The Annals of Probability
2017, Vol. 45, No. 6A, 3664–3696
DOI: 10.1214/16-AOP1146

UNIVERSALITY OF CUTOFF FOR THE ISING MODEL

BY EYAL LUBETZKY AND ALLAN SLY

New York University and University of California, Berkeley

On any locally-finite geometry, the stochastic Ising model is known to be contractive when the inverse-temperature β is small enough, via classical results of Dobrushin and of Holley in the 1970s. By a general principle proposed by Peres, the dynamics is then expected to exhibit cutoff. However, so far cutoff for the Ising model has been confirmed mainly for lattices, heavily relying on amenability and log Sobolev inequalities. Without these, cutoff was unknown at any fixed $\beta > 0$, no matter how small, even in basic examples such as the Ising model on a binary tree or a random regular graph.

We use the new framework of information percolation to show that, in any geometry, there is cutoff for the Ising model at high enough temperatures. Precisely, on any sequence of graphs with maximum degree d, the Ising model has cutoff provided that $\beta < \kappa/d$ for some absolute constant κ (a result which, up to the value of κ, is best possible). Moreover, the cutoff location is established as the time at which the sum of squared magnetizations drops to 1, and the cutoff window is $O(1)$, just as when $\beta = 0$.

Finally, the mixing time from almost every initial state is not more than a factor of $1 + \varepsilon_\beta$ faster then the worst one (with $\varepsilon_\beta \to 0$ as $\beta \to 0$), whereas the uniform starting state is at least $2 - \varepsilon_\beta$ times faster.

1. Introduction. Classical results going back to Dobrushin [12] and to Holley [15] in the early 1970s and continuing with the works of Dobrushin and Shlosman [13] and of Aizenman and Holley [1] show that, if G is any graph on n vertices with maximum degree d, the Glauber dynamics for the Ising model on G exhibits a rapid convergence to equilibrium in total-variation distance at high enough temperatures. Namely, if the inverse-temperature β is at most c_0/d for some absolute $c_0 > 0$, then the continuous-time dynamics is contractive, whence coupling techniques show that the total-variation mixing time is $O(\log n)$.

A known consequence of contraction is that the spectral gap of the dynamics is bounded away from 0, and so, by a general principle proposed by Peres in 2004 (addressing whether or not the product of the spectral gap and mixing time diverges with n), one expects the *cutoff phenomenon*[1] to occur. (For more on the cutoff phenomenon, discovered in the early 80s by Aldous and Diaconis, see [3, 7].) Concretely, Peres conjectured ([17], Conjecture 1, [18], Section 23.2) cutoff for

Received June 2015; revised July 2016.
MSC2010 subject classifications. 0J27, 82C20, 60K35, 60B10.
Key words and phrases. Cutoff phenomenon, Ising model, mixing times of Markov chains.

[1] Sharp transition in the L^1-distance of a finite Markov chain from equilibrium, dropping quickly from near 1 to near 0.

the Ising model on any sequence of transitive graphs when the mixing time is $O(\log n)$, and in particular in the range $\beta < c_0/d$ as above.

This universality principle, whereby cutoff should accompany high enough temperatures in any underlying geometry, is supported by the heuristic that at small enough β the model should qualitatively behave as if $\beta = 0$. The latter, equivalent to random walk on the hypercube, was one of the first examples of cutoff, established with an $O(1)$-cutoff window by Aldous [2], and refined in [4, 8]. Thus, one may further expect cutoff for the Ising model with an $O(1)$-window provided that β is small enough.

In contrast, cutoff for the Ising model has so far mainly been confirmed on $\mathbb{Z}^d$ [22, 23], via proofs that hinged on log-Sobolev inequalities (see [5, 9, 10, 32]) that are known to hold for the Ising model on the lattice [16, 26–29, 34, 35] as well as on the subexponential growth rate of balls in the lattice.

Even before requiring these powerful log-Sobolev inequalities, the restriction to subexponential growth rate automatically precluded the analysis of examples as basic as the Ising model on a binary tree at *any* small $\beta > 0$, or on an expander graph (e.g., a random regular graph), the hypercube, etc.

Here, using the framework of information percolation that we introduced in the companion paper [25], we confirm that on any sequence of graphs with maximum degree d, cutoff indeed occurs whenever βd is small enough, and with an $O(1)$-window (just as when $\beta = 0$). Furthermore, we analyze the effect of the initial state on the mixing time (e.g., a warm start of i.i.d. spins versus the all-plus starting state).

1.1. *Results*. Our first result establishes that, on any geometry, at high enough temperature there is cutoff within an $O(1)$-window around the point

$$t_{\mathfrak{m}} = \inf\Big\{t > 0 : \sum_v \mathfrak{m}_t(v)^2 \le 1\Big\}, \tag{1.1}$$

where $\mathfrak{m}_t(v)$ is the magnetization at a vertex $v \in V$ at time $t > 0$, that is,

$$\mathfrak{m}_t(v) = \mathbb{E}X_t^+(v), \tag{1.2}$$

with X_t^+ denoting the dynamics started from all-plus. Note that on a transitive graph (such as $\mathbb{Z}_n^d$), the point $t_{\mathfrak{m}}$ coincides with the time at which $\sum_v \mathfrak{m}_t(v)$ drops to a square-root of the volume, which has the intuitive interpretation that mixing occurs once the expected sum of spins in X_t^+ drops within the normal deviations in the Ising measure. However, it turns out that for general (nontransitive) geometries (such as trees) it is the sum of *squared* magnetizations $\sum_v \mathfrak{m}_t(v)^2$ that governs the mixing.

THEOREM 1. *There exist absolute constants $\kappa, C > 0$ such that the following holds. Let G be a graph on n vertices with maximum degree d. For any fixed*

$0 < \varepsilon < 1$ and large enough n, the continuous-time heat-bath Glauber dynamics for the Ising model on G with inverse-temperature $0 \le \beta < \kappa/d$ satisfies

$$t_{\mathrm{MIX}}(1-\varepsilon) \ge t_{\mathfrak{m}} - C\log(1/\varepsilon),$$
$$t_{\mathrm{MIX}}(\varepsilon) \le t_{\mathfrak{m}} + C\log(1/\varepsilon).$$

In particular, on any sequence of such graphs the dynamics has cutoff with an $O(1)$-window around $t_{\mathfrak{m}}$.

Apart from giving a first proof of cutoff for the Ising model on any tree/expander graph at $\beta > 0$, note that the above theorem allows the maximum degree d to depend on n in any way, and so it applies, for example, to the Ising model on the hypercube (with $d = \log_2 n$), a dense Erdős–Rényi graph $\mathcal{G}(n, \frac{1}{2})$, etc.

As mentioned above, the proof uses the new information percolation framework, which analyzes interactions between spins viewed as a percolation process in the space-time slab. As opposed to the application of this method in the companion paper [25] for the torus, various obstacles arise in the present setting due to the asymmetry between vertices and lack of amenability. Moreover, a naïve application of the method would require β to be as small as about d^{-d}, and carrying it up to κ/d (the correct dependence in d up to the value of κ) required several novel ingredients, notably using a discrete Fourier expansion (see Section 4.2) to prescribe update rules for the dynamics that would endow the resulting percolation clusters with a subcritical behavior.

Roughly put, the framework considers the dynamics at a designated time around $t_{\mathfrak{m}}$, and for each site develops the history of updates that led to its final spin (tracing back branching to its neighbors). The resulting "information percolation" clusters in the space-time slab are then categorized into three types:—RED (those surviving to time zero and nontrivially depending on the initial state), BLUE (those remaining which involve a unique "ancestor") and GREEN (all remaining clusters), as illustrated in Figure 1. The green clusters (which may exhibit complicated dependencies but are independent of the initial state) are taken out of the equation via conditioning, leaving behind a competition between blue clusters (whose ancestor vertices are i.i.d. uniform spins by symmetry) and red clusters. Controlling the latter, namely an exponential moment of their cumulative size, then establishes mixing.

Overall, the information percolation framework allows one to reduce challenging problems involving mixing and cutoff for the Ising model into simpler and tractable problems on subcritical percolation.

It is natural to ask about extension such as boundary conditions, external fields or other spin systems. While the present arguments makes use of the symmetry between plus and minus spins in the Ising model, we expect that the results should generalize to the case of boundary conditions or external fields. Conversely adapting these methods to the antiferromagnetic Ising model or the Potts model likely

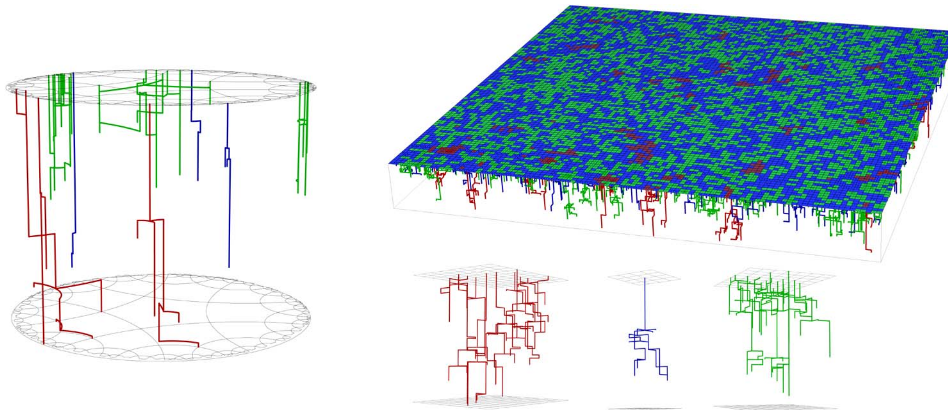

FIG. 1. *Information percolation clusters for the stochastic Ising model on two geometries*: *hyperbolic graph* (*left*, *showing largest* 3 *clusters of each type*) *and the lattice* $\mathbb{Z}_{100}^2$ (*right*). *A cluster is red if it survives to time* 0, *blue if it dies out and is the history of a single vertex*, *and green o/w*.

requires additional new ideas as the monotonicity of the Ising model is crucial in matching the upper and lower bounds. The specific challenge is that in the Potts model a grand coupling will couple more slowly than a coupling of two states.

Furthermore, by analyzing not only on the size of the red clusters, but rather *where* these hit the initial state at time zero, this framework opens the door to understanding the effect of the starting configuration on the mixing time (where sharp results on total-variation mixing for the Ising model were only applicable to worst-case starting states, usually via coupling techniques).

Our next result demonstrates this by comparing the worst-case mixing time [which is matched by the all-plus starting state up to an additive $O(1)$-term] with a typical starting configuration, and finally with the uniform starting configuration, that is, each site is initialized by an independent uniform ± 1 spin. Informally, we show that the uniform starting state is roughly at least twice faster compared to all-plus, but perhaps surprisingly, almost every deterministic starting state is about as slow as the worst one.

Formally, if $\mu_t^{(x_0)}$ is the distribution of the dynamics at time t started from x_0 then $t_{\text{MIX}}^{(x_0)}(\varepsilon)$ is the minimal t for which $\mu_t^{(x_0)}$ is within distance ε from equilibrium, and $t_{\text{MIX}}^{(\text{U})}(\varepsilon)$ is the analogue for the average $2^{-n}\sum_{x_0}\mu_t^{(x_0)}$ (i.e., the *annealed* version, as opposed to the *quenched* $t_{\text{MIX}}^{(X_0)}$ for a uniform X_0).

THEOREM 2. *Consider continuous-time heat-bath Glauber dynamics for the Ising model on an n-vertex graph G with maximum degree at most some fixed* $d>0$, *and define* $t_{\mathfrak{m}}$ *as in* (1.1). *For every* $\varepsilon>0$, *there exists* $\beta_0>0$ *such that the following hold for any* $0<\beta<\beta_0$ *and any fixed* $0<\alpha<1$ *at large enough* n:

1. (*Annealed*) *Uniform initial state*: $t_{\text{MIX}}^{(\text{U})}(\alpha)\le(\frac{1}{2}+\varepsilon)t_{\mathfrak{m}}$.

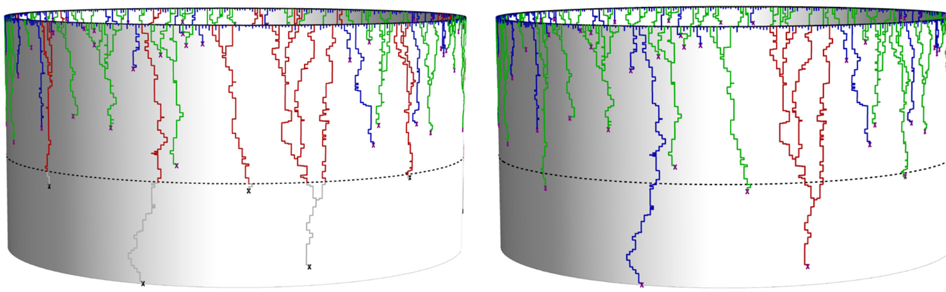

FIG. 2. *Flavor of information percolation for analyzing random initial states in* 1D *Ising model*: *On the left*, *the standard framework* (*red clusters are those reaching* $t = 0$) *for worst-case analysis*. *On the right*, *red clusters are redefined as those coalescing* below $t = 0$ *for the annealed analysis*.

2. (*Quenched*) *Deterministic initial state*: $t_{\mathrm{MIX}}^{(x_0)}(\alpha) \geq (1-\varepsilon)t_{\mathfrak{m}}$ *for asymptotically almost every* x_0, *while* $t_{\mathrm{MIX}}^{(+)}(\alpha) \sim t_{\mathfrak{m}}$.

The delicate part in the proof of the above theorem is comparing the distribution at time t directly to the Ising measure. One often bypasses this point by coupling the distributions started at worst-case states; here, however, that would fail as we are analyzing the dynamics well before these distributions can couple with high probability. Instead (and as demonstrated in the companion paper for analyzing the effect of initial states in the 1D Ising model), we appeal to the *Coupling From The Past* method [31].

Rather than developing the information percolation clusters until reaching time zero, we continue until time $-\infty$, letting all clusters eventually die. The beautiful Coupling From The Past argument implies that, if we ignore the initial state altogether, the final configuration would be a perfect simulation of the Ising measure. Thus, the natural coupling of the information percolation clusters allows one to compare the dynamics with the Ising measure, simply by considering the effect of replacing the spins generated along the interval $(-\infty, 0]$ by those of the initial state.

Specifically for the annealed analysis, even if a cluster survives to time zero (and beyond) it might still be perfectly coupled to the stationary measure, for example, a singleton strand (and more generally, a blue cluster) would receive a uniform spin both from the Ising measure and from the random initial state. Hence, we modify the framework by redefining red clusters as those in which *at least two branches* of the cluster reach time zero, then proceed to merge in the interval $(-\infty, 0)$, as illustrated in Figure 2. It is this factor of 2 that eventually transforms into the factor of $2 - \varepsilon$ improvement in the mixing time. It seems reasonable to expect asymptotic factor is exactly 2 for β positive and sufficiently which has been shown on the cycle [25] but the current analysis is not precise enough to establish it. Of course, for $\beta = 0$, the annealed measure is already mixed.

Organization. The rest of this paper is organized as follows. In Section 2, we give the formal definitions of the above described framework, including several modification needed here (e.g., custom update rules to be derived from a Fourier expansion) and two lemmas analyzing the information percolation clusters. In Section 3, we prove the cutoff result in Theorem 1 modulo these technical lemmas, which are proved in Section 4. The final section, Section 5, is devoted to the effect of the initial states on mixing and the proof of Theorem 2.

2. Information percolation for the Ising model.

2.1. *Preliminaries.* In what follows, we set up standard notation for analyzing the mixing of Glauber dynamics for the Ising model; see [22, 25] and the references therein for additional information.

Mixing time and cutoff. Let (X_t) be an ergodic finite Markov chain with stationary measure π. An important gauge in MCMC theory for measuring the convergence of a Markov chain to stationarity is its total-variation mixing time. Denoted $t_{\mathrm{MIX}}(\varepsilon)$ for a precision parameter $0 < \varepsilon < 1$, it is defined as

$$t_{\mathrm{MIX}}(\varepsilon) \stackrel{\triangle}{=} \inf\Big\{t : \max_{x_0 \in \Omega} \|\mathbb{P}_{x_0}(X_t \in \cdot) - \pi\|_{\mathrm{TV}} \le \varepsilon\Big\},$$

where here and in what follows $\mathbb{P}_{x_0}$ denotes the probability given $X_0 = x_0$, and the total-variation distance $\|\cdot\|_{\mathrm{TV}}$ between two probability measures ν_1, ν_2 on a finite space Ω is given by

$$\|\nu_1 - \nu_2\|_{\mathrm{TV}} = \max_{A \subset \Omega} \big|\nu_1(A) - \nu_2(A)\big| = \frac{1}{2} \sum_{\sigma \in \Omega} \big|\nu_1(\sigma) - \nu_2(\sigma)\big|,$$

that is, half the L^1-distance between the two measures.

Addressing the role of the parameter ε, the cutoff phenomenon is essentially the case where the choice of any fixed ε does not affect the asymptotics of $t_{\mathrm{MIX}}(\varepsilon)$ as the system size tends to infinity. Formally, a family of ergodic finite Markov chains (X_t), indexed by an implicit parameter n, is said to exhibit *cutoff* (a concept going back to the pioneering works [2, 11]) iff the following sharp transition in its convergence to stationarity occurs:

$$\lim_{n \to \infty} \frac{t_{\mathrm{MIX}}(\varepsilon)}{t_{\mathrm{MIX}}(1-\varepsilon)} = 1 \qquad \text{for any } 0 < \varepsilon < 1. \tag{2.1}$$

That is, $t_{\mathrm{MIX}}(\alpha) = (1 + o(1)) t_{\mathrm{MIX}}(\beta)$ for any fixed $0 < \alpha < \beta < 1$. The *cutoff window* addresses the rate of convergence in (2.1): a sequence $w_n = o(t_{\mathrm{MIX}}(e^{-1}))$ is a cutoff window if $t_{\mathrm{MIX}}(\varepsilon) = t_{\mathrm{MIX}}(1 - \varepsilon) + O(w_n)$ holds for any $0 < \varepsilon < 1$ with an implicit constant that may depend on ε. Equivalently, if t_n and w_n are sequences with $w_n = o(t_n)$, we say that a sequence of chains exhibits cutoff at t_n with window w_n if

$$\begin{cases} \lim\limits_{\gamma\to\infty} \liminf\limits_{n\to\infty} \max\limits_{x_0\in\Omega} \|\mathbb{P}_{x_0}(X_{t_n-\gamma w_n} \in \cdot) - \pi\|_{\mathrm{TV}} = 1, \\ \lim\limits_{\gamma\to\infty} \limsup\limits_{n\to\infty} \max\limits_{x_0\in\Omega} \|\mathbb{P}_{x_0}(X_{t_n+\gamma w_n} \in \cdot) - \pi\|_{\mathrm{TV}} = 0. \end{cases}$$

Verifying cutoff is often quite challenging, for example, even for simple random walk on an expander graph, no examples were known prior to [20, 21] (while this had been conjectured for almost all such graphs), and to date there is no known transitive example (while conjectured to hold for all transitive expanders).

Glauber dynamics for the Ising model. Let G be a finite graph with vertex-set V and edge-set E. The Ising model on G is a distribution over the set $\Omega = \{\pm 1\}^V$ of possible configurations, each corresponding to an assignment of plus/minus spins to the sites in V. The probability of $\sigma \in \Omega$ is given by

$$\pi(\sigma) = Z^{-1} e^{\beta \sum_{uv\in E} \sigma(u)\sigma(v)}, \tag{2.2}$$

where the normalizer $Z = Z(\beta, h)$ is the partition function. The parameter β is the inverse-temperature, which we always to take to be nonnegative (ferromagnetic). These definitions extend to infinite locally finite graphs (see, e.g., [19, 26]).

The Glauber dynamics for the Ising model (the *Stochastic Ising* model) is a family of continuous-time Markov chains on the state space Ω, reversible w.r.t. the Ising measure π, given by the generator

$$(\mathscr{L} f)(\sigma) = \sum_u c(u,\sigma)\big(f(\sigma^u) - f(\sigma)\big), \tag{2.3}$$

where σ^u for $u \in V$ is the configuration σ with the spin at the vertex u flipped. We will focus on the two most notable examples of Glauber dynamics, each having an intuitive and useful graphical interpretation where each site experiences updates via an associated i.i.d. rate-one Poisson clock:

(i) *Metropolis*: flip $\sigma(u)$ if the new state σ^u has a lower energy (i.e., $\pi(\sigma^u) \geq \pi(\sigma)$), otherwise perform the flip with probability $\pi(\sigma^u)/\pi(\sigma)$. This corresponds to $c(u,\sigma) = \exp(2\beta\sigma(u)\sum_{v\sim u}\sigma(y)) \wedge 1$.

(ii) *Heat-bath*: erase $\sigma(u)$ and replace it with a sample from the conditional distribution given the spins at its neighboring sites. This corresponds to $c(u,\sigma) = 1/[1+\exp(-2\beta\sigma(u)\sum_{v\sim u}\sigma(v))]$.

It is easy to verify that these chains are indeed ergodic and reversible w.r.t. the Ising distribution π. While our main results were all stated for the heat-bath chain, we note that by using the same approach one can infer the analogous statements for the Metropolis chain.

Until recently, sharp mixing results for this dynamics were obtained in relatively few cases, with cutoff only known for the complete graph [6, 17] prior to the works [22, 23].

2.2. *Red, green and blue information percolation clusters*. We begin by describing the basic setting of the framework (cf. [24]) which will be enhanced in Section 2.3 to support the setting of Theorem 1 (where the underlying geometry may feature an exponential growth rate and we consider $\beta < \kappa/d$). This can be viewed as an extension of Harris' graphical representation [14] (cf. [19], Section III.6)—an instrumental tool in analyzing voter models and contact processes—in the context of the Ising model, where our understanding of clusters associated with particle interactions in the space-time slab is limited.

The *update sequence* of the Glauber dynamics along an interval $(t_0, t_1]$ is the set of tuples of the form (J, U, τ), where $t_0 < \tau \le t_1$ is the update time, $J \in V$ is the site to be updated and U is a uniform unit variable. Given this update sequence, X_{t_1} is a deterministic function of X_{t_0}, right-continuous w.r.t. t_1.

We call a given update (J, U, τ) an *oblivious update* iff $U \le \theta$ for

$$\theta = \theta_{\beta,d} := 1 - \tanh(\beta d), \tag{2.4}$$

since in that situation one can update the spin at J to plus/minus with equal probability (i.e., with probability $\theta/2$ each) independently of the spins at the neighbors of the vertex J, and a properly chosen rule for the case $U > \theta$ legally extends this protocol to the Glauber dynamics.

Consider some designated target time $t_\star$ for analyzing the spin distribution of the dynamics on G. The *update history* of $X_{t_\star}(v)$ going back to time t, denoted $\mathscr{H}_v(t)$, is a subset $A \times \{t\}$ of the space-time slab $V \times \{t\}$, such that one we can determine $X_{t_\star}(v)$ from the update sequence and spin-set $X_t(A)$. The most basic way of defining $\{\mathscr{H}_v(t) : 0 \le t \le t_\star\}$ is as follows:

- List the updates in reverse chronological order as $\{(J_i, U_i, t_i)\}_{i \ge 1}$ (i.e., $t_i > t_{i+1}$ for all i), and initialize the update history by $\mathscr{H}_v(t) = \{v\}$ for all $t \in [t_1, t_\star]$.
- In step $i \ge 1$, process the update (J_i, U_i, t_i) to determine $\mathscr{H}_v(t)$ for $t \in [t_{i+1}, t_i)$:
  - If $J_i \notin \mathscr{H}_v(t_i)$ then the history is unchanged, that is, $\mathscr{H}_v(t) = \mathscr{H}_v(t_i)$ for all $t \in [t_{i+1}, t_i)$.
  - If $J_i \in \mathscr{H}_v(t_i)$ but $U_i \le \theta$ then J_i is removed, that is, $\mathscr{H}_v(t) = \mathscr{H}_v(t_i) \setminus \{J_i\}$ for all $t \in [t_{i+1}, t_i)$.
  - Otherwise, replace J_i by its neighbors $N(J_i)$, that is, $\mathscr{H}_v(t) = \mathscr{H}_v(t_i) \cup N(J_i) \setminus \{J_i\}$ for all $t \in [t_{i+1}, t_i)$.

The *information percolation clusters* are the transitive closure on the vertex set V induced by a relation of u and v if $\mathscr{H}_u(t) \cap \mathscr{H}_v(t) \ne \varnothing$ for some $t \ge 0$. Denote by $\mathcal{C}_v$ the cluster containing $v \in V$.

We will also consider clusters in the context of the full space-time slab. The cluster of a point $(w, r) \in V \times [0, t_\star]$, denoted $\mathcal{X}_{w,r}$, is the connected component of $\bigcup\{\mathscr{H}_v(t) : v \in V, 0 \le t \le t_\star\}$ that contains (w, r). (Thus, the cluster $\mathcal{C}_v$ is identified with the intersection of $\mathcal{X}_{v,t_\star}$ with the slab $V \times \{t_\star\}$.) For any $A \subset V$, we use the notation $\mathscr{H}_A(t) = \bigcup_{v \in A} \mathscr{H}_v(t)$, as well as $\mathscr{H}_A(t_1, t_2) = \bigcup_{t_1 \le t \le t_2} \mathscr{H}_A(t)$ (both cases describing subsets of V).

The clusters are classified into three classes (identifying for this purpose $\mathcal{C}_v$ and $\mathcal{X}_{v,t_\star}$) as follows:

- A cluster $\mathcal{C}$ is RED if, given the update sequence, its final state $X_{t_\star}(\mathcal{C})$ is a nontrivial function of the initial configuration X_0; in particular, its history must survive to time zero ($\mathscr{H}_\mathcal{C}(0) \neq \varnothing$).
- A cluster $\mathcal{C}$ is BLUE if it is a singleton—that is, $\mathcal{C} = \{v\}$ for some $v \in V$—whose history does not survive to time zero ($\mathscr{H}_v(0) = \varnothing$).
- Every other cluster $\mathcal{C}$ is GREEN.

Note that if a cluster is blue then its single spin at time $t_\star$ does not depend on the initial state X_0, and so, by symmetry, it is a uniform ± 1 spin. (While a green cluster is similarly independent of X_0, as multiple update histories intersect, the distribution of its spin set $X_{t_\star}(\mathcal{C})$ may become quite nontrivial.)

Let us briefly mention the structure of the clusters in a couple of special cases. When $\beta = 0$, since all updates are oblivious, all clusters are singletons and so none are green while the red vertices are those which recipe no updates. On the cycle for $\beta > 0$, updates can be constructed in such a way that the set of histories of the clusters form a collection of coalescing killed random walks (see [25]). Those which survive to time 0 are red while those the which are killed before merging are blue and those which merge and are subsequently killed are green.

Let V_{RED} denote the union of the red clusters, and let $\mathscr{H}_{\mathrm{RED}}$ be the its collective history—the union of $\mathscr{H}_v(t)$ for all $v \in V_{\mathrm{RED}}$ and $0 \le t \le t_\star$ (with analogous definitions for blue/green).

A beautiful short lemma of Miller and Peres [30] shows that, if a measure μ on $\{\pm 1\}^V$ is given by sampling a variable $R \subset V$ and using an arbitrary law for its spins and a product of Bernoulli($\frac{1}{2}$) for $V \setminus R$, then the L^2-distance of μ from the uniform measure is at most $\mathbb{E}2^{|R \cap R'|} - 1$ for i.i.d. copies R, R'. (See Lemma 3.1 below; also see [25], Lemma 4.3, for a generalization of this to a product of general measures, which becomes imperative for the information percolation framework at β near criticality.) Applied to our setting, if we condition on $\mathscr{H}_{\mathrm{GREEN}}$ and look at the spins of $V \setminus V_{\mathrm{GREEN}}$ then V_{RED} can assume the role of the variable R, as the remaining blue clusters are a product of Bernoulli($\frac{1}{2}$) variables.

In this conditional space, since the law of the spins of V_{GREEN}, albeit potentially complicated, is independent of the initial state, we can safely project the configurations on $V \setminus V_{\mathrm{GREEN}}$ without it increasing the total-variation distance between the distributions started at the two extreme states. Hence, a sharp upper bound on worst-case mixing will follow by showing for this exponential moment

$$\mathbb{E}\big[2^{|V_{\mathrm{RED}} \cap V'_{\mathrm{RED}}|} \mid \mathscr{H}_{\mathrm{GREEN}}\big] \to 1 \qquad \text{in probability as } n \to \infty, \tag{2.5}$$

by coupling the distribution of the dynamics at time $t_\star$ from any initial state to the uniform measure. Finally, with the green clusters out of the picture by the conditioning (which has its own toll, forcing various updates along history so that no other cluster would intersect with those nor become green), we can bound the probability that a subset of sites would become a red cluster by its ratio with the probability of all sites being blue clusters. Being red entails connecting the subset in the space-time slab, hence the exponential decay needed for (2.5).

2.3. *Enhancements of the framework*: *Custom update rules and modified last unit interval*. We will consider the information percolation clusters developed as above from the designated time

$$t_\star = t_{\mathfrak{m}} + s_\star \qquad \text{for } s_\star = C \log(1/\varepsilon),$$

where $C > 0$ will be specified later, and $\varepsilon > 0$ is the parameter for the mixing time. However, instead of the standard procedure of developing the history, where an update at v either deletes it from the history (via an oblivious update) or replaces it by its set of neighbors $N(v)$, we will allow v to be replaced (with varying probabilities) by any subset of its neighbors, in the following way.

Recall that an update of the form $(J, U, t) \in V \times [0, 1] \times [0, t_\star]$ results in replacing the spin at J at time t by some deterministic function $\Upsilon(x, U)$, where $x = \sum_{u \in N(J)} X_t(u)$. We introduce the notion of a *generalized* update rule in order to establish the right dependence on β in Theorem 1 by having rules updates which typically only observe a small number of neighboring vertices. The updates are of the form (J, A, U, t) where (J, U, t) is as before and the additional variable $A \subset [d]$ corresponds to a subset of the neighbors of vertex J. The new update rule exposes the spins $\{\sigma_1, \dots, \sigma_{|A|}\}$ of these neighbors at time t, then generates the new spin at J via $\Phi_A(\sigma_1, \dots, \sigma_{|A|}, U)$.

With this generalized update rule, one unfolds the update history of a vertex $\{\mathscr{H}_v(t) : 0 \le t \le t_\star\}$ as before, with the one difference that an update (J_i, A_i, U_i, t_i) for which $J_i \in \mathscr{H}_v(t_i)$ now results in $\mathscr{H}_v(t) = \mathscr{H}_v(t_i) \cup A_i \setminus \{J_i\}$ for all $t \in [t_{i+1}, t_i)$. The functions $\{\Phi_A : A \subset [d]\}$, as well as the probability distribution over the subsets $A \subset [d]$ to be exposed, will be derived from a discrete Fourier expansion of the original rule Υ (see Lemma 4.1), so that the new update procedure would, on one hand, couple with the Glauber dynamics, and on the other, generally use small subsets A which endows our percolation clusters with a subcritical behavior giving a better dependence on β.

A final ingredient needed for coping with the arbitrary underlying geometry is a modification of the update history, denoted by $\hat{\mathscr{H}}$: in the modified version, every vertex $v \in V$ receives an (extra) update at time $t_\star$, and no vertex is removed from the history along the unit interval $(t_\star - 1, t_\star]$. Since we do not remove vertices, there is no issue of ordering of the vertices. Note that this construction no longer corresponds to the dynamics, it generates a set of information percolation clusters which dominate the original histories. The purpose of this construction is that forbidding vertices to die in the first unit interval will be useful in the context of conditioning on other clusters. We will write $\hat{\mathcal{C}}$, $\hat{\mathcal{X}}$, as well as $\hat{\mathscr{H}}_A(t)$ etc. for the corresponding notation w.r.t. the modified history $\hat{\mathscr{H}}$.

We end this section with two results on the information percolation clusters—Lemmas 2.1 and 2.2—which will be central in the proof of Theorem 1. The proofs of these lemmas are postponed to Section 4.

Let us denote the collective history of the complement of a set $A \subset V$ as

$$\mathscr{H}_A^- = \{\mathscr{H}_v(t) : v \notin A, t \le t_\star\}.$$

As explained following the definition of the three cluster types, at the heart of the matter is estimating an exponential moment of the size of the red clusters given $\mathscr{H}_{\mathrm{GREEN}}$, the joint history of all green clusters. To this end, we consider the following conditional probability that a subset A is a red cluster. Define

$$\Psi_A = \sup_H \mathbb{P}\big(A \in \mathrm{RED} \mid \mathscr{H}_A^- = H, \{A \in \mathrm{RED}\} \cup \{A \subset V_{\mathrm{BLUE}}\}\big), \tag{2.6}$$

noting that, toward estimating the probability of $A \in \mathrm{RED}$, the effect of conditioning on $\mathscr{H}_A^-$ amounts to requiring that $\mathscr{H}_A$ must not intersect $\mathscr{H}_A^-$. We should note that the event $A \in \mathrm{RED}$ means that A is a single red cluster, not a collection of them.

LEMMA 2.1. *If $\beta < 1/(5d)$ then for any $A \subset V$ and $v \in A$,*

$$\Psi_A \le 2^{|A|}\mathbb{E}\bigg[\mathbb{1}_{\{A \subset \hat{\mathcal{C}}_v\}} e^{\hat{\tau}_v} \sum_w \mathbb{1}_{\{w \in \hat{\mathscr{H}}_A(t_\star - \hat{\tau}_v, t_\star)\}} \mathfrak{m}_{t_\star}(w)\bigg],$$

where $\hat{\tau}_v$ is the time it takes the history of $\hat{\mathcal{C}}_v$ to first coalesce into a single point (if at all), that is,

$$\hat{\tau}_v = \min\{t \ge 1 : |\hat{\mathscr{H}}_{\hat{\mathcal{C}}_v}(t_\star - t)| = 1\} \wedge t_\star. \tag{2.7}$$

It is worthwhile noting in the context of the parameter $\hat{\tau}_v$ that, when developing the update history backward in time, $\hat{\tau}_v$ is not a stopping time, since $\hat{\mathcal{C}}_v$ is affected by any potential coalescence points for $t < t_\star - \hat{\tau}_v$; instead, one can determine $\hat{\tau}_v$ as soon as $\hat{\mathscr{H}}_{\hat{\mathcal{C}}_v}(t) = \varnothing$. Also observe that $\hat{\tau}_v = 1$ iff $|\hat{\mathcal{C}}_v| = 1$. Finally, the coalescence point w at time $t = t_\star - \hat{\tau}_v$ (when $t > 0$) need not belong to $\hat{\mathscr{H}}_v$, for example, we may have $\hat{\mathscr{H}}_v(t) = \varnothing$ while $w \in \hat{\mathscr{H}}_u$ for some $u \ne v$ whose history intersected that of v at time $t' > t$.

The subcritical nature of the information percolation clusters (prompted by our modified update functions Φ_A) allows one to control exponential moments of the cluster sizes, as in the following lemma.

LEMMA 2.2. *Fix $0 < \eta < 1$ and $\lambda > 0$. There exist constants $\kappa, \gamma > 0$ such that the following holds. For any point (w_0, t_0) in the space-time slab $V \times (0, t_\star]$, if $\beta < \kappa/d$ then*

$$\mathbb{E}\big[\exp\big(\eta\mathfrak{L}(\hat{\mathcal{X}}_{w_0,t_0}) + \lambda|\hat{\mathscr{H}}_{\hat{\mathcal{X}}_{w_0,t_0}}|\big)\big] < \gamma,$$

where

$$\mathfrak{L}(\hat{\mathcal{X}}) = \sum_{u \in V} \int_0^{t_\star} \mathbb{1}_{\{(u,t) \in \hat{\mathcal{X}}\}} \, dt.$$

The above lemma, the proof of which follows standard arguments from percolation theory, will be applied for absolute constants η and λ in the proof of Theorem 1 (any $1/2 < \eta < 1$ and $\lambda > \log 8$ would do), leading to the absolute constant κ in the statement of that theorem. The above formulation will be important in the context of Theorem 2, where one requires η that may be very close to 1 (as a function of ε from the statement of that theorem) and λ that depends on the maximum degree.

3. Cutoff with constant window from a worst starting state. In this section we prove Theorem 1 via the framework defined in Section 2. As is often the case in proofs of cutoff, the upper bound will require the lion's share of the efforts.

3.1. *Upper bound modulo Lemmas* 2.1 *and* 2.2. Define the coupling distance $\bar{d}_{\mathrm{TV}}(t)$ to be

$$\bar{d}_{\mathrm{TV}}(t) = \max_{x_0, y_0} \| \mathbb{P}_{x_0}(X_t \in \cdot) - \mathbb{P}_{y_0}(X_t \in \cdot) \|_{\mathrm{TV}}$$

(so that $\frac{1}{2}\bar{d}_{\mathrm{TV}}(t) \le \max_{x_0} \|\mathbb{P}_{x_0}(X_t \in \cdot) - \pi(\cdot)\|_{\mathrm{TV}} \le \bar{d}_{\mathrm{TV}}(t)$), and observe that

$$\begin{aligned}
\bar{d}_{\mathrm{TV}}(t) &\le \mathbb{E}\Big[\max_{x_0, y_0} \| \mathbb{P}_{x_0}(X_t \in \cdot \mid \mathscr{H}_{\mathrm{GREEN}}) - \mathbb{P}_{y_0}(X_t \in \cdot \mid \mathscr{H}_{\mathrm{GREEN}}) \|_{\mathrm{TV}}\Big] \\
&\le \sup_{\mathscr{H}_{\mathrm{GREEN}}} \max_{x_0, y_0} \big\| \mathbb{P}_{x_0}\big(X_t(V \setminus V_{\mathrm{GREEN}}) \in \cdot \mid \mathscr{H}_{\mathrm{GREEN}}\big) \\
&\qquad - \mathbb{P}_{y_0}\big(X_t(V \setminus V_{\mathrm{GREEN}}) \in \cdot \mid \mathscr{H}_{\mathrm{GREEN}}\big) \big\|_{\mathrm{TV}},
\end{aligned}$$

where the first inequality follows by Jensen's inequality and the second follows since $X_t(V_{\mathrm{GREEN}})$ is independent of the initial condition and so taking a projection onto $V \setminus V_{\mathrm{GREEN}}$ does not change the total-variation distance between the distributions started at x_0 and y_0. Thus,

$$\begin{aligned}
\bar{d}_{\mathrm{TV}}(t) \le 2 \sup_{\mathscr{H}_{\mathrm{GREEN}}} \max_{x_0} \big\| \mathbb{P}_{x_0}\big(X_t(V \setminus V_{\mathrm{GREEN}}) \in \cdot \mid \mathscr{H}_{\mathrm{GREEN}}\big) \\
- \nu_{V \setminus V_{\mathrm{GREEN}}} \big\|_{\mathrm{TV}},
\end{aligned} \tag{3.1}$$

where ν_A is the uniform measure on configurations on the sites in A. At this point, we appeal to the exponential-moment bound of [30], the short proof of which is included here for completeness.

LEMMA 3.1 ([30]). *Let $\Omega = \{\pm 1\}^V$ for a finite set V. For each $S \subset V$, let φ_S be a measure on $\{\pm 1\}^S$. Let ν be the uniform measure on Ω, and let μ be* *the measure on Ω obtained by sampling a subset $S \subset V$ via some measure $\tilde{\mu}$, generating the spins of S via φ_S, and finally sampling $V \setminus S$ uniformly. Then*

$$\|\mu - \nu\|_{L^2(\nu)}^2 \le \mathbb{E}[2^{|S\cap S'|}] - 1,$$

where the variables S and S' are i.i.d. with law $\tilde{\mu}$.

PROOF. Write $n = |V|$, and let x_S $(S \subset V)$ denote the projection of x onto S. With this notation, by definition of the $L^2(\nu)$ metric (see, e.g., [33]) one has that $\|\mu - \nu\|_{L^2(\nu)}^2 + 1 = \int |\mu/\nu - 1|^2 \, d\nu + 1$ equals

$$\sum_{x\in\Omega} \frac{\mu^2(x)}{\nu(x)} = 2^n \sum_{x\in\Omega} \sum_S \tilde{\mu}(S) \frac{\varphi_S(x_S)}{2^{n-|S|}} \sum_{S'} \tilde{\mu}(S') \frac{\varphi_{S'}(x_{S'})}{2^{n-|S'|}}$$

by the definition of μ. Since $\sum_x \varphi_S(x_S)\varphi_{S'}(x_{S'}) \le 2^{n-|S\cup S'|}$ it then follows that

$$\sum_{x\in\Omega} \frac{\mu^2(x)}{\nu(x)} \le \sum_{S,S'} 2^{|S|+|S'|-|S\cup S'|} \tilde{\mu}(S)\tilde{\mu}(S') = \sum_{S,S'} 2^{|S\cap S'|} \tilde{\mu}(S)\tilde{\mu}(S'). \qquad \square$$

REMARK 3.2. In the special case where the distribution φ_S is a point-mass on all-plus for every S, the single inequality in the above proof is an equality (since then $\sum_x \varphi_S(x_S)\varphi_{S'}(x_{S'}) = \#\{x : x_{S\cup S'} \equiv 1\}$) and so in that situation the L^2-distance $\|\mu - \nu\|_{L^2(\nu)}^2$ is precisely equal to $\mathbb{E}[2^{|S\cap S'|}] - 1$.

For example, consider Glauber dynamics for an n-vertex graph at $\beta = 0$ (i.e., continuous-time lazy random walk on the hypercube $\{\pm 1\}^n$) starting (say) from all-plus, and let S be the set of coordinates which were not updated: here $\mathbb{P}(v \in S) = e^{-t}$ at time t, and $\|\mathbb{P}(X_t^+ \in \cdot) - \nu\|_{L^2(\nu)}^2 = (1 + e^{-2t})^n - 1$.

Applying the above lemma to the right-hand side of (3.1), while recalling that any two measures μ and ν on a finite probability space satisfy $\|\mu - \nu\|_{\mathrm{TV}} = \frac{1}{2}\|\mu - \nu\|_{L^1(\nu)} \le \frac{1}{2}\|\mu - \nu\|_{L^2(\nu)}$, we find that

$$\bar{d}_{\mathrm{TV}}(t_\star) \le \Big(\sup_{\mathscr{H}_{\mathrm{GREEN}}} \mathbb{E}[2^{|V_{\mathrm{RED}} \cap V_{\mathrm{RED}'}|} \mid \mathscr{H}_{\mathrm{GREEN}}] - 1\Big)^{1/2}, \tag{3.2}$$

where V_{RED} and $V_{\mathrm{RED}'}$ are i.i.d. copies of the variable $\bigcup\{v \in V : \mathcal{C}_v \in \mathrm{RED}\}$.

Let $\{Y_{A,A'} : A, A' \subset V\}$ be a family of independent indicators satisfying

$$\mathbb{P}(Y_{A,A'} = 1) = \Psi_A \Psi_{A'} \qquad \text{for any } A, A' \subset V. \tag{3.3}$$

We claim that it is possible to couple the conditional distribution of $(V_{\mathrm{RED}}, V_{\mathrm{RED}'})$ given $\mathscr{H}_{\mathrm{GREEN}}$ to the variables $Y_{A,A'}$ in such a way that

$$|V_{\mathrm{RED}} \cap V_{\mathrm{RED}'}| \le \sum_{A\cap A' \ne \varnothing} |A \cup A'| Y_{A,A'}.$$

To do so, let $\{(A_l, A'_l)\}_{l\ge 1}$ denote all pairs of intersecting subsets ($A, A' \subset V \setminus V_{\mathrm{GREEN}}$ with $A \cap A' \neq \varnothing$) arbitrarily ordered, associate each pair with a variable R_l initially set to 0, then process these in order:

- If (A_l, A'_l) is such that, for some $j < l$, one has $R_j = 1$ and either $A_j \cap A_l \neq \varnothing$ or $A'_j \cap A'_l \neq \varnothing$, then skip this pair (keeping $R_l = 0$).
- Otherwise, set R_l to the indicator of $\{A_l \in \mathrm{RED}, A'_l \in \mathrm{RED}'\}$.

The claim is that $\mathbb{P}(R_l = 1 \mid \mathcal{F}_{l-1}) \le \mathbb{P}(Y_{A_l, A'_l} = 1)$ for all l, where $\mathcal{F}_l$ denotes the natural filtration associated to the above process. Indeed, consider some (A_l, A'_l) for which we are about to set R_l to the value of $\mathbb{1}_{\{A_l \in \mathrm{RED}, A'_l \in \mathrm{RED}'\}}$, and take any A_j ($j < l$) such that $A_j \cap A_l \neq \varnothing$ and $\mathbb{1}_{\{A_j \in \mathrm{RED}, A'_j \in \mathrm{RED}'\}}$ was revealed (and necessarily found to be zero, by definition of the above process). The supremum over $\mathscr{H}^-_{A_l}$ in the definition of Ψ_{A_l} implies that we need only consider the information $\mathcal{F}_{l-1}$ offers on $\mathscr{H}_{A_l}$:

- If $A_j \cap A_l \neq A_l$, then the event $\{A_j \in \mathrm{RED}\}$ does not intersect the event $\{A_l \in \mathrm{RED}\} \cup \{A_l \subset V_{\mathrm{BLUE}}\}$ (on which we condition in Ψ_{A_l}) as it requires A_j to be a full red cluster (so a strict subset of A_j cannot belong to a separate red cluster, nor can it contain any blue singleton).
- If $A_j = A_l$, conditioning on $\{A_j \in \mathrm{RED}, A'_j \in \mathrm{RED}'\}^c$ will not increase the probability of $\{A_l \in \mathrm{RED}\}$.

Either way, $\mathbb{P}(A_l \in \mathrm{RED} \mid \mathcal{F}_{l-1}) \le \Psi_{A_l}$. Similarly, $\mathbb{P}(A'_l \in \mathrm{RED}' \mid \mathcal{F}_{l-1}, \mathbb{1}_{\{A_l \in \mathrm{RED}\}}) \le \Psi_{A'_l}$, and together these inequalities support the desired coupling, since if $v \in V_{\mathrm{RED}} \cap V_{\mathrm{RED}'}$ then there is some l for which $v \in A_l \cup A'_l$ and $A_l \in \mathrm{RED}$, $A'_l \in \mathrm{RED}'$, in which case every A_j intersecting A_l nontrivially will receive $R_j = 0$ (it cannot be red) and the first j with $A_j = A_l$ to receive $R_j = 1$ will account for v in $A_j \cup A'_j$.

Relaxing $|A \cup A'|$ into $|A| + |A'|$ (which will be convenient for factorization), we get

$$\begin{aligned}\sup_{\mathscr{H}_{\mathrm{GREEN}}} \mathbb{E}\big[2^{|V_{\mathrm{RED}} \cap V_{\mathrm{RED}'}|} \mid \mathscr{H}_{\mathrm{GREEN}}\big] &\le \mathbb{E}\big[2^{\sum_{A \cap A' \neq \varnothing} (|A|+|A'|) Y_{A,A'}}\big] \\ &= \prod_{A \cap A' \neq \varnothing} \mathbb{E}\big[2^{(|A|+|A'|) Y_{A,A'}}\big],\end{aligned}$$

with the equality due to the independence of the $Y_{A,A'}$'s. By the definition of these indicators in (3.3), this last expression is at most

$$\prod_v \prod_{\substack{A, A' \\ v \in A \cap A'}} \big((2^{|A|+|A'|} - 1)\Psi_A \Psi_{A'} + 1\big) \le \exp\bigg[\sum_v \bigg(\sum_{A \ni v} 2^{|A|} \Psi_A\bigg)^2\bigg],$$

and so, revisiting (3.2), we conclude that

$$\bar{d}_{\mathrm{TV}}(t_\star)^2 \le \biggl(\exp\biggl[\sum_v \Bigl(\sum_{A\ni v} 2^{|A|}\Psi_A\Bigr)^2\biggr] - 1\biggr) \wedge 1 \le 2\sum_v \Bigl(\sum_{A\ni v} 2^{|A|}\Psi_A\Bigr)^2, \tag{3.4}$$

where we used that $e^x - 1 \le 2x$ for $x \in [0,1]$. We have thus reduced the upper bound in Theorem 1 into showing that the right-hand side of (3.4) is at most ε if $s_\star = C\log(1/\varepsilon)$ for some large enough $C = C(\beta)$.

Plugging the bound on Ψ_A from Lemma 2.1 shows that the sum in the right-hand side of (3.4) is at most

$$\sum_v \biggl(\sum_{A\ni v} 4^{|A|}\mathbb{E}\biggl[\mathbb{1}_{\{A\subset \hat{\mathcal{C}}_v\}} e^{\hat{\tau}_v} \sum_w \mathbb{1}_{\{w\in \hat{\mathscr{H}}_A(t_\star - \hat{\tau}_v, t_\star)\}} \mathfrak{m}_{t_\star}(w)\biggr]\biggr)^2.$$

In each of the two sums over $A \ni v$, we can specify the size of $\hat{\mathcal{C}}_v$, and then relax $\{w \in \hat{\mathscr{H}}_A(t_\star - \hat{\tau}_v, t_\star)\}$ into $\{w \in \hat{\mathscr{H}}_{\hat{\mathcal{C}}_v}\}$ (thus permitting all $2^{|\hat{\mathcal{C}}_v|}$ subsets to play the role of A); thus, the last display is at most

$$\begin{aligned} &\sum_v \sum_{k,k'} \sum_{w,w'} 8^k \mathbb{E}\bigl[\mathbb{1}\{|\hat{\mathcal{C}}_v| = k, w \in \hat{\mathscr{H}}_{\hat{\mathcal{C}}_v}\} e^{\hat{\tau}_v}\mathfrak{m}_{t_\star}(w)\bigr] \\ &\qquad \cdot 8^{k'}\mathbb{E}\bigl[\mathbb{1}\{|\hat{\mathcal{C}}_v| = k', w' \in \hat{\mathscr{H}}_{\hat{\mathcal{C}}_v}\} e^{\hat{\tau}_v}\mathfrak{m}_{t_\star}(w')\bigr]. \end{aligned} \tag{3.5}$$

Denoting the indicators above by $\Xi(v,w,k)$ and $\Xi(v,w',k')$, respectively, and using the fact that

$$\sum_{w,w'} \mathfrak{m}_t(w)\mathfrak{m}_t(w') \le \frac{1}{2}\sum_{w,w'}\bigl(\mathfrak{m}_t(w)^2 + \mathfrak{m}_t(w')^2\bigr) = \sum_{w,w'} \mathfrak{m}_t(w)^2$$

in (3.5) culminates in the following bound on sum in the right-hand side of (3.4):

$$\begin{aligned} \sum_v \Bigl(\sum_{A\ni v} 2^{|A|}\Psi_A\Bigr)^2 &\le \sum_w \mathfrak{m}_{t_\star}(w)^2 \sum_k \sum_v \mathbb{E}\bigl[8^k\,\Xi(v,w,k)e^{\hat{\tau}_v}\bigr] \\ &\qquad \times \sum_{k'} \mathbb{E}\biggl[e^{\hat{\tau}_v}\sum_{w'} 8^{k'}\,\Xi(v,w',k')\biggr]. \end{aligned} \tag{3.6}$$

For the summation over k' in (3.6), we combine the facts that $\hat{\tau}_v \le \frac{1}{2}\mathfrak{L}(\hat{\mathscr{H}}_{\hat{\mathcal{C}}_v}(t_\star - \hat{\tau}_v, t_\star)) + 1 \le \frac{1}{2}\mathfrak{L}(\hat{\mathscr{H}}_{\hat{\mathcal{C}}_v}) + 1$ (either $|\hat{\mathcal{C}}_v| = 1$ and then $\hat{\tau}_v = 1$, or $|\hat{\mathcal{C}}_v| \ge 2$ whence at least two strands survive for a period of $\hat{\tau}_v$), that at most $|\hat{\mathscr{H}}_{\hat{\mathcal{C}}_v}|$ choices for w' support $\Xi(v,w',k') = 1$ and that $\sum_{k'}\Xi(v,w',k') \le 1$, to get

$$\sum_{k'} \mathbb{E}\biggl[e^{\hat{\tau}_v}\sum_{w'} 8^{k'}\,\Xi(v,w',k')\biggr] \le \mathbb{E}\bigl[|\hat{\mathscr{H}}_{\hat{\mathcal{C}}_v}| 8^{|\hat{\mathscr{H}}_{\hat{\mathcal{C}}_v}|} e^{\frac{1}{2}\mathfrak{L}(\hat{\mathscr{H}}_{\hat{\mathcal{C}}_v})+1}\bigr] \le \gamma_1 \tag{3.7}$$

for some absolute constant $\gamma_1 > 0$, where the last inequality applied Lemma 2.2.

Next, to treat the summation over k in (3.6), recall that $\hat{\mathcal{X}}_{w,r}$ for $(w,r) \in V \times [0,t_\star]$ is the information percolation cluster containing the point (w,r) in the space-time slab (i.e., the cluster is exposed from time r instead of time $t_\star$ and the process of developing it moves both forward and backward in time). Further, write $\hat{\mathcal{X}}^+_{w,r} = \lim_{t\to r^+} \hat{\mathcal{X}}_{w,t}$ and $\hat{\mathcal{X}}^-_{w,r} = \lim_{t\to r^-} \hat{\mathcal{X}}_{w,t}$.

Note that whenever $\Xi(v,w,k) = 1$, necessarily $(v,t_\star) \in \hat{\mathcal{X}}^-_{w,r}$ for some $r \in \Pi_w$, where Π_w records the update times for the vertex w (always including $t_\star$, by definition of $\hat{\mathscr{H}}$). Indeed, if $w \in \hat{\mathscr{H}}_{\hat{\mathcal{C}}_v}(t_\star - \hat{\tau}_v, t_\star)$ then by definition we can find some $q \in (t_\star - \hat{\tau}_v, t_\star)$ such that (w,q) shares the same information percolation cluster as $(v,t_\star)$. Furthermore, if r is the earliest update of w after time q then the cluster of (w,t) for any $t \in (q,r)$ will contain (w,q), and thus $(v,t_\star)$ as-well. (It is for this reason that we addressed $\hat{\mathcal{X}}^-_{w,r}$, in case the update at (w,r) should cut its information percolation cluster from (w,q).) For that r, we further have $t_\star - r \le \hat{\tau}_v \le \frac{1}{2}\mathfrak{L}(\hat{\mathcal{X}}^-_{w,r}) + 1$, and so

$$\begin{aligned}
&\sum_k \sum_v \mathbb{E}\big[8^k \Xi(v,w,k) e^{\hat{\tau}_v}\big] \\
&\quad \le \mathbb{E}\Bigg[\sum_{r\in\Pi_w} |\hat{\mathscr{H}}_{\hat{\mathcal{X}}^-_{w,r}}| 8^{|\hat{\mathscr{H}}_{\hat{\mathcal{X}}^-_{w,r}}|} e^{\frac{1}{2}\mathfrak{L}(\hat{\mathcal{X}}^-_{w,r})} \mathbb{1}_{\{\frac{1}{2}\mathfrak{L}(\hat{\mathcal{X}}^-_{w,r}) \ge t_\star - r - 1\}}\Bigg] \\
&\quad \le \mathbb{E}\Bigg[\sum_{r\in\Pi_w} |\hat{\mathscr{H}}_{\hat{\mathcal{X}}^-_{w,r}}| 8^{|\hat{\mathscr{H}}_{\hat{\mathcal{X}}^-_{w,r}}|} e^{\frac{3}{4}\mathfrak{L}(\hat{\mathcal{X}}^-_{w,r}) - \frac{1}{2}(t_\star - r - 1)}\Bigg],
\end{aligned}$$

which, recalling that Π_w is the union of $\{t_\star\}$ and a rate-1 Poisson process, is at most

$$\begin{aligned}
&\mathbb{E}\big[|\hat{\mathscr{H}}_{\hat{\mathcal{C}}_w}| 8^{|\hat{\mathscr{H}}_{\hat{\mathcal{C}}_w}|} e^{\frac{3}{4}\mathfrak{L}(\hat{\mathscr{H}}_{\hat{\mathcal{C}}_w}) + \frac{1}{2}}\big] \\
&\qquad + \int_0^{t_\star} \mathbb{E}\big[|\hat{\mathscr{H}}_{\hat{\mathcal{X}}^-_{w,r}}| 8^{|\hat{\mathscr{H}}_{\hat{\mathcal{X}}^-_{w,r}}|} e^{\frac{3}{4}\mathfrak{L}(\hat{\mathcal{X}}^-_{w,r}) - \frac{1}{2}(t_\star - r - 1)} \mid r \in \Pi_w\big]\, dr \\
&\quad \le \sqrt{e}\gamma_2 \bigg[1 + \int_0^{t_\star} e^{-\frac{1}{2}(t_\star - r)}\, dr\bigg] \le 5\gamma_2
\end{aligned}$$

for some absolute constant $\gamma_2 > 0$, using Lemma 2.2 (with γ_2 from that lemma) for the first inequality.

Substituting the last two displays together with (3.7) in (3.6), while recalling (3.4), finally gives

$$\bar{d}_{\mathrm{TV}}(t_\star)^2 \le 10\gamma_1\gamma_2 \sum_w \mathfrak{m}_{t_\star}(w)^2. \tag{3.8}$$

The proof will be concluded with the help of the next simple claim that establishes a submultiplicative bound for the second moment of the magnetization.

CLAIM 3.3. *For any $t, s > 0$, we have*

$$e^{-2s} \le \frac{\sum_w \mathfrak{m}_{t+s}(w)^2}{\sum_w \mathfrak{m}_t(w)^2} \le e^{-2(1-\beta d)s}.$$

PROOF. The lower bound follows from the straightforward fact that $\mathfrak{m}_{t+s}(w) \ge e^{-s}\mathfrak{m}_t(w)$ for any $s, t > 0$ and w, since the probability of observing no updates to w along the interval $(t, t+s)$ (thus maintaining the magnetization without a change) is e^{-s}. It therefore remains to prove the upper bound.

When a vertex v is updated if σ is the sum of the spins of its neighbors, then v is set to 1 with probability $\frac{1}{2} + \frac{1}{2}\tanh(\beta\sigma)$ and -1 otherwise. Thus, since $\frac{d}{dx}\tanh(x) \le 1$ for all $x \in \mathbb{R}$ after an update of v at time t,

$$\begin{aligned}
&\mathbb{P}\big(X_t^+(v) = 1\big) - \mathbb{P}\big(X_t^-(v) = 1\big) \\
&\quad = \frac{1}{2}\mathbb{E}\Big[\tanh\Big(\beta \sum_{w \sim v} X_t^+(w)\Big) - \tanh\Big(\beta \sum_{w \sim v} X_t^-(w)\Big)\Big] \\
&\quad \le \frac{\beta}{2}\mathbb{E}\Big[\sum_{w \sim v} X_t^+(w) - X_t^-(w)\Big] = \beta \sum_{w \sim v} \mathfrak{m}_t(w),
\end{aligned}$$

and so since updates arrive at rate 1, $\frac{d}{dt}\mathfrak{m}_t(v) \le \beta \sum_{w \sim v} \mathfrak{m}_t(w) - \mathfrak{m}_t(v)$. Hence,

$$\begin{aligned}
\frac{d}{dt}\sum_v \mathfrak{m}_t(v)^2 &= 2\sum_v \mathfrak{m}_t(v)\frac{d}{dt}\mathfrak{m}_t(v) \\
&\le -2\sum_v \mathfrak{m}_t(v)^2 + 2\beta \sum_v \mathfrak{m}_t(v) \sum_{w \sim v} \mathfrak{m}_t(w),
\end{aligned}$$

and using $\mathfrak{m}_t(v)\mathfrak{m}_t(w) \le \frac{1}{2}(\mathfrak{m}_t(v)^2 + \mathfrak{m}_t(w)^2)$ it follows that

$$\frac{d}{dt}\sum_v \mathfrak{m}_t(v)^2 \le -2(1-\beta d)\sum_v \mathfrak{m}_t(v)^2,$$

which implies the desired upper bound. $\square$

Recalling that $t_\star = t_\mathfrak{m} + s_\star$, we apply the above claim for $t = t_\mathfrak{m}$ (at which point $\sum_w \mathfrak{m}_{t_\mathfrak{m}}(w)^2 = 1$ by definition) and $s = s_\star$ to find that $\sum_w \mathfrak{m}_{t_\star}(w)^2 \le \exp(-2(1-\beta d)s_\star) \le \exp(-s_\star)$, with the last inequality via $\beta d \le \frac{1}{2}$. By (3.8) (keeping in mind that γ_1 and γ_2 are absolute constants), this implies that $\bar{d}_{\mathrm{TV}}(t_\star) \le \varepsilon$ if we take $s_\star \ge C\log(1/\varepsilon)$ for some absolute constant $C > 0$, as required. $\square$

3.2. *Lower bound.* We now estimate the correlation of two vertices at an arbitrary time.

CLAIM 3.4. *There exist absolute constants $\kappa, \gamma > 0$ such that, for any initial state x_0, if $\beta < \kappa/d$ then*

$$\sum_u \mathrm{Cov}_{x_0}\big(X_t(u), X_t(v)\big) \le \gamma \qquad \textit{for any } t > 0 \textit{ and } v \in V.$$

PROOF. Let X_t' and X_t'' be two independent copies of the dynamics started from x_0. We claim that we may couple X_t with X_t' and X_t'' so that, on the event $\{u \notin \mathcal{C}_v\}$, we have $X_t(u) = X_t'(v)$ and $X_t(v) = X_t''(v)$. Couple the updates of X_t and X_t' inside the history $\mathscr{H}_u$. On the event $\{u \notin \mathcal{C}_v\}$ couple the updates of X_t and X_t'' inside the history $\mathscr{H}_v$. This is permissable since these histories are disjoint and the histories are measurable with respect to the updates inside them. Hence,

$$\begin{aligned}\mathbb{E}\big[X_t(u)X_t(v)\big] &= \mathbb{E}\big[X_t'(u)X_t''(v) + \big(X_t(u)X_t(v) - X_t'(u)X_t''(v)\big)\mathbb{1}_{\{u\in\mathcal{C}_v\}}\big] \\ &\le \mathbb{E}\big[X_t'(u)\big]\mathbb{E}\big[X_t''(v)\big] + 2\mathbb{P}(u \in \mathcal{C}_v).\end{aligned}$$

It follows that $\mathrm{Cov}(X_t(u), X_t(v)) \le 2\mathbb{P}(u \in \mathcal{C}_v) \le 2\mathbb{P}(u \in \hat{\mathcal{C}}_v)$, and so

$$\sum_u \mathrm{Cov}\big(X_t(u), X_t(v)\big) \le 2\mathbb{E}|\hat{\mathcal{C}}_v| \le \gamma,$$

with the final equality thanks to Lemma 2.2. $\square$

We are now ready to prove the lower bound on the mixing time in Theorem 1. To this end, we use the magnetization to generate a distinguishing statistic at time $t_\star^- = t_\mathfrak{m} - s_\star$, given by

$$f(\sigma) = \sum_{v\in V} \mathfrak{m}_{t_\star^-}(v)\sigma(v).$$

Putting $Y = f(X_{t_\star^-}^+)$ for the dynamics started from all-plus and $Y' = f(\sigma)$ with σ drawn from the Ising distribution π, we combine Claim 3.3 with the fact that $\sum_v \mathfrak{m}_{t_\mathfrak{m}}(v)^2 = 1$ (by definition) to get

$$\mathbb{E}Y = \sum_v \mathfrak{m}_{t_\star^-}(v)^2 \ge e^{2(1-\beta d)s_\star} \sum_v \mathfrak{m}_{t_\mathfrak{m}}(v)^2 = e^{2(1-\beta d)s_\star} \ge e^{s_\star} \tag{3.9}$$

(the last inequality using $\beta d \le \frac{1}{2}$), whereas $\mathbb{E}Y' = 0$ (as $\mathbb{E}[\sigma(v)] = 0$ for any v).

For the variance estimate, observe that

$$\begin{aligned}\mathrm{Var}(Y) &= \sum_{u,v} \mathfrak{m}_{t_\star^-}(u)\mathfrak{m}_{t_\star^-}(v)\,\mathrm{Cov}\big(X_{t_\star^-}^+(u), X_{t_\star^-}^+(v)\big) \\ &\le \frac{1}{2}\sum_{u,v}\big(\mathfrak{m}_{t_\star^-}(u)^2 + \mathfrak{m}_{t_\star^-}(v)^2\big)\,\mathrm{Cov}\big(X_{t_\star^-}^+(u), X_{t_\star^-}^+(v)\big) \\ &\le \gamma \sum_v \mathfrak{m}_{t_\star^-}(v)^2 = \gamma\mathbb{E}Y,\end{aligned}$$

using Claim 3.4 for the inequality in the last line. Furthermore, since the law of X_t converges as $t \to \infty$ to that of σ, for any $v \in V$ we have

$$\sum_u \operatorname{Cov}(\sigma(u), \sigma(v)) = \lim_{t\to\infty} \sum_u \operatorname{Cov}(X_t(u), X_t(v)) \le \gamma,$$

and so the same calculation in the above estimate for $\operatorname{Var}(Y)$ shows that

$$\operatorname{Var}(Y') \le \gamma \mathbb{E}Y.$$

Altogether, by Chebyshev's inequality,

$$\mathbb{P}\left(Y \ge \frac{2}{3}\mathbb{E}Y\right) \ge 1 - 9\gamma/\mathbb{E}Y,$$

whereas

$$\mathbb{P}\left(Y' \le \frac{1}{3}\mathbb{E}Y\right) \ge 1 - 9\gamma/\mathbb{E}Y.$$

Recalling (3.9), the expression $9\gamma/\mathbb{E}Y$ can be made less than $\varepsilon/2$ by choosing $s_\star \ge C\log(1/\varepsilon)$ for some absolute constant $C > 0$, we have that

$$\|\mathbb{P}_+[X_{t_\star^-} \in \cdot] - \pi\|_{\mathrm{TV}} \ge \mathbb{P}\left[Y \ge \frac{1}{2}\mathbb{E}Y\right] - \mathbb{P}\left[Y' \ge \frac{1}{2}\mathbb{E}Y\right] \ge 1 - \varepsilon,$$

concluding the proof of the lower bound. $\square$

4. Analysis of percolation clusters.

4.1. *Red clusters*: *Proof of Lemma* 2.1. In estimating Ψ_A, we need to understand the effect of conditioning on the fact that either $A \in \text{RED}$ or $A \subset V_{\text{BLUE}}$, together with the collective history of every $v \notin A$.

For a given subset $S \subset V$, let RED_S^* denote the red clusters that arise when exposing the joint histories of $\mathscr{H}_S$ (as opposed to all the histories $\mathscr{H}_V$). Note the events $\{A \in \text{RED}\}$ and $\{A \in \text{RED}_A^*\} \cap \{\mathscr{H}_A \cap \mathscr{H}_A^- = \varnothing\}$ are identical since the event of A being a red cluster means the histories of A satisfy the properties of being a red cluster and they do not intersect the histories of any vertices in the complement of A. Similarly, define BLUE_S^*, and by the same reasoning $\{A \subset V_{\text{BLUE}}\} = \{A \subset V_{\text{BLUE}_A^*}\} \cap \{\mathscr{H}_A \cap \mathscr{H}_A^- = \varnothing\}$.

Next, given $\mathscr{H}_A^- = \mathcal{X}$, we need to ensure that the component of A avoids $\mathcal{X}$. This can be complicated by the fact that $\mathcal{X}$ could include parts of A for times before $t_\star$ and so the event may be very unlikely. To take care of this possibility, let $s_u = s_u(\mathcal{X}) = \max\{s : (u, s) \in \mathcal{X}\}$ be the latest most time at which $\mathcal{X}$ contains $u \in A$. This implies that any u with $s_u \le t_\star$ must receive an update in the interval $(s_u, t_\star]$ in order to avoid $\mathcal{X}$. We set $A' = \{u \in A : s_u > t_\star - 1\}$ and define $\mathcal{U} = \mathcal{U}(A', \{s_u\}_{u\in A'})$

as the event that each vertex v in A' received an update in $(t_\star - 1, t_\star]$. We find $\mathbb{P}(A \in \text{RED} \mid \mathscr{H}_A^- = \mathcal{X}, \{A \in \text{RED}\} \cup \{A \subset V_{\text{BLUE}}\})$ to be equal to

$$\frac{\mathbb{P}(A \in \text{RED}^*_A, \mathscr{H}_A \cap \mathcal{X} = \varnothing, \mathcal{U} \mid \mathscr{H}_A^- = \mathcal{X})}{\mathbb{P}(\{A \in \text{RED}^*_A\} \cup \{A \subset V_{\text{BLUE}^*_A}\}, \mathscr{H}_A \cap \mathcal{X} = \varnothing, \mathcal{U} \mid \mathscr{H}_A^- = \mathcal{X})}.$$

Since the event $\{A \in \text{RED}^*_A\} \cap \{\mathscr{H}_A \cap \subset \mathcal{X} = \varnothing\} \cap \mathcal{U}$ is measurable in the σ-field of updates within the space-time slab complement of $\mathcal{X}$ (thanks to the specifically defined notion of RED^*_A), it is independent of $\{\mathscr{H}_A^- = \mathcal{X}\}$, and the same applies to the event $(\{A \in \text{RED}^*_A\} \cup \{A \subset V_{\text{BLUE}^*_A}\}) \cap \{\mathscr{H}_A \cap \subset \mathcal{X} = \varnothing\} \cap \mathcal{U}$. Thus, the above expression equals

$$\frac{\mathbb{P}(A \in \text{RED}^*_A, \mathscr{H}_A \cap \mathcal{X} = \varnothing \mid \mathcal{U})}{\mathbb{P}(\{A \in \text{RED}^*_A\} \cup \{A \subset V_{\text{BLUE}^*_A}\}, \mathscr{H}_A \cap \mathcal{X} = \varnothing \mid \mathcal{U})}.$$

The numerator is at most $\mathbb{P}(A \in \text{RED}^*_A \mid \mathcal{U})$. As for the denominator, it is at least the probability that, in the space conditioned on $\mathcal{U}$, every $u \in A$ gets updated in the interval $(s_u \vee t_\star - 1, t_\star]$ and the last such update (i.e., the first we expose when revealing $\mathscr{H}_u$) is oblivious (implying $A \subset V_{\text{BLUE}^*_A}$)—which is $\theta^{|A|}(1 - 1/e)^{|A \setminus A'|}$. As this is at least $e^{-|A|}$ for small enough β [recall the definition of θ in (2.4)],

$$\Psi_A \le e^{|A|} \mathbb{P}(A \in \text{RED}^*_A \mid \mathcal{U}). \tag{4.1}$$

Recall that in order for A to form a complete red cluster, the update histories $\{\mathscr{H}_u : u \in A\}$ must belong to the same connected component of the space-time slab, and moreover, the configuration of A at time $t_\star$ must be a nontrivial function of the initial configuration. Thus, either the histories $\{\mathscr{H}_u : u \in A\}$ coalesce to a single point w at some time $1 \le T < t_\star$—and then the spin there must depend nontrivially on the initial state, that is, $X_T^+(w) \neq X_T^-(w)$—or the histories for all $u \in A$ all join into one cluster along $(0, t_\star]$ and at least one of these survives to time 0. (The same would be true if we did not restrict the coalescence time to be at least 1, yet in this way the conditioning on $\mathcal{U}$, which only pertains to updates along the interval $(t_\star - 1, t_\star]$, does not cause any complications.) For the latter, we denote by $\mathcal{J}(a, b)$ the event that the histories join in the interval (a, b), and for the former we let

$$\tau' = \min\{t \ge 1 : |\mathscr{H}_A(t_\star - t)| = 1\} \wedge t_\star, T = t_\star - \tau',$$

and note that the variable τ' is a stopping time w.r.t. the natural filtration associated with exposing the update histories backward from time $t_\star$; indeed, in contrast to a definition of τ_v analogous to (2.7)—asking for $\{\mathscr{H}_u : u \in \mathcal{C}_v\}$ to coalesce to a single point—here one only requires this for $\{\mathscr{H}_u : u \in A\}$ (whereas $\mathcal{C}_v$ may be affected by the histories along $(0, T]$ as these may admit additional vertices to it). With this notation, we deduce from the above discussion that

$$\mathbb{P}(A \in \text{RED}^*_A \mid \mathcal{U}) \le \mathbb{P}\Big(\bigcup_w \{\mathcal{J}(T, t_\star), w \in \mathscr{H}_A(T), X_T^+(w) \neq X_T^-(w)\} \Bigm| \mathcal{U}\Big).$$

[If $T = 0$ and $A \in \mathrm{RED}_A^*$ then $\mathscr{H}_A(0) \neq \varnothing$, whence $X_0^+(w) \neq X_0^-(w)$ trivially holds for any $w \in \mathscr{H}_A(0)$.] By conditioning on T as well as on $\mathscr{H}_A(T, t_\star)$, the first two events on the right-hand side become measurable, while the event $X_T^+(w) \neq X_T^-(w)$ only depends on the histories along $(0, T]$ and satisfies

$$\mathbb{P}\big(X_T^+(w) \neq X_T^-(w) \mid T, \mathscr{H}_A(T, t_\star)\big) = \mathfrak{m}_T(w) \le e^{t_\star - T}\mathfrak{m}_{t_\star}(w),$$

where the final inequality used the fact, mentioned in the proof of Claim 3.3, that $\mathfrak{m}_{t+s}(w) \ge e^{-s}\mathfrak{m}_t(w)$ for any $s, t > 0$ and w, as the probability of no updates to w along the interval $(t, t+s)$ (maintaining the magnetization without a change) is e^{-s}. Now, averaging over this conditional space yields

$$\begin{aligned}\mathbb{P}\big(A \in \mathrm{RED}_A^* \mid \mathcal{U}\big) &\le \mathbb{E}\bigg[\sum_w \mathbb{1}_{\{\mathcal{J}(T,t_\star)\}}\mathbb{1}_{\{w\in\mathscr{H}_A(T)\}}e^{t_\star - T}\mathfrak{m}_{t_\star}(w) \,\bigg|\, \mathcal{U}\bigg] \\ &\le \mathbb{E}\bigg[\sum_w \mathbb{1}_{\{A\subset\mathcal{C}_v\}}\mathbb{1}_{\{w\in\mathscr{H}_A(t_\star-\tau',t_\star)\}}e^{\tau'}\mathfrak{m}_{t_\star}(w) \,\bigg|\, \mathcal{U}\bigg],\end{aligned}$$

where we increased the event $\mathcal{J}(T, t_\star)$ (the joining of $\mathscr{H}_A$ along $(T, t_\star]$) into $A \subset \mathcal{C}_v$ (valid for any $v \in A$) as well as the event $\{w \in \mathscr{H}_A(T)\}$ into $\{w \in \mathscr{H}_A(T, t_\star)\}$, and finally plugged in that $T = t_\star - \tau'$. Since by definition $\tau' \le \tau_v = \min\{t \ge 1 : |\mathscr{H}_{\mathcal{C}_v}(t_\star - t)| = 1\} \wedge t_\star$ on the event $A \subset \mathcal{C}_v$, we conclude that

$$\mathbb{P}\big(A \in \mathrm{RED}_A^* \mid \mathcal{U}\big) \le \mathbb{E}\bigg[\sum_w \mathbb{1}_{\{A\subset\mathcal{C}_v\}}\mathbb{1}_{\{w\in\mathscr{H}_A(t_\star-\tau_v,t_\star)\}}e^{\tau_v}\mathfrak{m}_{t_\star}(w) \,\bigg|\, \mathcal{U}\bigg]. \tag{4.2}$$

The final step is to eliminate the conditioning on $\mathcal{U}$ using the modified update history $\hat{\mathscr{H}}$, which we recall does not remove vertices from the history along the unit interval $(t_\star - 1, t_\star]$ and grants each vertex an automatic update at time $t_\star$. As such, $\mathscr{H}_u(t) \subset \hat{\mathscr{H}}_u(t)$ for any vertex u and time t.

We claim that each of the terms in the right-hand side of (4.2) is increasing in the percolation space-time slab (i.e., they can only increase when adding connections to the update histories). Indeed, this trivially holds for $\{A \subset \mathcal{C}_v\}$; the variable τ_v is increasing as it may take only longer for $\mathcal{C}_v$ to coalesce to a single point; finally, as the interval $(t_\star - \tau_v, t_\star]$ does not decrease and neither does $\mathscr{H}_A$ along it, the event $\{w \in \mathscr{H}_A(t_\star - \tau_v, t_\star)\}$ is also increasing.

Therefore, if we do not remove vertices from the update history along $(t_\star - 1, t_\star]$ then the right-hand side of (4.2) could only increase. Further, observe that, as long as no vertices are removed from the history along that unit interval, the connected components of the update history at time $t_\star - 1$ remain exactly the same were we to modify the update times of any vertex there, while keeping them within that unit interval. In particular, should a vertex at all be updated in that period, we can move its latest update time to $t_\star$.

In this version of the update history (retaining all vertices in the given unit interval, and letting the latest most update, if it is in that interval, be performed at time $t_\star$), the effect of conditioning on $\mathcal{U}$ in that every $u \in A'$ receives an update at time $t_\star$. The fact that $\mathrm{Po}(\lambda \mid \cdot \geq 1) \preccurlyeq \mathrm{Po}(\lambda) + 1$ for any $\lambda > 0$ [as the ratio $\mathbb{P}(\mathrm{Po}(\lambda) = k)/\mathbb{P}(\mathrm{Po}(\lambda) > k)$ is monotone increasing in k] now implies [taking $\lambda \in (0, 1)$] that the number of updates that any $u \in A'$ receives along $(t_\star - 1, t_\star]$ conditioned on $\mathcal{U}$ as part of $\mathscr{H}$ is stochastically dominated by the corresponding number of updates as part of $\hat{\mathscr{H}}$.

Altogether we conclude that the right-hand side of (4.2) can be increased to yield

$$\mathbb{P}\big(A \in \mathrm{RED}_A^* \mid \mathcal{U}\big) \leq \mathbb{E}\Big[\sum_w \mathbb{1}_{\{A \subset \hat{\mathcal{C}}_v\}} \mathbb{1}_{\{w \in \hat{\mathscr{H}}_A(t_\star - \hat{\tau}_v, t_\star)\}} e^{\hat{\tau}_v} \mathfrak{m}_{t_\star}(w)\Big],$$

and combining this with (4.1) completes the proof. $\square$

4.2. *Discrete Fourier expansion for the update rules.* The following lemma, which constructs the modified update rules Φ_A (as described in Section 2), will play a key role in the proof of Lemma 2.2.

LEMMA 4.1. *For every $\varepsilon > 0$, there exists some $\kappa > 0$ such that the following holds provided $\beta d < \kappa$. For any $r \leq d$, there are nonnegative reals $\{p_{k,r} : k = 0, \dots, r\}$ satisfying*

$$p_{0,r} \geq 1 - \varepsilon, \quad \sum_k \binom{r}{k} p_{k,r} = 1, \quad \textit{and} \quad \binom{r}{k} p_{k,r} \leq D_0(2\beta r)^k \quad \textit{for all } k, \tag{4.3}$$

where D_0 is an absolute constant, such that the Glauber dynamics can be coupled to an update function Φ that selects a subset $A \subset [r]$ of the neighbors of a degree-r vertex with probability $p_{|A|,r}$ and applies to it a symmetric monotone boolean function Φ_A [i.e., $\Phi_A(-x) = -\Phi_A(x)$ and $\Phi_A(x)$ is increasing in x].

PROOF. Setting

$$f(x) = \frac{1}{2}\big(\tanh(x) + 1\big) = \frac{e^x}{e^x + e^{-x}}$$

we have that the Glauber dynamics update function at a given site with neighbors $\sigma_1, \dots, \sigma_r$ assigns it a new spin of 1 with probability $f(\beta \sum_{i=1}^r \sigma_i)$. Writing the Taylor series expansion $f(x) = \sum_{\ell=0}^\infty B_\ell x^\ell$, with

$$B_\ell = \frac{1}{\ell!} \frac{d^\ell f}{dx^\ell}(0),$$

and so, bearing in mind that $\tanh(z)$ has no singularities in the open disc of radius $\pi/2$ around 0 in $\mathbb{C}$ and thus $\sum B_\ell$ converges absolutely,

$$B_0 = B_1 = 1/2 \quad \text{and} \quad \sum |B_\ell| = B \qquad \text{for some absolute constant } B > 0.$$

Next, since $\sigma_i \in \{\pm 1\}$ the power series is multi-linear in σ_i, whence we can write

$$\Big(\beta \sum \sigma_i\Big)^{\ell} = \sum_{\substack{A \subset [r] \\ |A| \le \ell}} C_{\ell,A} \prod_{i \in A} \sigma_i = \sum_{k=1}^{\ell \wedge r} C_{\ell,k} \sum_{|A|=k} \prod_{i \in A} \sigma_i,$$

where we used that the nonnegative coefficient $C_{\ell,A}$ depends by symmetry on $|A|$ rather than A itself, thus we can write $C_{\ell,k}$ for $|A| = k$. (Note that for $\ell = 1$ we have $C_{1,1} = \beta$.)

Now, for any particular $k \le \ell \wedge r$, we can put $\sigma_1 = \cdots = \sigma_r = 1$ to find that

$$(\beta r)^{\ell} = \sum_{i=0}^{\ell \wedge r} \sum_{|A|=i} C_{\ell,i} \ge \sum_{|A|=k} C_{\ell,k} = \binom{r}{k} C_{\ell,k},$$

and so

$$0 \le C_{\ell,k} \le \frac{(\beta r)^{\ell}}{\binom{r}{k}}. \tag{4.4}$$

Therefore, letting

$$C_k = \sum_{\ell=k}^{\infty} C_{\ell,k} B_{\ell} \qquad \text{for } k \ge 1$$

and recalling that $\sum |B_\ell| = B$, we see that

$$|C_k| \le B \sum_{\ell \ge k} \frac{(\beta r)^{\ell}}{\binom{r}{k}} \le 2B \frac{(\beta r)^{k}}{\binom{r}{k}}, \tag{4.5}$$

with the last inequality valid as long as $\beta r \le 1/2$.

We now define $p_{k,r}$ as follows:

$$p_{k,r} = \begin{cases} 2|C_k|(k+1) & k \ge 2, \\ 2\Big(C_1 - \sum_{\substack{A' \ni 1 \\ |A'| \ge 2}} |C_{|A'|}|\Big) & k = 1, \\ 1 - \sum_{k \ge 1} \binom{r}{k} p_{k,r} & k = 0. \end{cases} \tag{4.6}$$

Our first step in verifying that this definition satisfies (4.6) is to show that $0 < p_{1,r} < 1$. For the upper bound, using (4.5) we have $p_{1,r} \le 2|C_1| \le 4B\beta < 1$ for β small enough. For the lower bound, observe that since $B_1 = 1/2$, $C_{1,1} = \beta$ and $C_{\ell,1} \le (\beta r)^{\ell}/r$ using (4.4),

$$C_1 \ge \frac{\beta}{2} - \sum_{\ell=2}^{\infty} C_{\ell,1} |B_\ell| \ge \frac{\beta}{2} - \frac{B}{r} \sum_{\ell \ge 2} (\beta r)^{\ell} \ge \beta\Big(\frac{1}{2} - 2\beta r B\Big) > \beta/4. \tag{4.7}$$

as long as $\beta < 1/(4rB)$. On the other hand, again appealing to (4.5),

$$\begin{aligned}\sum_{\substack{A'\ni 1\\|A'|\geq 2}} |C_{|A'|}| &= \sum_{k=2}^{r}\binom{r-1}{k-1}|C_k| \leq 2B\sum_{k=2}^{r}\frac{k}{r}(\beta r)^k\\ &= 2B\beta\sum_{k=2}^{r}k(\beta r)^{k-1} \leq \beta/8\end{aligned}\tag{4.8}$$

provided βr is sufficiently small. Combining the last two displays yields $p_{1,r} \geq \beta/8$.

Next, we wish to verify that $\binom{r}{k}p_{k,r} \leq D_0(2\beta r)^k$ for some absolute constant D_0 and all k. Let $D_0 = 4B$ and note that for $k = 0$ the sought inequality is trivial since $D_0 > 1$ (recall $B \geq B_0 + B_1 = 1$) whereas $p_{0,r} < 1$ (we have shown that $p_{1,r} > 0$ and clearly $p_{k,r} \geq 0$ for all $k \geq 2$). For $k = 1$, we again recall from (4.5) that $rp_{1,r} \leq 2r|C_1| \leq 4\beta rB < D_0(2\beta r)$, and similarly, for $k \geq 2$ we have

$$\binom{r}{k}p_{k,r} = 2\binom{r}{k}|C_k|(k+1) \leq 4B(k+1)(\beta r)^k \leq 4B(2\beta r)^k = D_0(2\beta r)^k.$$

For any sufficiently small βr, this of course also shows that $p_{k,r} \leq 1$ for all k, as well as the final fact that $p_{0,d} \geq 1 - \varepsilon$ since

$$\sum_{k\geq 1}\binom{r}{k}p_{k,r} \leq D_0\sum_{k\geq 1}(2\beta r)^k < 4\beta r D_0 < \varepsilon\tag{4.9}$$

for a small enough βr.

Having established that desired properties for $\{p_{k,r} : 0 \leq k \leq r\}$, define the new update function Φ which will examine a random subset A of the r neighbors of a vertex, selected with probability $p_{|A|,r}$ (giving a proper distribution over the subsets of $[r]$ since $\sum_k \binom{r}{k}p_{k,r} = 1$ as shown above), then apply the following function Φ_A to determine the probability of a plus update given $\sigma_A = \{\sigma_i : i \in A\}$:

$$\Phi_A(\sigma_A) = \begin{cases}\dfrac{1}{2} & A = \varnothing,\\[1ex] \dfrac{1}{2} + \dfrac{1}{2}\sigma_i & A = \{i\},\\[1ex] \dfrac{1}{2} + \dfrac{1}{2(|A|+1)}\Big[\sum_{i\in A}\sigma_i + \operatorname{sign}(C_{|A|})\prod_{i\in A}\sigma_i\Big] & |A| \geq 2.\end{cases}\tag{4.10}$$

In order to establish that Φ can be coupled to the Glauber dynamics, we need to show that $f(\beta\sum_{i=1}^{r}\sigma_i)$ identifies with $\mathbb{E}[\Phi(\sigma_1,\ldots,\sigma_r)]$ over all inputs $\{\sigma_i\}$. Since $B_0 = 1/2$, we must show that $\mathbb{E}[\Phi] - 1/2$ is equal to $\sum_{\ell=1}^{\infty} B_\ell(\beta\sum\sigma_i)^\ell$.

Indeed,

$$\mathbb{E}[\Phi] - \frac{1}{2} = \sum_i \sigma_i \Big(C_1 - \sum_{\substack{A' \ni i \\ |A'| \geq 2}} C_{|A'|} \Big) + \sum_{|A| \geq 2} |C_{|A|}| \Big(\sum_{i \in A} \sigma_i + \operatorname{sign}(C_{|A|}) \prod_{i \in A} \sigma_i \Big)$$

$$= \sum_{|A| \geq 1} C_{|A|} \prod_{i \in A} \sigma_i = \sum_{\ell=1}^{\infty} \sum_{k=1}^{\ell \wedge r} \sum_{|A|=k} C_{\ell,k} B_\ell \prod_{i \in A} \sigma_i = \sum_{\ell=1}^{\infty} B_\ell \Big(\beta \sum \sigma_i \Big)^\ell,$$

with the last two equalities following from the definition of C_k and $C_{\ell,k}$. This completes the proof. $\square$

4.3. *Exponential decay of cluster sizes*: *Proof of Lemma* 2.2. Using the update rule from Lemma 4.1, the probability that an update of a vertex v of degree $r \leq d$ will examine precisely k of its neighbors is

$$\binom{r}{k} p_{k,r} \leq D_0 (2\beta r)^k \leq D_0 (2\beta d)^k,$$

with the inequality thanks to (4.3). The probability that a given neighbor of v, with degree some $r' \leq d$, receives an update in which it examines both v and $k-1$ additional neighbors is at most

$$\max_{r' \leq d} p_{k,r'} \binom{r'-1}{k-1} \leq \max_{r' \leq d} \frac{k}{r'} D_0 \big(2\beta r'\big)^k = \frac{1}{d} D_0 (3\beta d)^k,$$

using that $x^{1/x} \leq e^{1/e} < 3/2$ for all $x \geq 2$. Hence, the rate at which the history of the vertex v expands to k additional vertices along the time interval $(0, t_\star)$ is at most

$$D_0(1 + r/d)(3\beta d)^k \leq 2 D_0 (3\beta d)^k.$$

By the same reasoning, the extra update at time $t_\star$ that is applied to v in $\hat{\mathscr{H}}$ connects it to k of its neighbors ($k = 0, \dots, r$) with probability at most $D_0(2\beta d)^k$, while each of its r neighbors contributes at most k new points with probability at most $D_0(3\beta d)^k / d$.

We now develop the cluster of the vertex (w_0, t_0) in the space-time slab by exploring the branch at w_0, both forward and backward in time, examining which connections it has to new vertices—either through its own updates or through those which examine it—until it terminates via oblivious updates in both directions. We then repeat this process with one of the points discovered in the exploration process (arbitrary chosen), until all such points are exhausted and the cluster is completely revealed.

Let Y_m denote the number of vertices explored in this way after iteration m [i.e., Y_1 is the number of vertices discovered via the branch incident to (w_0, t_0), etc.], and let Z_m be the total length of edges in the time dimension [i.e., (z, a), (z, b) for

$z \in V$ and $0 \le a < b \le t_\star]$ explored by then. We can stochastically dominate these by a process $(\bar{Y}_m, \bar{Z}_m) \succeq (Y_m, Z_m)$ given as follows.

First, for the length variable, we apply Lemma 4.1 with $\varepsilon = (1-\eta)/4$, and put

$$\bar{Z}_0 = 0,$$
$$\bar{Z}_m = \bar{Z}_{m-1} + W_m \qquad \text{where } W_m \sim 1 + \Gamma(2, 1-\varepsilon),$$

with the gamma variable $\Gamma(2, 1-\varepsilon)$ measuring the time until the explored branch terminates (in both ends) using the key estimate $p_{0,r} \ge 1 - \varepsilon$ from Lemma 4.1, translated by 1 to account for the unit interval $(t_\star - 1, t_\star]$ in which vertices are not removed from $\hat{\mathscr{H}}$.

For the vertex count variable, with the above discussion in mind, observe that conditioned on W_m the number of new vertices exposed along the new branch is dominated by $\sum_{k=1}^d V_m^{(k)}$, in which

$$V_m^{(k)} \sim k\,\mathrm{Po}\big(2D_0(3\beta d)^k W_m\big) \qquad (k = 1, \dots, d)$$

are mutually independent, while the extra update at time $t_\star$ (should the branch extend to that time) introduces at most $\sum_{k=0}^d \hat{V}_m^{(k)}$ additional vertices, where all $V_m^{(k)}$ and $\hat{V}_m^{(k)}$ are independent, given by

$$\mathbb{P}\big(\hat{V}_m^{(0)} = j\big) \le D_0(2\beta d)^j, \qquad \mathbb{P}\big(\hat{V}_m^{(k)} = j\big) \le D_0(3\beta d)^j/d \qquad (k = 1, \dots, d).$$

Therefore, with this notation, we write

$$\bar{Y}_0 = 1,$$
$$\bar{Y}_m = \bar{Y}_{m-1} + U_m \qquad \text{where } U_m = \sum_{k=1}^d V_m^{(k)} + \sum_{k=0}^d \hat{V}_m^{(k)}.$$

Letting $\tau \ge 1$ be the iteration after which the exploration process exhausts all new vertices [so $\tau = 1$ iff both ends of the branch of (v_0, t_0) terminated before introducing any new vertices to the cluster], we wish to show that

$$\mathbb{E}\big[\exp(\eta \bar{Z}_\tau + \lambda \bar{Y}_\tau)\big] \le \gamma \tag{4.11}$$

for $\gamma(\lambda, \eta) < \infty$. We may assume without loss of generality—recalling that $\varepsilon = (1-\eta)/4 \le \frac{1}{4}$—that

$$\lambda \ge 4\log(1/\varepsilon), \tag{4.12}$$

as the left-hand side of (4.11) is monotone increasing in λ. Observe that as long as $3\beta d e^\lambda < 1/2$ we have

$$\mathbb{E}\big[\exp\big(\lambda \hat{V}_m^{(0)}\big)\big] \le 1 + D_0 \sum_{k \ge 1} \big(2\beta d e^\lambda\big)^k \le 1 + 4D_0 \beta d e^\lambda$$

as well as

$$\prod_{k=1}^{d}\mathbb{E}\big[\exp\big(\lambda \hat{V}_m^{(k)}\big)\big] \le 1 + 2D_0\sum_{k\ge 1}\big(3\beta d e^{\lambda}\big)^k \le 1 + 12 D_0 \beta d e^{\lambda},$$

and similarly,

$$\begin{aligned}\prod_{k=1}^{d}\mathbb{E}\big[\exp\big(\lambda V_m^{(k)}\big)\mid W_m\big] &= \exp\Bigg[2W_m D_0\sum_{k=1}^{d}\big(e^{\lambda k}-1\big)(3\beta d)^k\Bigg]\\ &\le \exp\big[12 W_m D_0 \beta d e^{\lambda}\big].\end{aligned}$$

We can further assume that

$$h(\lambda) := D_0\beta d e^{\lambda} \quad \text{satisfies} \quad 12h(\lambda) < (1-\eta)/2 = 1 - 2\varepsilon - \eta,$$

achievable by letting βd be sufficiently small. With this notation,

$$\mathbb{E}\big[e^{\lambda U_m + \eta W_m} \mid W_m\big] \le e^{(12h(\lambda)+\eta)W_m + 16h(\lambda)},$$

and upon taking expectation over W_m, having $12h(\lambda) + \eta < 1 - 2\varepsilon$ implies that the moment-generating function of the gamma distribution will only contribute a polynomial factor, giving that

$$\mathbb{E}\big[e^{\lambda U_m + \eta W_m}\big] \le e^{28h(\lambda)+\eta}\bigg(1 - \frac{12h(\lambda)+\eta}{1-\varepsilon}\bigg)^{-2} \le \varepsilon^{-2} e^{28h(\lambda)}. \tag{4.13}$$

Combining this with our definition of $\bar{Y}_m = 1 + \sum_{i=1}^{m} U_i$ and $\bar{Z}_m = \sum_{i=1}^{m} W_i$, we find that

$$\begin{aligned}\mathbb{E}\big[e^{\lambda \bar{Y}_\tau + \eta \bar{Z}_\tau}\big] &= \mathbb{E}\Bigg[\sum_{m=1}^{\infty} e^{-\lambda m + 2\lambda \bar{Y}_m + \eta \bar{Z}_m}\mathbb{1}_{\{\tau = m\}}\Bigg] \le \sum_{m=1}^{\infty} e^{-\lambda m}\mathbb{E}\big[e^{2\lambda \bar{Y}_m + \eta \bar{Z}_m}\big]\\ &= \sum_{m=1}^{\infty} e^{\lambda(2-m)}\big(\mathbb{E}\big[e^{2\lambda U_1 + \eta W_1}\big]\big)^m,\end{aligned}$$

which, recalling (4.13) and plugging in the expression for $h(2\lambda)$, is at most

$$\begin{aligned}&e^{2\lambda}\sum_{m=1}^{\infty}\exp\bigg[m\bigg(-\lambda + 2\log\bigg(\frac{4}{1-\eta}\bigg) + 28h(2\lambda)\bigg)\bigg]\\ &\qquad \le e^{2\lambda}\sum_{m=1}^{\infty}\exp\big[m\big(-\lambda/2 + 28h(2\lambda)\big)\big] = \gamma < \infty\end{aligned}$$

using (4.12) for the first inequality and, say, that $28h(2\lambda) \le \lambda/3$ (achieved by taking βd small enough) for the second one. [Note that $\gamma = \gamma(\lambda, \eta)$, as the assumption (4.12) introduces a dependence on η.] This establishes (4.11) and thereby completes the proof. $\square$

5. The effect of initial conditions on mixing. In this section, we consider random initial conditions (both quenched and annealed), and prove Theorem 2. The first observation is that, thanks to Theorem 1, the worst-case mixing time satisfies

$$t_{\mathrm{MIX}}(\alpha) = t_{\mathfrak{m}} + O(1) \qquad \text{for any fixed } 0 < \alpha < 1,$$

with $t_{\mathfrak{m}}$ as defined in (1.1), and moreover, the same holds for $t_{\mathrm{MIX}}^{(+)}(\alpha)$, the mixing time started from all-plus. By Claim 3.3, we have $\frac{1}{2}\log n \le t_{\mathfrak{m}} \le (\frac{1}{2} + \varepsilon_\beta)\log n$ with $\varepsilon_\beta = \beta d/(2 - 2\beta d)$ vanishing as $\beta \downarrow 0$. Thus, we may prove the bounds on the annealed/quenched mixing times when replacing $t_{\mathfrak{m}}$ by $\frac{1}{2}\log n$.

5.1. *Annealed analysis.* As mentioned in the Introduction, rather than comparing two worst case boundary conditions we will compare a random one directly with the stationary distribution: By considering updates in the range $t \in (-\infty, t_{\mathfrak{m}}]$ we can use the coupling from the past construction to generate a coupling with the stationary distribution. Let X_t denote the process started from uniform initial conditions at time 0 and let Y_t be the process generated by coupling from the past.

The information percolation clusters of V will now be defined as the connected components of the graph on the vertex set V where (u, v) is an edge iff $\mathscr{H}_u(t) \cap \mathscr{H}_v(t) \neq \emptyset$ for some $-\infty < t \le t_{\mathfrak{m}}$ (in contrast to the previous definition where we had $0 < t \le t_{\mathfrak{m}}$). The notion of being a red cluster is redefined to be any $\mathcal{C}_v$ such that $|\bigcup_{u \in \mathcal{C}_v} \mathscr{H}_u(t')| \ge 2$ for all $0 \le t' \le t_{\mathfrak{m}}$. Blue clusters will be defined as before and green clusters will again be the remaining clusters. We claim that we can couple the spins at time $t_{\mathfrak{m}}$ of all nonred clusters. Indeed, if a cluster $\mathcal{C}_v$ is not red, then there is some time $t' > 0$ such that $|\bigcup_{u \in \mathcal{C}_v} \mathscr{H}_u(t')| = 1$. Call this vertex w. By symmetry, both $X_{t'}(w)$ and $Y_{t'}(w)$ are equally likely to be plus or minus and so we may couple them to be equal independently of the spins of the other clusters. We may then also couple the spins in that cluster to be the same in both X_t and Y_t to be equal for all $t > t'$. Thus, the configurations will agree outside of the red clusters.

Let $\mathfrak{W}(A)$ denote the size of the smallest connected set of vertices (animal) containing A. In a graph of maximum degree d, the number of trees of size k containing the vertex v is bounded above by $(ed)^k$, and hence the number of animals A containing a specified vertex with $\mathfrak{W}(A) = k$ is at most $(ed)^k$.

LEMMA 5.1. *For any $d, C, \varepsilon > 0$ there exists $\beta_0 > 0$ such that the following holds for large enough n. If $0 < \beta < \beta_0$ and $t_\star = (\frac{1}{4} + \varepsilon)\log n$ then for any A,*

$$\sup_{\mathscr{H}_A^-} \mathbb{P}\big(A \in \textsc{Red} \mid \mathscr{H}_A^-, \{A \in \textsc{Red}\} \cup \{A \subset V_{\textsc{Blue}}\}\big) \le \frac{1}{\sqrt{n}\log n} e^{-C\mathfrak{W}(A)}.$$

PROOF. The line of reasoning in the proof of Lemma 2.1 establishing equation (4.1) remains unchanged with the modified definition of red clusters so again we have

$$\mathbb{P}\big(A \in \text{RED} \mid \mathscr{H}_A^-, \{A \in \text{RED}\} \cup \{A \subset V_{\text{BLUE}}\}\big) \le 2^{|A|}\mathbb{P}(A \in \text{RED} \mid \mathcal{U}),$$

where $\mathcal{U}$ is unchanged from the proof of Lemma 2.1. Then for any $v \in A$,

$$\mathbb{P}(A \in \text{RED} \mid \mathcal{U}) \le \mathbb{P}\big(|\mathscr{H}_{\mathcal{C}_v}| \ge \mathfrak{W}(A), \mathfrak{L}(\mathcal{X}_{v,0}) \ge 2t_\star \mid \mathcal{U}\big),$$

since the total length of a red cluster must be at least $2t_\star$ and it must contain at least $\mathfrak{W}(A)$ vertices. Both $|\mathscr{H}_{\mathcal{C}_v}|$ and $\mathfrak{L}(\mathcal{X}_{v,0})$ are increasing in the component sizes, and so, by the same monotonicity argument as Lemma 2.1, we have that

$$\begin{aligned}&\mathbb{P}\big(A \in \text{RED} \mid \mathscr{H}_A^-, \{A \in \text{RED}\} \cup \{A \subset V_{\text{BLUE}}\}\big)\\ &\qquad \le 2^{|A|}\mathbb{P}\big(|\hat{\mathscr{H}}_{\hat{\mathcal{C}}_v}| \ge \mathfrak{W}(A), \mathfrak{L}(\hat{\mathcal{X}}_{v,0}) \ge 2t_\star\big).\end{aligned}$$

Taking $\lambda = \log 2 + C$ and $\frac{1}{4}(\frac{1}{4}+\varepsilon)^{-1} < \eta < 1$ in Lemma 2.2 then shows that, for β_0 small enough,

$$\begin{aligned}&\mathbb{P}\big(|\hat{\mathscr{H}}_{\hat{\mathcal{C}}_v}| \ge \mathfrak{W}(A), \mathfrak{L}(\hat{\mathcal{X}}_{v,0}) \ge 2t_\star\big)\\ &\qquad \le \frac{\mathbb{E}[\exp(\eta\mathfrak{L}(\hat{\mathcal{X}}_{v,0}) + \lambda|\hat{\mathscr{H}}_{\hat{\mathcal{C}}_v}|)]}{\exp(2\eta t_\star + \lambda\mathfrak{W}(A))}\\ &\qquad \le \gamma \exp\left(-2\eta\left(\frac{1}{4}+\varepsilon\right)\log n - (\log 2 + C)\mathfrak{W}(A)\right)\\ &\qquad \le \frac{1}{\sqrt{n}\log n}2^{-\mathfrak{W}(A)}e^{-C\mathfrak{W}(A)}\end{aligned}$$

(with room, as we could have replaced the $\sqrt{n}\log n$ by some $n^{1/2+\varepsilon'}$), which completes the proof. □

We now establish an upper bound on $t_{\text{MIX}}^{(\text{U})}$, the mixing time starting from the uniform distribution.

PROPOSITION 5.2. *For any $d, \varepsilon > 0$, there exists $\beta_0 > 0$ such that the following holds. If $0 < \beta < \beta_0$ and $t_\star = (\frac{1}{4}+\varepsilon)\log n$, then $\|\mathbb{P}(X_{t_\star} \in \cdot) - \pi\|_{\text{TV}} \to 0$ as $n \to \infty$.*

PROOF. Having coupled X_t and Y_t as described above we have that

$$\begin{aligned}&\big\|\mathbb{P}(X_{t_\star} \in \cdot) - \mathbb{P}(Y_{t_\star} \in \cdot)\big\|_{\text{TV}}\\ &\qquad \le \mathbb{E}\big[\big\|\mathbb{P}\big(X_{t_\star}(V \setminus V_{\text{GREEN}}) \in \cdot \mid \mathscr{H}_{\text{GREEN}}\big) - \nu_{V\setminus V_{\text{GREEN}}}\big\|_{\text{TV}}\big]\\ &\qquad\quad + \mathbb{E}\big[\big\|\mathbb{P}\big(Y_{t_\star}(V \setminus V_{\text{GREEN}}) \in \cdot \mid \mathscr{H}_{\text{GREEN}}\big) - \nu_{V\setminus V_{\text{GREEN}}}\big\|_{\text{TV}}\big],\end{aligned}$$

where ν_A is the uniform measure on the configurations on A. Similar to the argument used to derive equation (3.2), we find that

$$\|\mathbb{P}(X_{t_\star} \in \cdot) - \mathbb{P}(Y_{t_\star} \in \cdot)\|_{\mathrm{TV}} \le \Bigl(\sup_{\mathscr{H}_{\mathrm{GREEN}}} \mathbb{E}[2^{|V_{\mathrm{RED}} \cap V_{\mathrm{RED}'}|} \mid \mathscr{H}_{\mathrm{GREEN}}] - 1\Bigr)^{1/2}.$$

With the same coupling as in the proof of Theorem 1, analogously to equation (3.4), we have

$$\|\mathbb{P}(X_{t_\star} \in \cdot) - \mathbb{P}(Y_{t_\star} \in \cdot)\|_{\mathrm{TV}}^2 \le 2\sum_v \Bigl(\sum_{A \ni v} 2^{|A|}\Psi_A\Bigr)^2. \tag{5.1}$$

Applying Lemma 5.1 with $C = \lceil \log(4ed)\rceil$ [while recalling that $\#\{A \ni v : \mathfrak{W}(A) = k\} \le (ed)^k$], we get

$$\sum_{A \ni v} 2^{|A|}\Psi_A \le \sum_k \sum_{\substack{A:\mathfrak{W}(A)=k \\ v \in A}} \frac{2^k e^{-Ck}}{\sqrt{n}\log n} \le \sum_k \frac{(2ed)^k e^{-Ck}}{\sqrt{n}\log n} \le \frac{1}{\sqrt{n}\log n},$$

provided that $\beta > \beta_0$ with β_0 from that lemma. It follows that

$$\|\mathbb{P}(X_{t_\star} \in \cdot) - \mathbb{P}(Y_{t_\star} \in \cdot)\|_{\mathrm{TV}} \le O(\log^{-2} n),$$

and in particular $\|\mathbb{P}(X_{t_\star} \in \cdot) - \pi\|_{\mathrm{TV}} = o(1)$, as required. $\square$

REMARK 5.3. In the above proof, one could instead carry the analysis as in the proof of Theorem 1 (partitioning the event in Lemma 2.1 according to the events $\{|\hat{\mathcal{C}}_v| = k\}$ when estimating the sum over $v \ni A$), that way replacing the factor of $(ed)^k$ lattice animals by 2^k subsets of $\hat{\mathcal{C}}_v$. Consequently, the statement of Proposition 5.2 remains valid for any $\beta < c_0/d$ where c_0 depends on ε but not on d.

5.2. *Quenched analysis.* Here, we show that the mixing time from a typical random initial state is at most a factor of $1 + \varepsilon_\beta$ faster than that from the worst starting state. As before, let X_t be started from a uniformly chosen initial state X_0 and let Y_t be started from the stationary distribution π.

PROPOSITION 5.4. *Let* $t_\star^- = \frac{1}{2}\log n - w_n$ *for some* $w_n \uparrow \infty$. *Then* $\|\mathbb{P}_{X_0}(X_{t_\star^-} \in \cdot) - \pi\|_{\mathrm{TV}} \xrightarrow{p} 1$ *as* $n \to \infty$.

PROOF. Note that by the monotonicity of the update rules, for any update history of u, the spin at u is a monotone function of X_0. With probability e^{-t}, the vertex u is never updated in which case $X_t(u) = X_0(u)$. Since by symmetry $\mathbb{E}[X_t(u)] = 0$, it follows that

$$\mathbb{E}\bigl[X_t(u) \mid X_0(u) = +1\bigr] \ge e^{-t}, \qquad \mathbb{E}\bigl[X_t(u) \mid X_0(u) = -1\bigr] \le -e^{-t}.$$

Thus we have that $\mathbb{E}[X_0(u)X_t(u)] \geq e^{-t}$ and so

$$\mathbb{E}\bigg[\sum_u X_0(u)X_{t_\star^-}(u)\bigg] \geq ne^{-t_\star^-} = \sqrt{n}e^{w_n}.$$

Let $\mathcal{E}_{u,v}$ be the event that $u \in \mathcal{C}_v$ or $v \in \mathscr{H}_u(0)$ or $u \in \mathscr{H}_v(0)$ for the history developed from time $t_\star^-$. Similar to Claim 3.4, let X_t' and X_t'' be two independent copies of the dynamics. By exploring the histories, we may couple X_t with X_t' and X_t'' so that, on the event $\mathcal{E}_{u,v}^c$, the history of v in X_t is equal to the history of v in X_t' and the history of u in X_t is equal to the history of u in X_t''. Hence,

$$\mathbb{E}\big[X_0(u)X_{t_\star^-}(u)X_0(v)X_{t_\star^-}(v)\big] \leq \mathbb{E}\big[X_0'(u)X_{t_\star^-}'(u)\big]\mathbb{E}\big[X_0''(v)X_{t_\star^-}''(v)\big] + 2\mathbb{P}(\mathcal{E}_{u,v}),$$

yielding $\mathrm{Cov}(X_0(u)X_{t_\star^-}(u), X_0(v)X_{t_\star^-}(v)) \leq 2\mathbb{P}(\mathcal{E}_{u,v})$. By Lemma 2.2,

$$\sum_u \mathrm{Cov}\big(X_0(u)X_{t_\star^-}(u), X_0(v)X_{t_\star^-}(v)\big) \leq c_1.$$

and so

$$\mathrm{Var}\bigg(\sum_u X_0(u)X_{t_\star^-}(u)\bigg) \leq c_1 n.$$

Thus, by Chebyshev's inequality we infer that

$$\mathbb{P}\bigg(\sum_u X_0(u)X_{t_\star^-}(u) > \frac{1}{2}\sqrt{n}e^{w_n}\bigg) \geq 1 - O\big(e^{-2w_n}\big),$$

and so by Markov's inequality,

$$\begin{aligned}(5.2)\qquad &\mathbb{P}\bigg(\mathbb{P}\bigg(\sum_u X_0(u)X_{t_\star^-}(u) > \frac{1}{2}\sqrt{n}e^{w_n} \;\Big|\; X_0\bigg) \geq 1 - e^{-w_n}\bigg)\\ &\quad \geq 1 - O\big(e^{-w_n}\big) \to 1.\end{aligned}$$

By the exponential decay of correlations of Y and the fact that it is independent of X, we have that

$$\mathrm{Var}\bigg(\sum_u X_0(u)Y_{t_\star^-}(u) \;\Big|\; X_0\bigg) \leq c_2 n$$

for some $c_2 > 0$. Thus, since $\mathbb{E}[\sum_u X_0(u)Y_{t_\star^-}(u) \mid X_0] = 0$, it follows that

$$(5.3)\qquad \mathbb{P}\bigg(\sum_u X_0(u)Y_{t_\star^-}(u) > \frac{1}{2}\sqrt{n}e^{w_n} \;\Big|\; X_0\bigg) = O\big(e^{-2w_n}\big) \to 0$$

uniformly in X_0. Comparing equations (5.2) and (5.3) completes the result. □

Acknowledgments. We thank Yuval Peres and Perla Sousi for useful comments and to the referee for their thorough reading and helpful comments.

REFERENCES

[1] AIZENMAN, M. and HOLLEY, R. (1987). Rapid convergence to equilibrium of stochastic Ising models in the Dobrushin Shlosman regime. In *Percolation Theory and Ergodic Theory of Infinite Particle Systems* (*Minneapolis*, *Minn*., 1984–1985). *IMA Vol*. *Math*. *Appl*. **8** 1–11. Springer, New York. MR0894538

[2] ALDOUS, D. (1983). Random walks on finite groups and rapidly mixing Markov chains. In *Seminar on Probability*, *XVII*. *Lecture Notes in Math*. **986** 243–297. Springer, Berlin. MR0770418

[3] ALDOUS, D. and DIACONIS, P. (1986). Shuffling cards and stopping times. *Amer. Math*. *Monthly* **93** 333–348. MR0841111

[4] DIACONIS, P. and SHAHSHAHANI, M. (1987). Time to reach stationarity in the Bernoulli–Laplace diffusion model. *SIAM J*. *Math*. *Anal*. **18** 208–218. MR0871832

[5] DIACONIS, P. and SALOFF-COSTE, L. (1993). Comparison theorems for reversible Markov chains. *Ann*. *Appl*. *Probab.* **3** 696–730. MR1233621

[6] DING, J., LUBETZKY, E. and PERES, Y. (2009). The mixing time evolution of Glauber dynamics for the mean-field Ising model. *Comm*. *Math*. *Phys*. **289** 725–764.

[7] DIACONIS, P. (1996). The cutoff phenomenon in finite Markov chains. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **93** 1659–1664.

[8] DIACONIS, P., GRAHAM, R. L. and MORRISON, J. A. (1990). Asymptotic analysis of a random walk on a hypercube with many dimensions. *Random Structures Algorithms* **1** 51–72. MR1068491

[9] DIACONIS, P. and SALOFF-COSTE, L. (1993). Comparison techniques for random walk on finite groups. *Ann*. *Probab.* **21** 2131–2156. MR1245303

[10] DIACONIS, P. and SALOFF-COSTE, L. (1996). Logarithmic Sobolev inequalities for finite Markov chains. *Ann*. *Appl*. *Probab.* **6** 695–750. MR1410112

[11] DIACONIS, P. and SHAHSHAHANI, M. (1981). Generating a random permutation with random transpositions. *Z. Wahrsch*. *Verw*. *Gebiete* **57** 159–179. MR0626813

[12] DOBRUSHIN, R. L. (1971). Markov processes with a large number of locally interacting components: Existence of a limit process and its ergodicity [in Russian]. *Problemy Peredachi Informatsii* **7** 70–87.

[13] DOBRUSHIN, R. L. and SHLOSMAN, S. B. (1985). Constructive criterion for the uniqueness of Gibbs field. In *Statistical Physics and Dynamical Systems* (*Köszeg*, 1984). *Progress in Probability* **10** 347–370. Birkhäuser, Boston, MA.

[14] HARRIS, T. E. (1978). Additive set-valued Markov processes and graphical methods. *Ann*. *Probab.* **6** 355–378. MR0488377

[15] HOLLEY, R. A. (1972). An ergodic theorem for interacting systems with attractive interactions. *Z. Wahrsch*. *Verw*. *Gebiete* **24** 325–334.

[16] HOLLEY, R. A. and STROOCK, D. W. (1987). Logarithmic Sobolev inequalities and stochastic Ising models. *J*. *Stat*. *Phys*. **46** 1159–1194.

[17] LEVIN, D. A., LUCZAK, M. and PERES, Y. (2010). Glauber dynamics for the mean-field Ising model: Cut-off, critical power law, and metastability. *Probab. Theory Related Fields* **146** 223–265. MR2550363

[18] LEVIN, D. A., PERES, Y. and WILMER, E. L. (2009). *Markov Chains and Mixing Times*. Amer. Math. Soc., Providence, RI. MR2466937

[19] LIGGETT, T. M. (2005). *Interacting Particle Systems*. Springer, Berlin.

[20] LUBETZKY, E. and SLY, A. (2010). Cutoff phenomena for random walks on random regular graphs. *Duke Math*. *J*. **153** 475–510. MR2667423

[21] LUBETZKY, E. and SLY, A. (2011). Explicit expanders with cutoff phenomena. *Electron*. *J*. *Probab.* **16** 419–435. MR2774096

[22] LUBETZKY, E. and SLY, A. (2013). Cutoff for the Ising model on the lattice. *Invent. Math.* **191** 719–755. MR3020173

[23] LUBETZKY, E. and SLY, A. (2014). Cutoff for general spin systems with arbitrary boundary conditions. *Comm. Pure Appl. Math.* **67** 982–1027. MR3193965

[24] LUBETZKY, E. and SLY, A. (2015). An exposition to information percolation for the Ising model. *Ann. Fac. Sci. Toulouse Math.* (6) **24** 745–761. MR3434254

[25] LUBETZKY, E. and SLY, A. (2016). Information percolation and cutoff for the stochastic Ising model. *J. Amer. Math. Soc.* **29** 729–774. MR3486171

[26] MARTINELLI, F. (1999). Lectures on Glauber dynamics for discrete spin models. In *Lectures on Probability Theory and Statistics* (*Saint-Flour*, 1997). *Lecture Notes in Math.* **1717** 93–191. Springer, Berlin.

[27] MARTINELLI, F. and OLIVIERI, E. (1994). Approach to equilibrium of Glauber dynamics in the one phase region. I. The attractive case. *Comm. Math. Phys.* **161** 447–486. MR1269387

[28] MARTINELLI, F. and OLIVIERI, E. (1994). Approach to equilibrium of Glauber dynamics in the one phase region. II. The general case. *Comm. Math. Phys.* **161** 487–514.

[29] MARTINELLI, F., OLIVIERI, E. and SCHONMANN, R. H. (1994). For 2-D lattice spin systems weak mixing implies strong mixing. *Comm. Math. Phys.* **165** 33–47. MR1298940

[30] MILLER, J. and PERES, Y. (2012). Uniformity of the uncovered set of random walk and cutoff for lamplighter chains. *Ann. Probab.* **40** 535–577. MR2952084

[31] PROPP, J. G. and WILSON, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. In *Proceedings of the Seventh International Conference on Random Structures and Algorithms* (*Atlanta*, *GA*, 1995). *Random Structures Algorithms* **9** 223–252.

[32] SALOFF-COSTE, L. (1997). Lectures on finite Markov chains. In *Lectures on Probability Theory and Statistics* (*Saint-Flour*, 1996). *Lecture Notes in Math.* **1665** 301–413. Springer, Berlin. MR1490046

[33] SALOFF-COSTE, L. (2004). Random walks on finite groups. In *Probability on Discrete Structures*. *Encyclopaedia Math. Sci.* **110** 263–346. Springer, Berlin. MR2023654

[34] STROOCK, D. W. and ZEGARLIŃSKI, B. (1992). The equivalence of the logarithmic Sobolev inequality and the Dobrushin–Shlosman mixing condition. *Comm. Math. Phys.* **144** 303–323. MR1152374

[35] STROOCK, D. W. and ZEGARLIŃSKI, B. (1992). The logarithmic Sobolev inequality for discrete spin systems on a lattice. *Comm. Math. Phys.* **149** 175–193. MR1182416

COURANT INSTITUTE
OF MATHEMATICAL SCIENCES
NEW YORK UNIVERSITY
251 MERCER STREET
NEW YORK, NEW YORK 10012
USA
E-MAIL: eyal@courant.nyu.edu

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, BERKELEY
BERKELEY, CALIFORNIA 94720
USA
E-MAIL: sly@stat.berkeley.edu