and since $\rho_{z_j}(\eta^{(j-1)},\eta^{(j)})) = \rho_{z_j}(\eta,\xi)$ and $\rho_\Delta = \sum_{y\in\Delta}\rho_y$, we get:

$$\begin{aligned}
\rho_\Delta(F_S(Q)) &\le \sum_{y\in\Delta}\left(1-\frac{w_{B(y)}}{w_S}+\frac{I_{\leftarrow y}}{w_S}\right)\rho_y(Q)+\sum_{y\in\Phi(B(\Delta))\setminus\Delta}\frac{I_{\leftarrow y}}{w_S}\rho_y(Q)\\
&\le \left(1-\frac{\min_{y\in\Delta}\left\{w_{B(y)}-I_{\leftarrow y}\right\}}{w_S}\right)\rho_\Delta(Q)+\frac{\max_{y\in\Phi(B(\Delta))\setminus\Delta}\left\{I_{\leftarrow y}\right\}}{w_S}\rho_{\Phi(B(\Delta))\setminus\Delta}(Q)\\
&\le \left(1-\frac{\mathrm{MIN}+\mathrm{MAX}}{w_S}+\frac{\mathrm{MAX}}{w_S}\right)\rho_\Delta(Q)+\frac{\mathrm{MAX}}{w_S}\rho_{\Phi(B(\Delta))\setminus\Delta}(Q)\\
&= \left(1-\frac{\mathrm{MIN}+\mathrm{MAX}}{w_S}\right)\rho_\Delta(Q)+\frac{\mathrm{MAX}}{w_S}\rho_{\Phi(B(\Delta))}(Q),
\end{aligned}$$

where we used the fact that $\rho_\Delta+\rho_{\Phi(B(\Delta))\setminus\Delta}=\rho_{\Phi(B(\Delta))}$. □

From here onwards the proof of Theorem 3.4 continues in a very similar manner to that of the proof of Theorem 2.5, using Lemma 3.5 iteratively to improve the bounds on average distances in regions.

Proof of Theorem 3.4: For the δ given in the theorem, let $t_m = \lceil \left(w_{B(\Lambda_m)} \,/\, \min_{y\in\Lambda_m}\left\{w_{B(y)}\right\}\right)\ln\frac{1}{\delta}\rceil$. We will show that for any coupling Q, every $0\le k\le m$, $x\in\Lambda_{m-k}$, and $t\ge kt_m$,

$$\rho_{\Lambda_{m-k}}(F^t_{B(\Lambda_m)}(Q)) \;\le\; c\alpha^k, \tag{13}$$

where c and α are as defined in Theorem 3.4. The theorem follows from (13) as explained next. Take any coupling Q of $\gamma^\sigma_{\Lambda_{m+1}}$ and $\gamma^\tau_{\Lambda_{m+1}}$. Then, as we already explained in the proof of Theorem 3.2, for every t, $Q' = F^t_{B(\Lambda_m)}(Q)$ is also a coupling of $\gamma^\sigma_{\Lambda_{m+1}}$ and $\gamma^\tau_{\Lambda_{m+1}}$. Thus, by setting $t = mt_m$ we get a coupling Q' for which $\rho_{\Lambda_0}(Q')\le c\alpha^m$, as required since $\Lambda_0=\Lambda$.

We go on to prove (13). The idea of the proof is that once we have established a bound for the average total distance in Λ_{m-k}, we can improve on this bound for the average total distance in Λ_{m-k-1} by updating a random block.

The formal proof proceeds by induction on k. The base case ($k=0$) is clear since $\rho_{\Lambda_m}(Q)\le\max_{\eta,\xi}\rho_{\Lambda_m}(\eta,\xi)\le c$ by definition of c. We assume (13) for k and *show* for $k+1$. We have to show that, for every $t\ge(k+1)t_m$, we have $\rho_{\Lambda_{m-k-1}}(F^t_{B(\Lambda_m)}(Q))\le c\alpha^{k+1}$. Since $\Phi(B(\Lambda_{m-k-1}))=\Lambda_{m-k}$, we can use the induction hypothesis together with Lemma 3.5 to get that for every $t>kt_m$,

$$\rho_{\Lambda_{m-k-1}}(F^t_{B(\Lambda_m)}(Q)) \;\le\; \left(1-\frac{\mathrm{MAX}+\mathrm{MIN}}{w_{B(\Lambda_m)}}\right)\rho_{\Lambda_{m-k-1}}(F^{t-1}_{B(\Lambda_m)}(Q))+\frac{\mathrm{MAX}}{w_{B(\Lambda_m)}}c\alpha^k,$$

where $\mathrm{MAX}=\max_{y\in\Lambda_{m-k}}\left\{I_{\leftarrow y}\right\}$ and $\mathrm{MIN}=\min_{y\in\Lambda_{m-k}}\left\{w_{B(y)}-I_{\leftarrow y}\right\}$. Therefore,

$$\rho_{\Lambda_{m-k-1}}(F^t_{B(\Lambda_m)}(Q))-\frac{\mathrm{MAX}}{\mathrm{MAX}+\mathrm{MIN}}c\alpha^k \;\le$$
$$\left(1-\frac{\mathrm{MAX}+\mathrm{MIN}}{w_{B(\Lambda_m)}}\right)\left[\rho_{\Lambda_{m-k-1}}(F^{t-1}_{B(\Lambda_m)}(Q))-\frac{\mathrm{MAX}}{\mathrm{MAX}+\mathrm{MIN}}c\alpha^k\right].$$

Notice that $\min_{y\in\Lambda_{m-k}}\left\{w_{B(y)}\right\}\le\mathrm{MAX}+\mathrm{MIN}\le\max_{y\in\Lambda_{m-k}}\left\{w_{B(y)}\right\}$. In particular, this means that the factor $(1-\frac{\mathrm{MAX}+\mathrm{MIN}}{w_{B(\Lambda_m)}})\ge 0$. Now, since by the induction hypothesis $\rho_{\Lambda_{m-k-1}}(F^{kt_m}_{B(\Lambda_m)}(Q))\le\rho_{\Lambda_{m-k}}(F^{kt_m}_{B(\Lambda_m)}(Q))\le c\alpha^k$, then for all $t\ge kt_m$,

$$\begin{aligned}
\rho_{\Lambda_{m-k-1}}(F^t_{B(\Lambda_m)}(Q)) &\le \frac{\mathrm{MAX}}{\mathrm{MAX}+\mathrm{MIN}}c\alpha^k+\left(1-\frac{\mathrm{MAX}+\mathrm{MIN}}{w_{B(\Lambda_m)}}\right)^{t-kt_m}c\alpha^k\\
&\le \frac{\mathrm{MAX}}{\mathrm{MAX}+\mathrm{MIN}}c\alpha^k+\left(1-\frac{\min_{y\in\Lambda_{m-k}}\left\{w_{B(y)}\right\}}{w_{B(\Lambda_m)}}\right)^{t-kt_m}c\alpha^k.
\end{aligned}$$

In particular, for all $t \geq (k+1)t_m$,

$$\rho_{\Lambda_{m-k-1}}(F^t_{B(\Lambda_m)}(Q)) \;\leq\; \frac{\mathrm{MAX}}{\mathrm{MAX}+\mathrm{MIN}} c\alpha^k + \delta c\alpha^k \;\leq\; c\alpha^{k+1}.$$

This concludes the proof of (13) and thus completes the proof of Theorem 3.4. □

Remark: Notice that if the diameter of all blocks Θ_i is bounded by some constant r then the region Λ_m as defined in Section 3.1 is included in the ball of radius mr around Λ. Thus, if the conditions in our theorems are satisfied by coupled update rules with blocks of bounded diameter, then not only is the Gibbs measure unique, but in fact the influence of a boundary condition in finite volume distributions on an inner subset Λ is exponentially small in the distance of Λ from the boundary.

4 Extensions

4.1 Extending the model

The conditions in Theorems 2.5 and 2.7 are applicable in more general settings as well. First, the requirement that $\kappa^\eta_{\Theta_i}$ (the result of updating the block Θ_i) depends only on Θ_i and its boundary can be relaxed to dependency on sites within a bounded radius r from Θ_i. The definition of the sequence $\{\Lambda_m\}$ is then adapted to this setting by letting $\Lambda_{m+1} = \beta(\Lambda_m) \cup \partial_r\beta(\Lambda_m)$, where $\partial_r\Lambda$ stands for the set of sites outside Λ that are within distance r from Λ. The rest of the statements and the proofs follow unchanged. Using update rules that depend on sites within distance r is useful when the models have finite range interactions rather than just nearest-neighbor interactions, i.e, potentials are defined for every subset of diameter at most r rather than just single sites and edges, which is the case $r = 1$.

A second observation is that we can omit the restriction that the spin space $\mathcal{S}$ is finite, and instead work with a measurable space $\mathcal{S}$ equiped with a σ-algebra $\mathcal{B}$ of subsets of $\mathcal{S}$. In this case, the metrics ρ_x are required to be measurable functions w.r.t. $\mathcal{B}\times\mathcal{B}$. Up to minor notational and language issues involving infinite spaces, our proofs carry through to this setting except that it may no longer be possible to derive an upper bound on the total variation distance of two distributions when projected onto $\mathcal{S}^\Lambda$ from $\rho_\Lambda(Q)$ as we did in (5). However, the rest of our discussion leading to (5) is still valid. In particular, under our conditions, for arbitrary τ and σ there exists a coupling of $\gamma^\tau_{\Lambda_m}$ and $\gamma^\sigma_{\Lambda_m}$ for which $\rho_\Lambda(Q)$ is exponentially small in m. This means that the Kantorovich-Rubinstein-Ornstein-Vasserstein (KROV) distance (see, e.g., [5] for a definition) between the two distributions w.r.t. ρ_Λ is exponentially small in m and in particular, that the limits of the two sequences of distributions as $m \to \infty$ are the same, i.e., the Gibbs measure is unique.

4.2 Markov chains

As was shown in [2] and [20], conditions like the ones in Theorems 2.5 and 2.7 give upper bounds on the mixing times of the corresponding Markov chains. In this setting the graph G is finite, and thus there is always a unique Gibbs measure μ. Notice that now the collection of blocks $\{\Theta_i\}$ specified by a coupled update rule is finite as well. Let $n = |V|$, and $m = \sum_i w_i \,/\, \min_x w_{B(x)}$. Given a specification γ and a coupled update rule for γ, the corresponding Markov chain is the one whose steps are defined as follows. First, a block Θ_i is chosen at random from the finite set S, where S is the whole collection of blocks and the probability to choose Θ_i is proportional to w_i. Then, the configuration in Θ_i is updated according to κ^τ_i, where τ is the current configuration. By definition, μ is a stationary measure of this Markov chain. If μ is the unique stationary measure we can discuss the mixing time $\tau(\epsilon)$ of the chain, defined as the number of steps required to get within total variation distance ϵ from μ starting from an arbitrary configuration.

When the influence *on* a site is small as in Theorem 2.5, i.e., $\max_x \left\{ I_{x\leftarrow} \,/\, w_{B(x)} \right\} < 1$, then μ is necessarily the unique stationary distribution of the above Markov chain and the mixing time $\tau(\epsilon) \leq \frac{m}{c} \log(\frac{nD_1}{\epsilon})$, where $c = 1 - \max_x \left\{ I_{x\leftarrow} \,/\, w_{B(x)} \right\}$ and $D_1 = \frac{\max_{x\in V, s_1,s_2 \in \mathcal{S}} \rho_x(s_1,s_2)}{\min_{x,s_1\neq s_2} \rho_x(s_1,s_2)}$. The proof of this fact is based on Lemma 3.3 in a similar way to the proof of Theorem 2.5 by giving an upper bound on the distance at any site when running a coupled process. Specifically, for any initial coupling Q (of two arbitrary distributions), let $\delta_t = \max_x F_S^t(Q)$. Then Lemma 3.3 yields $\delta_{t+1} \leq (1-\frac{c}{m})\delta_t$. Since $\delta_0 \leq \max_{x,s_1,s_2} \rho_x(s_1,s_2)$, we conclude that the probability that two coupled configurations on G disagree after t steps is at most $nD_1(1-\frac{c}{m})^t$, and in particular, the mixing time $\tau(\epsilon) \leq \frac{m}{c} \log(\frac{nD_1}{\epsilon})$. Furthermore, the fact that the variation distance goes to zero at least as fast as $(1-\frac{c}{m})^t$ means that the *spectral gap* of the Markov chain is at least $\frac{c}{m}$.[§] We also note that we can get better bounds for the *projected mixing time* (see [8] for precise definition). Specifically, when we require that the probability of disagreement in a subset (rather than the whole graph) be ϵ, we obtain the same bound on the mixing time as above, but with n replaced by the size of the subset and the minimum in the denominator of D_1 taken only over sites x that belong to the subset.

When the influence *of* a site is small in the Markov chain setting, we can use the standard path coupling argument as in [2] to bound the mixing time. Specifically, if $\max_y \left\{ I_{\leftarrow y} \,/\, w_{B(y)} \right\} < 1$ then μ is necessarily the unique stationary distribution of the Markov chain and the mixing time $\tau(\epsilon) \leq \frac{m}{c} \log(\frac{D_2}{\epsilon})$, where $c = 1 - \max_y \left\{ I_{\leftarrow y} \,/\, w_{B(y)} \right\}$ and $D_2 = \frac{\max_{\eta,\xi} \rho_V(\eta,\xi)}{\min_{\eta\neq\xi} \rho_V(\eta,\xi)}$. This bound can also be derived from Lemma 3.5 by using the fact that $V = \Phi(B(V))$ to get that $\rho_V(F_S^{t+1}(Q)) \leq (1-\frac{c}{m})\rho_V(F_S^t(Q))$. Since $\rho_V(Q) \leq \max_{\eta,\xi} \rho_v(\eta,\xi)$, we get that the probability that the two coupled configurations disagree after t steps is at most $D_2(1-\frac{c}{m})^t$, and in particular, the mixing time $\tau(\epsilon) \leq \frac{m}{c} \log(\frac{D_2}{\epsilon})$. Again, the fact that the variation distance goes to zero at least as fast as $(1-\frac{c}{m})^t$ means that the spectral gap is at least $\frac{c}{m}$. Finally, as a side remark, we note that $D_2 \geq n$ since $\max_{\eta,\xi} \rho_V(\eta,\xi) = \sum_x \max_{s_1,s_2} \rho_x(s_1,s_2)$ while $\min_{\eta\neq\xi} \rho_V(\eta,\xi) = \min_{x,s_1,s_2} \rho_x(s_1,s_2)$.

In many situations, the finite graph G is in fact a subset of an infinite graph. For example, take any spin system on the integer lattice $\mathbb{Z}^d$. We can then consider a Markov chain for sampling from γ_Λ^τ , where τ is any boundary condition and Λ is the regular box of side length L centered at 0. It is important to notice that the fact that the condition in Theorem 2.5 (or 2.7) holds for the infinite graph does not necessarily imply that the corresponding condition (with the same parameters) holds when we consider the Markov chain on the finite subset Λ. The reason is that some of the blocks Θ_i may intersect Λ without being included in it. Therefore, these blocks cannot be updated in the Markov chain. But then, if we omit these blocks, the condition need not necessarily continue to hold. For example, it may be the case that the influence on a site x near the boundary of Λ is increased because the blocks that include x but are not contained in Λ ("ignored blocks") contribute less to the distance at x than an average block does. This issue was already addressed in [6] and [20], where stronger conditions were given which require that the influence bounds still hold even if instead of updating an entire block Θ_i we update any subset of it; this allows one to add $\Theta_i \cap \Lambda$ to the blocks used by the Markov chain and retain the original condition on total influence. We skip the details here. These stronger conditions also imply what is often called *strong spatial mixing*, where correlations decay exponentially with the distance from the portion of the boundary where the two boundary configurations differ rather than with the distance from the boundary as a whole. See [6, 20, 14] for more on stronger conditions of this type.

Finally, we mention that the fact that similar conditions imply both the uniqueness of the Gibbs measure and optimal mixing time of the corresponding Markov chain is part of a more general framework of relationships between spatial and temporal mixing properties [20, 14, 3, 13, 8, 1]. For

[§]The spectral gap of a Markov chain with transition matrix P is defined as $1-|\lambda_2|$, where λ_2 is the second eigenvalue of P. It is well known (see, e.g., Equation (1.2.5) in [19]) that for any matrix norm $\|\cdot\|$, $\lim_{t\to\infty} \|P^t - P^\infty\|^{1/t} = |\lambda_2|$, where P^∞ is the matrix whose rows are all equal to the stationary distribution μ of P. Hence, if for every initial configuration of the chain the variation distance from μ goes to zero at least as fast as $(1-\frac{c}{m})^t$ then clearly $|\lambda_2| \leq (1-\frac{c}{m})$.

example, optimal projected mixing as mentioned above implies uniqueness of the Gibbs measure (and an exponential decay with m of the influence of any boundary condition outside Λ_m, as in Theorem 3.2) [8].

5 Applications

In this section we illustrate the use of the conditions given in Theorems 2.5 and 2.7 by carrying out the appropriate calculations for a few specific models in specific ranges of their parameters, thus showing uniqueness of the Gibbs measure for these models in the appropriate ranges. Although we do not extend the previously known range of parameters for which the Gibbs measure is unique, we do extend the range for which finite size conditions of the Dobrushin type hold. In addition, the examples given here shed additional light on our two conditions and the differences between them, and might also serve as guiding examples for readers seeking to establish uniqueness of the Gibbs measure for other models by applying Theorem 2.5 or Theorem 2.7.

The following notation is used in all our examples. Recall that one of the ingredients that needs to be specified in our conditions is a collection of metrics ρ. All the examples we mention in this paper use a collection of metrics of the form $\rho_x = u_x \rho_\delta$, where $u_x \in \mathbb{R}^+$ is a weight associated with site x, and ρ_δ is the metric that assigns 1 to any pair of distinct spins and 0 to a pair of identical spins. In particular, for a coupling Q, $\rho_x(Q)$ is exactly u_x times the probability (under Q) that the two coupled spins at x differ. From here onwards a collection of metrics will be specified by determining the set of weights u_x, and implicitly setting $\rho_x = u_x \rho_\delta$.

5.1 Colorings of biregular bipartite graphs

We start with an example that emphasizes the differences between the two conditions. These differences are better clarified when the matrix of influences is not symmetric, i.e., the influence of site y on site x is not the same as that of x on y. An example of a model where this symmetry is broken is the model of colorings on a tree with alternating branching degrees. In the colorings model (or antiferromagnetic Potts model at zero temperature) the set of spins $\mathcal{S} = \{1, \ldots, q\}$, where we view each of these values as a distinct color. The edge potentials $U_{\{x,y\}}$ assign infinite energy to pairs of the same color and zero energy otherwise, and the single-site potentials U_x are identically zero. Thus, feasible configurations correspond to proper colorings of the sites, and the finite region Gibbs distributions are uniform over proper colorings that agree with the boundary condition. We consider colorings on a tree with alternating branching degrees, i.e., the underlying graph is the infinite rooted tree in which vertices at even distance from the root have b_1 children, vertices at odd distance from the root have b_2 children, and $b_1 \neq b_2$. The parameters of the model are thus q and (b_1, b_2). We apply our conditions to establish that, for $q > b_1 + b_2 + 2$, the Gibbs measure is unique. We give two proofs, one using Theorem 2.5 and the other using Theorem 2.7. We note that uniqueness for colorings on a tree is known to hold for a wider range of parameters. For example, when $b_1 = b_2$ (the tree is regular of degree $b_1 + 1$), the Gibbs measure is unique if and only if $q > b_1 + 1$ [11]. However, the range of parameters for which we show uniqueness here is still larger than that given by the original Dobrushin condition, which is $q > 2(\max\{b_1, b_2\} + 1)$. (The calculation using the original Dobrushin condition was first done in [18], and implies uniqueness for any underlying graph and $q > 2 \max_x \deg(x)$, where $\deg(x)$ stands for the number of neighbors of x).

In order to use our theorems we need to specify a coupled update rule for the model. For this example, we use the simple update rule in which each block is a distinct single site, the weights of the blocks are uniform (e.g., all 1) [¶], and where updates are done according to the heat-bath

[¶] When each block is a single site (or more generally, when each site is included in exactly one block), allowing general

rule. We identify a block Θ_i with the site x it consists of and write κ_x^τ in place of κ_i^τ. Notice that for the colorings model, a heat-bath update means that under κ_x^τ, the color at x is chosen uniformly at random from the set of colors not assigned to neighbors of x under τ. Notice also that for $q > \max\{b_1, b_2\} + 1$ (i.e., when the number of colors is larger than the maximum degree of the graph — as is the case for the range of q we consider), κ_x^τ as above is well defined even if τ is infeasible, as required.

In order to complete the specification of our coupled update rule we have to specify how to couple two updates starting from two configurations that disagree at exactly one site, i.e., we need to specify $K_x(\eta, \xi)$ for pairs η and ξ that differ at exactly one site y. We need only specify this coupling when y is a neighbor of x or $y = x$ because otherwise the coupling is required to be the one in which the spins at x agree with probability 1. We set $K_x(\eta, \xi)$ to be a coupling that minimizes the probability of disagreement between the spins at x. When $y = x$ this simply means the coupling in which the two resulting configurations agree with certainty (since we use a heat-bath update, $\kappa_x^\eta = \kappa_x^\xi$ when η and ξ differ only at x). When y is a neighbor of x, and when the number of colors available for the update of the spin at x under η is the same as under ξ, the optimal coupling is described as follows. Suppose that under both configurations the number of available colors at x is a, and w.l.o.g. that $\eta_y = 1$ and $\xi_y = 2$. Then, $K_x(\eta, \xi)$ assigns probability $1/a$ to the pair of configurations in which x is colored 2 and 1 respectively, and for each of the other $a - 1$ available colors s, $K_x(\eta, \xi)$ assigns probability $1/a$ to the pair of configurations in which both spins at x are colored s. Thus, the probability of disagreement is $1/a$. The coupling $K_x(\eta, \xi)$ takes a similar form when the number of available colors at x under η is not the same as under ξ (this number may differ by one) so that in either case, the probability of disagreement at x is $1/a$, where a is the number of available colors at x under the configuration for which the number of available colors is smaller. In particular, the probability of disagreement is at most $1/(q - \deg(x))$ (with equality for at least one pair η, ξ).

The final ingredient we need to specify is the collection of metrics ρ. Since we use a collection of the form $\rho_x = u_x \rho_\delta$, we only need to specify the weights u_x. Although we use non-uniform weights in order to show uniqueness for the range of parameters mentioned before, it is instructive to first consider the case in which the u_x are uniformly set to 1, which is the setting in the original Dobrushin condition. Under this setting, $I_{x \leftarrow y} = 1/(q - \deg(x))$ since $\rho_x(K_x(\eta, \xi))$ is simply the probability of disagreement at x under the coupling $K_x(\eta, \xi)$. Thus, $\sup_x I_{x\leftarrow} = \sup_x \frac{\deg(x)}{q - \deg(x)} = \max\left\{\frac{b_1+1}{q-b_1-1}, \frac{b_2+1}{q-b_2-1}\right\}$. Recall that for the collection of blocks we use, $w_{B(x)} = 1$ for every x since each site is covered by exactly one block whose weight is 1. Thus, using Theorem 2.5 (or equivalently, the original Dobrushin condition) we get that the Gibbs measure is unique in the range of parameters that satisfy $\max\left\{\frac{b_1+1}{q-b_1-1}, \frac{b_2+1}{q-b_2-1}\right\} < 1$, i.e., for $q > 2(\max\{b_1, b_2\} + 1)$.

We pause to observe that the colorings model with the above choices of update rule and collection of metrics is a good example of the fact that influence *on* and *of* a site may differ. First, since neighboring sites have different degrees, $I_{x \leftarrow y} \neq I_{y \leftarrow x}$. Furthermore, for a site y with b_1 children, the total influence of y is $I_{\leftarrow y} = \frac{b_1+1}{q-b_2-1}$. (This is because y has $b_1 + 1$ neighbors, and the influence of y on each is $1 \,/\, (q - b_2 - 1)$ because the degree of each neighbor is $b_2 + 1$.) Thus, the maximum total influence is $\sup_y I_{\leftarrow y} = \max\left\{\frac{b_1+1}{q-b_2-1}, \frac{b_2+1}{q-b_1-1}\right\}$, which is less than the maximum total influence on a site when $b_1 \neq b_2$. Notice also that we cannot use the above uniform collection of metrics in order to show uniqueness using the condition based on total influence of a site (Theorem 2.7) because this condition requires that the collection of metrics be bounded, and a uniform collection

sets of weights $\{w_i\}$ does not add any generality to our conditions, i.e., for any coupled update rule using a collection of blocks of the above type, the satisfiability of the conditions in Theorems 2.5 and 2.7 is unaffected when changing the set of weights to be uniformly 1. This is because, when each site is covered by exactly one block, the quantity $I_{x\leftarrow} \,/\, w_{B(x)}$ in Theorem 2.5 is independent of the choice of weights, and the condition in Theorem 2.7 is not affected if we absorb the weights w_i into the collection of metrics ρ.

is clearly not bounded.

We now go on to establish uniqueness of the Gibbs measure for $q > b_1 + b_2 + 2$, improving on the range of parameters for which uniqueness is obtained using the original Dobrushin condition. We give two proofs of this fact, using Theorems 2.5 and 2.7 respectively. For each proof we use a different set of weights u_x. We start with the condition based on the total influence on a site (Theorem 2.5). As is apparent from the analysis of the setting in which u_x is uniform, there is room for improvement since the total influence on sites of the larger degree is larger than on those of the smaller degree. We give the two types of sites different weights in order to balance the total influence they get. Let $u_x = \sqrt{\frac{q-\deg(x)}{\deg(x)}}$. This yields $I_{x\leftarrow y} = \frac{u_x}{u_y(q-\deg(x))} = \sqrt{\frac{\deg(y)}{(q-\deg(y))(q-\deg(x))\deg(x)}}$, and therefore, $I_{x\leftarrow} = \sqrt{\frac{(b_1+1)(b_2+1)}{(q-b_1-1)(q-b_2-1)}}$ for every x. Thus, using Theorem 2.5, the Gibbs measure is unique in the range of parameters for which the last expression is < 1, i.e., for $q > b_1 + b_2 + 2$.

We now give the second proof of uniqueness for the above range of parameters, this time using the condition based on the total influence of a site (Theorem 2.7). In order to use this condition, we have to set the weights u_x so that they yield a summable collection of metrics, i.e., $\sum_x u_x$ has to be finite. In addition, we optimize the weights to minimize the maximum total influence of a site, i.e., we balance the total influence of different sites. In the resulting choice of weights, u_x depends on (and is determined by) the distance of x from the root of the tree. We thus write u_ℓ for the weight of a site at distance ℓ from the root. Set $u_{2\ell} = [(1+\epsilon)b_1b_2]^{-\ell}\sqrt{\frac{q-b_1-1}{b_1+1}}$ and $u_{2\ell+1} = \frac{1}{b_1}[(1+\epsilon)b_1b_2]^{-\ell}\sqrt{\frac{q-b_2-1}{b_2+1}}$, where $\epsilon > 0$ is a small enough constant to be determined later. Clearly, $\sum_x u_x$ is finite because the total weight at level ℓ is proportional to $(1+\epsilon)^{-\lfloor \ell/2 \rfloor}$. We go on to calculate the influence of a site under this choice of weights. Consider a site y at distance 2ℓ from the root. This site influences its parent as well as its b_1 children. The probability of disagreement under the relevant coupling is $1/(q-b_2-1)$ for both the parent and the children, but observe that the weight of the parent differs from that of the children. Specifically,

$$\begin{aligned} I_{\leftarrow y} &= \frac{u_{2\ell-1}}{u_{2\ell}(q-b_2-1)} + b_1\frac{u_{2\ell+1}}{u_{2\ell}(q-b_2-1)} \\ &= \frac{(1+\epsilon)b_2+1}{q-b_2-1}\sqrt{\frac{(b_1+1)(q-b_2-1)}{(q-b_1-1)(b_2+1)}} \le (1+\epsilon)\sqrt{\frac{(b_1+1)(b_2+1)}{(q-b_1-1)(q-b_2-1)}} \ . \end{aligned}$$

For y at distance $2\ell+1$ from the root, a similar calculation gives the slightly better bound $I_{\leftarrow y} \le \sqrt{\frac{(b_1+1)(b_2+1)}{(q-b_1-1)(q-b_2-1)}}$. Applying Theorem 2.7, we conclude that the Gibbs measure is unique in the range of parameters for which there exists $\epsilon > 0$ such that $(1+\epsilon)\sqrt{\frac{(b_1+1)(b_2+1)}{(q-b_1-1)(q-b_2-1)}} < 1$, i.e., for $q > b_1 + b_2 + 2$.

We conclude this subsection by observing that the result obtained here (uniqueness for $q > b_1 + b_2 + 2$) holds for any (b_1+1, b_2+1)-biregular bipartite graph, i.e., for any bipartite graph on vertex set $V_1 \cup V_2$ in which vertices in V_1, V_2 have degrees b_1, b_2 respectively. Indeed, if we examine the first of the two proofs we gave for the tree (the one using total influence *on* a site), we see that the only structure of the graph that we used is that for any site, either its degree is b_1+1 and all its neighbors are of degree b_2+1, or its degree is b_2+1 and all its neighbors are of degree b_1+1. This property holds for any bipartite graph of the above type. We are unaware of any literature discussing colorings of biregular bipartite graphs, and we believe the above bound to be the best known for general graphs of this type (specifically, for those that are not trees). The only previously known bound available for graphs of this type is the one obtained from the original Dobrushin condition (which holds for any graph). As mentioned before, this bound is $q > 2\sup_x \deg(x) = 2(\max\{b_1, b_2\}+1)$, and our bound improves on this for $b_1 \neq b_2$.

5.2 Ising model on a regular tree

The next model we discuss is the Ising model (as defined in Example 1) on the infinite b-ary tree (in which every vertex except the root has degree $b+1$). Recall that in this model the parameters are the inverse temperature β, the external field h, and the branching degree of the tree b. In this model, the range of parameters for which the Gibbs measure is unique is known exactly [17, 9]. Specifically, there exists a critical temperature $\beta_0(b) = \frac{1}{2}\ln(\frac{b+1}{b-1})$ such that for $\beta \leq \beta_0$ the Gibbs measure is unique for all external fields h. For $\beta > \beta_0$, there exists a known critical value $h_c(\beta, b) > 0$ such that the Gibbs measure is unique if $|h| > h_c$, and there are multiple Gibbs measures if $|h| \leq h_c$. We will show that both our conditions hold throughout the supercritical regime (i.e., for $\beta > \beta_0$ and arbitrary h, or $\beta \leq \beta_0$ and $|h| > h_c$), an evidence of the tightness of our conditions. We compare this with the range of parameters for which the original Dobrushin condition holds, which is $\beta > \frac{1}{2}\ln(\frac{b+2}{b})$ and arbitrary h. The update rule we use in our proofs is a heat-bath update of a finite subtree. This illustrates another new feature of our conditions, i.e., allowing updates of finite subsets on any graph. As part of our discussion, we also show that the respective restrictions on the collections of metrics in Theorems 2.5 and 2.7 are both necessary. This is done by giving a coupled update rule and collections of metrics that satisfy the conditions in these theorems except for the respective restrictions on the collection of metrics, for some $\beta > \beta_c$ and $h = 0$, i.e., in the regime where there are multiple Gibbs measures.

5.2.1 Single-site updates

We start by considering the heat-bath update rule on single sites and the uniform collection of metrics (where $u_x = 1$ for all x), which is the setting in the original Dobrushin condition. Recall that a heat-bath update simply means that $\kappa_x^\tau = \gamma_x^\tau$. For η and ξ that differ at a single site y, we set $K_x(\eta, \xi)$ as the optimal coupling of γ_x^η and γ_x^ξ, i.e., $\rho_x(K_x(\eta,\xi)) = \|\gamma_x^\eta - \gamma_x^\xi\|_x$. Notice that if $y = x$ then $K_x(\eta, \xi) = 0$ because then γ_x^η and γ_x^ξ are the same. If y is a neighbor of x, it is well known (e.g., [1, 15]) that $\|\gamma_x^\eta - \gamma_x^\xi\|_x \leq \frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}}$ (with equality if the spins of the neighbors of x other than y are divided equally between pluses and minuses). Thus, $I_{x \leftarrow y} \leq \frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}}$ (with equality if the degree of x is odd), and the total influence on any site x is $I_{x\leftarrow} \leq (b+1)\frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}}$. Using Theorem 2.5 (or, equivalently, the original Dobrushin condition), this immediately establishes uniqueness of the Gibbs measure for β such that $(b+1)\frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}} < 1$, i.e., for $\beta < \frac{1}{2}\ln(\frac{b+2}{b})$, and arbitrary h. For the same range of parameters, it is also easy to see that the dual condition in Theorem 2.7 holds for the same coupled update rule, but setting $u_x = [(1+\epsilon)b]^{-|x|}$, where $|x|$ stands for the distance of x from the root of the tree and ϵ is a small enough constant (this is needed in order for the collection of metrics to be summable).

We now use the simple coupled update rule described above in order to show that the restrictions imposed on the collection of metrics used in Theorems 2.5 and 2.7 respectively are necessary. We start with Theorem 2.5. Consider the collection of metrics resulting from setting $u_x = (\sqrt{b})^{|x|}$. This is clearly not a bounded collection because u_x grows to infinity with the distance of x from the root of the tree. Since now the weight of a site is $\sqrt{b}$ times the weight of its parent, it is easy to see that for every x, $I_{x\leftarrow} \leq [b \cdot \frac{1}{\sqrt{b}} + 1 \cdot \sqrt{b}]\frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}} = 2\sqrt{b}\frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}}$. Thus, for this choice of weights, $\sup_x I_{x\leftarrow} < 1$ for $\beta < \frac{1}{2}\ln(\frac{2\sqrt{b}+1}{2\sqrt{b}-1})$. However, since $\frac{1}{2}\ln(\frac{2\sqrt{b}+1}{2\sqrt{b}-1}) > \frac{1}{2}\ln(\frac{b+1}{b-1})$ for $b > 4$, for this choice of weights the range of parameters for which the condition holds includes values for which the Gibbs measure is not unique. We thus conclude that the requirement that the collection of metrics be bounded is necessary for Theorem 2.5 to hold.

We go on to consider Theorem 2.7, and show that the requirement that the collection of metrics be summable is necessary for this Theorem to hold. Consider the same coupled update rule as above, but set the weights $u_x = (\sqrt{b})^{-|x|}$. Although this collection is bounded, it is not summable

because the total weight of sites at distance ℓ from the root is $(\sqrt{b})^\ell$, which goes to infinity with ℓ. A calculation similar to that in the previous paragraph gives that $\sup_y I_{\leftarrow y} < 2\sqrt{b}\frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}}$. Hence, as before, for this choice of weights $\sup_y I_{\leftarrow y} < 1$ for some values of β for which the Gibbs measure is not unique. We thus conclude that the requirement that the collection of metrics is summable is necessary in Theorem 2.7.

A slight modification of the last example shows that the requirement that $\sup_y w_{B(y)}$ is finite is also necessary in Theorem 2.7. Recall that so far, we only used collections of blocks in which the weights were uniformly set to 1. In fact, in the rest of the applications in this paper we continue to only use collections of this type, except here, where we wish to demonstrate the necessity of the restriction that $\sup_y w_{B(y)}$ is finite. Thus, consider the coupled update rule from the last two paragraphs, with metric weights $u_x = [(1+\epsilon)b]^{-|x|}$ (so the collection of metrics is summable), except that now the weight of a block $w_x = [(1+\epsilon)\sqrt{b}]^{|x|}$. Since each site y is included in the unique block $\Theta_y = \{y\}$, $B(y)$ consists of only this block and $w_{B(y)} = w_y$. Thus, the above choice of block weights violates the requirement that $\sup_y w_{B(y)}$ is finite because w_y grows to infinity with $|y|$. In addition, it is easy to see that the quantity $\sup_y \left\{ I_{\leftarrow y} \,/\, w_{B(y)} \right\}$ remains exactly the same as in the example of the previous paragraph because the product $w_x u_x$ is unchanged for all x, and since the coupled update rule is the same (up to the change of weights). We thus conclude as in the previous paragraph that the requirement that $\sup_y w_{B(y)}$ is finite is necessary in Theorem 2.7.

5.2.2 Sharp uniqueness bounds using larger blocks

We now go on to show that both our conditions hold throughout the supercritical regime of parameters by considering updates of finite sub-trees.

Theorem 5.1 *In the Ising model on the regular b-ary tree, for the following regimes of parameters there exist a coupled update rule and a collection of metrics that satisfy the condition in Theorem 2.5:*

(i) *$\beta < \beta_0$ and arbitrary h;*

(ii) *$\beta \geq \beta_0$ and $|h| > h_c(\beta)$.*

Furthermore, for the same regimes of parameters, there exists a collection of metrics that together with the above coupled update rule satisfies the condition in Theorem 2.7.

Proof: We will only give the proof for regime (i). The proof for regime (ii) goes by a similar but slightly more involved argument, which we sketch at the end. The coupled update rule we use is based on a heat-bath update of a finite size complete subtree. Thus, the collection of blocks is constructed as follows. For every site z, let Θ_z be the complete subtree of height $\ell - 1$ rooted at z, where ℓ is a (large enough) constant to be determined later. Notice that Θ_z consists of ℓ levels (including the level of z itself). The collection of blocks includes Θ_z for every z, plus the $\ell - 1$ blocks which are the complete subtrees of height $0, 1, \ldots, \ell - 2$ respectively, rooted at the root of the original infinite tree (for convenience, we think of these extra blocks as subtrees rooted at imaginary ancestors of the root of the original tree). The addition of the extra blocks guarantees that every site is covered by exactly ℓ blocks. As usual, the weight of every block is set to 1.

As before, we write κ_z^η for the distribution resulting from an update of Θ_z. Since we use a heat-bath update, $\kappa_z^\eta = \gamma_{\Theta_z}^\eta$. We need to specify the coupling $K_z(\eta, \xi)$, where η and ξ differ at exactly one site $y \in \Theta_z \cup \partial\Theta_z$. If $y \in \Theta_z$ then $K_z(\eta, \xi)$ is defined as the coupling in which the two configurations agree with certainty. For $y \in \partial\Theta_z$ we use the optimal coupling as constructed in [1] and later in [15]. In particular, this coupling is constructed recursively along paths of the tree, such that for every $x \in \Theta_z$, the probability of disagreement at x under $K_z(\eta, \xi)$ is $\leq (\frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}})^{|x-y|}$, where $|x - y|$ stands for the graph distance between x and y.

Consider now the uniform collection of metrics where $u_x = 1$ for every x. We note that for this choice of weights, if $x \in \Theta_z$ then $\rho_x(K_z(\eta,\xi)) \leq (\frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}})^{|x-y|}$, for η and ξ that disagree only at $y \in \partial\Theta_z$. We go on to calculate the influence a site y has on site x. First, observe that if $|y-x| > \ell$ then $I_{x \leftarrow y} = 0$. If $|y-x| \leq \ell$ (and $y \neq x$) then there is exactly one block Θ_z through which y influences x. This is because y is on the boundary of exactly $b+1$ blocks, namely the b blocks rooted at each of the children of y, and the block rooted at the ancestor ℓ levels above y. Thus, if x is an ancestor of y, then y influences x through Θ_z if and only if z is the ancestor ℓ levels above y. If x is a descendent of y, then y influences x through Θ_z if and only if z is the child of y that is the ancestor of x (and $z = x$ if x is an immediate child of y). We conclude that $I_{x \leftarrow y} \leq (\frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}})^{|x-y|}$. Thus, for every x,

$$I_{x\leftarrow} \;=\; \sum_y I_{x \leftarrow y} \;\leq\; \frac{b+1}{b} \sum_{j=1}^{\ell} \left(b \frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}} \right)^j \;.$$

Now, since each site is included in exactly ℓ blocks then $w_{B(x)} = \ell$ for every x. Thus, the above coupled update rule satisfies the condition in Theorem 2.5 if $I_{x\leftarrow} < \ell$ for every x. However, for $\beta < \frac{1}{2}\ln(\frac{b+1}{b-1})$, $\frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}} < \frac{1}{b}$. Thus, for this range of β, $I_{x\leftarrow}$ is bounded by a constant independent of ℓ, and hence $I_{x\leftarrow} < \ell$ for a large enough ℓ (depending on β), as required.

We go on to show that for the same range of parameters of the Ising model, there exists a collection of metrics which together with the above coupled update rule satisfies the condition in Theorem 2.7. Here we need to have a summable collection of metrics, and for this purpose we set $u_x = [(1+\epsilon)b]^{-|x|}$, where $\epsilon > 0$ is a small enough constant (which may depend on (β, b) but not on ℓ) to be set later. Let us calculate $I_{\leftarrow y}$ for this collection of metrics. As before, it is enough to show that $I_{\leftarrow y}$ is bounded by a constant independent of ℓ. First, notice that for the above choice of weights, the total weight of the sites below y is at most a constant times u_y, so the contribution to $I_{\leftarrow y}$ of the blocks immediately below y is bounded by a constant even if the spins of all sites included in these blocks disagree with certainty. We still need to consider the block above y. Let z be the ancestor ℓ levels above y. We need to show that $\rho_{\Theta_z}(K_z(\eta,\xi))$ is bounded by u_y times a constant independent of ℓ, for every η and ξ that differ only at y. Since the coupling we use was constructed recursively (see [1, 15]), a disagreement at a site $x \in \Theta_z$ can occur only if all the sites on the path from y to x have disagreeing spins. Combining this with the fact that for every site x, the total weight of sites below x is at most a constant times u_x, we see that $\rho_{\Theta_z}(K_z(\eta,\xi))$ is bounded by a constant times the average total distance along the path from y to z. It is therefore enough to show that the average total distance along this path is at most a constant (independent of ℓ) times u_y. Notice that the weight of an ancestor j levels above y is $u_y[(1+\epsilon)b]^j$, but that the probability of disagreement at that site under $K_z(\eta,\xi)$ is at most $(\frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}})^j$. Thus, the above distance along the path is at most $u_y \sum_{j=1}^{\ell} [(1+\epsilon) b \frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}}]^j$, which is bounded by a constant times u_y if $\beta < \frac{1}{2}\ln(\frac{b+1}{b-1})$ and ϵ is small enough (such that $(1+\epsilon)\frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}} < \frac{1}{b}$), as required.

We conclude with a few comments about the proof for regime (ii). Notice that in the proof for regime (i) (for both conditions) the crucial properties we used were that for η and ξ that differ only at y, the probability of disagreement at $x \in \Theta_z$ under $K_z(\eta,\xi))$ is at most $(\frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}})^{|x-y|}$, and that $\frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}} < \frac{1}{b}$ for β as in regime (i). In regime (ii), the latter bound no longer holds. However, using methods that are outside the scope of this paper, it is possible to show that for supercritical values of the parameters as in regime (ii), there exist constants c and $\gamma < \frac{1}{b}$ such that the above probability of disagreement is at most $c\gamma^{|x-y|}$. From there onwards the proof for both conditions proceeds as for regime (i). For reading on the methods used for bounding the above probability by $c\gamma^{|x-y|}$, we refer to [16], where these methods were used to show that in the hard-core model, $\gamma \leq \frac{1}{b}$ throughout the uniqueness regime of that model. $\square$

Remark: The fact that our conditions hold throughout the uniqueness regime (except at the critical point) is not specific to the Ising model and holds for a number of other models on the regular b-ary tree. In particular, the heat-bath update of complete subtrees of height $\ell-1$ for large enough ℓ satisfies our conditions throughout the uniqueness regime of a number of other models on a regular tree. These include the hard-core model, the colorings model and the ferromagnetic Potts model. See [16] for the relevant calculations.

5.3 Independent sets of graphs of subexponential growth

In this subsection we discuss an update rule for the hard-core (independent sets) model (defined in Example 2) that was presented and analyzed in [7] in the context of Markov chains, where it was shown to mix in time $O(n \log n)$ for $\lambda < \frac{2}{\Delta-2}$, where Δ is the maximum degree of the underlying graph. We put this analysis in the context of our conditions, showing that they are satisfied by the above update rule for the same range of parameters if the underlying graph is of *subexponential growth*. The fact that the Gibbs measure is unique for this range of parameters on graphs of subexponential growth is not new because there is an independent argument (see, e.g., [8]) that states that if a model on a graph of subexponential growth admits a Markov chain that uses bounded diameter blocks (as does the one in [7]) and mixes in $O(n \log n)$ time, then the Gibbs measure is unique. Our motivation for discussing this update rule is twofold. First, it is an example of an interesting update rule that is more sophisticated than heat-bath. Second, it illustrates how an analysis that was carried out in the context of Markov chains in order to establish $O(n \log n)$ mixing time can also be used in order to show that our conditions hold w.r.t. the same update rule and choice of parameters (but only for graphs of subexponential growth).

Before going on to our analysis, we mention some other known bounds for the hard-core model. For general graphs, the best known bound is that achieved by the original Dobrushin condition, which establishes uniqueness for $\lambda < \frac{1}{\Delta-1}$. For the special case in which the underlying graph is the square integer lattice $\mathbb{Z}^2$ (which is, of course, of subexponential growth), the best known bound [12] is that the Gibbs measure is unique for $\lambda < 1.185$. The proof in [12] is computer-assisted and uses the Dobrushin-Shlosman condition, i.e., a special case of Theorem 2.7 above, where updates are heat-bath of $k \times k$ squares for some k. When the underlying graph is a regular tree (obviously not of subexponential growth), the uniqueness regime is completely known. Specifically, on a regular tree, the Gibbs measure is unique if and only if $\lambda \leq \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}}$.

We start our analysis with a general discussion of how to convert an analysis of the type carried out in [7] to our setting. The analysis in [7] is based on path coupling, where a coupling of an update is given for every pair of current configurations that differ at exactly one site, and it is shown that the average Hamming distance between the two resulting configurations is strictly less than 1, i.e., the distance decreases. Translating to our notation, using Hamming distance is equivalent to setting $u_x = 1$ for every x, and the fact that the distance decreases in every step is equivalent to $I_{\leftarrow y} - w_{B(y)} < 0$ for every site y, or equivalently, $I_{\leftarrow y} \,/\, w_{B(y)} < 1$. Thus, this coupled update rule satisfies the condition in Theorem 2.7, except that the (uniform) collection of metrics is not summable. However, as we already explained in the remark following Theorem 2.7, if the update rule uses blocks of bounded diameter and if the underlying graph is of sub-exponential growth, then the uniform metric can be modified to be summable while still maintaining $I_{\leftarrow y} \,/\, w_{B(y)} < 1$ for every site y. (Recall that a graph is said to be of subexponential growth if the volume of balls in the graph grows subexponentially with their radius, or equivalently, if there exists a vertex x_0 such that for every $\epsilon > 0$, $\sum_y (1+\epsilon)^{-|x_0-y|}$ is finite, where $|x_0 - y|$ stands for the graph distance between x_0 and y.)

From here onwards we repeat the description and analysis w.r.t. the uniform collection of metrics of the update rule given in [7], but we do it using our terminology. Recall that in the hard-core model a configuration specifies a subset of occupied sites. It is therefore useful to write $\eta \cup \{x\}$ for the configuration in which the subset of occupied sites is as in η, except that x is also occupied.

Similarly, $\eta \setminus \{x\}$ stands for the configuration in which x is not occupied. The collection of blocks we use for the update rule is as follows: there is a block Θ_z for every site z, and Θ_z consists of z and all neighbors of z. As usual, the weight of each block is set to 1. As before, since blocks are indexed by sites, we write κ_z^τ for the distribution of an update of Θ_z in current configuration τ. The result of an update of Θ_z depends on the current configuration on the neighbors of z. Specifically, κ_z^τ is defined as the distribution resulting from the following update of Θ_z:

- If all the neighbors of z are unoccupied under τ: with probability $\frac{\lambda}{1+\lambda}$ the resulting configuration is $\tau \cup \{z\}$, and with probability $\frac{1}{1+\lambda}$ the resulting configuration is $\tau \setminus \{z\}$.
- If at least two of the neighbors of z are occupied under τ: the resulting configuration is deterministically set to $\tau \setminus \{z\}$.
- If exactly one neighbor of z is occupied under τ, say this is x: with probability $\frac{\lambda}{4(1+\lambda)}$ the resulting configuration is $(\tau \setminus \{x\}) \cup \{z\}$, and with probability $1 - \frac{\lambda}{4(1+\lambda)}$, the resulting configuration is $\tau \setminus \{z\}$.

Notice that κ_z^τ is defined for all τ, not just feasible ones. (In [7], the update was defined only for feasible current configurations τ[‖].) It is easy to verify that $\gamma^\tau_{\Theta_z}$ is stationary w.r.t. κ_z for every feasible τ (since κ_z is reversible w.r.t. $\gamma^\tau_{\Theta_z}$ for every feasible τ).

From the definition above it is easy to see that κ_z^τ depends neither on the spin of z itself nor on the configuration on $\partial\Theta_z$, i.e., it depends only on the configuration of the neighbors of z. With that in mind, we go on to define the coupling $K_z(\eta, \xi)$ for pairs (η, ξ) the differ only at y, where $y \in \Theta_z \cup \partial\Theta_z$. Since κ_z^τ does not depend on τ_z or on $\tau_{\partial\Theta_z}$, in case $y = z$ or $y \in \partial\Theta_z$ we define $K_z(\eta, \xi)$ as the coupling in which the two configurations agree on Θ_z with probability 1. When $y \in \Theta_z$ and $y \neq z$, i.e., y is a neighbor of z, $K_z(\eta, \xi)$ is defined as follows. Recall that η and ξ agree on all sites other than y and assume w.l.o.g. that y is occupied under η and unoccupied under ξ. $K_z(\eta, \xi)$ is then defined according to the number of neighbors of z other than y that are occupied:

- If two or more of the neighbors of z other than y are occupied then both κ_z^η and κ_z^η are deterministic, so there is a unique coupling of these two distributions. This is the coupling in which, with probability one, the resulting pair of configurations is $(\eta \setminus \{z\}, \xi \setminus \{z\})$. Notice that in this case $\rho_{\Theta_z}(K_z(\eta, \xi)) = 1$.
- If exactly one neighbor other than y is occupied (say, x is the occupied neighbor) then κ_z^η is still deterministic so there is a unique coupling of κ_z^η and κ_z^ξ, the one in which with probability $1 - \frac{\lambda}{4(1+\lambda)}$ the resulting pair of configurations is $(\eta \setminus \{z\}, \xi \setminus \{z\})$, and with probability $\frac{\lambda}{4(1+\lambda)}$ the resulting pair is $(\eta \setminus \{z\}, (\xi \setminus \{x\}) \cup \{z\})$. Notice that in this case $\rho_{\Theta_z}(K_z(\eta, \xi)) = 1 - \frac{\lambda}{4(1+\lambda)} + 3\frac{\lambda}{4(1+\lambda)} = 1 + \frac{\lambda}{2(1+\lambda)}$.
- If all neighbors of z other than y are unoccupied then $K_z(\eta, \xi)$ is the following coupling. With probability $\frac{1}{1+\lambda}$ the resulting pair is $(\eta \setminus \{z\}, \xi \setminus \{z\})$, with probability $\frac{3\lambda}{4(1+\lambda)}$ the resulting pair is $(\eta \setminus \{z\}, \xi \cup \{z\})$, and with probability $\frac{\lambda}{4(1+\lambda)}$ the resulting pair is $((\eta \setminus \{y\}) \cup \{z\}, \xi \cup \{z\})$. Notice that $(\eta \setminus \{y\}) \cup \{z\} = \xi \cup \{z\}$ and hence $\rho_{\Theta_z}(K_z(\eta, \xi)) = \frac{1}{1+\lambda} + 2\frac{3\lambda}{4(1+\lambda)} = 1 + \frac{\lambda}{2(1+\lambda)}$ in this case.

We conclude that $\rho_{\Theta_z}(K_z(\eta, \xi)) \leq 1 + \frac{\lambda}{2(1+\lambda)}$ if y is a neighbor of z, and $\rho_{\Theta_z}(K_z(\eta, \xi)) = 0$ otherwise. In particular, $I_{\leftarrow y} = \deg(y)(1 + \frac{\lambda}{2(1+\lambda)})$.

[‖] Strictly speaking, for the hard-core model, it is possible to slightly modify our construction of the path coupling in Section 3.2 so that it would be enough to define the update rule (and the couplings K_z) only for feasible configurations. Nevertheless, we define the update rule for any current configuration τ so that we can use the general form of our theorems.

Now, since y is included in $\deg(y)+1$ blocks (which are Θ_y, and Θ_z for every neighbor z of y), $w_{B(y)} = \deg(y)+1$. Thus, $I_{\leftarrow y} \,/\, w_{B(y)} = \deg(y)(1+\frac{\lambda}{2(1+\lambda)})/(\deg(y)+1)$, and $\sup_y I_{\leftarrow y} \,/\, w_{B(y)} < 1$ if $\frac{\deg(y)\lambda}{2(1+\lambda)} < 1$ for every y, i.e., if $\lambda < \frac{2}{\Delta-2}$. Hence, using Theorem 2.7 and the explanation at the beginning of this subsection, the hard-core model with activity parameter λ on graphs of subexponential growth of maximum degree Δ admits a unique Gibbs measure for $\lambda < \frac{2}{\Delta-2}$.

Remark: Notice that for the update rule described in this subsection, $I_{\leftarrow y} < \sum_x I_{x \leftarrow y}$. In other words, it was crucial for our analysis that in the definition of $I_{\leftarrow y}$, the quantification over configurations is taken only once before summing over x, rather than quantifying separately for each x. To see this, recall our analysis of the distance $\rho_{\Theta_z}(K_z(\eta,\xi))$, and notice that for every pair of configurations (η,ξ), there can be at most one site x other than z and y for which $\rho_x(K_z(\eta,\xi)) > 0$ (specifically, this can only happen if x is the unique neighbor of z other than y that is occupied). However, when calculating $\sum_x I_{x \leftarrow y}$ we need to consider the worst pair of configurations for each x separately, and hence the coupling K_z contributes to the distance at all neighbors x of z. The last observation follows from the fact that for each x, we have to consider the pair of configurations in which x is the unique occupied neighbor of z other than y. When one considers the total influence on a site w.r.t. the above update rule, a similar issue arises. In particular, $I_{y\leftarrow} = \sum_x I_{y \leftarrow x} > I_{\leftarrow y}$, so we cannot use the same update rule in order to establish uniqueness of the Gibbs measure using Theorem 2.5 (which would apply to any underlying graph) for the same range of parameters. In fact, w.r.t. the above update rule, the condition based on the total influence on a site holds for an even smaller range of parameter values then the single-site Dobrushin condition.

References

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

[2] R. BUBLEY and M.E. DYER, "Path coupling: a technique for proving rapid mixing in markov chains," *Proc. of the 38th IEEE Symp. on Foundations of Computer Science*, (1997), pp 223–231.

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

[4] R.L. DOBRUSHIN, "Prescribing a system of random variables by the help of conditional distributions," *Theory Prob. and its Appl.* **15** (1970), pp. 469–497.

[5] R.L. DOBRUSHIN and S.B. SHLOSMAN, "Constructive criterion for the uniqueness of a Gibbs field," in: J. FRITZ, A. JAFFE, D. SZASZ, *Statistical mechanics and dynamical systems*, Birkhauser, Boston (1985), pp. 347–370.

[6] R.L. DOBRUSHIN and S.B. SHLOSMAN, "Completely Analytical Gibbs Fields," in: J. FRITZ, A. JAFFE, D. SZASZ, *Statistical mechanics and dynamical systems*, Birkhauser, Boston (1985), pp. 371–403.

[7] M. DYER and C. GREENHILL, "On Markov chains for independent sets," *J. Algorithms* **35** (2000), pp. 17–49.

[8] M. DYER, A. SINCLAIR, E. VIGODA and D. WEITZ, "Mixing in time and space for lattice spin systems: A combinatorial view," *Random Structures and Algorithms* **24** (2004), pp. 461–479.

[9] H.-O. GEORGII, *Gibbs Measures and Phase Transitions*, de Gruyter, Berlin (1988).

[10] H.-O. GEORGII, O.HÄGGSTRÖM and C. MAES, "The random geometry of equilibrium phases," *Phase Transitions and Critical Phenomena*, Volume 18 (C. Domb and J.L. Lebowitz, eds), Academic Press, London (2001), pp. 1–142.

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

[12] A.B. KIRILLOV, D.C. RADULESCU and D.F. STYER, "Vassertein distances in two-state systems," *J. Stat. Phys.* **56** (1989), pp. 931–937.

[13] F. MARTINELLI, "Lectures on Glauber dynamics for discrete spin models," *Lectures on probability theory and statistics (Saint-Flour, 1997)*, Lecture notes in Math. **1717**, Springer, Berlin (1998), pp. 93–191.

[14] F. MARTINELLI and E. OLIVIERI, "Approach to equilibrium of Glauber dynamics in the one phase region I: The attractive case," *Comm. Math. Phys.* **161** (1994), pp. 447–486.

[15] F. MARTINELLI, A. SINCLAIR and D. WEITZ, "Glauber dynamics on trees: Boundary conditions and mixing time," *Comm. Math. Phys.* **250** (2004), pp. 301–334.

[16] F. MARTINELLI, A. SINCLAIR and D. WEITZ, "Fast mixing for independent sets, colorings and other models on trees," *Proc. 15th ACM-SIAM Symp. on Discrete Algorithms*, 2004, pp. 456–465.

[17] C.J. PRESTON, *Gibbs States on Countable Sets*, Cambridge Univ. Press, (1974).

[18] J. SALAS and A.D. SOKAL, "Absence of phase transition for antiferromagnetic Potts models via the Dobrushin uniqueness theorem," *J. Stat. Phys.* **86** (1997), pp. 551-579.

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

[20] D.W. STROOCK and B. ZEGARLINSKI, "The Logarithmic Sobolev Inequality for Discrete Spin Systems on a Lattice," *Comm. Math. Phys.* **149** (1992), pp. 175–194.

[21] E. VIGODA, "Improved bounds for sampling colorings," *Proc. 40th IEEE Symp. on Foundations of Computer Science* (1999), pp. 51–59.