INFORMATION PERCOLATION AND CUTOFF
FOR THE STOCHASTIC ISING MODEL

EYAL LUBETZKY AND ALLAN SLY

ABSTRACT. We introduce a new framework for analyzing Glauber dynamics for the Ising model. The traditional approach for obtaining sharp mixing results has been to appeal to estimates on spatial properties of the stationary measure from within a multi-scale analysis of the dynamics. Here we propose to study these simultaneously by examining "information percolation" clusters in the space-time slab.

Using this framework, we obtain new results for the Ising model on $(\mathbb{Z}/n\mathbb{Z})^d$ throughout the high temperature regime: total-variation mixing exhibits cutoff with an $O(1)$-window around the time at which the magnetization is the square-root of the volume. (Previously, cutoff in the full high temperature regime was only known in dimensions $d \leq 2$, and only with an $O(\log \log n)$-window.)

Furthermore, the new framework opens the door to understanding the effect of the initial state on the mixing time. We demonstrate this on the 1D Ising model, showing that starting from the uniform ("disordered") initial distribution asymptotically halves the mixing time, whereas almost every deterministic starting state is asymptotically as bad as starting from the ("ordered") all-plus state.

1. INTRODUCTION

Glauber dynamics is one of the most common methods of sampling from the high-temperature Ising model (notable flavors are Metropolis-Hastings or Heat-bath dynamics), and at the same time provides a natural model for its evolution from any given initial configuration.

We introduce a new framework for analyzing the Glauber dynamics via "information percolation" clusters in the space-time slab, a unified approach to studying spin-spin correlations in $\mathbb{Z}^d$ over time (depicted in Fig. 1–2 and described in §1.2). Using this framework, we make progress on the following.

(i) *High-temperature vs. infinite-temperature:* it is believed that when the inverse-temperature β is below the critical β_c, the dynamics behaves qualitatively as if $\beta = 0$ (the spins evolve independently). In the latter case, the continuous-time dynamics exhibits the *cutoff phenomenon*[1] with an $O(1)$-window as shown by Aldous [2] and refined in [6, 11]; thus, the above paradigm suggests cutoff at any $\beta < \beta_c$. Indeed, it was conjectured by Peres in 2004 (see [16, Conjecture 1] and [17, Question 8, p316]) that cutoff occurs whenever there is $O(\log n)$-mixing[2]. Moreover, one expects the cutoff window to be $O(1)$.

Best-known results on $\mathbb{Z}^d$: cutoff for the Ising model in the full high-temperature regime $\beta < \beta_c$ was only confirmed in dimensions $d \leq 2$ ([20]), and only with a bound of $O(\log \log n)$ on the cutoff window.

(ii) *Warm start (random, disordered) vs. cold start (ordered)*: within the extensive physics literature offering numerical experiments for spin systems, it is common to find Monte Carlo simulations at high temperature started at a random (warm) initial state where spins are i.i.d. ("disordered"); cf. [15,31]. A natural question is whether this accelerates the mixing for the Ising model, and if so by how much.

Best-known results on $\mathbb{Z}^d$: none to our knowledge — sharp upper bounds on total-variation mixing for the Ising model were only applicable to worst-case starting states (usually via coupling techniques).

---

[1]sharp transition in the L^1-distance of a finite Markov chain from equilibrium, dropping quickly from near 1 to near 0.
[2]this pertains $\beta < \beta_c$, since at $\beta = \beta_c$ the mixing time for the Ising model on $(\mathbb{Z}/n\mathbb{Z})^d$ is at least polynomial in n by results of Aizenman and Holley [1,13] (see [13, Theorem 3.3]) whereas for $\beta > \beta_c$ it is exponential in n^{d-1} (cf. [23]).
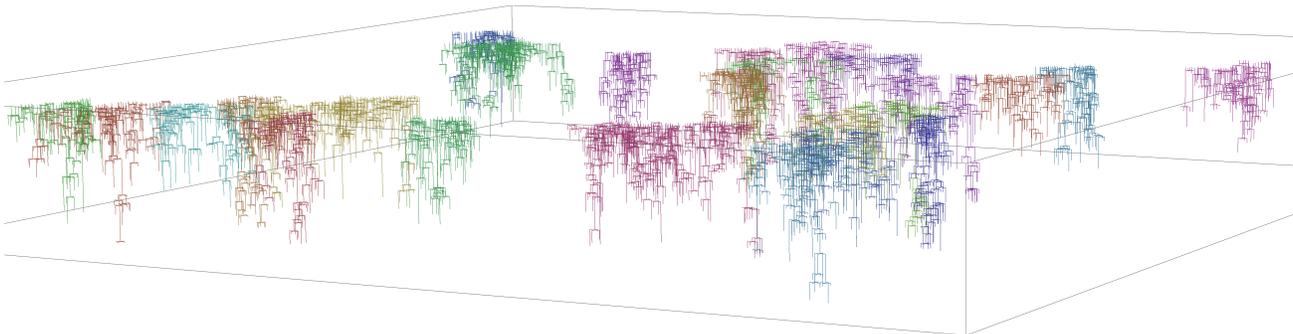
FIGURE 1. Information percolation clusters for the 2D stochastic Ising model:
showing the largest 25 clusters on a $\{1, \ldots, 200\}^2 \times [0, 20]$ space-time slab.

The cutoff phenomenon plays a role also in the second question above: indeed, whenever there is cutoff, one can compare the effect of different initial states x_0 on the asymptotics of the corresponding mixing time $t_{\mathrm{MIX}}^{(x_0)}(\varepsilon)$ independently of ε, the distance within which we wish to approach equilibrium. (For more on the cutoff phenomenon, discovered in the early 80's by Aldous and Diaconis, see [3, 5].)

1.1. **Results.** Our first main result confirms the above conjecture by Peres that Glauber dynamics for the Ising model on $\Lambda = (\mathbb{Z}/n\mathbb{Z})^d$, in *any* dimension d, exhibits cutoff in the full high temperature regime $\beta < \beta_c$. Moreover, we establish cutoff within an $O(1)$-window (as conjectured) around the point

$$t_{\mathfrak{m}} = \inf \left\{ t > 0 \ : \ \mathfrak{m}_t(v) \leq 1/\sqrt{|\Lambda|} \right\}, \tag{1.1}$$

where $\mathfrak{m}_t(v)$ is the magnetization at the vertex $v \in \Lambda$ at time $t > 0$, i.e.,

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

with (X_t^+) being the dynamics started from the all-plus starting state (by translational invariance we may write $\mathfrak{m}_t$ for brevity). Intuitively, at time $t_{\mathfrak{m}}$ the expected value of $\sum_v X_t^+(v)$ becomes $O(\sqrt{|\Lambda|})$, within the normal deviations of the Ising distribution, and expect mixing to occur. For instance, in the special case $\beta = 0$ we have $\mathfrak{m}_t = e^{-t}$ and so $t_{\mathfrak{m}} = \frac{1}{2} \log |\Lambda|$, the known cutoff location from [3, 6, 11].

Theorem 1. *Let $d \geq 1$ and let β_c be the critical inverse-temperature for the Ising model on $\mathbb{Z}^d$. Consider continuous-time Glauber dynamics for the Ising model on the torus $(\mathbb{Z}/n\mathbb{Z})^d$. Then for any inverse-temperature $\beta < \beta_c$ the dynamics exhibits cutoff at $t_{\mathfrak{m}}$ as given in (1.1) with an $O(1)$-window. Moreover, there exists $C = C_{\beta,d} > 0$ so that for any fixed $0 < \varepsilon < 1$ and large n,*

$$t_{\mathrm{MIX}}(1 - \varepsilon) \geq t_{\mathfrak{m}} - C \log(1/\varepsilon) \,,$$

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

This improves on [20] in two ways: (a) A prerequisite for the previous method of proving cutoff for the Ising model on lattices (and all of its extensions) was that the stationary measure would satisfy the decay-of-correlation condition known as *strong spatial mixing*, valid in the full high temperature regime for $d \leq 2$. However, for $d \geq 3$ it is known to hold only for β small enough[3]; Theorem 1 removes this limitation and covers all $\beta < \beta_c$ (see also Theorem 3 below). (b) A main ingredient in the previous proofs was a reduction of L^1-mixing to very-fine L^2-mixing of sub-cubes of poly-logarithmic size, which was achieved via log-Sobolev inequalities in time $O(\log \log n)$. This led to a sub-optimal $O(\log \log n)$ bound on the cutoff window, which we now improve to the conjectured $O(1)$-window.

---

[3]At low temperatures on $\mathbb{Z}^d$ (see the discussion above Theorem 3) there might not be *strong* spatial mixing despite an exponential decay-of-correlations (*weak* spatial mixing); however, one expects to have strong spatial mixing for all $\beta < \beta_c$.
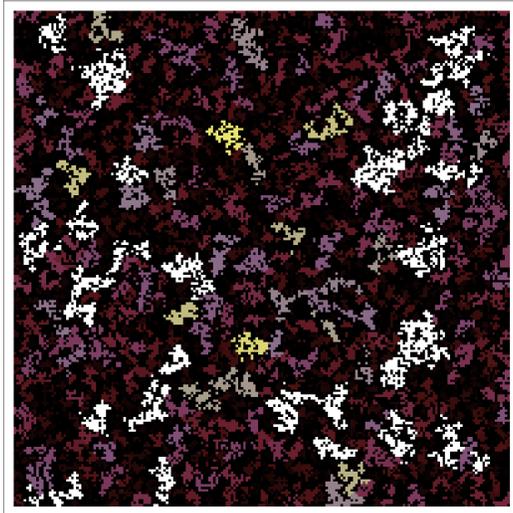
FIGURE 2. Top view of information percolation clusters for the 2D stochastic Ising model: sites of a 200×200 square color-coded by their cluster size (increasing from red to white).

The lower bound on the $t_{\mathrm{MIX}}(1-\varepsilon)$ in Theorem 1 is realized from the all-plus starting configuration; hence, for any $d \geq 1$ this is (as expected) the worst-case starting state up to an additive $O(1)$-term:

$$t_{\mathrm{MIX}}^{(+)}(\varepsilon) = t_{\mathfrak{m}} + O\big(\log(1/\varepsilon)\big) \quad \text{for any } \beta < \beta_c \text{ and } 0 < \varepsilon < 1. \tag{1.3}$$

This brings us to the aforementioned question of understanding the mixing from specific initial states. Here the new methods can be used to give sharp bounds, and in particular to compare the warm start using the uniform (i.i.d.) distribution to various deterministic initial states. We next demonstrate this on the 1D Ising model (we treat higher dimensions, and more generally any bounded-degree geometry, in the companion paper [22]), where, informally, we show that

- The uniform starting distribution is asymptotically *twice faster* than the worst-case all-plus;
- Almost all deterministic initial states are asymptotically *as bad* as the worst-case all-plus.

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

Theorem 2. *Fix any $\beta > 0$ and $0 < \varepsilon < 1$, and consider continuous-time Glauber dynamics for the Ising model on $(\mathbb{Z}/nZ)$. Letting $t_{\mathfrak{m}} = \frac{1}{2\theta} \log n$ with $\theta = 1 - \tanh(2\beta)$, the following hold:*

1. (Annealed) Starting from a uniform initial distribution: $t_{\mathrm{MIX}}^{(\mathrm{U})}(\varepsilon) \sim \frac{1}{2} t_{\mathfrak{m}}$.

2. (Quenched) Starting from a deterministic initial state: $t_{\mathrm{MIX}}^{(x_0)}(\varepsilon) \sim t_{\mathrm{MIX}}^{(+)}(\varepsilon) \sim t_{\mathfrak{m}}$ for almost every x_0.

Unlike the proof of Theorem 1, which coupled the distributions started at worst-case states, in order to analyze the uniform initial state one is forced to compare the distribution at time t directly to the stationary measure. This delicate step is achieved via the *Coupling From The Past* method [28].

Remark. The bound on $t_{\mathrm{MIX}}^{(x_0)}(\varepsilon)$ applies not only to a typical starting state X_0, but to any *deterministic* X_0 which satisfies that $1/\sum_v (\mathbb{E}X_{t_{\mathfrak{m}}}(v))^2$ is sub-polynomial in n — e.g., $O((\log n)^{100})$ — a condition that can be expressed via $X_0(Y_{t_{\mathfrak{m}}})$ where Y_t is continuous-time random walk on $\mathbb{Z}_n$; see Proposition 6.5.

As noted earlier, the new framework relaxes the *strong spatial mixing* hypothesis from previous works into *weak spatial mixing* (i.e., exponential decay-of-correlation, valid for all $\beta < \beta_c$ in any dimension). This has consequences also for *low temperatures*: there it is strongly believed that in dimension $d \geq 3$ (see [23, §5] and [4]) under certain non-zero external magnetic field (some fixed $h \neq 0$ for all sites) there would be weak but not strong spatial mixing. Using the periodic boundary conditions to preclude boundary effects, our arguments remain valid also in this situation, and again we obtain cutoff:

Theorem 3 (low temperature with external field). *The conclusion of Theorem 1 holds in $(\mathbb{Z}/n\mathbb{Z})^d$ for any large enough fixed inverse-temperature β in the presence of a non-zero external magnetic field.*

We now discuss extensions of the framework towards showing *universality of cutoff*, whereby the cutoff phenomenon — believed to be widespread, despite having been rigorously shown only in relatively few cases — is not specific to the underlying geometry of the spin system, but instead occurs always at high temperatures (following the intuition behind the aforementioned conjecture of Peres from 2004). Specializing this general principle to the Ising model, one expects the following to hold:

> *On any locally finite geometry the Ising model should exhibit cutoff at high temperature*
> *(i.e., cutoff always occurs for $\beta < c_d$ where c_d depends only on the maximum degree d).*

The prior technology for establishing cutoff for the Ising model fell well short of proving such a result. Indeed, the approach in [20], as well as its generalization in [21], contained two major provisos:

 (i) heavy reliance on log-Sobolev constants to provide sharp L^2-bounds on local mixing (see [7–9, 29]); the required log-Sobolev bounds can in general be highly nontrivial to verify (see [14, 23–26, 32, 33]).
 (ii) an assumption on the geometry that the growth rate of balls (neighborhoods) is sub-exponential; while satisfied on lattices (linear growth rate), this rules out trees, random graphs, expanders, etc.

Demonstrating these limitations is the fact that the required log-Sobolev inequalities for the Ising model were established essentially only on lattices and regular trees, whereas on the latter (say, a binary tree) it was unknown whether the Ising model exhibits cutoff at any small $\beta > 0$, due to the second proviso.

In contrast with this, the above mentioned paradigm instead says that, at high enough temperatures, cutoff should occur without necessitating log-Sobolev inequalities, geometric expansion properties, etc. Using the new framework of information percolation we can now obtain such a result. Define the non-transitive analogue of the cutoff-location $t_{\mathfrak{m}}$ from (1.1) to be

$$t_{\mathfrak{m}} = \inf \left\{ t > 0 \ : \ \sum_v \mathfrak{m}_t(v)^2 \leq 1 \right\}, \tag{1.4}$$

with $\mathfrak{m}_t(v) = \mathbb{E} X_t^+(v)$ as in (1.2). The proof of the following theorem — which, apart from the necessary adaptation of the framework to deal with a non-transitive geometry, required several novel ingredients to obtain the correct dependence of β on the maximal degree — appears in a companion paper [22].

Theorem 4. *There exists an absolute constant $\kappa > 0$ so that the following holds. Let G be a graph on n vertices with maximum degree at most d. For any fixed $0 < \varepsilon < 1$ and large enough n, the continuous-time Glauber dynamics for the Ising model on G with inverse-temperature $0 \leq \beta < \kappa/d$ has*

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

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

The companion paper further extends Theorem 2 to any bounded-degree graph at high temperature: the mixing time is at least $(1 - \varepsilon_\beta) t_{\mathfrak{m}}$ from almost every deterministic initial state x_0, yet from a uniform initial distribution it is at most $(\frac{1}{2} + \varepsilon_\beta) t_{\mathfrak{m}}$, where ε_β can be made arbitrarily small for β small enough.
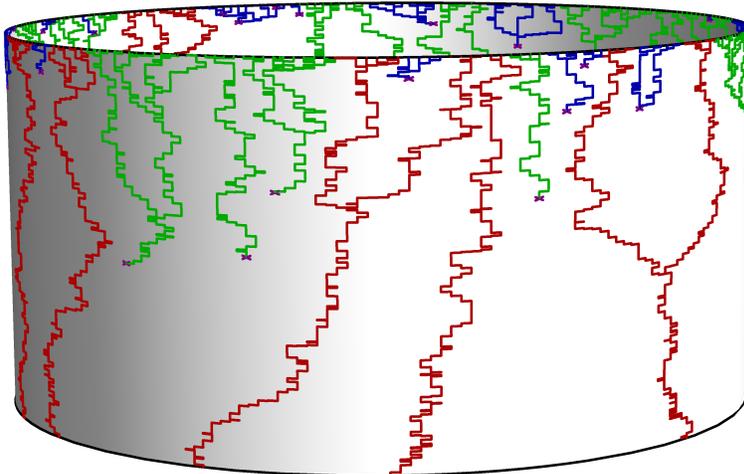
FIGURE 3. Information percolation clusters in Glauber dynamics for the 1D Ising model: red (reaching time zero), blue (dying out quickly) and green clusters on $n = 256$ sites.

In summary, on any locally-finite geometry (following Theorems 1–2 for $\mathbb{Z}^d$) one roughly has that (1) the time needed to couple the dynamics from the extreme initial states, X_t^+ and X_t^-, via the *monotone coupling* (a standard upper bound on the mixing time) overestimates t_{MIX} by a factor of 2; (2) the worst-case mixing time t_{MIX}, which is asymptotically the same as when starting from almost every deterministic state, is *another* factor of 2 worse compared to starting from the uniform distribution.

1.2. **Methods: red, green and blue information percolation clusters.** The traditional approach for obtaining sharp mixing results for the Ising model has been two-fold: one would first derive certain properties of the stationary Ising measure (ranging from as fundamental as strong spatial mixing to as proprietary as interface fluctuations under specific boundary conditions); these static properties would then drive a dynamical multi-scaled analysis (e.g., recursion via block-dynamics/censoring); see [23].

We propose to analyze the spatial and temporal aspects of the Glauber dynamics simultaneously by tracking the update process for the Ising model on $(\mathbb{Z}/n\mathbb{Z})^d$ in the $(d+1)$-dimensional space-time slab. Following is an outline of the approach for heat-bath dynamics[4]; formal definitions of the framework (which is valid for a class of Glauber dynamics that also includes, e.g., Metropolis) will be given in §2.

As a first step, we wish to formulate the dynamics (X_t) so that the update process, viewed backward in time, would behave as subcritical percolation in $(\mathbb{Z}/n\mathbb{Z})^d \times \mathbb{R}_+$; crucially, establishing this subcritical behavior will build on the classical fact that the site magnetization $\mathfrak{m}_t$ (defined in (1.2)), decays exponentially fast to 0 (the proof of which uses the monotonicity of the Ising model; see Lemma 2.1). Recall that each site of $(\mathbb{Z}/n\mathbb{Z})^d$ is updated via a Poisson point process, whereby every update utilizes an independent unit variable to dictate the new spin, and the probability of both plus and minus is bounded away from 0 for any fixed $\beta > 0$ even when all neighbors have the opposing spin. Hence, we can say that with probability $\theta > 0$ bounded away from 0 (explicitly given in (2.4)), the site is updated to a ± 1 fair coin flip independently of the spins at its neighbors, to be referred to as an *oblivious update*.

Clusters definition. For simplicity, we first give a basic definition that will be useful only for small β. Going backward in time from a given site v at time t, we reveal the *update history* affecting $X_t(v)$: in case of an oblivious update we "kill" the branch, and otherwise we split it into its neighbors, continuing until all sites die out or reach time 0 (see Figure 1). The final cluster then allows one to recover $X_t(v)$ given the unit variables for the updates and the intersections of the cluster with the initial state x_0.

---

[4]A single-site heat-bath update replaces a spin by a sample from the Ising measure conditioned on all other spins.

Note that the dependencies in the Ising measure show up in this procedure when update histories of different sites at time t merge into a single cluster, turning the spins at time t into a complicated function of the update variables and the initial state. Of course, since the probability of an oblivious update θ goes to 1 as $\beta \to 0$, for a small enough β the aforementioned branching process is indeed subcritical, and so these clusters should have an exponential tail (see Figure 2). For β close to the critical point in lattices, this is no longer the case, and one needs to refine the definition of an information percolation cluster — roughly, it is the subset of the update history that the designated spin truly depends on (e.g., in the original procedure above, an update can cause the function determining $X_t(v)$ to become independent of another site in the cluster, whence the latter is removed without being directly updated).

The motivation behind studying these clusters is the following. Picture a typical cluster as a single strand, linking between "sausages" of branches that split and quickly dye out. If this strand dies before reaching time 0 then the spin atop would be uniform, and otherwise, starting e.g. from all-plus, that spin would be plus. Therefore, our definition of the cutoff time t_{m} has that about $\sqrt{|\Lambda|}$ of the sites reach time 0; in this way, most sites are independent of the initial state, and so X_t would be well mixed. Further seen now is the role of the initial state x_0, opening the door to non-worst-case analysis: one can analyze the distribution of the spins atop a cluster in terms of its intersection with x_0 at time 0.

Red, green and blue clusters. To quantify the above, we classify the clusters into three types: informally,
- a cluster is BLUE if it dies out very quickly both in space and in time;
- a cluster is RED if the initial state affects the spins atop;
- a cluster is GREEN in all other situations.

(See §2 for formal definitions, and Figure 3 for an illustration of these for the Ising model on $\mathbb{Z}/n\mathbb{Z}$.) Once we condition on the green clusters (to be thought of as having a negligible effect on mixing), what remains is a competition between red clusters — embodying the dependence on the initial state x_0 — and blue ones, the projection on which is just a product measure (independent of x_0). Then, one wants to establish that red clusters are uncommon and "lost within a sea of blue clusters". This is achieved via a simple yet insightful lemma of Miller and Peres [27], bounding the total-variation distance in terms of a certain exponential moment; in our case, an exponential of the intersection of the set of vertices in RED clusters between two i.i.d. instances of the dynamics. Our main task — naturally becoming increasingly more delicate as β approaches β_c — will be to bound this exponential moment, by showing that each red set behaves essentially as a uniformly chosen subset of size $O(e^{-cs}\sqrt{|\Lambda|})$ at time $t_{\mathsf{m}} + s$; thus, the exponential moment will approach 1 as $s \to \infty$, implying mixing.

Flavors of the framework. Adaptations of the general framework above can be used in different settings:
- To tackle arbitrary graphs at high enough temperatures (Theorem 4), a blue cluster is one that dies out before reaching the bottom (time 0) and has a singleton spin at the top (the target time t), and a red cluster is one where the spins at the top have a nontrivial dependence on the initial state x_0.
- For lattices at any $\beta < \beta_c$, the branching processes encountered are not sufficiently subcritical, and one needs to boost them via a phase in which (roughly) some of the oblivious updates are deferred, only to be sprinkled at the end of the analysis. This entails a more complicated definition of blue clusters, referring to whether history dies out quickly enough from the end of that special phase, whereas red clusters remain defined as ones where the top spins are affected by the initial state x_0.
- For random initial states (Theorem 2) we define a red cluster as one in which the intersection with x_0 is of size at least 2 and coalesces to a single point *before* time 0 under *Coupling From The Past*. The fact that pairs of sites surviving to time 0 are now the dominant term (as opposed to singletons) explains the factor of 2 between the annealed/worst-case settings (cf. the two parts of Theorem 2).

1.3. **Organization.** The rest of this paper is organized as follows. In §2 we give the formal definitions of the above described framework, while §3 contains the modification of the general framework tailored to lattices up to the critical point, including three lemmas analyzing the information percolation clusters. In §4 we prove the cutoff results in Theorems 1 and 3 modulo these technical lemmas, which are subsequently proved in §5. The final section, §6, is devoted to the analysis of non-worst-case initial states (random vs. deterministic, annealed vs. quenched) and the proof of Theorem 2.

2. Framework of information percolation

2.1. **Preliminaries.** In what follows we set up standard notation for analyzing the mixing of Glauber dynamics for the Ising model; see [20] and its references for additional information and background.

Mixing time and the cutoff phenomenon. The total-variation distance between two probability measures ν_1, ν_2 on a finite space Ω — one of the most important gauges in MCMC theory for measuring the convergence of a Markov chain to stationarity — is defined as

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

i.e., half the L^1-distance between the two measures. Let (X_t) be an ergodic finite Markov chain with stationary measure π. Its total-variation mixing-time, denoted $t_{\mathrm{MIX}}(\varepsilon)$ for $0 < \varepsilon < 1$, is defined to be

$$t_{\mathrm{MIX}}(\varepsilon) \triangleq \inf \left\{ t \; : \; \max_{x_0 \in \Omega} \|\mathbb{P}_{x_0}(X_t \in \cdot) - \pi\|_{\mathrm{TV}} \leq \varepsilon \right\},$$

where here and in what follows $\mathbb{P}_{x_0}$ denotes the probability given $X_0 = x_0$. A family of ergodic finite Markov chains (X_t), indexed by an implicit parameter n, is said to exhibit *cutoff* (this concept going back to the works [2, 10]) iff the following sharp transition in its convergence to stationarity occurs:

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

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

$$\begin{cases} \lim_{\gamma \to \infty} \liminf_{n \to \infty} \max_{x_0 \in \Omega} \|\mathbb{P}_{x_0}(X_{t_n - \gamma w_n} \in \cdot) - \pi\|_{\mathrm{TV}} = 1, \\ \lim_{\gamma \to \infty} \limsup_{n \to \infty} \max_{x_0 \in \Omega} \|\mathbb{P}_{x_0}(X_{t_n + \gamma w_n} \in \cdot) - \pi\|_{\mathrm{TV}} = 0. \end{cases}$$

Verifying cutoff is often quite challenging, e.g., even for the simple random walk on a bounded-degree graph, no examples were known prior to [19], while this had been conjectured for almost all such graphs.

Glauber dynamics for the Ising model. Let G be a finite graph G 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

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

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 non-negative (ferromagnetic), and h is the external field, taken to be 0 unless stated otherwise. These definitions extend to infinite locally finite graphs (see, e.g., [18, 23]).
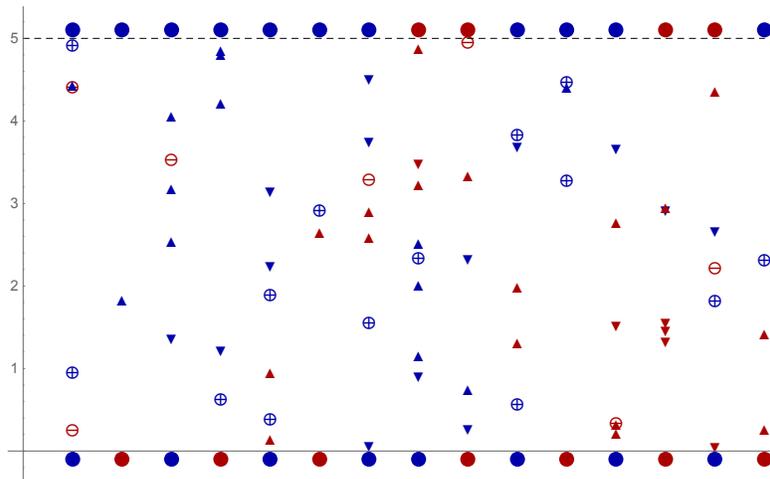
FIGURE 4. Heat-bath dynamics for the 1D Ising model, marking oblivious updates by $\oplus$ and $\ominus$ and non-oblivious updates by $\blacktriangle$ and $\blacktriangledown$ to denote $\sigma(i) \mapsto \sigma(i-1) \wedge \sigma(i+1)$ and $\sigma(i) \mapsto \sigma(i-1) \vee \sigma(i+1)$.

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

$$(\mathscr{L}f)(\sigma) = \sum_{u \in \Lambda} c(u, \sigma) \left(f(\sigma^u) - f(\sigma) \right) \tag{2.3}$$

where σ^u is the configuration σ with the spin at u flipped and $c(u, \sigma)$ is the rate of flipping (cf. [18]). We focus on the two most notable examples of Glauber dynamics, each having an intuitive and useful graphical interpretation where each site receives 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 $c(u, \sigma) = \exp\left(2\beta\sigma(u) \sum_{v \sim u} \sigma(y)\right) \wedge 1$.

(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 $c(u, \sigma) = 1/\left[1 + \exp\left(-2\beta\sigma(u) \sum_{v \sim u} \sigma(v)\right)\right]$.

It is easy to verify that these chains are indeed ergodic and reversible w.r.t. the Ising distribution π. Until recently, sharp mixing results for this dynamics were obtained in relatively few cases, with cutoff only known for the complete graph [12, 16] prior to the works [20, 21].

2.2. **Update history and support.** The *update sequence* along an interval $(t_0, t_1]$ is a set of tuples (J, U, T), where $t_0 < T \leq t_1$ is the update time, $J \in \Lambda$ 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. (For instance, in heat-bath Glauber dynamics, if s is the sum of spins at the neighbors of J at time T then the update (J, U, T) results in a minus spin if $U \leq \frac{1}{2}(1 - \tanh(\beta s))$, and in a plus spin otherwise.)

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

$$\theta = \theta_{\beta,d} := 1 - \tanh(\Delta\beta) \qquad \text{where } \Delta = 2d \text{ is the vertex degree}, \tag{2.4}$$

since in that situation one can update the spin at J to plus/minus with equal probability (that is, with probability $\theta/2$ each via the same U) 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. (For instance, in heat-bath Glauber dynamics, the update is oblivious if $U \leq 1 - \theta/2$ or $U \geq 1 - \theta/2$, corresponding to minus and plus updates, respectively; see Figure 4 for an example in the case $d = 1$.)

The following functions will be used to unfold the update history of a set A at time t_2 to time $t_1 < t_2$:

- *The update function* $\mathscr{F}_{\mathrm{UPD}}(A, t_1, t_2)$: the random set that, given the update sequence along the interval $(t_1, t_2]$, contains every site $u \in \Lambda$ that A "reaches" through the updates in reverse chronological order; that is, every $u \in \Lambda$ such that there exists a subsequence of the updates (J_i, U_i, T_i) with increasing T_i's in the interval $(t_1, t_2]$, such that $J_1, J_2, \ldots$ is a path in Λ that connects u to some vertex in A.
- *The update support function* $\mathscr{F}_{\mathrm{SUP}}(A, t_1, t_2)$: the random set whose value, given the update sequence along the interval $(t_1, t_2]$, is the *update support* of $X_{t_2}(A)$ as a function of X_{t_1}; that is, it is the minimal subset $S \subset \Lambda$ which determines the spins of A given the update sequence (this concept from [20] extends more generally to random mapping representations of Markov chains, see Definition 4.1).

The following lemma establishes the exponential decay of both these update functions for any $\beta < \beta_c$. Of these, $\mathscr{F}_{\mathrm{SUP}}$ is tied to the magnetization $\mathfrak{m}_t$ whose exponential decay, as mentioned in §1.2, in a sense characterizes the one phase region $\beta < \beta_c$ and serves as a keystone to our analysis of the subcritical nature of the information percolation clusters. Here and in what follows, for a subset $A \subset \mathbb{Z}^d$ and $r > 0$, let $B(A, r)$ denote the set of all sites in $\mathbb{Z}^d$ with L^∞ distance at most r from A.

Lemma 2.1. *The update functions for the Ising model on $\Lambda = (\mathbb{Z}/n\mathbb{Z})^d$ satisfy the following for any $\beta < \beta_c$. There exist some constant $c_{\beta, d} > 0$ such that for any $\Lambda' \subset \Lambda$, any vertex $v \in \Lambda'$ and any $h > 0$,*

$$\mathbb{P}(\mathscr{F}_{\mathrm{SUP}}(v, t - h, t) \neq \emptyset) = \mathfrak{m}_h \leq 2 e^{-c_{\beta, d} h} \tag{2.5}$$

with $\mathfrak{m}_h = \mathfrak{m}_h(v)$ as defined in (1.1), whereas for $\ell > 20 d h$,

$$\mathbb{P}(\mathscr{F}_{\mathrm{UPD}}(v, t - h, t) \not\subset B(v, \ell)) \leq e^{-\ell}. \tag{2.6}$$

Proof. The left-hand equality in (2.5) is by definition, whereas the right-hand inequality was derived from the weak spatial mixing property of the Ising model using the monotonicity of the model in the seminal works of Martinelli and Olivieri [24, 25] (see Theorem 3.1 in [24] as well as Theorem 4.1 in [23]); we note that this is the main point where our arguments rely on the monotonicity of the Ising model. As it was shown in [13, Theorem 2.3] that $\lim_{t \to \infty} \frac{-1}{t} \log \mathfrak{m}_t = \mathtt{gap}$ where $\mathtt{gap}$ is the smallest positive eigenvalue of the generator of the dynamics, this is equivalent to having $\mathtt{gap}$ be bounded away from 0.

We therefore turn our attention to (2.6), which is a consequence of the finite speed of information flow vs. the amenability of lattices. Let $\mathcal{W}$ denote the set of sequences of vertices

$$\mathcal{W} = \left\{ \tilde{w} = (w_1, w_2, \ldots, w_\ell) : w_1 = v, \|w_{i-1} - w_i\|_1 = 1 \right\}.$$

For $\mathscr{F}_{\mathrm{UPD}}(v, t - h, t) \not\subset B(v, \ell)$ to hold there must be some $w \in \mathcal{W}$ and a sequence $t > t_1 > \ldots > t_\ell > t - h$ so that vertex w_i was updated at time t_i. If this event holds call it $M_{\tilde{w}}$. It is easy to see that

$$\mathbb{P}(M_{\tilde{w}}) = \mathbb{P}(\mathrm{Po}(h) \geq \ell) \leq e^{-\ell(\log(\ell/h) - 1)},$$

where the last transition is by Bennet's inequality. By a union bound over $\mathcal{W}$ we have that for $\ell > 20 d h$,

$$\mathbb{P}\left(\mathscr{F}_{\mathrm{UPD}}(v, t - h, t) \not\subset B(v, \ell)\right) \leq (2d)^\ell \exp(-\ell(\log(\ell/h) - 1)) \leq e^{-\ell},$$

thus establishing (2.6) and completing the proof. ∎

2.3. **Red, green and blue clusters.** In what follows, we describe the basic setting of the framework, which will be enhanced in §3 to support all $\beta < \beta_c$. Consider some designated target time $t_\star$ for analyzing the distribution of the dynamics on $\Lambda = (\mathbb{Z}/n\mathbb{Z})^d$. The *update support* of $X_{t_\star}(v)$ at time t is

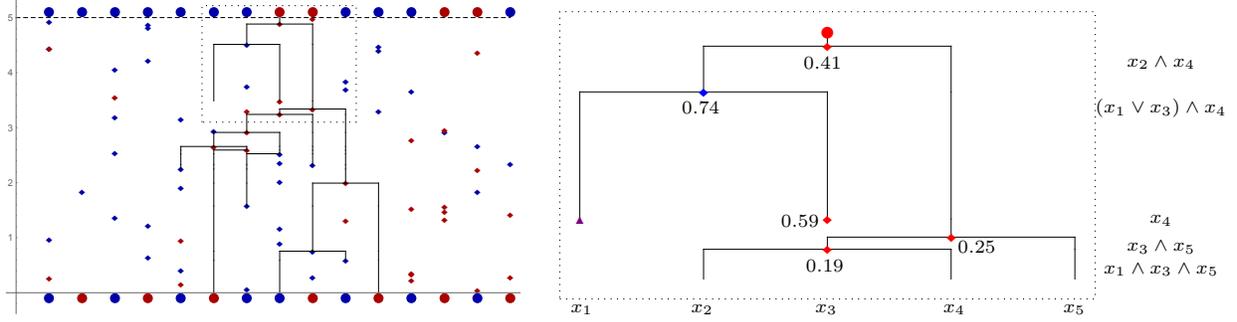$$\mathscr{H}_v(t) = \mathscr{F}_{\mathrm{SUP}}(v, t, t_\star),$$

FIGURE 5. Update support for heat-bath dynamics for the 1D Ising model at $\beta = 0.4$ ($\theta \approx 0.34$); zoomed-in part shows the update history with the root spin as a deterministic function of the leaves.
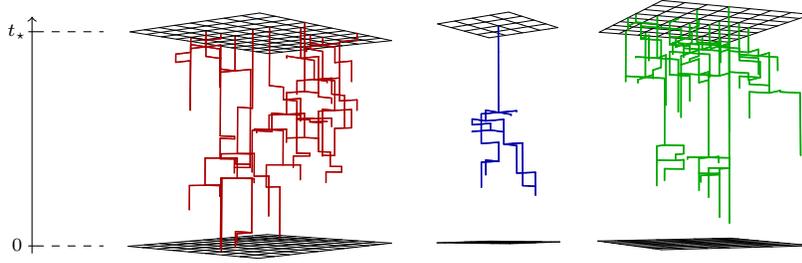

FIGURE 6. Red, blue and green information percolation clusters as per Definition 2.3.

i.e., the minimum subset of sites whose spins at time t determine $X_{t_\star}(v)$ given the updates along $(t, t_\star]$. Developing $\{\mathscr{H}_v(t) : t \leq t_\star\}$ backward in time, started at time $t_\star$, gives rise to a subgraph $\mathscr{H}_v$ of the space-time slab $\Lambda \times [0, t_\star]$, where we connect (u, t) with (u, t') (a temporal edge) if $u \in \mathscr{H}_v(t)$ and there are no updates along $(t', t]$, and connect (u, t) with (u', t) (a spatial edge) when $u \in \mathscr{H}_v(t)$, $u' \notin \mathscr{H}_v(t)$ and $u' \in \mathscr{H}_v(t - \delta)$ for any small enough $\delta > 0$ due to an update at (u, t) (see Figure 5).

Remark 2.2. An oblivious update at (u, t) clearly removes u from $\mathscr{H}_v(t - \delta)$; however, the support may also shrink due to non-oblivious updates: the zoomed-in update history in Figure 5 shows x_1, x_3 being removed from $\mathscr{H}_v(t)$ due to the update $x_3 \mapsto x_2 \vee x_4$, as the entire function then collapses to x_4.

The *information percolation clusters* are the connected components in the space-time slab $\Lambda \times [0, t_\star]$ of the aforementioned subgraphs $\{\mathscr{H}_v : v \in \Lambda\}$.

Definition 2.3. An information percolation cluster is marked RED if it has a nonempty intersection with the bottom slab $\Lambda \times \{0\}$; it is BLUE if it does not intersect the bottom slab and has a singleton in the top slab, $v \times \{t_\star\}$ for some $v \in \Lambda$; all other clusters are classified as GREEN. (See Figure 6.)

Observe that if a cluster is blue then the distribution of its singleton at the top does not depend on the initial state x_0; hence, by symmetry, it is $(\frac{1}{2}, \frac{1}{2})$ plus/minus.

Let Λ_{RED} denote the union of the red clusters, and let $\mathscr{H}_{\text{RED}}$ be the its collective history — the union of $\mathscr{H}_v(t)$ for all $v \in \Lambda_{\text{RED}}$ and $0 \leq t < t_\star$ (with analogous definitions for blue/green). A beautiful short lemma of Miller and Peres [27] shows that, if a measure μ on Ω is constructed by sampling a random variable $R \subset \Lambda$ and using an arbitrary law for its spins and a product of Bernoulli($\frac{1}{2}$) for $\Lambda \setminus R$, then the L^2-distance of μ from the uniform measure is bounded by $\mathbb{E}2^{|R \cap R'|} - 1$ for two i.i.d. copies R, R'. (See Lemma 4.3 below for a generalization of this, as we will have a product of complicated measures.) Applied to our setting, if we condition on $\mathscr{H}_{\text{GREEN}}$ and look at the spins of $\Lambda \setminus \Lambda_{\text{GREEN}}$ then Λ_{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 Λ_{GREEN}, albeit potentially complicated, is independent of the initial state, we can safely project the configurations on $\Lambda \setminus \Lambda_{\text{GREEN}}$ without 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

$$\mathbb{E}\left[2^{|\Lambda_{\text{RED}} \cap \Lambda'_{\text{RED}}|} \mid \mathscr{H}_{\text{GREEN}}\right] \to 1 \quad \text{in probability as } n \to \infty, \tag{2.7}$$

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 (2.7).

Example 2.4 (Red, green and blue clusters in the 1D Ising model). Consider the relatively simple special case of $\Lambda = \mathbb{Z}/n\mathbb{Z}$ to illustrate the approach outlined above. Here, since the vertex degree is 2, an update either writes a new spin independently of the neighbors (with probability θ) or, by symmetry, it takes the spin of a uniformly chosen neighbor. Thus, the update history from any vertex v is simply a continuous-time simple random walk that moves at rate $1 - \theta$ and dies at rate θ; the collection of these for all $v \in \Lambda$ forms coalescing (but never splitting) histories (recall Figure 3).

The probability that $\mathscr{F}_{\text{SUP}}(v, 0, t) \neq \emptyset$ (the history of $X_t(v)$ is nontrivially supported on the bottom of the space-time slab) is therefore $e^{-\theta t}$, which becomes $1/\sqrt{n}$ once we take $t = t_{\mathfrak{m}} = \frac{1}{2\theta} \log n$. If we ignore the conditioning on the green clusters (which poses a technical difficulty for the analysis— as the red and blue histories must avoid them—but does not change the overall behavior by much), then $\mathbb{P}(v \in \Lambda_{\text{RED}} \cap \Lambda'_{\text{RED}}) = \mathbb{P}(\mathscr{H}_v(0) \neq \emptyset)^2 = e^{-2\theta t_\star}$ by the independence of the copies $\Lambda_{\text{RED}}, \Lambda'_{\text{RED}}$. Furthermore, if the events $\{v \in \Lambda_{\text{RED}} \cap \Lambda'_{\text{RED}}\}_{v \in \Lambda}$ were mutually independent (of course they are not, yet the intuition is still correct) then $\mathbb{E}[2^{|\Lambda_{\text{RED}} \cap \Lambda'_{\text{RED}}|}] = \mathbb{E}\left[\prod_v (1 + \mathbb{1}_{\{v \in \Lambda_{\text{RED}} \cap \Lambda'_{\text{RED}}\}})\right]$ would translate into

$$\prod_v \mathbb{E}\left[1 + \mathbb{1}_{\{v \in \Lambda_{\text{RED}} \cap \Lambda'_{\text{RED}}\}}\right] = \left(1 + e^{-2\theta t_\star}\right)^n \leq \exp\left(n e^{-2\theta t_\star}\right),$$

which for $t_\star = t_{\mathfrak{m}} + s$ is at most $\exp(e^{-2\theta s})$. As we increase the constant $s > 0$ this last quantity approaches 1, from which the desired upper bound on the mixing time will follow via (2.7).

The above example demonstrated (modulo conditioning on $\mathscr{H}_{\text{GREEN}}$ and dependencies between sites) how this framework can yield sharp upper bounds on mixing when the update history corresponds to a subcritical branching process. However, in dimension $d \geq 2$, this stops being the case midway through the high temperature regime in lattices, and in §3 we describe the additional ideas that are needed to extend the framework to all $\beta < \beta_c$.
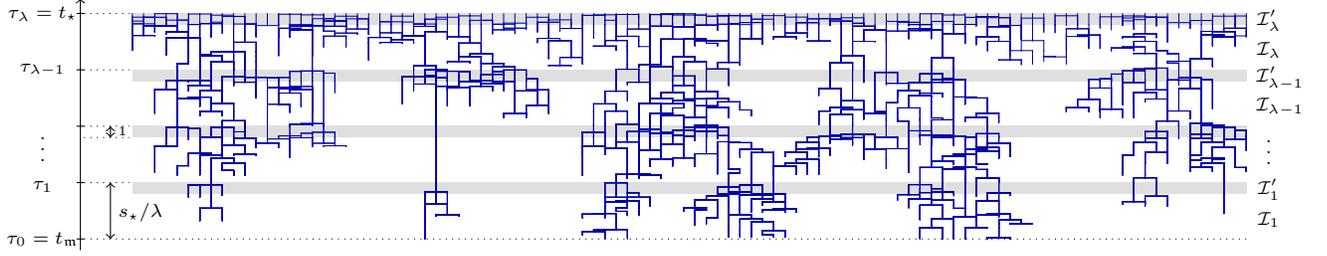
FIGURE 7. Regular phases ($\mathcal{I}_k$) and deferred phases ($\mathcal{I}'_k$) of the update history in the range $t > t_{\mathfrak{m}}$: no vertices are removed from the support along the deferred phases (marked by shaded regions).

3. Enhancements for the lattice up to criticality

To extend the framework to all $\beta < \beta_c$ we will modify the definition of the support at time t for $t > t_{\mathfrak{m}}$, as well as introduce new notions in the space-time slab both for $t > t_{\mathfrak{m}}$ and for $t < t_{\mathfrak{m}}$ within the cutoff window. These are described in §3.1 and §3.2, resp., along with three key lemmas (Lemmas 3.1–3.3) whose proofs are postponed to §5.

3.1. **Post mixing analysis: percolation components.** Let $\lambda > 0$ be some large enough integer, and let $s_\star > 0$ denote some larger constant to be set last. As illustrated in Figure 7, set

$$t_\star = t_{\mathfrak{m}} + s_\star \,,$$

for $k = 0, 1, \ldots, \lambda$ let

$$\tau_k = t_{\mathfrak{m}} + k s_\star / \lambda \,,$$

and partition each interval $(\tau_{k-1}, \tau_k]$ for $k = 1, \ldots, \lambda$ into the subintervals

$$\mathcal{I}_k = (\tau_{k-1}, \tau_k - 1] \,, \qquad \mathcal{I}'_k = (\tau_k - 1, \tau_k] \,.$$

We refer to $\mathcal{I}_k$ as a *regular phase* and to $\mathcal{I}'_k$ as a *deferred phase*.

Definition (the support $\mathscr{H}_v(t)$ for $t > t_{\mathfrak{m}}$)**.** Starting from time $t_\star = t_{\mathfrak{m}} + s_\star$ and going backwards to time $t_{\mathfrak{m}}$ we develop the history of a vertex $v \in V$ rooted at time $t_\star$ as follows:

- Regular phases ($\mathcal{I}_k$ for $k = 1, \ldots, \lambda$): For any $\tau_{k-1} < t \leq \tau_k - 1$,

$$\mathscr{H}_v(t) = \mathscr{F}_{\text{SUP}}(\mathscr{H}_v(\tau_k - 1), t, \tau_k - 1) \,.$$

  Note that an oblivious update at time t to some $w \in \mathscr{H}_v(t)$ will cause it to be removed from the corresponding support (so $w \notin \mathscr{H}_v(t - \delta)$ for any small enough $\delta > 0$), while a non-oblivious update replaces it by a subset of its neighbors. We stress that w may become irrelevant (thus ejected from the support) due to an update to some other, potentially distant, vertex z (see Remark 2.2).
- Deferred phases ($\mathcal{I}'_k$ for $k = 1, \ldots, \lambda$): For any $\tau_k - 1 < t \leq \tau_k$,

$$\mathscr{H}_v(t) = \mathscr{F}_{\text{UPD}}(\mathscr{H}_v(\tau_k), t, \tau_k) \,.$$

  Here vertices do not leave the support: an update to $w \in \mathscr{H}_v(t)$ adds its $2d$ neighbors to $\mathscr{H}_v(t - \delta)$.

Recalling the form (J_i, U_i, T_i) of updates (see §2.2), let the *undeferred randomness* $\mathcal{U}$ be the updates along $(t_{\mathfrak{m}}, t_\star]$ excluding the uniform unit variables U_i when $T_i \in \cup_k \mathcal{I}'_k$, and let the *deferred randomness* $\mathcal{U}'$ denote this set of excluded uniform unit variables (corresponding to updates in the deferred phases).
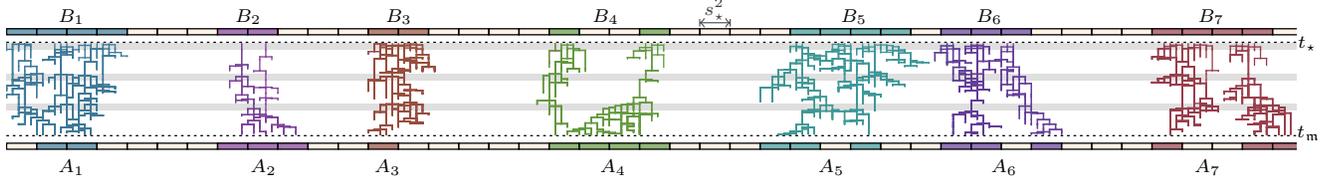
FIGURE 8. Components $\mathcal{A} = \{A_i\}$ and $\mathcal{B} = \{B_i\}$ corresponding to $\{\Upsilon_i\}$. Each Υ_i joins vertices of Υ via *history intersection*; Υ_4 and Υ_7 also join vertices via *final proximity* and *initial proximity*, resp.

Remark. Observe that this definition of $\{\mathscr{H}_v(t) : t > t_{\mathfrak{m}}\}$ is a function of the undeferred randomness $\mathcal{U}$ alone (as the deferred phases $\mathcal{I}'_k$ involved $\mathscr{F}_{\text{UPD}}$ as opposed to $\mathscr{F}_{\text{SUP}}$); thus, $X_{t_\star}$ may be obtained from $X_{t_{\mathfrak{m}}}$ by first exposing $\mathcal{U}$, then incorporating the deferred randomness $\mathcal{U}'$ along the deferred phases $\mathcal{I}'_k$.

Remark. The goal behind introducing the deferred phases $\mathcal{I}'_k$ is to boost the subcritical behavior of the support $\mathscr{H}_v(t)$ towards an analog of the exponential moment in (2.7). In what follows we will describe how the set of sites with $\mathscr{H}_v(t_{\mathfrak{m}}) \neq \emptyset$ are partitioned into components (according to proximity and intersection of their histories); roughly put, by exposing $\mathcal{U}$ but not $\mathcal{U}'$ one can identify, for each set of vertices B in such a component, a time t in which $\mathscr{H}_B(t)$ is suitably "thin" (we will refer to it as a "cut-set") so that — recalling that a branch of the history is killed at rate $\theta > 0$ via oblivious updates — one obtains a good lower bound on the probability of arriving at any configuration for the spin-set $\mathscr{H}_B(t)$ (which then determines $X_{t_\star}(B)$) once the undeferred randomness $\mathcal{U}'$ is incorporated.

Blocks of sites and components of blocks. Partition $\mathbb{Z}^d$ into boxes of side-length $s_\star^2$, referred to as *blocks*. We define *block components*, composed of subsets of blocks, as follow (see Figure 8).

Definition (Block components)**.** Given the undeferred update sequence $\mathcal{U}$, we say that $u \sim v$ if one of the following conditions holds:
   (1) History intersection: $\mathscr{H}_u(t) \cap \mathscr{H}_v(t) \neq \emptyset$ for some $t \in (t_{\mathfrak{m}}, t_\star]$.
   (2) Initial proximity: u, v belong to the same block or to adjacent ones.
   (3) Final proximity: there exist $u' \in \mathscr{H}_u(t_{\mathfrak{m}})$ and $v' \in \mathscr{H}_v(t_{\mathfrak{m}})$ belonging to the same block or to adjacent ones.
Let $\Upsilon = \{v : \mathscr{H}_v(t_{\mathfrak{m}}) \neq \emptyset\}$ be the vertices whose history reaches $t_{\mathfrak{m}}$. We partition Υ into components $\{\Upsilon_i\}$ via the transitive closure of $\sim$. Let $\mathscr{H}_{\Upsilon_i}(t) = \cup_{v \in \Upsilon_i} \mathscr{H}_v(t)$, let A_i be the minimal set of blocks covering $\mathscr{H}_{\Upsilon_i}(t_{\mathfrak{m}})$ and let B_i be the minimal set of blocks covering $\mathscr{H}_{\Upsilon_i}(t_\star) = \Upsilon_i$. The collection of all components $\{A_i\}$ is denoted $\mathcal{A} = \mathcal{A}(\mathcal{U})$ and the collection of all components $\{B_i\}$ is denoted $\mathcal{B} = \mathcal{B}(\mathcal{U})$.

We now state a bound on the probability for witnessing a given set of blocks. In what follows, let $\mathfrak{W}(S)$ denote the size, in blocks, of the minimal lattice animal[5] containing the block-set S. Further let $\{R \leftrightsquigarrow S\}$ denote the event for some i the block-sets (R, S) satisfy $R = A_i \in \mathcal{A}$ and $S = B_i \in \mathcal{B}$; i.e., R, S correspond to the same component, some Υ_i, as the minimal block covers of $\mathscr{H}_{\Upsilon_i}(t)$ at $t = t_{\mathfrak{m}}, t_\star$.

Lemma 3.1. *Let $\beta < \beta_c$. There exist constants $c(\beta, d), \lambda_0(\beta, d)$ such that, if $\lambda > \lambda_0$ and $s_\star > \lambda^2$, then for every collection of pairs of block-sets $\{(R_i, S_i)\}$,*

$$\mathbb{P}\left(\bigcap_i \{R_i \leftrightsquigarrow S_i\}\right) \leq \exp\left[-c\frac{s_\star}{\lambda} \sum_i \mathfrak{W}(R_i \cup S_i)\right].$$

---

[5] A lattice animal is a connected subset of sites in the lattice.

Cut-sets of components. The cut-set of a component A_i is defined as follows. For $k = 1, \dots, \lambda$ let

$$\chi_{i,k} = \mathscr{H}_{\Upsilon_i}(\tau_k)\,,$$
$$\Xi_{i,k} = \prod_{v \in \chi_{i,k}} \left(\tfrac{1}{4}\theta T_{v,k}\right)$$

where $\theta = 1 - \tanh(2d\beta)$ is the oblivious update probability, and $T_{v,k}$ is the time elapsed since the last update to the vertex v within the deferred phase $\mathcal{I}'_k$ until τ_k, the end of that interval. That is, $T_{v,k} = (\tau_k - t) \wedge 1$ for the maximum time $t < \tau_k$ at which there was an update to v. With this notation, the *cut-set* of A_i is the pair (k_i, χ_i) where k_i is the value of $1 \leq k \leq \lambda$ minimizing $\Xi_{i,k}$ and $\chi_i = \chi_{i,k_i}$. The following lemma will be used to estimate these cut-sets.

Lemma 3.2. *Let $\beta < \beta_c$. Let S be a set of blocks, let $\chi_k(S) = \mathscr{H}_S(\tau_k)$ and $\Xi_k(S) = \prod_{v \in \chi_k(S)} \left(\tfrac{1}{4}\theta T_{v,k}\right)$ where $T_{v,k}$ is the time that elapsed since the last update to the vertex v within the deferred phase $\mathcal{I}'_k$ until τ_k. If λ is large enough in terms of β, d and $s_\star$ is large enough in terms of λ then*

$$\mathbb{E}\left[\min_k \left\{(\Xi_k(S))^{-4} : 1 \leq k \leq \lambda\right\}\right] \leq 2^{\lambda+3} e^{|S|}\,.$$

3.2. **Pre mixing analysis: percolation clusters.** Going backwards from time $t_{\mathfrak{m}}$ to time 0, the history is defined in the same way as it was in the regular phases (see §3.1); that is, for any $0 < t \leq t_{\mathfrak{m}}$,

$$\mathscr{H}_v(t) = \mathscr{F}_{\text{SUP}}(\mathscr{H}_v(t_{\mathfrak{m}}), t, t_{\mathfrak{m}})\,.$$

Further recall that the set of sites $\Upsilon = \{v : \mathscr{H}_v(t_{\mathfrak{m}}) \neq \emptyset\}$ were partitioned into components Υ_i (see §3.1), and for each i we let A_i be the minimal block-set covering $\mathscr{H}_{\Upsilon_i}(t_{\mathfrak{m}})$.

Definition (Information percolation clusters)**.** We write $A_i \sim A_j$ if the supports of these components satisfy either one of the following conditions (recall that $B(V, r) = \{x \in \mathbb{Z}^d : \min_{v \in V} \|x - v\|_\infty < r\}$).
  (1) Intersection: $\mathscr{F}_{\text{SUP}}(A_i, t, t_{\mathfrak{m}}) \cap \mathscr{F}_{\text{SUP}}(A_j, t, t_{\mathfrak{m}}) \neq \emptyset$ for some $0 \leq t < t_{\mathfrak{m}}$.
  (2) Early proximity: $\mathscr{F}_{\text{SUP}}(A_i, t, t_{\mathfrak{m}}) \cap B(A_j, s_\star^2/3) \neq \emptyset$ for some $t_{\mathfrak{m}} - s_\star \leq t \leq t_{\mathfrak{m}}$, or the analogous statement when the roles of A_i, A_j are reversed.

We partition $\mathcal{A} = \{A_i\}$ into clusters $\mathcal{C}^{(1)}, \mathcal{C}^{(2)}, \dots$ according to the transitive closure of the above relation, and then classify these clusters into three color groups:
  • BLUE: a cluster $\mathcal{C}^{(k)}$ consisting of a single A_i (for some $i = i(k)$) which dies out within the interval $(t_{\mathfrak{m}} - s_\star, t_{\mathfrak{m}}]$ without exiting the ball of radius $s_\star^2/3$ around A_i:

$$\mathcal{C}^{(k)} = \{A_i\}\,, \quad \bigcup_{v \in A_i} \mathscr{F}_{\text{SUP}}(v, t_{\mathfrak{m}} - s_\star, t_{\mathfrak{m}}) = \emptyset\,, \quad \bigcup_{t > t_{\mathfrak{m}} - s_\star,\, v \in A_i} \mathscr{F}_{\text{SUP}}(v, t, t_{\mathfrak{m}}) \subset B(A_i, s_\star^2/3)\,.$$

  • RED: a cluster $\mathcal{C}^{(k)}$ containing a vertex whose history reaches time 0:

$$\bigcup_{v \in A_i \in \mathcal{C}^{(k)}} \mathscr{F}_{\text{SUP}}(v, 0, t_{\mathfrak{m}}) \neq \emptyset\,.$$

  • GREEN: all other clusters (neither red nor blue).

Let $\mathcal{A}_{\text{RED}}$ be the set of components whose cluster is red, and let $\mathscr{H}_{\text{RED}}$ be the collective history of all $v \in \mathcal{A}_{\text{RED}}$ going backwards from time $t_{\mathfrak{m}}$, i.e.,

$$\mathscr{H}_{\text{RED}} = \left\{\mathscr{H}_v(t) \,:\, v \in \mathcal{A}_{\text{RED}}\,,\, t \leq t_{\mathfrak{m}}\right\},$$

setting the corresponding notation for blue and green clusters analogously.

For a generic collection of blocks $\mathcal{C}$, the collective history of all $v \notin \mathcal{C}$ is defined as

$$\mathscr{H}_{\mathcal{C}}^{-} = \left\{ \mathscr{H}_v(t) : v \notin \cup \{A_i \in \mathcal{C}\} \,,\, t \leq t_{\mathfrak{m}} \right\},$$

and we say $\mathscr{H}_{\mathcal{C}}^{-}$ is $\mathcal{C}$-*compatible* if there is a positive probability that $\mathcal{C}$ is a cluster conditioned on $\mathscr{H}_{\mathcal{C}}^{-}$.

Central to the proof will be to understand the conditional probability of a set of blocks $\mathcal{C}$ to be a single red cluster as opposed to a collection of blue ones (having ruled out green clusters by conditioning on $\mathscr{H}_{\text{GREEN}}$) given the undeferred randomness $\mathcal{U}$ and the history up to time $t_{\mathfrak{m}}$ of all the other vertices:

$$\Psi_{\mathcal{C}} = \sup_{\mathcal{X}} \mathbb{P} \left(\mathcal{C} \in \text{RED} \mid \mathscr{H}_{\mathcal{C}}^{-} = \mathcal{X} \,,\, \{\mathcal{C} \in \text{RED}\} \cup \{\mathcal{C} \subset \text{BLUE}\} \,,\, \mathcal{U} \right) \qquad (3.1)$$

The next lemma bounds $\Psi_{\mathcal{C}}$ in terms of the lattice animals for $\mathcal{C}$ and the individual A_i's; note that the dependence of this estimate for $\Psi_{\mathcal{C}}$ on $\mathcal{U}$ is through the geometry of the components A_i. Here and in what follows we let $|x|^{+} = x \wedge 0$ denote the positive part of x.

Lemma 3.3. *Let $\beta < \beta_c$. There exists $c(\beta, d), s_0(\beta, d) > 0$ such that, for any $s_\star > s_0$, any large enough n and every $\mathcal{C} \subset \mathcal{A}$, the quantity $\Psi_{\mathcal{C}}$ from (3.1) satisfies*

$$\Psi_{\mathcal{C}} \leq \frac{s_\star^{4d}}{\sqrt{|\Lambda|}} e^{4 \sum_i |A_i| - cs_\star \left| \mathfrak{W}(\mathcal{C}) - \sum_{A_i \in \mathcal{C}} \mathfrak{W}(A_i) \right|^{+}}. \qquad (3.2)$$

4. CUTOFF WITH A CONSTANT WINDOW

4.1. Upper bound modulo Lemmas 3.1–3.3.

Let $\mathcal{U}$ be the undeferred randomness along $(t_{\mathfrak{m}}, t_\star]$ (the update sequence excluding the uniform unit variables of updates in the deferred phases $\cup_{k=1}^{\lambda} \mathcal{I}_k'$). Let $\bar{d}(t, \mathcal{U})$ be the coupling time conditioned on this update sequence, that is,

$$\bar{d}(t, \mathcal{U}) = \sup_{x_0, y_0} \left\| \mathbb{P}_{x_0}(X_t \in \cdot \mid \mathcal{U}) - \mathbb{P}_{y_0}(X_t \in \cdot \mid \mathcal{U}) \right\|_{\text{TV}}.$$

Towards an upper bound on $\bar{d}(t, \mathcal{U})$ (which will involve several additional definitions; see (4.4) below), our starting point would be to consider the notion of the *support of a random map*, first introduced in [20]. Its following formulation in a more general framework appears in [21]. Let K be a transition kernel of a finite Markov chain. A *random mapping representation* for K is a pair (g, W) where g is a deterministic map and W is a random variable such that $\mathbb{P}(g(x, W) = y) = K(x, y)$ for all x, y in the state space of K. It is well-known that such a representation always exists.

Definition 4.1 (*Support of a random mapping representation*)**.** Let K be a Markov chain on a state space Σ^{Λ} for some finite sets Σ and Λ. Let (g, W) be a random mapping representation for K. The *support* corresponding to g for a given value of W is the minimum subset $\Lambda_W \subset \Lambda$ such that $g(\cdot, W)$ is determined by $x(\Lambda_W)$ for any x, i.e.,

$$g(x, W) = f_W(x(\Lambda_W)) \text{ for some } f_W : \Sigma^{\Lambda_W} \to \Sigma^{\Lambda} \text{ and all } x.$$

That is, $v \in \Lambda_W$ if and only if there exist $x, x' \in \Sigma^{\Lambda}$ differing only at v such that $g(x, W) \neq g(x', W)$.

Lemma 4.2 ([21, Lemma 3.3])**.** *Let K be a finite Markov chain and let (g, W) be a random mapping representation for it. Denote by Λ_W the support of W w.r.t. g as per Definition 4.1. Then for any distributions φ, ψ on the state space of K,*

$$\left\| \varphi K - \psi K \right\|_{\text{TV}} \leq \int \left\| \varphi|_{\Lambda_W} - \psi|_{\Lambda_W} \right\|_{\text{TV}} d\mathbb{P}(W).$$

To relate this to our context of seeking an upper bound for $\bar{d}(t, \mathcal{U})$, recall that (as remarked following the definition of the regular and deferred phases in §3) $\mathscr{H}_v(t)$ for $t > t_\mathfrak{m}$ is a function of the undeferred randomness $\mathcal{U}$ alone. Hence, both the A_i's and their corresponding cut-sets (k_i, χ_i) are completely determined from $\mathcal{U}$. Letting Z_χ denote the joint distribution of $X_{\tau_{k_i}}(\chi_i)$ for all the components A_i, we can view $(X_{t_\star} \in \cdot \mid \mathcal{U})$ as a random function of $(\cup_i X_{\tau_{k_i}}(\chi_i) \mid \mathcal{U})$ whose randomness arises from the deferred updates $\mathcal{U}'$ (using which every $X_{t_\star}(v)$ for $v \in \Upsilon_i$ can be deduced from $X_{\tau_{k_i}}(\chi_i), \mathcal{U}$, while $X_{t_\star}(v)$ for $v \notin \Upsilon$ is completely determined by $\mathcal{U}$). It then follows from Lemma 4.2 that

$$\bar{d}(t_\star, \mathcal{U}) \leq \mathbb{E}\left[\sup_{x_0, y_0} \|\mathbb{P}_{x_0}(Z_\chi \in \cdot \mid \mathcal{U}) - \mathbb{P}_{y_0}(Z_\chi \in \cdot \mid \mathcal{U})\|_{\mathrm{TV}} \;\Big|\; \mathcal{U} \right]. \tag{4.1}$$

Conditioning on $\mathscr{H}_{\mathrm{GREEN}}$ in the main term in (4.1), then taking expectation,

$$\sup_{x_0, y_0} \|\mathbb{P}_{x_0}(Z_\chi \in \cdot \mid \mathcal{U}) - \mathbb{P}_{y_0}(Z_\chi \in \cdot \mid \mathcal{U})\|_{\mathrm{TV}}$$

$$\leq \sup_{\mathscr{H}_{\mathrm{GREEN}}} \sup_{x_0, y_0} \left\| \mathbb{P}_{x_0}\left(Z_\chi' \in \cdot \mid \mathscr{H}_{\mathrm{GREEN}}, \mathcal{U}\right) - \mathbb{P}_{y_0}\left(Z_\chi' \in \cdot \mid \mathscr{H}_{\mathrm{GREEN}}, \mathcal{U}\right) \right\|_{\mathrm{TV}} \tag{4.2}$$

where Z_χ' is the joint distribution of $X_{\tau_{k_i}}(\chi_i)$ for $A_i \notin \mathcal{A}_{\mathrm{GREEN}}$, i.e., the projection onto cut-sets of blue or red components. (The inequality above replaced the expectation over $\mathscr{H}_{\mathrm{GREEN}}$ by a supremum, then used the fact that the values of $\{X_{\tau_{k_i}}(\chi_i) : A_i \in \mathcal{A}_{\mathrm{GREEN}}\}$ are independent of the initial condition, and so taking a projection onto the complement spin-set does not change the total-variation distance.)

Now let ν_i be the distribution of the spins at the cut-set of A_i when further conditioning that A_i is blue, i.e.,

$$\nu_i = \left(X_{\tau_{k_i}}(\chi_i) \in \cdot \;\Big|\; \mathscr{H}_{\mathrm{GREEN}}, \mathcal{U}, A_i \in \mathcal{A}_{\mathrm{BLUE}} \right),$$

and further set

$$\nu_i^* = \min_x \nu_i(x), \quad \nu = \prod_{i: A_i \notin \mathcal{A}_{\mathrm{GREEN}}} \nu_i.$$

The right-hand side of (4.2) is then clearly at most

$$2 \sup_{\mathscr{H}_{\mathrm{GREEN}}} \sup_{x_0} \left\| \mathbb{P}_{x_0}\left(Z_\chi' \in \cdot \mid \mathscr{H}_{\mathrm{GREEN}}, \mathcal{U}\right) - \nu \right\|_{\mathrm{TV}}$$

$$\leq 2 \sup_{\mathscr{H}_{\mathrm{GREEN}}} \sup_{x_0} \left[\left\| \mathbb{P}_{x_0}\left(Z_\chi' \in \cdot \mid \mathscr{H}_{\mathrm{GREEN}}, \mathcal{U}\right) - \nu \right\|_{L^2(\nu)} \wedge 1 \right]. \tag{4.3}$$

At this point we wish to appeal to the following lemma — which generalizes [27, Proposition 3.2], via the exact same proof, from unbiased coin flips to a general distribution — bounding the L^2-distance in terms of an exponential moment of the intersection between two i.i.d. configurations.

Lemma 4.3. *Let $\{\Lambda_i : i \in \mathcal{I}\}$ be a partition of Λ, and let ν_i $(i \in \mathcal{I})$ be a measure on $\{\pm 1\}^{\Lambda_i}$. For each $S \subset \mathcal{I}$, let φ_S be a measure on $\{\pm 1\}^{\cup_{i \in S} \Lambda_i}$. Let μ be a measure on configurations in $\Omega = \{\pm 1\}^\Lambda$ obtained by sampling a subset $S \subset \mathcal{I}$ via some measure $\tilde{\mu}$, then sampling $\cup_{i \in S} \Lambda_i$ via φ_S and setting each Λ_i for $i \notin S$ via an independent sample of ν_i. Letting $\nu = \prod_{i \in \mathcal{I}} \nu_i$,*

$$\|\mu - \nu\|_{L^2(\nu)}^2 \leq \left[\sum_{S, S'} \left(\prod_{i \in S \cap S'} \min_{x_i \in \{\pm 1\}^{\Lambda_i}} \nu_i(x_i) \right)^{-1} \tilde{\mu}(S) \tilde{\mu}(S') \right] - 1.$$

Proof. For any $S \subset \mathcal{I}$, let x_S denote the projection of x onto $\cup_{i \in S} \Lambda_i$. With this notation, by definition of the $L^2(\nu)$ metric (see, e.g., [30]) one has that $\|\mu - \nu\|^2_{L^2(\nu)} + 1 = \int |\mu/\nu - 1|^2 d\nu + 1$ is equal to

$$\sum_{x \in \Omega} \frac{\mu^2(x)}{\nu(x)} = \sum_{x \in \Omega} \frac{1}{\prod \nu_i(x_i)} \sum_{S,S'} \tilde{\mu}(S)\tilde{\mu}(S')\varphi_S(x_S)\varphi_{S'}(x_{S'}) \prod_{i \notin S} \nu_i(x_i) \prod_{i \notin S'} \nu_i(x_i)$$

by the definition of μ. This can in turn be rewritten as

$$\sum_{S,S'} \left(\sum_{x_{S \cap S'}} \frac{\varphi_S(x_{S \cap S'})\varphi_{S'}(x_{S \cap S'})}{\prod_{i \in S \cap S'} \nu_i(x_i)} \right) \left(\sum_{x_{(S \cup S')^c}} \prod_{i \in (S \cup S')^c} \nu_i(x_i) \right)$$

$$\cdot \left(\sum_{x_{S \setminus S'}} \varphi_S(x_{S \setminus S'} \mid x_{S \cap S'}) \right) \left(\sum_{x_{S' \setminus S}} \varphi_{S'}(x_{S' \setminus S} \mid x_{S' \cap S}) \right) \tilde{\mu}(S)\tilde{\mu}(S')$$

$$= \sum_{S,S'} \left(\sum_{x_{S \cap S'}} \frac{\varphi_S(x_{S \cap S'})\varphi_{S'}(x_{S \cap S'})}{\prod_{i \in S \cap S'} \nu_i(x_i)} \right) \tilde{\mu}(S)\tilde{\mu}(S') \,,$$

which is at most

$$\sum_{S,S'} \left(\prod_{i \in S \cap S'} \min_{x_i \in \{\pm 1\}^{\Lambda_i}} \nu_i(x_i) \right)^{-1} \left(\sum_{x_{S \cap S'}} \varphi_S(x_{S \cap S'})\varphi_{S'}(x_{S \cap S'}) \right) \tilde{\mu}(S)\tilde{\mu}(S')$$

$$\leq \sum_{S,S'} \left(\prod_{i \in S \cap S'} \min_{x_i \in \{\pm 1\}^{\Lambda_i}} \nu_i(x_i) \right)^{-1} \tilde{\mu}(S)\tilde{\mu}(S') \,. \qquad \blacksquare$$

Applying the above lemma to the quantity featured in (4.3) yields

$$\left\| \mathbb{P}_{x_0}\left(Z'_\chi \in \cdot \mid \mathcal{H}_{\text{GREEN}}, \mathcal{U} \right) - \nu \right\|^2_{L^2(\nu)} \leq \mathbb{E}\left[\prod_{A_i \in \mathcal{A}_{\text{RED}} \cap \mathcal{A}_{\text{RED}'}} \frac{1}{\nu^*_i} \; \middle| \; \mathcal{H}_{\text{GREEN}}, \mathcal{U} \right] - 1 \,,$$

where $\mathcal{A}_{\text{RED}}$ and $\mathcal{A}_{\text{RED}'}$ are two i.i.d. samples conditioned on $\mathcal{H}_{\text{GREEN}}$ and $\mathcal{U}$. Combining the last inequality with (4.1),(4.2) and (4.3), we conclude the following.

$$\bar{d}(t_\star, \mathcal{U}) \leq 2 \sup_{\mathcal{H}_{\text{GREEN}}} \sup_{x_0} \left[\left(\mathbb{E}\left[\prod_{A_i \in \mathcal{A}_{\text{RED}} \cap \mathcal{A}_{\text{RED}'}} \frac{1}{\nu^*_i} \; \middle| \; \mathcal{H}_{\text{GREEN}}, \mathcal{U} \right] - 1 \right)^{\frac{1}{2}} \wedge 1 \right]. \tag{4.4}$$

Note that the expectation above is only w.r.t. the update sequence along the interval $(0, t_{\mathsf{m}}]$. Indeed, the variables $\mathcal{A}_{\text{RED}}$ and $\mathcal{A}_{\text{RED}'}$ do not depend on the deferred randomness $\mathcal{U}'$, which in turn is embodied in the measures ν_i (and consequently, the values ν^*_i).

The expectation in the right-hand side of (4.4) is treated by the following lemma.

Lemma 4.4. *Let $\beta > \beta_c$, let $\mathcal{A}_{\text{RED}}$ and $\mathcal{A}_{\text{RED}'}$ denote the collection of components within red clusters in two independent instances of the dynamics, and define $\Psi_{\mathcal{C}}$ as in (3.1). Then*

$$\mathbb{E}\left[\prod_{A_i \in \mathcal{A}_{\text{RED}} \cap \mathcal{A}_{\text{RED}'}} \frac{1}{\nu^*_i} \; \middle| \; \mathcal{H}_{\text{GREEN}}, \mathcal{U} \right] \leq \exp\left[\sum_{\mathcal{C} \cap \mathcal{C}' \neq \emptyset} \Psi_{\mathcal{C}} \Psi_{\mathcal{C}'} \prod_{A_j \in \mathcal{C}} \frac{1}{\nu^*_j} \prod_{A_j \in \mathcal{C}'} \frac{1}{\nu^*_j} \right].$$

One should emphasize the dependence of the bound given by this lemma on $\mathcal{H}_{\text{GREEN}}$ and $\mathcal{U}$: the dependence on $\mathcal{H}_{\text{GREEN}}$ was eliminated thanks to the supremum in the definition of $\Psi_{\mathcal{C}}$. On the other hand, both $\Psi_{\mathcal{C}}$ and ν^*_j still depend on $\mathcal{U}$.

Proof of Lemma 4.4. We first claim that, if $\{Y_{\mathcal{C}}\}$ is a family of independent indicators given by

$$\mathbb{P}(Y_{\mathcal{C}} = 1) = \Psi_{\mathcal{C}},$$

then, conditioned on $\mathscr{H}_{\text{GREEN}}$, one can couple the distribution of $\mathcal{A}_{\text{RED}}$ to $\{Y_{\mathcal{C}} : \mathcal{C} \subset \mathcal{A} \setminus \mathcal{A}_{\text{GREEN}}\}$ in such a way that

$$\{\mathcal{C} : \mathcal{C} \in \text{RED}\} \subset \{\mathcal{C} : Y_{\mathcal{C}} = 1\}. \tag{4.5}$$

To see this, order all $\mathcal{C} \subset \mathcal{A} \setminus \mathcal{A}_{\text{GREEN}}$ arbitrarily as $\{\mathcal{C}_l\}_{l \geq 1}$ and let $\mathcal{F}_l$ correspond to the filtration that successively reveals $\mathbb{1}_{\{\mathcal{C}_l \in \text{RED}\}}$. Then $\mathbb{P}(\mathcal{C}_l \in \text{RED} \mid \mathcal{F}_{l-1}) \leq \Psi_{\mathcal{C}_l}$ since

$$\mathbb{P}(\mathcal{C}_l \in \text{RED} \mid \mathcal{F}_{l-1}) \leq \sup_{\mathscr{X}} \mathbb{P}\left(\mathcal{C}_l \in \text{RED} \mid \mathcal{F}_{l-1}, \{\mathcal{C}_l \in \text{RED}\} \cup \{\mathcal{C}_l \subset \text{BLUE}\}, \mathscr{H}_{\mathcal{C}_l}^- = \mathscr{X}\right),$$

and in the new conditional space the variables $\{\mathcal{C}_j \in \text{RED}\}_{j<l}$ are measurable since

   (i) the event $\{\mathcal{C}_j \in \text{RED}\}$ for any $\mathcal{C}_j$ disjoint from $\mathcal{C}_l$ is determined by the conditioning on $\mathscr{H}_{\mathcal{C}_l}^-$;

   (ii) any $\mathcal{C}_j$ nontrivially intersecting $\mathcal{C}_l$ is not red under the conditioning on $\{\mathcal{C}_l \in \text{RED}\} \cup \{\mathcal{C}_l \subset \text{BLUE}\}$.

This establishes (4.5).

   Consequently, our next claim is that for a family $\{Y_{\mathcal{C},\mathcal{C}'}\}$ of independent indicators given by

$$\mathbb{P}(Y_{\mathcal{C},\mathcal{C}'} = 1) = \Psi_{\mathcal{C}} \Psi_{\mathcal{C}'} \quad \text{for any } \mathcal{C}, \mathcal{C}' \subset \mathcal{A} \setminus \mathcal{A}_{\text{GREEN}},$$

one can couple the conditional distributions of $\mathcal{A}_{\text{RED}}$ and $\mathcal{A}_{\text{RED}'}$ given $\mathscr{H}_{\text{GREEN}}$ in such a way that

$$\prod_{\substack{A_i \in \mathcal{A}_{\text{RED}} \cap \mathcal{A}_{\text{RED}'}}} \frac{1}{\nu_i^*} \leq \prod_{\substack{\mathcal{C},\mathcal{C}' \subset \mathcal{A} \setminus \mathcal{A}_{\text{GREEN}} \\ \mathcal{C} \cap \mathcal{C}' \neq \emptyset}} \left(1 + Y_{\mathcal{C},\mathcal{C}'} \left(\prod_{A_j \in \mathcal{C}} \frac{1}{\nu_j^*} \prod_{A_j \in \mathcal{C}'} \frac{1}{\nu_j^*} - 1\right)\right). \tag{4.6}$$

To see this, take $\{Y_{\mathcal{C}}\}$ achieving (4.5) and $\{Y_{\mathcal{C}}'\}$ achieving its analog for RED'. Letting $\{(\mathcal{C}_l, \mathcal{C}_l')\}_{l \geq 1}$ be an arbitrary ordering of all pairs of potential clusters that intersect ($\mathcal{C}, \mathcal{C}' \subset \mathcal{A} \setminus \mathcal{A}_{\text{GREEN}}$ with $\mathcal{C} \cap \mathcal{C}' \neq \emptyset$), associate each pair with a variable R_l initially set to 0, then process them sequentially:

   • If $(\mathcal{C}_l, \mathcal{C}_l')$ is such that for some $j < l$ we have $R_j = 1$ and either $\mathcal{C}_j \cap \mathcal{C}_l \neq \emptyset$ or $\mathcal{C}_j' \cap \mathcal{C}_l' \neq \emptyset$, then we skip this pair (keeping $R_l = 0$).

   • Otherwise, we set R_l to be the indicator of $\{\mathcal{C}_l \in \text{RED}, \mathcal{C}_l' \in \text{RED}'\}$.

Observe that, if $\mathcal{F}_l$ denote the natural filtration corresponding to this process, then for all l we have

$$\mathbb{P}(R_l = 1 \mid \mathcal{F}_{l-1}) \leq \mathbb{P}(Y_{\mathcal{C}_l} = 1, Y_{\mathcal{C}_l'} = 1) = \Psi_{\mathcal{C}} \Psi_{\mathcal{C}'},$$

since testing if $R_l = 1$ means that we received only negative information on $\{Y_{\mathcal{C}_l} = 1\}$ and $\{Y_{\mathcal{C}_l'} = 1\}$; this implies the existence of a coupling in which $\{l : R_l = 1\} \subset \{l : Y_{\mathcal{C}_l,\mathcal{C}_l'} = 1\}$. Hence, if $A_i \in \mathcal{A}_{\text{RED}} \cap \mathcal{A}_{\text{RED}'}$ then $A_i \in \mathcal{C}_l \cap \mathcal{C}_l'$ for some l where $Y_{\mathcal{C}_l} = Y_{\mathcal{C}_l'} = 1$, so either $Y_{\mathcal{C}_l,\mathcal{C}_l'} = 1$, or else $Y_{\mathcal{C}_j,\mathcal{C}_j'} = 1$ for a previous pair $(\mathcal{C}_j, \mathcal{C}_j')$ in which $\mathcal{C}_j = \mathcal{C}_l$ or $\mathcal{C}_j' = \mathcal{C}_l'$ (a nontrivial intersection in either coordinate will not yield a red cluster). Either way, the term $1/\nu_i^*$ is accounted for in the right-hand of (4.6), and (4.6) follows.

   Taking expectations in (4.6) within the conditional space given $\mathscr{H}_{\text{GREEN}}, \mathcal{U}$, and using the definition (and independence) of the $Y_{\mathcal{C},\mathcal{C}'}$'s, we find that

$$\mathbb{E}\left[\prod_{A_i \in \mathcal{A}_{\text{RED}} \cap \mathcal{A}_{\text{RED}'}} \frac{1}{\nu_i^*} \;\middle|\; \mathscr{H}_{\text{GREEN}}, \mathcal{U}\right] \leq \prod_{\mathcal{C} \cap \mathcal{C}' \neq \emptyset} \left(1 + \Psi_{\mathcal{C}} \Psi_{\mathcal{C}'} \prod_{A_j \in \mathcal{C}} \frac{1}{\nu_j^*} \prod_{A_j \in \mathcal{C}'} \frac{1}{\nu_j^*}\right)$$

$$\leq \exp\left[\sum_{\mathcal{C} \cap \mathcal{C}' \neq \emptyset} \Psi_{\mathcal{C}} \Psi_{\mathcal{C}'} \prod_{A_j \in \mathcal{C}} \frac{1}{\nu_j^*} \prod_{A_j \in \mathcal{C}'} \frac{1}{\nu_j^*}\right]. \qquad \blacksquare$$

Corollary 4.5. *Let $\beta > \beta_c$. With the above defined $\Psi_{\mathcal{C}}$ and Ξ_j's we have*

$$\bar{d}(t_\star) \leq 4 \left(\mathbb{E} \left[\sum_{\mathcal{C} \cap \mathcal{C}' \neq \emptyset} \Psi_{\mathcal{C}} \Psi_{\mathcal{C}'} \prod_{A_j \in \mathcal{C}} \frac{1}{\Xi_j} \prod_{A_j \in \mathcal{C}'} \frac{1}{\Xi_j} \right] \right)^{1/2}. \tag{4.7}$$

Proof of Corollary 4.5. Plugging the bound from Lemma 4.4 into (4.4), then integrating over the undeferred randomness $\mathcal{U}$, produces an upper bound on the total-variation distance at time $t_\star$:

$$\bar{d}(t_\star) \leq 2 \mathbb{E} \left[\left(\exp \left[\sum_{\mathcal{C} \cap \mathcal{C}' \neq \emptyset} \Psi_{\mathcal{C}} \Psi_{\mathcal{C}'} \prod_{A_j \in \mathcal{C}} \frac{1}{\Xi_j} \prod_{A_j \in \mathcal{C}'} \frac{1}{\Xi_j} \right] - 1 \right)^{1/2} \wedge 1 \right],$$

where $\mathbb{E}$ denotes expectation w.r.t. $\mathcal{U}$, and we used the observation that $\nu_j^* \geq \Xi_j$ by construction. Indeed, ν_j^* denotes the minimal measure of a configuration of the spins in the cut-set χ_j of a blue component A_j given $\mathcal{U}, \mathscr{H}_{\text{GREEN}}$ at time τ_{k_j} (where k_j is the index of the phase optimizing the choice of the cut-set). Clearly, any particular configuration $\eta \in \{\pm 1\}^{\chi_j}$ can occur at time τ_{k_j} if *every* $x \in \chi_j$ were to receive an *oblivious deferred update* — with the appropriate new spin of η_x — before its first splitting point in the deferred phase $\mathcal{I}'_{k_j}$. Since oblivious updates occur at rate θ, this event has probability at least $\frac{1}{2}(1 - \exp(-\theta T_x)) \geq \frac{1}{4}\theta T_x$ where T_x is the length of the interval between τ_{k_j} and the first update to x in $\mathcal{I}'_{k_j}$, and the inequality used $1 - e^{-x} \geq x - x^2/2 \geq x/2$ for $x \in [0,1]$ (with $x = \theta T_x \leq 1$). The independence of the deferred updates therefore shows that $\nu_j^* \geq \Xi_j$.

Since $\sqrt{e^x - 1} \leq 2\sqrt{x}$ for $x \in [0,1]$, the inequality $(\sqrt{e^x - 1} \wedge 1) \leq 2\sqrt{x}$ holds for all $x \geq 0$; thus, Jensen's inequality allows us to derive (4.7) from the last display, as required. ∎

It now remains to show that the expectation over $\mathcal{U}$ on the right-hand side of (4.7) is at most $\varepsilon(s_\star)$ for some $\varepsilon(s_\star) > 0$ that is exponentially small in $s_\star$, which will be achieved by the following lemma.

Lemma 4.6. *Let $\beta > \beta_c$. With the above defined $\Psi_{\mathcal{C}}$ and Ξ_j's we have*

$$\mathbb{E} \left[\sum_{\mathcal{C} \cap \mathcal{C}' \neq \emptyset} \Psi_{\mathcal{C}} \Psi_{\mathcal{C}'} \prod_{A_i \in \mathcal{C}} \frac{1}{\Xi_i} \prod_{A_i \in \mathcal{C}'} \frac{1}{\Xi_i} \right] \leq e^{-\frac{1}{5} c s_\star / \lambda}.$$

Proof of Lemma 4.6. We begin by breaking up the sum over potential clusters $\mathcal{C}, \mathcal{C}'$ in the left-hand side of the sought inequality as follows: first, we will root a single component $A \in \mathcal{C} \cap \mathcal{C}'$; second, we will enumerate over the partition of these clusters into components: $\mathcal{C} = \{A_{i_j}\}$ and $\mathcal{C}' = \{A_{i_k}\}$; finally, we will sum over the the block-sets $\{B_i\}$ that are the counterparts (via $\mathcal{U}$) at time $t_\star$ to the block-sets $\{A_i\}$ at time $t_{\mathfrak{m}}$. Noting that the event $\{A_i \leftrightsquigarrow B_i\}$ — testing the consistency of $\{A_i\}$ and $\{B_i\}$ — is $\mathcal{U}$-measurable, we have

$$\mathbb{E} \left[\sum_{\mathcal{C} \cap \mathcal{C}' \neq \emptyset} \Psi_{\mathcal{C}} \Psi_{\mathcal{C}'} \prod_{A_j \in \mathcal{C}} \frac{1}{\Xi_j} \prod_{A_j \in \mathcal{C}'} \frac{1}{\Xi_j} \right]$$

$$\leq \sum_A \sum_{\substack{\mathcal{C} = \{A_{i_j}\} \ni A \\ \mathcal{C}' = \{A_{i_k}\} \ni A}} \sum_{\substack{\{B_j\} \\ \{B_k'\}}} \mathbb{E} \left[\left(\prod_j \mathbb{1}_{\{A_{i_j} \leftrightsquigarrow B_j\}} \frac{1}{\Xi_{i_j}} \right) \Psi_{\mathcal{C}} \left(\prod_k \mathbb{1}_{\{A_{i_k} \leftrightsquigarrow B_k'\}} \frac{1}{\Xi_{i_k}} \right) \Psi_{\mathcal{C}'} \right]. \tag{4.8}$$

Recall from (3.1) that Lemma 3.3 provides us with an upper bound on $\Psi_{\mathcal{C}}$ in terms of the components $\{A_{i_j}\}$ of $\mathcal{C}$ and uniformly over $\mathcal{U}$. Letting $\bar{\Psi}_{\{A_{i_j}\}}$ denote this bound (i.e., the right-hand side of (3.2))

for brevity, we can therefore deduce that the expectation in the last display is at most

$$\mathbb{E}\left[\left(\prod_j \mathbb{1}_{\{A_{i_j}\leftrightsquigarrow B_j\}}\frac{1}{\Xi_{i_j}}\right)\left(\prod_k \mathbb{1}_{\{A_{i_k}\leftrightsquigarrow B'_k\}}\frac{1}{\Xi_{i_k}}\right)\right]\bar{\Psi}_{\{A_{i_j}\}}\bar{\Psi}_{\{A_{i_k}\}}.$$

Hölder's inequality now implies that the last expectation is at most

$$\left[\mathbb{P}\left(\bigcap_j\{A_{i_j}\leftrightsquigarrow B_j\}\right)\mathbb{E}\left[\prod_j\frac{1}{\Xi_{i_j}^4}\right]\mathbb{P}\left(\bigcap_k\{A_{i_k}\leftrightsquigarrow B'_k\}\right)\mathbb{E}\left[\prod_k\frac{1}{\Xi_{i_k}^4}\right]\right]^{\frac14},$$

and when incorporating the last two steps in (4.8) it becomes possible to factorize the terms involving $\mathcal{C},\mathcal{C}'$ and altogether obtain that

$$\mathbb{E}\left[\sum_{\mathcal{C}\cap\mathcal{C}'\neq\emptyset}\Psi_{\mathcal{C}}\Psi_{\mathcal{C}'}\prod_{A_i\in\mathcal{C}}\frac{1}{\Xi_i}\prod_{A_i\in\mathcal{C}'}\frac{1}{\Xi_i}\right]\leq\sum_A\left[\sum_{\substack{\mathcal{C}=\{A_{i_j}\}\\A\in\mathcal{C}}}\sum_{\{B_j\}}\mathbb{P}\left(\bigcap_j\{A_{i_j}\leftrightsquigarrow B_j\}\right)^{\frac14}\mathbb{E}\left[\prod_j\frac{1}{\Xi_{i_j}^4}\right]^{\frac14}\bar{\Psi}_{\{A_{i_j}\}}\right]^2.$$

$$(4.9)$$

The term $\mathbb{P}(\bigcap_j\{A_{i_j}\leftrightsquigarrow B_j\})$ is bounded via Lemma 3.1. The term $\mathbb{E}[\prod_j\Xi_{i_j}^{-4}]$ is bounded via Lemma 3.2 using the observation that one can always restrict the choice of phases for the cut-sets (only worsening our bound) to be the same for all the components, whence $\prod_j\Xi_{i_j}$ identifies with a single variable Ξ whose source block-set at time $t_\star$ is $\cup_j B_j$. Finally, $\bar{\Psi}_{\{A_{i_j}\}}$ corresponds to the right-hand side of (3.2) from Lemma 3.3, in which we may decrease the exponent by a factor λ (only relaxing the bound as $\lambda>1$). Altogether, for some $c=c(\beta,d)>0$ (taken as $\frac{c_1}{4}\wedge c_2$ where c_1,c_2 are the constants from Lemmas 3.1 and 3.3, respectively) the last expression is at most

$$\sum_A\left(\sum_{\substack{\mathcal{C}=\{A_{i_j}\}\\A\in\mathcal{C}}}\sum_{\{B_j\}}\frac{2^{\lambda+3}s_\star^{4d}}{\sqrt{|\Lambda|}}e^{-c(s_\star/\lambda)\left(\left|\mathfrak{W}(\mathcal{C})-\sum\mathfrak{W}(A_{i_j})\right|^++\sum\mathfrak{W}(A_{i_j}\cup B_j)\right)+4\sum|A_{i_j}|+\frac14\sum|B_j|}\right)^2.\quad(4.10)$$

It is easy to see that since $|\mathfrak{W}(\mathcal{C})-\sum_j\mathfrak{W}(A_{i_j})|^+\geq\frac12(\mathfrak{W}(\mathcal{C})-\sum_j\mathfrak{W}(A_{i_j}\cup B_j))$, we have that

$$\sum_j\mathfrak{W}(A_{i_j}\cup B_j)+\left|\mathfrak{W}(\mathcal{C})-\sum_j\mathfrak{W}(A_{i_j})\right|^+\geq\frac12\mathfrak{W}(\mathcal{C})+\frac12\sum_j\mathfrak{W}(A_{i_j}\cup B_j).$$

Since each of the summands $|A_{i_j}|$ and $|B_j|$ in the exponent above is readily canceled by $\mathfrak{W}(A_{i_j}\cup B_j)$, we deduce that if $cs_\star/\lambda$ is large enough then (4.10) is at most

$$\frac{2^{2\lambda+6}s_\star^{8d}}{|\Lambda|}\sum_A\left(\sum_{\substack{\mathcal{C}=\{A_{i_j}\}\\A\in\mathcal{C}}}\sum_{\{B_j\}}\exp\left[-\frac{c}{4}\frac{s_\star}{\lambda}\mathfrak{W}\left(\mathcal{C}\cup\bigcup_j B_j\right)\right]\right)^2.\quad(4.11)$$

Now, the number of different lattice animals containing κ blocks and rooted at a given block X is easily seen to be at most $(2d)^{2(\kappa-1)}$, since these correspond to trees on κ vertices containing a given point in $\mathbb{Z}^d$, and one can enumerate over such trees by traveling along there edges via a depth-first-search: beginning with $2d$ options for the first edge from the root, each additional edge has at most $2d$ options (at most $2d-1$ new vertices plus one edge for backtracking, where backtracking at the root is regarded

as terminating the tree). The bound on the number of rooted trees (and hence the number of rooted lattice animals) now follows from the fact that each edge is traversed precisely twice in this manner.

Next, enumerate over collections of blocks $\{A_{i_j}, B_j\}$ with $\mathfrak{W}(\mathcal{C}) = \mathcal{S}$ and $\sum_j \mathfrak{W}(A_{i_j} \cup B_j) = \mathcal{R}$:

- There are at most $(2d)^{2(\mathcal{S}-1)}$ ways to choose $\mathcal{C}$ containing A by the above lattice animal bounds.
- There are at most $2^{\mathcal{S}}$ choices of blocks $D_j \in \mathcal{C}$ so that each $D_j \in A_{i_j}$ will be in a distinct A_{i_j}.
- There are at most $2^{\mathcal{R}}$ choices of r_j representing $\mathfrak{W}(A_{i_j} \cup B_j)$ since $\sum_j r_j = \mathcal{R}$.
- For each j there are at most $(2d)^{2(r_j-1)}$ choices of minimal lattice animals of size r_j rooted at D_j which will contain $A_{i_j} \cup B_j$. Together, this is at most $(2d)^{\sum_j 2(r-1)} \leq 2^{\mathcal{R}}$.
- For each lattice animal there are 2^{r_j} ways to assign the vertices to be either in or not in A_{i_j} and 2^{r_j} choices to be either in or not in B_j. In total this gives another $4^{\mathcal{R}}$ choices.

Altogether, we have that the number of choices of the $\{A_{i_j}, B_j\}$ is at most $2^{\mathcal{S}} 8^{\mathcal{R}} (2d)^{2(\mathcal{S}+\mathcal{R})}$. Thus,

$$\sum_{\substack{\mathcal{C}=\{A_{i_j}\} \\ A \in \mathcal{C}}} \sum_{\{B_j\}} \exp\left[-\frac{c\, s_\star}{4\, \lambda}\left(\mathfrak{W}(\mathcal{C}) + \sum_j \mathfrak{W}(A_{i_j} \cup B_j) \right) \right] \leq \sum_{\mathcal{S}, \mathcal{R} \geq 1} e^{-\frac{c}{4}\frac{s_\star}{\lambda}(\mathcal{S}+\mathcal{R})} 2^{\mathcal{S}} 8^{\mathcal{R}} (2d)^{2(\mathcal{S}+\mathcal{R})} \leq e^{-\frac{c}{5}\frac{s_\star}{\lambda}}$$

provided $s_\star$ is large enough compared to d. Plugging this in (4.11) finally gives

$$\mathbb{E}\left[\sum_{\mathcal{C} \cap \mathcal{C}' \neq \emptyset} \Psi_{\mathcal{C}} \Psi_{\mathcal{C}'} \prod_{A_i \in \mathcal{C}} \frac{1}{\Xi_i} \prod_{A_i \in \mathcal{C}'} \frac{1}{\Xi_i} \right] \leq \frac{2^{2\lambda+6} s_\star^{8d}}{|\Lambda|} \sum_A e^{-\frac{2}{5} c s_\star/\lambda} \leq e^{-\frac{1}{5} c s_\star/\lambda},$$

where the last inequality holds whenever, e.g., $s_\star \geq \lambda^2$ and λ is large enough in terms of β, d. ∎

Combining Corollary 4.5 and Lemma 4.6 shows that $\bar{d}(t_\star) \leq 4 \exp[-\frac{1}{10} c s_\star/\lambda]$, and so in particular, once we fix λ larger than some $\lambda_0(\beta, d)$, the total-variation distance at time $t_\star$ will decrease in $s_\star$ as $O(\exp[-c' s_\star])$ for some $c'(\beta, d) > 0$, concluding the proof. ∎

4.2. **Lower bound on the mixing time.** We begin with two simple lemmas, establishing exponential decay for the magnetization in time and for the correlation between spins in X_t in space.

Lemma 4.7. *There exist $c_1(\beta, d)$ and $c_2(\beta, d)$ such that for all $0 < h < t$,*

$$\mathfrak{m}_t \leq c_1 e^{-c_2 h} \mathfrak{m}_{t-h}.$$

Proof. By Lemma 2.1,

$$\mathbb{E}\big|\mathscr{F}_{\mathrm{UPD}}(v, t-h, t)\big|^2 \leq |B(v, 20dh)|^2 + \sum_{k=20dh}^{n} k^{2d} \mathbb{P}\left(\mathscr{F}_{\mathrm{SUP}}(v, t_{\mathfrak{m}} - h, t_{\mathfrak{m}}) \not\subset B(v, k) \right)$$

$$\leq |B(v, 20dh)|^2 + \sum_{k=20dh}^{n} k^{2d} e^{-ck} = O(h^{2d}).$$

Then by Cauchy-Schwarz,

$$\mathbb{E}\big|\mathscr{F}_{\mathrm{SUP}}(v, t-h, t)\big| \leq \mathbb{E}\left[\big|\mathscr{F}_{\mathrm{UPD}}(v, t-h, t)\big| \mathbb{1}_{\{\mathscr{F}_{\mathrm{SUP}}(v, t-h, t) \neq \emptyset\}} \right]$$

$$\leq \left(\mathbb{E}\left[\big|\mathscr{F}_{\mathrm{UPD}}(v, t-h, t)\big|^2 \right] \mathbb{P}(\mathscr{F}_{\mathrm{SUP}}(v, t-h, t) \neq \emptyset) \right)^{1/2} \leq O(h^d e^{-ch/2}).$$

If $\mathscr{F}_{\mathrm{SUP}}(v,0,t) \neq \emptyset$ then $\mathscr{F}_{\mathrm{SUP}}(u,t-h,t) \neq \emptyset$ for some $u \in \mathscr{F}_{\mathrm{UPD}}(v,t-h,t)$. Using the translational invariance of the torus (which implies that all vertices have the same magnetization),

$$\mathfrak{m}_t = \mathbb{P}(\mathscr{F}_{\mathrm{SUP}}(v,0,t) \neq \emptyset) \leq \mathbb{E}\left[\left|\mathscr{F}_{\mathrm{UPD}}(v,t-h,t)\right|\right] \mathbb{P}(\mathscr{F}_{\mathrm{SUP}}(v,t-h,t) \neq \emptyset) \leq O(h^d e^{-ch/2})\mathfrak{m}_{t-h},$$

as claimed.   ∎

Lemma 4.8. *There exist $c_1(\beta,d), c_2(\beta,d) > 0$ so that starting from any initial condition X_0,*

$$\mathrm{Cov}(X_t(u), X_t(v)) \leq c_1 \exp(-c_2|u-v|) \quad \text{for any } t > 0 \text{ and } u,v \in \Lambda.$$

Proof. Let $\mathcal{E}$ denote the event that the supports of u and v intersect, that is

$$\mathcal{E} = \left\{ \bigcup_{0<t'<t} \left(\mathscr{F}_{\mathrm{SUP}}(v,t',t) \cap \mathscr{F}_{\mathrm{SUP}}(v,t',t) \right) = \emptyset \right\}.$$

Let X_t' and X_t'' be two independent copies of the dynamics. By exploring the histories of the support we may couple X_t with X_t' and X_t'' so that on the event $\mathcal{E}$ 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,

$$\mathbb{E}\left[X_t(v)X_t(u)\right] = \mathbb{E}\left[X_t'(v)X_t''(u) + \left(X_t(v)X_t(u) - X_t'(v)X_t''(u)\right)\mathbb{1}_{\mathcal{E}}\right]$$
$$\leq \mathbb{E}\left[X_t'(v)\right] \mathbb{E}\left[X_t''(u)\right] + 2\mathbb{P}(\mathcal{E})$$

and so $\mathrm{Cov}(X_t(v), X_t(u)) \leq 2\mathbb{P}(\mathcal{E})$. Define the event

$$K_{v,r} = \left\{ \mathscr{F}_{\mathrm{SUP}}\left(v, t - \tfrac{r}{40ed}, t\right) = \emptyset, \ \mathscr{F}_{\mathrm{UPD}}\left(v, t - \tfrac{r}{40ed}, t\right) \subset B\left(v, \tfrac{r}{2}\right) \right\}.$$

By Lemma 2.1,

$$\mathbb{P}(K_{v,r}) \geq 1 - \exp(-c'\tfrac{r}{40ed}).$$

If $K_{v,|u-v|}$ and $K_{u,|u-v|}$ both hold then the histories of u and v do not intersect and so

$$\mathbb{P}(\mathcal{E}) \leq \mathbb{P}(K_{v,|u-v|}^c \cup K_{u,|u-v|}^c) \leq 2\exp\left(-c'\tfrac{|u-v|}{40ed}\right).$$

This completes the proof, as it implies that

$$\mathrm{Cov}(X_t(v), X_t(u)) \leq 2\mathbb{P}(\mathcal{E}) \leq 4\exp\left(-c'\tfrac{|u-v|}{40ed}\right).$$ ∎

We are now ready to prove the lower bound for the mixing time. To lower bound the total variation distance at time $t_{\mathfrak{m}} - h$ we take the magnetization as a distinguishing statistics. By Lemma 4.7,

$$\mathbb{E}\left[\sum_{v \in \Lambda} X_{t_{\mathfrak{m}}-h}^+(v)\right] = |\Lambda|\mathfrak{m}_{t_{\mathfrak{m}}-h} \geq c_1 e^{c_2 h}|\Lambda|\mathfrak{m}_{t_{\mathfrak{m}}} = c_1 e^{c_2 h}\sqrt{|\Lambda|},$$

while Lemma 4.8 implies that

$$\mathrm{Var}\left(\sum_{v \in \Lambda} X_{t_{\mathfrak{m}}-h}^+(v)\right) = \sum_{u,v \in \Lambda} \mathrm{Cov}(X_t^+(u), X_t^+(v)) = |\Lambda| \sum_{u \in \Lambda} c_1 e^{-c_2|u-v|} \leq c'|\Lambda|$$

for some $c' = c'(\beta,d) > 0$. By Chebyshev's inequality,

$$\mathbb{P}\left(\sum_{v \in \Lambda} X_{t_{\mathfrak{m}}-h}^+(v) > \frac{1}{2}|\Lambda|\mathfrak{m}_{t_{\mathfrak{m}}-h}\right) \geq 1 - \frac{\mathrm{Var}\left(\sum_{v \in \Lambda} X_{t_{\mathfrak{m}}-h}^+(v)\right)}{\frac{1}{2}\mathbb{E}\left[\sum_{v \in \Lambda} X_{t_{\mathfrak{m}}-h}^+(v)\right]} \geq 1 - c'e^{-2c_2 h}.$$

Now if σ is a configuration drawn from the stationary distribution then $\mathbb{E}[\sum_{v \in \Lambda} \sigma(v)] = 0$, and since X_t^+ converges in distribution to the stationary distribution,

$$\mathrm{Var}\left(\sum_{v \in \Lambda} \sigma(v)\right) = \lim_{t \to \infty} \mathrm{Var}\left(\sum_{v \in \Lambda} X_t^+(v)\right) \le c'|\Lambda|.$$

Hence, by Chebyshev's inequality, the probability that the magnetization is at least $\frac{1}{2}|\Lambda|\mathfrak{m}_{t_{\mathfrak{m}}-h}$ satisfies

$$\mathbb{P}\left(\sum_{v \in \Lambda} \sigma(v) > \frac{1}{2}|\Lambda|\mathfrak{m}_{t_{\mathfrak{m}}-h}\right) \le c'e^{-2c_2 h}.$$

Thus, considering this as the distinguishing characteristic yields

$$d_{\mathrm{TV}}(X_t^+, \pi) \ge \mathbb{P}\left(\sum_{v \in \Lambda} X_{t_{\mathfrak{m}}-h}^+(v) > \tfrac{1}{2}|\Lambda|\mathfrak{m}_{t_{\mathfrak{m}}-h}\right) - \mathbb{P}\left(\sum_{v \in \Lambda} \sigma(v) > \tfrac{1}{2}|\Lambda|\mathfrak{m}_{t_{\mathfrak{m}}-h}\right) \ge 1 - 2c'e^{-2c_2 h},$$

concluding the proof of the lower bound. ■

5. Analysis of percolation components and clusters

5.1. Percolation component structure: Proof of Lemma 3.1.

To each $v \in \Lambda$, associate the *column* $Q_v = B(v, s_\star^{3/2}) \times (t_{\mathfrak{m}}, t_\star]$ in the space-time slab $\mathbb{Z}^d \times (t_{\mathfrak{m}}, t_\star]$. Recall that the history of vertices gives rises to edges in the above space-time slab as per the description in §3. Namely, if at time t there is a non-oblivious update at site x we mark up to $2d$ intervals $[(x,t),(y,t)]$ for $x \sim y$, and if a site x is born at time t' and dies at time t'' we mark the interval $[(x,t),(x,t'')]$. Given these marked intervals, we say that a column Q_v is *exceptional* if it contains one of the following:

- *spatial crossing*: a path connecting (x,t) to (y,t') for $|x - y| \ge \frac{1}{2}s_\star^{3/2}$ and some $t, t' \in (t_{\mathfrak{m}}, t_\star]$.
- *temporal crossing*: a path connecting $(v, t_\star)$ to $B(v, s_\star^{3/2}) \times \{t_{\mathfrak{m}}\}$.

Eq. (2.6) from Lemma 2.1 tells us that, even if all phases were deferred (i.e., the update support were ignored and vertices would never die) then the probability of witnessing a spatial crossing of length $s_\star^{3/2}$ starting from a given site x during a time interval of $s_\star$ is at most $\exp(-s_\star^{3/2})$ provided that $s_\star > (20d)^2$. In lieu of such a spatial crossing, the number of points reachable from $(v, t_\star)$ at time $\tau_k - 1 = t_\star - 1$ (marking the transition between the deferred phase $\mathcal{I}_k'$ and the regular phase $\mathcal{I}_k$) is $O(s_\star^{3d/2})$. By Eq. (2.5) from that same lemma, there exists some $c_1 = c_1(\beta, d) > 0$ so that the probability that the history of a given u would survive the interval $\mathcal{I}_k$ is at most $2\exp(-c_1 \frac{s_\star}{\lambda})$. A union bound now shows that, overall, the probability that Q_v is exceptional is $O(s_\star^{3d/2} \exp(-c_1 \frac{s_\star}{\lambda}))$, which is at most $\exp(-\frac{c_1}{2} \frac{s_\star}{\lambda})$ if, say, $s_\star \ge \lambda^2$ and λ is large enough in terms of β, d.

Consider now the collection of block-set pairs $\{(R_i, S_i)\}$. If $R_i \leftrightsquigarrow S_i$ on account of some component Υ_{j_i} at times $t_{\mathfrak{m}}$ and $t_\star$ (i.e., Υ_{j_i} is minimally covered by S_i while $\mathscr{H}_{\Upsilon_{j_i}}(t_{\mathfrak{m}})$ is minimally covered by R_i) then every block $S \in S_i$ contains some $v \in \Upsilon_{j_i}$ such that $(v, t_\star)$ is connected by a path (arising from the aforementioned marked intervals) to $(R_i, t_{\mathfrak{m}})$ and every R_i contains some $w \in \mathscr{H}_{\Upsilon_{j_i}}(t_{\mathfrak{m}})$ such that $(w, t_{\mathfrak{m}})$ is connected to $(S_i, t_\star)$. Moreover, the set of blocks traversed by these paths necessarily forms a lattice animal (by our definition of the component Υ_{j_i} via the equivalence relation on blocks according to intersecting histories or adjacency at times $t_{\mathfrak{m}}$ or $t_\star$). We claim that for any block X in this lattice animal, either X contains some vertex v such that Q_v is exceptional, or one of its $2d$ neighboring blocks does (and belongs to the lattice animal). Indeed, take $x \in X$ such that there is a path P from some

$v \in S_i$ to some $w \in R_i$ going through x (such a path exists by the construction of the lattice animal). If P is contained in $B(x, \frac{1}{2}s_\star^{3/2}) \times (t_{\mathfrak{m}}, t_\star]$, and hence also in $B(v, s_\star^{3/2})$, then it gives rise to a temporal crossing in Q_v and v belongs either to a neighboring block of X or to X itself. Otherwise, P visits both x and some $y \in \partial B(x, \frac{1}{2}s_\star^{3/2})$ and in doing so gives rise to a spatial crossing in Q_x, as claimed.

It follows that if (R_i, S_i) are the blocks corresponding to the components Υ_{j_i} for all i then there are pairwise disjoint lattice animals, with $m_i \geq \mathfrak{W}(R_i \cup S_i)$ blocks each (recall that $\mathfrak{W}(S)$ is the smallest number of blocks in a lattice animal containing S), such that each block either contains some v for which Q_v is exceptional, or it has a neighboring block with such a vertex v. Therefore, by going through the blocks in the lattice animals according to an arbitrary ordering, one can find a subset S of at least $\sum m_i/(2d+1)$ blocks, such that each block in S contains a vertex with an exceptional column. Similarly, we can arrive at a subset $S' \subset S$ of size at least $\sum m_i/(2d+1)^2$ such that every pair of blocks in it has distance (in blocks) at least 2. Since the event that Q_v is exceptional depends only on the updates within $B(v, s_\star^{3/2})$, the distances between the blocks in S' ensure that the events of containing such a vertex v are mutually independent. Hence, the probability that a given collection of lattice animals complies with the event $\{R_i \leftrightsquigarrow S_i\}$ for all i is at most $\exp[-\frac{c_1}{2}\frac{s_\star}{\lambda}(2d+1)^{-2}\sum m_i]$, or $\exp(-c_2\frac{s_\star}{\lambda}\sum m_i)$ for $c_2 = c_2(\beta, d)$.

Finally, recall from the discussion below (4.11) that the number of different lattice animals containing m blocks and rooted at a given block is at most $(2d)^{2(m-1)}$. Combined with the preceding discussion, using $m_i \geq \mathfrak{W}(R_i \cup S_i)$ we find that

$$\mathbb{P}(\cap_i\{R_i \leftrightsquigarrow S_i\}) \leq \prod_i \sum_{m_i \geq \mathfrak{W}(R_i \cup S_i)} (2d)^{2m_i} e^{-c_2\frac{s_\star}{\lambda}m_i} \leq \exp\left[-\frac{c_2}{2}\frac{s_\star}{\lambda}\sum_i \mathfrak{W}(R_i \cup S_i)\right]$$

if for instance $s_\star \geq 4\lambda\log(2d)/c_2$, readily guaranteed when $s_\star \geq \lambda^2$ for any λ that is sufficiently large in terms of β, d. ∎

5.2. **Cut-sets estimates: Proof of Lemma 3.2.** Partition the space-time slab $\Lambda \times (t_{\mathfrak{m}}, t_\star]$ into *cubes* of the form $Q \times (t, t+r]$ where r is some large integer to be later specified (its value will depend only on β and d) and $Q \subset \mathbb{Z}^d$ is a box of side-length r^2. We will refer to $Q^+ \times (t, t+r]$ for $Q^+ := B(Q, r^{3/2})$ as the corresponding *extended cube*. Let us first focus on some regular phase $\mathcal{I}_k$. Similar to the argument from the proof of Lemma 3.1 (yet modified slightly), we will say that a given cube $Q \times (t, t+r]$ is *exceptional* if one of the following conditions is met:

- *spatial crossing*: the cube has a path connecting (x, t') to (y, t'') for some $x, y \in Q$ such that $|x - y| \geq r^{3/2}$.
- *temporal crossing*: the extended cube has a path connecting $(x, t+r)$ to (y, t) for some $x, y \in Q^+$.

As before, the probability that a given cube contains a spatial crossing is $O(r^{2d}\exp(-r^{3/2}))$ provided that $r > (20d)^2$, by the bound from Eq. (2.6). Similarly, the probability of the aforementioned temporal crossing within the regular phase is $O(r^{2d}\exp(-c_1 r))$ for some $c_1 = c_1(\beta, d) > 0$ by Eq. (2.5). Combining the two, the probability that a cube is exceptional is at most $\exp(-c_2 r)$ for some $c_2 = c_2(\beta, d) > 0$ if r is a large enough in terms of β, d.

Next, break the time interval $\mathcal{I}_k = (\tau_{k-1}, \tau_k - 1]$ into length-r subintervals $\mathcal{I}_{k,1}, \ldots, \mathcal{I}_{k,m}$ (so that $m = \frac{s_\star}{(\lambda-1)r}$) in reverse chronological order, i.e.,

$$\mathcal{I}_{k,l} = (\tau_k - lr - 1, \tau_k - (l-1)r - 1] \qquad (l = 1, \ldots, m).$$

Further let $\mathcal{I}_{k,0} = \tau_k - 1$, and let $Y_{k,l}$ for $l = 0, \ldots, m$ count the number of cubes $Q \times \mathcal{I}_{k,l}$ (boxes when $l = 0$) that, at time $t = \tau_k - (l-1)r - 1$ (the end of the subinterval $\mathcal{I}_{k,l}$), intersect the history of S.

Our next goal is now to bound the exponential moments of $Y_{k,m}$, the number of cubes intersecting the history of S at time τ_{k-1}, which will be achieved by the following claim:

Claim 5.1. *For any $k = 1, \ldots, \lambda$, the above defined variables $(Y_{k,l})$ satisfy*

$$\mathbb{E}\left[e^{aY_{k,m}}\right] \leq 1 + \left(\tfrac{3}{4}\right)^m \exp\left[\left(\tfrac{2}{3}\right)^m aY_{k,0}\right] \quad \text{for any } 0 < a < \tfrac{1}{3}. \tag{5.1}$$

Proof. Throughout the proof of the claim we drop the subscript k from the $Y_{k,l}$'s and simply write (Y_l).

If $v \in Q \times \tau_k - (l-1)r - 1$ belongs to the history-line, we can trace its origin in the cube $Q \times \mathcal{I}_{k,l-1}$ and necessarily either that cube is exceptional or one of its $2d$ neighbors is (as otherwise there will not be a path from v making it to time $t_\star$). Hence, $Y_{l+1} \leq (2d+1)X_{l+1}$, where X_{l+1} counts the number of exceptional cubes in the k-th subinterval. Moreover, starting from Y_l cubes covering the history, the set of exceptional cubes counted by X_{l+1} is comprised of Y_l lattice animals — each rooted at one of those Y_l cubes. So, if $Y_l = a$ for some integer a and we consider lattice animals of sizes $w_1, \ldots, w_a$ (cubes) for each of these, the number of configurations for these lattice animals would be at most $(2d)^{2\sum w_i}$ as was noted in the proof of Lemma 3.1. Out of these, we can always extract a subset of $(\sum w_i)/(2d+1)$ cubes which are pairwise non-adjacent, whereby the events of being exceptional are mutually independent.

Combining these ingredients, and setting $\delta = \tfrac{1}{2}c_2(2d+1)^{-2}$, if $\mathcal{F}_{k,l}$ is the σ-algebra generated by the updates in the subintervals $\mathcal{I}_{k,l'}$ for $l' \leq l$ then

$$\mathbb{E}\left[e^{\delta r Y_{l+1}} \mid \mathcal{F}_{k,l}\right] \leq \mathbb{E}\left[e^{(2d+1)\delta r X_{l+1}} \mid \mathcal{F}_{k,l}\right]$$

$$\leq \sum_{w_1,\ldots,w_{Y_l}} \exp\left[\left[(2d+1)\delta r + 2\log(2d) - c_2(2d+1)^{-1}r\right]\sum w_i\right]$$

$$= \left[\sum_w e^{-[(2d+1)\delta r - 2\log(2d)]w}\right]^{Y_l} \leq \left[\sum_w e^{-2d\delta rw}\right]^{Y_l} \leq \exp\left[e^{-\delta r}Y_l\right], \tag{5.2}$$

where the last two inequalities hold provided that δr is sufficiently large, i.e., when r a large enough function of β and d. In particular, by Markov's inequality this implies that for any $y > 0$,

$$\mathbb{P}\left(Y_{l+1} \geq y \mid \mathcal{F}_{k,l}\right) \leq \exp\left[e^{-\delta r}Y_l - \delta ry\right] \leq \exp\left[\tfrac{1}{4}Y_l - y\right],$$

provided that δr is large. This enables us to complement the bound in (5.2) when taking a small factor instead of δr; namely, for any $0 < a < 1$ we have

$$\mathbb{E}\left[e^{aY_{l+1}} \mid \mathcal{F}_{k,l}\right] = \int_0^\infty \mathbb{P}\left(e^{aY_{l+1}} \geq t \mid \mathcal{F}_{k,l}\right) \leq \int_0^\infty \left(1 \wedge \frac{\exp\left[\tfrac{1}{4}Y_l\right]}{t^{1/a}}\right)dt$$

$$= e^{\frac{a}{4}Y_l} + e^{\frac{1}{4}Y_l}\int_{e^{\frac{a}{4}Y_l}}^\infty t^{-\frac{1}{a}}dt = e^{\frac{a}{4}Y_l} + \frac{a}{1-a}e^{\frac{a}{4}Y_l} = \frac{e^{\frac{a}{4}Y_l}}{1-a}.$$

When $Y_l \geq 6$ we can upper bound the last exponent by $\exp(-\tfrac{3}{2}a + \tfrac{1}{2}aY_l)$ and get that for any $0 < a < \tfrac{1}{3}$,

$$\mathbb{E}\left[e^{aY_{l+1}} - 1 \mid \mathcal{F}_{k,l}, Y_l \geq 6\right] \leq \frac{e^{-\frac{3}{2}a}}{1-a}e^{\frac{a}{2}Y_l} - 1 \leq \frac{3}{4}\left(e^{\frac{2}{3}aY_l} - 1\right),$$

where the last inequality used $1 - a \geq \exp(-\tfrac{a}{1-a}) \geq \exp(-\tfrac{3}{2}a)$ for $0 < a < \tfrac{1}{3}$, followed by the fact that $e^{x/\alpha} - 1 \leq \alpha(e^{2x} - 1)$ for any $0 < \alpha \leq 1$ and $x \geq 0$ thanks to Jensen's inequality. On the other hand,

if $Y_l \leq 6$ then again by Jensen's inequality (now taking $\alpha = a/\delta r$) and Eq. (5.2),

$$\mathbb{E}\left[e^{aY_{l+1}} - 1 \mid \mathcal{F}_{k,l}, \, Y_l \leq 6\right] \leq \frac{a}{\delta r}\left(e^{e^{-\delta r}Y_l} - 1\right) \leq \frac{3}{4}\left(e^{\frac{2}{3}aY_l} - 1\right),$$

with the last inequality justified since $\exp(e^{-\delta r}Y_l) \leq \exp(6e^{-\delta r}) \leq 2$ for large δr, so its left-hand side is at most $a/\delta r \leq a/3$ (again for large δr), while using $Y_l \geq 1$ in its right-hand side (when $Y_l = 0$ both sides are 0) shows it is always at least $a/2$.

We have thus established the above relation for all values of Y_l; iterating it through the m subintervals of $\mathcal{I}_k$ yields (5.1), as required.   ∎

Moving our attention to the deferred phase $\mathcal{I}'_k$, here we would like to stochastically dominate the number of vertices in the history at any given time by a rescaled pure birth process $Z_{k,t}$ along a unit interval, where each particle adds $2d$ new ones at rate 1 (recall that by definition particles do not die in deferred phases, and their splitting rate is $1 + \tanh(-2d\beta) < 1$) and furthermore, every vertex receives an *extra update* at time τ_k. Indeed, these can only increase the size of the history at $\tau_k - 1$, which in turn can only increase the quantity $\exp(\sum_i 4\Xi_i)$ (by introducing additional cut-vertices in deferred phases further down the history) that we ultimately wish to bound.

Overestimating the splitting rate suffices for our purposes and simplifies the exposition. On the other hand, introducing the extra update at time τ_k plays a much more significant role: Let M_k denote the number of vertices in the history at the beginning of each phase $\mathcal{I}'_k$. By the discussion above, the variables M_k in our process dominate those in the original dynamics, and so $(\Xi_k) \succcurlyeq (\Xi_k^+)$ jointly, where

$$\Xi_k^+ = \prod_{v \in M_k} \left(\tfrac{1}{4}\theta T_{v,k}\right) \quad \text{for} \quad T_{v,k} \sim (\mathrm{Exp}(1) \wedge 1) \tag{5.3}$$

is the analog of Ξ_k in the modified process (the variable $T_{v,k}$ corresponding to what would be the update time to $v \in M_k$ nearest to τ_k in $\mathcal{I}'_k$ in lieu of the extra update at time τ_k). Crucially, thanks to the extra updates, Ξ_k^+ depends only on M_k and has no effect on the history going further back (and in particular on the M_j's for $j < k$). Therefore, we will (ultimately) condition on the values of all the M_k's, and thereafter the variables (Ξ_k^+) will be readily estimated, being conditionally independent.

Indexing the time $0 \leq t \leq 1$ of the process $Z_{k,t}$ in reverse chronological order along $\mathcal{I}'_k$ (identifying $t = 0$ and $t = 1$ with $\tau_k, \tau_k - 1$, resp.), the exponential moments of $Z_{k,1}$ can be estimated as follows.

Claim 5.2. *For any $k = 1, \ldots, \lambda$, the above defined variables $(Z_{k,t})$ satisfy*

$$\mathbb{E}\left[e^{a_1 Z_{k,1}} \mid Z_{k,0}\right] \leq \exp\left[2e^{2d}a_1 Z_{k,0}\right] \quad \text{for any } 0 < a_1 \leq \tfrac{1}{2d}\log\left(\tfrac{1}{1-e^{-3d}}\right). \tag{5.4}$$

Proof. Throughout the proof of the claim, put Z_t as short for $Z_{k,t}$ for brevity.

One easily sees that for any $\alpha > 0$,

$$\frac{d}{dt}\mathbb{E}\left[e^{\alpha(t)Z_t}\right] = \mathbb{E}\left[\left[\alpha'(t)Z_t + (e^{2d\alpha(t)} - 1)Z_t\right]e^{\alpha(t)Z_t}\right]$$

since $\frac{d}{dt}\mathbb{E}[e^{\alpha Z_t} \mid Z_t] = \lim_{h \to 0} e^{\alpha Z_t}(e^{2d\alpha} - 1)\mathbb{P}(Z_{t+h} \neq Z_t \mid Z_t)$ for fixed α. Taking $\alpha(t)$ to be the solution to $\alpha'(t) + \exp[2d\alpha(t)] - 1 = 0$, namely

$$\alpha(t) = \frac{1}{2d}\log\left(\frac{1}{1 - \zeta e^{-2dt}}\right) \quad \text{for } \zeta > 0,$$

we find that $\exp[\alpha(t)Z_t]$ is a martingale and in particular

$$\mathbb{E}\left[e^{\alpha(1)Z_1} \mid Z_0\right] = e^{\alpha(0)Z_0}. \tag{5.5}$$

Therefore, if we set

$$\zeta = e^{2d}\left(1 - e^{-2da_1}\right) \quad \text{for} \quad 0 < a_1 \leq \frac{1}{2d}\log\left(\frac{1}{1 - e^{-3d}}\right)$$

then $0 < \zeta \leq e^{-d}$ and so $\alpha(t)$ is real and decreasing along $[0,1]$ to $\alpha(1) = a_1$. For this choice of parameters we obtain that

$$\alpha(0) \leq \frac{1}{2d}\log\left(\frac{1}{1-\zeta}\right) \leq \frac{1}{2d}\frac{e^{2d}(1 - e^{-2da_1})}{1 - e^{-d}} \leq 2e^{2d}a_1,$$

using that $1/(1 - e^{-d}) \leq 2$ for any $d \geq 1$ and $1 - x \leq e^{-x}$ for $x > 0$. Overall, for any small enough a_1 in terms of d (as in the condition above, matching the one in (5.4)) we have by (5.5) that $\mathbb{E}[e^{a_1 Z_1} \mid Z_0] \leq \exp[2e^{2d}a_1 Z_0]$. ∎

Going through the regular phase will enable us to apply Claim 5.2 with a value of a_1 which is exponentially small in $s_\star$, let alone small enough in terms of d, easily satisfying the upper bound of roughly $e^{-3d}/2d$ from the condition in (5.4).

Putting together the analysis of the deferred and regular phases $\mathcal{I}'_k, \mathcal{I}_k$ in the last two claims, we can establish a recursion for M_k, the number of vertices in $\mathscr{H}_S(\tau_k)$. Using $Y_{k,0} \leq Z_{k,1}$ while $M_k \leq r^{2d}Y_{k,m}$ (by crudely taking the entire volume of each of the cubes that survived to that point), and recalling (5.1), gives

$$\mathbb{E}\left[e^{aM_k} \mid \mathcal{F}_{k+1}\right] \leq 1 + \left(\tfrac{3}{4}\right)^m \left(\exp\left[2(er)^{2d}\left(\tfrac{2}{3}\right)^m aM_{k+1}\right] - 1\right)$$

as long as $a < \frac{1}{3}r^{-2d}$ (to have $a' = r^{2d}a$ qualify for an application of (5.1)). Setting

$$\hat{a} = \tfrac{1}{4}r^{-2d} \tag{5.6}$$

and seeing as for large enough $s_\star$ (and therefore large enough m) compared to r and d, the pre-factor of M_{k+1} is at most $\left(\tfrac{3}{4}\right)^m$, we finally arrive at

$$\begin{aligned}
\mathbb{E}\left[e^{\hat{a}M_k} \mid \mathcal{F}_{k+1}\right] &\leq 1 + \left(\tfrac{3}{4}\right)^m \left(\exp\left[\left(\tfrac{3}{4}\right)^m \hat{a}M_{k+1}\right] - 1\right), \\
M_\lambda &\leq s_\star^{2d}|S|.
\end{aligned} \tag{5.7}$$

We will now utilize (5.7) for a bound on the probability that the median of the M_k's exceeds a given integer $b \geq 0$. More precisely, consider the event that the median of $\{M_0, M_1, \ldots, M_{\lambda-2}\}$, which we denote as $\mathrm{med}_{k<\lambda-1} M_k$, exceeds b (it will suffice for our purpose to consider this event — which excludes $M_{\lambda-1}$ before taking the median — and it is convenient to do so since M_λ was pre-given as input, and hence $\mathcal{I}_\lambda$ is exceptional compared to any other $\mathcal{I}_k$, where we have better control over M_k). To this end, notice that if $\max\{M_k : k < \lambda\} \leq \lambda b$ then the event $\{\mathrm{med}_{k<\lambda-1} M_k > b\}$ necessitates at least $(\lambda - 1)/2$ values of $1 \leq k \leq \lambda - 1$ for which $M_{k-1} \geq b$ even though $M_k \leq \lambda b$. Therefore,

$$\mathbb{P}\left(\mathrm{med}_{k<\lambda-1} M_k > b\right) \leq \mathbb{P}\left(\max_{k<\lambda} M_k \geq \lambda b\right) + 2^\lambda \left[\sup_{k<\lambda}\mathbb{P}\left(M_{k-1} > b \mid M_k \leq \lambda b\right)\right]^{\frac{\lambda-1}{2}}. \tag{5.8}$$

The first term in the right-hand side of (5.8) can be estimated via (5.7):

$$\mathbb{P}\left(\max_{k<\lambda} M_k \geq \lambda b\right) \leq \sum_{k<\lambda} \mathbb{P}\left(e^{\hat{a}M_k} \geq e^{\hat{a}\lambda b}\right) \leq \lambda \exp\left[-\lambda\hat{a}b + (\tfrac{3}{4})^m \hat{a}M_\lambda\right]$$

$$\leq \lambda \exp\left[-\lambda\hat{a}b + (\tfrac{3}{4})^m (\tfrac{s_\star}{r})^{2d}|S|\right].$$

Similarly, for the second term, we get from (5.7) that for any $k < \lambda$,

$$\mathbb{P}\left(M_{k-1} > b \mid M_k \leq \lambda b\right) \leq \exp\left[-\hat{a}b + (\tfrac{3}{4})^m \hat{a}\lambda b\right] \leq \exp\left[-\tfrac{1}{2}\hat{a}b\right]$$

provided that $s_\star$ (and hence m) is large enough in terms of λ (so $(\tfrac{3}{4})^m \lambda < \tfrac{1}{2}$). Plugging these two inequalities in (5.8), while using that $(\lambda-1)/2 > \lambda/3$ for λ large and $(\tfrac{3}{4})^m s_\star^{2d} \leq 1$ for $s_\star$ large enough in terms of r and λ, yields

$$\mathbb{P}\left(\underset{k<\lambda-1}{\mathrm{med}}\, M_k > b\right) \leq 2^{\lambda+1} \exp\left[-\tfrac{1}{6}\lambda\hat{a}b + |S|\right]. \tag{5.9}$$

The final step is to derive the desired upper bound on $\min_k\{(\Xi_k^+)^{-4}\}$ from the estimate (5.9) on the median of the M_k's. Write

$$\mathbb{E}\left[\min_k\left\{(\Xi_k^+)^{-4}\right\} \mid \{M_k\}, \underset{k<\lambda-1}{\mathrm{med}}\, M_k = b\right] = \int dt\, \mathbb{P}\left(\min_k\left\{(\Xi_k^+)^{-\frac{1}{2}}\right\} \geq t^{\frac{1}{8}} \mid \{M_k\}, \underset{k<\lambda-1}{\mathrm{med}}\, M_k = b\right), \tag{5.10}$$

consider some $t > 1$ and $b \geq 0$ and condition on the event $\mathrm{med}_{k<\lambda-1} M_k = b$. Revisiting (5.3), there are at least $\frac{\lambda-1}{2}$ values of $k \in \{0, \ldots, \lambda-1\}$ such that

$$\Xi_k^+ \succcurlyeq \prod_{j=1}^{b}\left(\tfrac{1}{4}\theta T_{v_j,k}\right) \quad \text{for i.i.d. } T_{v_j,k} \sim (\mathrm{Exp}(1) \wedge 1),$$

whence, by the independence of the $T_{v_j,k}$'s, if $T \sim \mathrm{Exp}(1) \wedge 1$ then

$$\mathbb{E}\left[(\Xi_k^+)^{-\frac{1}{2}}\right] \leq \left(\mathbb{E}\left[\frac{2}{\sqrt{\theta T}}\right]\right)^b.$$

The expectation above (involving a single T) is easily seen to be equal to

$$\int dx\, \mathbb{P}\left(T < \frac{4}{\theta x^2}\right) = O\left(\int \frac{dx}{\theta x^2}\right) < C_{\beta,d},$$

for some $C_{\beta,d} > 1$ depending only on θ. Hence, by Markov's inequality, under the above conditioning we have

$$\mathbb{P}\left((\Xi_k^+)^{-\frac{1}{2}} \geq t^{1/8}\right) \leq C_{\beta,d}^b t^{-1/8},$$

and already the first 10 (say) out of these $\frac{\lambda-1}{2}$ values of k show that

$$\mathbb{P}\left(\min_{k<\lambda-1}\left\{(\Xi_k^+)^{-\frac{1}{2}}\right\} \geq t^{1/8}\right) \leq (C_{\beta,d})^{10b} t^{-5/4}.$$

(Here we could replace 10 by any integer larger than 8, and it is convenient to use an absolute constant rather than a function of λ so as to keep the effect of the constant $C_{\beta,d}$ under control). Using (5.10) we find that

$$\mathbb{E}\left[\min_{k<\lambda-1}\{(\Xi_k^+)^{-4}\} \mid \{M_k\}, \operatorname*{med}_{k<\lambda-1} M_k = b\right] \leq \int dt \frac{(C_{\beta,d})^{10b}}{t^{5/4}} = 4(C_{\beta,d})^{10b},$$

and an integration with respect to $\mathbb{P}(\operatorname{med}_{k<\lambda-1} M_k = b)$ via (5.9) establishes that as long as, say, $\lambda\hat{a} > 100\log C_{\beta,d}$, we have

$$\mathbb{E}\left[\min_{k<\lambda-1}\{(\Xi_k^+)^{-4}\}\right] \leq 2^{\lambda+3}e^{|S|}.$$

In summary, the required result holds for a choice of $r = \lambda^{3d}$ provided that λ is large enough in terms of d, β (so r is large enough in terms of these as well, while (as we recall that $\hat{a} = \frac{1}{4}r^{-2d}$) in addition $\lambda\hat{a} = \frac{1}{4}r^d$ is large) and that $s_\star$ is then large enough in terms of λ (so $m \geq s_\star/(r\lambda)$ is large). For these choice, we may take, e.g., $s_\star \geq \lambda^{10d}$, whence $m \geq s\lambda^{-(3d+1)} \geq \sqrt{s}$ and all the requirements above are met for λ large enough in terms of d, β. ∎

5.3. Blue percolation clusters given the history of their exterior.

In this section we prove the following *lower bound* on the probability of a cluster to be BLUE given the update sequence $\mathcal{U}$ along the $(t_{\mathfrak{m}}, t_\star]$ and the complete history up to time $t_{\mathfrak{m}}$ of every vertex in its exterior:

Lemma 5.3. *There exists $s_0(\beta, d) > 0$ so that for any $s_\star > s_0$, any sufficiently large n and any $\mathcal{C} \subset \mathcal{A}$,*

$$\inf_{\mathscr{H}_{\mathcal{C}}^-} \mathbb{P}\left(\bigcap_{A_i \in \mathcal{C}}\{A_i \in \text{BLUE}\} \,\Big|\, \mathscr{H}_{\mathcal{C}}^-, \mathcal{U}\right) \geq e^{-\sum_{A_i \in \mathcal{C}}|A_i|},$$

where the infimum is over all $\mathcal{C}$-compatible histories.

Proof. Since $\mathscr{H}_{\mathcal{C}}^-$ is $\mathcal{C}$-compatible, the histories of all $A \in \mathcal{A} \setminus \mathcal{C}$ do not enter $B(\mathcal{C}, s_\star^2/3)$ before time $s_\star$. Therefore, it is enough to verify for all $A_i \in \mathcal{C}$ that $\cup_{t_{\mathfrak{m}}-s_\star<t<t_{\mathfrak{m}}}\mathscr{F}_{\text{SUP}}(A_i, t, t_{\mathfrak{m}}) \subset B(A_i, s_\star^2/3)$ and that $\mathscr{F}_{\text{SUP}}(A_i, t_{\mathfrak{m}} - s_\star, t_{\mathfrak{m}}) = \emptyset$. Since these events depend on disjoint updates and do not depend on $\mathcal{U}$,

$$\inf_{\mathscr{H}_{\mathcal{C}}^-} \mathbb{P}\left(\bigcap_{A \in \mathcal{C}}\{A_i \in \text{BLUE}\} \mid \mathscr{H}_{\mathcal{C}}^-\right) = \prod_{A_i \in \mathcal{C}} \mathbb{P}\left(\mathscr{F}_{\text{SUP}}(A_i, t_{\mathfrak{m}} - s_\star, t_{\mathfrak{m}}) = \emptyset, \bigcup_{t_{\mathfrak{m}}-s_\star<t<t_{\mathfrak{m}}}\mathscr{F}_{\text{SUP}}(A_i, t, t_{\mathfrak{m}}) \subset B(A_i, s_\star^2/3)\right),$$

and so we will treat the A_i's separately. For any A_i we cover $B(A_i, s_\star^2/3)$ with a set of tiles as follows. Let $0 = r_0 < r_1 < \ldots < r_\ell = n$ be such that $r_k - r_{k-1} \in \{s_\star^{4d}, 2s_\star^{4d}\}$. For each $u \in [s_\star^{4d}]^d$ and $k \in [\ell]^d$ denote

$$V_{k,u} := \{u_1 + r_{k_1-1} + 1, \ldots, u_1 + r_{k_1}\} \times \ldots \times \{u_d + r_{k_d-1} + 1, \ldots, u_d + r_{k_d}\}$$

where we embed $\{u_j + r_{k_j-1} + 1, \ldots, u_j + r_{k_j}\}$ into $\{1, \ldots, n\}$ modulo n. Let $\partial V_{k,u}$ denote the interior boundary of $V_{k,u}$, that is the subset of vertices of $V_{k,u}$ adjacent to a vertex in its complement. Then by construction

$$\frac{1}{|[s_\star^{4d}]^d|}\sum_{u\in[s_\star^{4d}]^d}\sum_{k\in[\ell]^d}|\partial V_{k,u} \cap B(A_i, s_\star^2/3)| \leq \frac{2d|B(A_i, s_\star^2/3)|}{s_\star^{4d}},$$

since in each vertex v and each coordinate i there are at most two choices of u_i for which v will be on the boundary of a block in coordinate i. Hence, it is possible for us to choose some $u \in [s_\star^{4d}]^d$

such that $\sum_{k \in [\ell]^d} |\partial V_{k,u} \cap B(A_i, s_\star^2/3)| \leq \frac{2d|B(A_i, s_\star^2/3)|}{s_\star^{4d}}$. Let $\mathcal{V}$ denote the set of tiles $V_{k,u}$ such that $V_{k,u} \cap B(A_i, s_\star^2/3) \neq \emptyset$. Each block of A_i is in at most 2^d tiles, so $|\mathcal{V}| \leq 2^d |A_i|$.

For each $V_k \in \mathcal{V}$, let $\tilde{V}_k$ denote an isomorphic copy of the graph induced by V_k disconnected from everything else together with a graph bijection $\varphi_k : \tilde{V}_k \to V_k$. Let $\tilde{\Lambda}_i = \cup_{V_k \in \mathcal{V}} \tilde{V}_k$ and let $\tilde{X}_t$ denote the Glauber dynamics on $\tilde{\Lambda}_i$ started from the all-plus configuration at time $t_{\mathfrak{m}} - s_\star$ and run until time $t_{\mathfrak{m}}$. Since the $\tilde{V}_k$ are disconnected, the projections of the chain onto each $\tilde{V}_k$ are independent. We define the update and support functions $\tilde{\mathscr{F}}_{\text{UPD}}$ and $\tilde{\mathscr{F}}_{\text{SUP}}$ analogously. Let $\tilde{\mathcal{E}}_k$ denote the event that for all $v \in \tilde{V}_k$ the following hold.

(1) The support function dies out by time $s_\star$, $\tilde{\mathscr{F}}_{\text{SUP}}(v, t_{\mathfrak{m}} - s_\star, t_{\mathfrak{m}}) = \emptyset$.
(2) The update function does not travel too far,

$$\tilde{\mathscr{F}}_{\text{UPD}}(v, t_{\mathfrak{m}} - s_\star, t_{\mathfrak{m}}) \subset \varphi_k^{-1}\left(V_k \cap B(v, s_\star^2/4)\right).$$

(3) All vertices have at most $10s_\star$ updates in the interval $[t_{\mathfrak{m}} - s_\star, t_{\mathfrak{m}}]$.

By Lemma 2.1 and the fact that the number of updates of a vertex in time $s_\star$ is $\text{Po}(s_\star)$,

$$\mathbb{P}(\tilde{\mathcal{E}}_k) \geq 1 - |\tilde{V}_k| C e^{-c s_\star} \geq \exp(2^{-d-1}),$$

for large enough $s_\star$.

Recall that we encode the dynamics X_t by a series of updates (J_i, U_i, T_i) for vertices $J_i \in \Lambda$, unit variables U_i and times T_i. If S_i is the sum of spins of the neighbors of J_i at time T_i, then the update sets the new spin of J_i to -1 if $U_i < \frac{e^{-S_i \beta}}{e^{-S_i \beta} + e^{S_i \beta}}$ and to $+1$ otherwise. We couple the updates of $\tilde{X}_t$ to those of X_t as follows. For $v \in \tilde{V}_k$ such that $\varphi_k(v) \in B(A_i, s_\star^2/3)$, we couple the update times, i.e., v has an update at time $t \in [t_{\mathfrak{m}} - s_\star, t_{\mathfrak{m}}]$ in $\tilde{X}$ if and only if $\psi_k(v)$ has one in X. Furthermore, if in addition $\varphi_k(v) \notin \partial V_k$ then we also couple the unit variable of the update. Otherwise (the case $\varphi_k(v) \in \partial V_k$), the unit variables of the updates are taken as independent.

Further recall that an update is oblivious if either $U_i \in [0, \frac{e^{-2d\beta}}{e^{-2d\beta} + e^{2d\beta}}]$ (the new spin is -1 irrespective of the neighbors of J_i) or $U_i \in [\frac{e^{2d\beta}}{e^{-2d\beta} + e^{2d\beta}}, 1]$ (similarly, the new spin is $+1$). Let $\mathcal{R}_k$ denote the event that all updates of $\varphi_k(v) \in \partial V_k \cap B(A_i, s_\star^2/3)$ are oblivious updates and that the updated values $\tilde{X}_t(v)$ and $X_t(\varphi(v))$ agree. This has probability $\frac{e^{-2d\beta}}{e^{-2d\beta} + e^{2d\beta}}$ for each update. Since on $\tilde{\mathcal{E}}_k$ there are at most $10s_\star |\partial V_k \cap B(A_i, s_\star^2/3)|$ updates on $\partial V_k \cap B(A_i, s_\star^2/3)$, we have that

$$\mathbb{P}(\mathcal{R}_k \mid \tilde{\mathcal{E}}_k) \geq \exp\left[-C_1 s_\star \left|\partial V_k \cap B\left(A_i, s_\star^2/3\right)\right|\right],$$

where $C_1 = 10 \log \frac{e^{-2d\beta} + e^{2d\beta}}{e^{-2d\beta}}$. Since these are independent for each k,

$$\mathbb{P}(\cap_{k \in \mathcal{V}}(\mathcal{R}_k \cap \tilde{\mathcal{E}}_k)) \geq \exp\left[-2^{-d-1}|\mathcal{V}| - C_1 s_\star \left|\partial V_k \cap B\left(A_i, s_\star^2/3\right)\right|\right]$$

$$\geq \exp\left(-\frac{1}{2}|A_i| - C_1 s_\star \frac{2d\left|B\left(A_i, s_\star^2/3\right)\right|}{s_\star^{4d}}\right) \geq \exp(-|A_i|),$$

provided that $s_\star$ is sufficiently large, as $|B(A_i, s_\star^2/3)| \leq \left(\frac{5}{3}\right)^d s_\star^{2d}|A_i|$ and $2^d|\mathcal{V}| \leq |A_i|$. By Eq. (??), to complete the lemma it therefore suffices to show that the event $\cap_{k \in \mathcal{V}}(\mathcal{R}_k \cap \tilde{\mathcal{E}}_k)$ implies

$$\mathscr{F}_{\text{SUP}}\left(A_i, t_{\mathfrak{m}} - s_\star, t_{\mathfrak{m}}\right) = \emptyset, \tag{5.11}$$

$$\bigcup_{t_{\mathfrak{m}}-s_\star < t < t_{\mathfrak{m}}} \mathscr{F}_{\text{SUP}}\left(A_i, t, t_{\mathfrak{m}}\right) \subset B\left(A_i, s_\star^2/3\right). \tag{5.12}$$

The updates on $\partial V_k \cap B(A_i, s_\star^2/3)$ are oblivious updates and hence do not examine the values of their neighbors on the event $\mathcal{R}_k$. Combining this with property (2) of the definition of $\tilde{\mathcal{E}}_k$ and the construction of the coupling implies that for $v \in \tilde{V}_k$ such that $\varphi_k(v) \in A_i$, the support of $\varphi_k(v)$ is contained in V_k. Hence, by the coupling it follows that

$$\mathscr{F}_{\text{SUP}}\left(\varphi_k(v), t, t_{\mathfrak{m}}\right) \subset \varphi_k\left(\tilde{\mathscr{F}}_{\text{UPD}}\left(v, t, t_{\mathfrak{m}}\right)\right) \subset V_k \cap B(\varphi_k(v), s_\star^2/3),$$

which implies (5.12). It remains to prove (5.11).

Knowing the updates of course allows one to determine the configuration at a later time from the configuration of an earlier time. We then define $\tilde{Y}_t^\eta(w)$ as follows. It is the the spin at time $t \in [t_{\mathfrak{m}} - s_\star, t_{\mathfrak{m}}]$ of the vertex $w \in \tilde{\mathscr{F}}_{\text{UPD}}(v, t, t_{\mathfrak{m}})$ generated from the Glauber dynamics with initial configuration η on $\tilde{\mathscr{F}}_{\text{UPD}}(v, t_{\mathfrak{m}} - s_\star, t_{\mathfrak{m}})$ at time $t_{\mathfrak{m}} - s_\star$ using the updates of $\tilde{X}_t$. Note that, by the definition of $\tilde{\mathscr{F}}_{\text{UPD}}$, these are the only initial values that need to be specified. Define Y_t^η in the same way except with the updates of X_t instead of $\tilde{X}_t$, where we take the domain of η to be $\varphi_k(\tilde{\mathscr{F}}_{\text{SUP}}(v, t_{\mathfrak{m}}-s_\star, t_{\mathfrak{m}}))$. As usual, $+$ and $-$ denote the all $+1$ and -1 initial conditions, respectively.

Since the initial condition for $\tilde{X}_t$ is all-plus, by the construction of the coupling for every time $t \in [t, t_{\mathfrak{m}}]$ and vertex $w \in \tilde{\mathscr{F}}_{\text{UPD}}(v, t, t_{\mathfrak{m}})$ we have that

$$\tilde{X}_t(w) = \tilde{Y}_t^+(w) = Y_t^+(\varphi_k(w)).$$

We claim that for all t and $w \in \tilde{\mathscr{F}}_{\text{UPD}}(v, t, t_{\mathfrak{m}})$, $\tilde{Y}_t^-(w) \leq Y_t^-(\varphi_k(w))$. This can be seen by induction applying the updates in turn. Let $\{(t_i, w_i)\}$ denote the set of updates in the update history of v in the interval $[t_{\mathfrak{m}} - s_\star, t_{\mathfrak{m}}]$ ordered so that $t_{\mathfrak{m}} - s_\star < t_1 < t_2 < \ldots < t_q < t_{\mathfrak{m}}$. For all updates with $w_i \in \tilde{V}_k \setminus \partial \tilde{V}_k$ this follows by the fact that the updates use the same unit variables, monotonicity of the update rule and the inductive assumption on the values of the neighbors. For updates $w_i \in \partial \tilde{V}_k$ note that

$$\tilde{Y}_{t_i}^-(w_i) \leq \tilde{Y}_{t_i}^+(w) = Y_{t_i}^+(\varphi_k(w_i)) = Y_{t_i}^-(\varphi_k(w_i)),$$

where the first inequality is by monotonicity while the final equality is by the fact that the boundary updates are oblivious ones. Hence, by induction, $\tilde{Y}_{t_{\mathfrak{m}}}^-(v) \leq Y_{t_{\mathfrak{m}}}^-(\varphi_k(v))$. We know that $\tilde{Y}_{t_{\mathfrak{m}}}^+(v) = \tilde{Y}_{t_{\mathfrak{m}}}^-(v)$ by the definition of the support and $\mathcal{E}_k$ and so combining the above results yields

$$Y_{t_{\mathfrak{m}}}^+(\varphi_k(w)) = \tilde{Y}_{t_{\mathfrak{m}}}^+(v) = \tilde{Y}_{t_{\mathfrak{m}}}^-(v) \leq Y_{t_{\mathfrak{m}}}^-(\varphi_k(v)) \leq Y_{t_{\mathfrak{m}}}^+(\varphi_k(v)),$$

so $Y_{t_{\mathfrak{m}}}^+(\varphi_k(v)) = Y_{t_{\mathfrak{m}}}^-(\varphi_k(v))$. This verifies (5.11), completing the proof. ∎

5.4. **Red percolation clusters given the history of their exterior.** This section is devoted to the proof of the following *upper bound* on the probability of a cluster to be RED given the update sequence $\mathcal{U}$ along the $(t_{\mathfrak{m}}, t_{\mathfrak{m}} + s_\star]$ and the history up to time $t_{\mathfrak{m}}$ of every vertex in its exterior:

For any cluster of components $\mathcal{C}$ and each $\ell \geq 1$ we define the relation

$$A_i \sim_\ell A_j \quad \text{iff} \quad B\left(A_i, s_\star^2\left(2^{\ell-2} + \tfrac{1}{4}\right)\right) \cap B\left(A_j, s_\star^2\left(2^{\ell-2} + \tfrac{1}{4}\right)\right) \neq \emptyset,$$

and extend the relation to an equivalence relation. Let V_ℓ denote the set of equivalence classes given by the equivalence relation and for each $v \in V_\ell$ let A_v denote the union of the components in v. We define $\mathfrak{L}$ to be the largest ℓ such that $|V_\ell| > 1$. We let $V_0 = \mathcal{C}$ be the set of A_i.

Claim 5.4. *For any cluster of components $\mathcal{C}$,*

$$\sum_{\ell=0}^{\mathfrak{L}} 2^\ell |V_\ell| \geq \mathfrak{W}(\cup_{A_i \in \mathcal{C}} A_i) - \sum_{A_i \in \mathcal{C}} \mathfrak{W}(A_i).$$

Proof. Let $\hat{A}_i$ denote a minimal lattice animal containing A_i so that $A_i \subset \hat{A}_i$ and $|\hat{A}_i| = \mathfrak{W}(A_i)$. We construct a lattice animal covering $\cup_i A_i$ by adding blocks to $\cup_i \hat{A}_i$ as follows. Starting with $\ell = 1$ we add the minimum number of blocks needed so that for all $v \in V_\ell$ all the $A_i \in v$ are connected together. By definition, if $A_i \sim_\ell A_j$ then these can be connected with at most $2^{\ell-1}$ blocks. Thus, after connecting together all sets of components at level $\ell - 1$ for each $v \in V_\ell$ we need to add at most $2^{\ell-1}(|\{v' \in V_{\ell-1} : v' \subset v\}| - 1)$ additional blocks to connect together all the components of V_ℓ. Summing over ℓ from 1 to $\mathfrak{L} + 1$ we add a total number of blocks of

$$\sum_{\ell=1}^{\mathfrak{L}+1} \sum_{v \in V_\ell} 2^{\ell-1}\left(|\{v' \in V_{\ell-1} : v' \subset v\}| - 1\right) = \sum_{\ell=1}^{\mathfrak{L}+1} 2^{\ell-1}(|V_{\ell-1}| - |V_\ell|) \leq \sum_{\ell=0}^{\mathfrak{L}} 2^\ell |V_\ell|.$$

Since adding $\sum_{\ell=0}^{\mathfrak{L}} 2^\ell |V_\ell|$ blocks to the $\hat{A}_i$ yields a connected component, the desired result follows. ∎

Lemma 5.5. *There exists $c(\beta, d), s_0(\beta, d) > 0$ such that, for any $s_\star > s_0$, any large enough n and every $\mathcal{C} \subset \mathcal{A}$, the quantity $\Psi_{\mathcal{C}}$ from (3.1) satisfies*

$$\mathbb{P}\left(\mathcal{C} \in \text{RED} \mid \mathscr{H}_{\mathcal{C}}^-, \mathcal{U}\right) \leq \frac{s_\star^{4d}}{\mathfrak{m}_{t_{\mathfrak{m}}}} e^{3\sum_i |A_i| - cs_\star \left|\mathfrak{W}(\mathcal{C}) - \sum_{A_i \in \mathcal{C}} \mathfrak{W}(A_i)\right|^+}. \tag{5.13}$$

Proof. The bound is trivial for histories which are not $\mathcal{C}$-compatible so we may restrict our attention to the supremum over $\mathcal{C}$-compatible histories. Denote the event $\mathcal{E}$ that the history of $\mathcal{C}$ does not intersect the history $\mathscr{H}_{\mathcal{C}}^-$. The set of clusters $\mathcal{A}$ depends only on $\mathcal{U}$ and this is the only dependence on $\mathcal{U}$ in the bound. Given $\mathcal{A}$, the partition into clusters and their colors depends only on the updates in $[0, t_{\mathfrak{m}}]$. Hence, we can view RED as a function $\text{RED}(\mathcal{A})$. We can extend this definition to any set of components and write $\text{RED}(\mathcal{A}')$ to denote the set of red clusters had the set of components instead been $\mathcal{A}'$. Now if $\mathcal{C} \in \text{RED}(\mathcal{A})$ then one also has $\mathcal{C} \in \text{RED}(\mathcal{C})$.

We let $\mathcal{E}(H)$ denote the event that the history of $\mathcal{C}$ does not intersect $H \subset \Lambda \times [0, t_{\mathfrak{m}}]$, a space-time slab. If $\mathcal{C}$ is a cluster then $\mathcal{E}(\mathscr{H}_{\mathcal{C}}^-)$ must hold. Exploring the history of $\mathcal{C}$ we see that it does not depend on the history of its complement, $\mathscr{H}_{\mathcal{C}}^-$, until the point at which they intersect (since they depend on disjoint updates) and hence

$$\mathbb{P}(\mathscr{H}_{\mathcal{C}} \in \cdot, \mathcal{E}(H) \mid \mathscr{H}_{\mathcal{C}}^- = H, \mathcal{U}) = \mathbb{P}(\mathscr{H}_{\mathcal{C}} \in \cdot, \mathcal{E}(H) \mid \mathcal{U}) \leq \mathbb{P}(\mathscr{H}_{\mathcal{C}} \in \cdot).$$

Since the event $\mathcal{C} \in \text{RED}(\mathcal{C})$ is $\mathscr{H}_{\mathcal{C}}$-measurable we have that,

$$\mathbb{P}(\mathcal{C} \in \text{RED} \mid \mathscr{H}_{\mathcal{C}}^-, \mathcal{U}) \leq \mathbb{P}(\mathcal{C} \in \text{RED}(\mathcal{C}), \mathcal{E}(\mathscr{H}_{\mathcal{C}}^-) \mid \mathscr{H}_{\mathcal{C}}^-, \mathcal{U}) \leq \mathbb{P}(\mathcal{C} \in \text{RED}(\mathcal{C})).$$

Next, we bound $\mathbb{P}(\mathcal{C} \in \text{RED}(\mathcal{C}))$. At least one vertex of $\mathcal{C}$ must have support at time 0 so by a union,

$$\mathbb{P}(\mathcal{C} \in \text{RED}(\mathcal{C})) = \mathbb{P}(\mathscr{F}_{\text{SUP}}(\cup_i A_i, 0, t_{\mathfrak{m}})) \leq s_\star^{2d} \mathfrak{m}_{t_{\mathfrak{m}}} \sum_i |A_i| = s_\star^{2d} \frac{1}{\sqrt{|\Lambda|}} \sum_i |A_i| ;$$

this implies (3.2) in the case $\mathfrak{W}(\cup_i A_i) - \sum_i \mathfrak{W}(A_i) \leq 0$, which we note includes the case $|\mathcal{C}| = 1$. We may thus restrict our attention to the case $\mathfrak{W}(\cup_i A_i) - \sum_i \mathfrak{W}(A_i) > 0$, in which $|\mathcal{C}| \geq 2$ and so $|V_{\mathfrak{L}}| \geq 2$.

Our approach will be to define a collection of events that must hold if $\mathcal{C} \in \text{RED}(\mathcal{C})$, one that depends on the structure of $\mathcal{C}$ and its above defined decomposition. For $\ell \leq \mathfrak{L}$ and $v \in V_\ell$ define the event $K_{v,\ell}$, which roughly says that the update set of A_v spreads unexpectedly quickly, as

$$K_{v,\ell} = \left\{ \exists u \in B(A_v, \tfrac{3}{7}2^\ell s_\star^2) \setminus B(A_v, \tfrac{2}{5}2^\ell s_\star^2) : \mathscr{F}_{\text{UPD}}(u, t_{\mathfrak{m}} - 2^\ell s_\star, t_{\mathfrak{m}}) \not\subset B(u, \tfrac{2^\ell}{100}s_\star^2) \right\},$$

and the event

$$J_{v,\ell} = \left\{ \mathscr{F}_{\text{SUP}}\left(B\left(A_v, \left(\tfrac{2}{5} + \tfrac{1}{100}\right)2^\ell s_\star^2\right), t_{\mathfrak{m}} - 2^\ell s_\star, t_{\mathfrak{m}} - 2^{\ell-1}s_\star\right) \neq \emptyset, K_{v,\ell}^c \right\},$$

which roughly says that the support of A_v lasts for a large time. Combined, we define $R_{v,\ell} = K_{v,\ell} \cup J_{v,\ell}$. For the final level $\ell = \mathfrak{L}$ we define a slight variant of these in terms of κ, some large positive constant to be fixed later,

$$\hat{K}_{v,\ell} = \left\{ \exists u \in B(A_v, \tfrac{3}{7}2^\ell s_\star^2) \setminus B(A_v, \tfrac{2}{5}2^\ell s_\star^2) : \mathscr{F}_{\text{UPD}}(u, t_{\mathfrak{m}} - \kappa 2^\ell s_\star, t_{\mathfrak{m}}) \not\subset B(u, \tfrac{2^\ell}{100}s_\star^2) \right\},$$

as well as

$$\hat{J}_{v,\ell} = \left\{ \mathscr{F}_{\text{SUP}}\left(B\left(A_v, \left(\tfrac{2}{5} + \tfrac{1}{100}\right)2^\ell s_\star^2\right), t_{\mathfrak{m}} - \kappa 2^\ell s_\star, t_{\mathfrak{m}} - 2^{\ell-1}s_\star\right) \neq \emptyset, \hat{K}_{v,\ell}^c \right\}$$

and $\hat{R}_{v,\ell} = \hat{K}_{v,\ell} \cup \hat{J}_{v,\ell}$. Finally, we need to consider events describing how the history connects to time 0. For $v \in V_{\mathfrak{L}}$, denote

$$\breve{J}_{v,\ell} = \left\{ \mathscr{F}_{\text{SUP}}\left(B\left(A_v, \left(\tfrac{2}{5} + \tfrac{1}{100}\right)2^\ell s_\star^2\right), 0, t_{\mathfrak{m}} - 2^{\ell-1}s_\star\right) \neq \emptyset, \hat{K}_{v,\ell}^c \right\}.$$

Define the set

$$\Gamma = \mathscr{F}_{\text{UPD}}\left(B(\cup_{A_i \in \mathcal{C}} A_i, 2^{\mathfrak{L}} s_\star^2), t_{\mathfrak{m}} - \kappa 2^{\mathfrak{L}} s_\star, t_{\mathfrak{m}}\right)$$

and the event

$$J_\Gamma = \left\{ \mathscr{F}_{\text{SUP}}\left(\Gamma, 0, t_{\mathfrak{m}} - \kappa 2^{\mathfrak{L}} s_\star\right) \neq \emptyset \right\}.$$

Claim 5.6. *If $\mathcal{C} \in \text{RED}(\mathcal{C})$ then the events $R_{v,\ell}$ and $\hat{R}_{v,\ell}$ hold for all $0 \leq \ell \leq \mathfrak{L}$ and $v \in V_\ell$. Furthermore, either $\cup_{v \in V_{\mathfrak{L}}} \breve{J}_{v,\mathfrak{L}}$ or $J_\Gamma \cap (\cup_{v \in V_{\mathfrak{L}}} \hat{K}_{v,\mathfrak{L}})$ hold.*

Proof of claim. If $\mathcal{C} \in \text{RED}(\mathcal{C})$ then the history of A_v must connect to the remainder of the component. This can occur either by the support of A_v and $\mathcal{C} \setminus A_v$ meeting in the interval $[t_{\mathfrak{m}} - 2^\ell s_\star, t_{\mathfrak{m}}]$ in which case $K_{v,\ell}$ holds. Alternatively, the support of $A_i \in A_v$ may enter $B(A_j, s_\star^2/3)$ for some $A_j \in \mathcal{C} \setminus A_v$ in the interval $[t_{\mathfrak{m}} - s_\star, t_{\mathfrak{m}}]$, in which case again $K_{v,\ell}$ holds. Another option is that the support of $A_j \in \mathcal{C} \setminus A_v$ enters $B(A_i, s_\star^2/3)$ for some $A_i \in A_v$ in the interval $[t_{\mathfrak{m}} - s_\star, t_{\mathfrak{m}}]$, whence again $K_{v,\ell}$ holds.

The final possibility is that the support survives until time $t_{\mathfrak{m}} - 2^\ell s_\star$ (i.e., $\mathscr{F}_{\text{SUP}}(A_v, , t_{\mathfrak{m}} - 2^\ell s_\star, t_{\mathfrak{m}}) \neq \emptyset$). In this case one of the following must hold:

- The support travels far in space by time $t_{\mathfrak{m}} - 2^{\ell-1}s_\star$ and

$$\mathscr{F}_{\text{SUP}}\left(A_v, t_{\mathfrak{m}} - 2^{\ell-1}s_\star, t_{\mathfrak{m}}\right) \not\subset B\left(A_v, \left(\tfrac{2}{5} + \tfrac{1}{100}\right)2^\ell s_\star^2\right),$$

  in which case $K_{v,\ell}$ holds.

- The support survives until time $t_{\mathfrak{m}} - 2^{\ell} s_{\star}$, given it did not travel too far by time $t_{\mathfrak{m}} - 2^{\ell-1} s_{\star}$, and

$$\mathscr{F}_{\text{SUP}}\left(B\big(A_v, (\tfrac{2}{5} + \tfrac{1}{100})2^{\ell} s_{\star}^2\big), t_{\mathfrak{m}} - 2^{\ell} s_{\star}, t_{\mathfrak{m}} - 2^{\ell-1} s_{\star}\right) \neq \emptyset,$$

  in which case $R_{v,\ell}$ holds.

This set of possibilities is exhaustive and completes the claim. The other claims follow similarly. ∎

We now observe that the events $R_{v,\ell}$ are independent as they depend on disjoint sets of updates. The event $K_{v,\ell}$ depends only on updates in the space time block

$$\left(B(A_v, \big(\tfrac{3}{7} + \tfrac{1}{100}\big)2^{\ell} s_{\star}^2) \setminus B(A_v, \big(\tfrac{2}{5} - \tfrac{1}{100}\big)2^{\ell} s_{\star}^2)\right) \times [t_{\mathfrak{m}} - 2^{\ell} s_{\star}, t_{\mathfrak{m}}]$$

while $J_{v,\ell}$ depends only on the same set (through the event $K_{v,\ell}^c$ in its definition) plus updates in

$$\left(B\big(A_v, \big(\tfrac{2}{5} + \tfrac{2}{100}\big)2^{\ell} s_{\star}^2\big)\right) \times [t_{\mathfrak{m}} - 2^{\ell} s_{\star}, t_{\mathfrak{m}} - 2^{\ell-1} s_{\star}].$$

Since these sets are disjoint for different v it follows that the $R_{v,\ell}$'s are independent. Similarly, the $\hat{R}_{v,\ell}$'s are independent for $\ell = \mathfrak{L}$ and independent of the $R_{v,\ell}$'s for $\ell < \mathfrak{L}$. The event $\breve{J}_{v,\ell}$ is also independent of the $R_{v,\ell}$'s and $\hat{R}_{v,\ell}$'s. The set Γ depends only on updates in $(\Lambda \setminus B(A_v, 2^{\ell} s_{\star}^2)) \times (t_{\mathfrak{m}} - \kappa 2^{\ell} s_{\star}, t_{\mathfrak{m}})$ and hence independent of all our constructed events except J_Γ, and J_Γ is independent of all our constructed events except $\breve{J}_{v,\ell}$.

We now estimate the probability of the above events using Lemma 2.1. Noting that

$$\left|B\big(A_v, \alpha 2^{\ell} s_{\star}^2\big)\right| \leq (1 + 2\alpha)^d 2^{\ell d} s_{\star}^{2d} |A_v| \leq e^{C(\ell + \log s_{\star}) + 2^{-\ell}|A_v|}$$

for some $c' > 0$, we have that

$$\mathbb{P}(K_{v,\ell}) \leq \left|B\big(A_v, \tfrac{3}{7} 2^{\ell} s_{\star}^2\big)\right| e^{-c2^{\ell} s_{\star}^2} \leq e^{-c' 2^{\ell} s_{\star}^2 + 2^{-\ell}|A_v|}$$

for large enough $s_{\star}$. Similarly,

$$\mathbb{P}(\hat{K}_{v,\ell}) \leq e^{-c' 2^{\ell} s_{\star}^2 + 2^{-\ell}|A_v|}.$$

Again by Lemma 2.1 it follows that

$$\mathbb{P}(J_{v,\ell}) \leq \left|B\big(A_v, \big(\tfrac{2}{5} + \tfrac{1}{100}\big) 2^{\ell} s_{\star}^2\big)\right| e^{-c2^{\ell} s_{\star}} \leq e^{-c' 2^{\ell} s_{\star} + 2^{-\ell}|A_v|},$$

and similarly

$$\mathbb{P}(\hat{J}_{v,\ell}) \leq \left|B\big(A_v, \big(\tfrac{2}{5} + \tfrac{1}{100}\big) 2^{\ell} s_{\star}^2\big)\right| e^{-c\kappa 2^{\ell} s_{\star}} \leq e^{-c' \kappa 2^{\ell} s_{\star} + 2^{-\ell}|A_v|}.$$

As $\mathbb{P}(\mathscr{F}_{\text{SUP}}(u, 0, t) \neq \emptyset) = \mathfrak{m}_t$ and $\mathfrak{m}_{t+h} \geq e^{-h}\mathfrak{m}_t$ we have that

$$\mathbb{P}(\breve{J}_{v,\mathfrak{L}}) \leq \left|B\big(A_v, \big(\tfrac{2}{5} + \tfrac{1}{100}\big) 2^{\mathfrak{L}} s_{\star}^2\big)\right| \mathfrak{m}_{t_{\mathfrak{m}} - 2^{\mathfrak{L}-1} s_{\star}} \leq \exp(2^{\mathfrak{L}} s_{\star} + 2^{-\ell}|A_v|)\mathfrak{m}_t.$$

For $h \geq 1/3$ and $s_{\star}$ large enough we have that

$$\mathbb{P}\left(\mathscr{F}_{\text{UPD}}(v, t_{\mathfrak{m}} - \kappa 2^{\ell} s_{\star}, t_{\mathfrak{m}}) \not\subset B\big(v, h2^{\ell} s_{\star}^2\big)\right) \leq e^{-s_{\star} 2^{\ell} h},$$

and so $\mathbb{E}|\mathscr{F}_{\mathrm{UPD}}(v, t_{\mathfrak{m}} - \kappa 2^\ell s_\star, t_{\mathfrak{m}})| \leq s_\star^{2d} 2^{\ell d}$. Hence,

$$\mathbb{E}|\Gamma| \leq \left| B\big(\cup_{A_i \in \mathcal{C}} A_i, 2^{\mathfrak{L}} s_\star^2 \big) \right| s_\star^{2d} 2^{\ell d} \leq \left(3 \cdot 2^{2\mathfrak{L}} s_\star^4 \right)^d \sum_i |A_i|.$$

Finally, by a union bound over $u \in \Gamma$,

$$\mathbb{P}\left(J_\Gamma \mid \Gamma \right) \leq \mathfrak{m}_{t_{\mathfrak{m}} - \kappa 2^{\mathfrak{L}} s_\star} |\Gamma| \leq \mathfrak{m}_{t_{\mathfrak{m}}} e^{\kappa 2^{\mathfrak{L}} s_\star} |\Gamma|.$$

Combining the above estimates with the claim and the independence of the events we have that

$$\mathbb{P}\left(\mathcal{C} \in \mathrm{RED}(\mathcal{C}) \right) \leq \mathbb{P}\left(\left(\bigcap_{\ell=0}^{\mathfrak{L}-1} \bigcap_{v \in V_\ell} R_{v,\ell} \right) \cap \left(\bigcup_{v \in V_{\mathfrak{L}}} \breve{J}_{v,\mathfrak{L}} \bigcap_{u \in V_{\mathfrak{L}} \setminus \{v\}} \hat{R}_{u,\mathfrak{L}} \right) \right)$$

$$+ \mathbb{P}\left(\left(\bigcap_{\ell=0}^{\mathfrak{L}-1} \bigcap_{v \in V_\ell} R_{v,\ell} \right) \cap \left(\bigcup_{v \in V_{\mathfrak{L}}} \hat{K}_{v,\mathfrak{L}} \bigcap_{u \in V_{\mathfrak{L}} \setminus \{v\}} \hat{R}_{u,\mathfrak{L}} \right) \cap J_\Gamma \right).$$

The right-hand side, in turn, is at most

$$\exp\left(-c' s_\star \sum_{\ell=0}^{\mathfrak{L}-1} 2^\ell |V_\ell| - c' \kappa 2^{\mathfrak{L}} (|V_{\mathfrak{L}}| - 1) + 2 \sum_i |A_i| \right) |V_{\mathfrak{L}}| \mathfrak{m}_{t_{\mathfrak{m}}} \left(e^{2^{\mathfrak{L}} s_\star} + e^{-c' 2^\ell s_\star^2} \mathbb{E}\Gamma e^{\kappa 2^\ell s_\star} \right)$$

$$\leq \frac{1}{\sqrt{|\Lambda|}} \exp\left(-c' s_\star \sum_{\ell=0}^{\mathfrak{L}} 2^\ell |V_\ell| + 3 \sum_i |A_i| \right),$$

where the final inequality comes from the fact that $|V_{\mathfrak{L}}| \geq 2$ and

$$|V_{\mathfrak{L}}| \left(e^{2^{2\mathfrak{L}} s_\star} + e^{-c' 2^\ell s_\star^2} (3 \cdot 2^{\mathfrak{L}} s_\star^4)^d e^{\kappa 2^\ell s_\star} \sum_i |A_i| \right) e^{-c'(\kappa-1) 2^{\mathfrak{L}} - \sum_i |A_i|} \leq 1,$$

provided that $c'(\kappa - 1) > 1$ and $s_\star$ is large. Combining this with Claim 5.4 gives (5.13), as required. ∎

5.5. **Proof of Lemma 3.3.** The desired bound in (3.2) is similar to the one we obtained in (5.13), with the difference being an extra conditioning on $\mathcal{G}_{\mathcal{C}} := \{\mathcal{C} \in \mathrm{RED}\} \cup \{\mathcal{C} \subset \mathrm{BLUE}\}$. Since $\{\mathcal{C} \in \mathrm{RED}\} \subset \mathcal{G}_{\mathcal{C}}$,

$$\mathbb{P}\left(\mathcal{C} \in \mathrm{RED} \mid \mathscr{H}_{\mathcal{C}}^-, \mathcal{G}_{\mathcal{C}}, \mathcal{U} \right) = \frac{\mathbb{P}\left(\mathcal{C} \in \mathrm{RED} \mid \mathscr{H}_{\mathcal{C}}^-, \mathcal{U} \right)}{\mathbb{P}\left(\mathcal{G}_{\mathcal{C}} \mid \mathscr{H}_{\mathcal{C}}^-, \mathcal{U} \right)} \leq \frac{\mathbb{P}\left(\mathcal{C} \in \mathrm{RED} \mid \mathscr{H}_{\mathcal{C}}^-, \mathcal{U} \right)}{\mathbb{P}(\bigcap_{A_i \in \mathcal{C}} \{A_i \in \mathrm{BLUE}\} \mid \mathscr{H}_{\mathcal{C}}^-, \mathcal{U})},$$

from which Eq. (3.2) follows after applying Lemmas 5.3 and 5.5 to the numerator and denominator of the right-hand side, respectively. ∎

6. RANDOM VS. DETERMINISTIC INITIAL STATES

In this section we consider the Ising model on the cycle $\mathbb{Z}/n\mathbb{Z}$ for small $\beta > 0$ with a random initial state, both quenched and annealed. Rather than comparing two worst case initial states we will compare a random one directly with the stationary distribution using coupling from the past. Recall that for the 1D Ising model we can give a special update rule: with probability $\theta = \theta_{\beta,d} = 1 - \tanh(2\beta)$ update to a uniformly random value and with probability $1 - \theta$ copy the spin of a random neighbor. The history of a vertex is simply a continuous time random walk which takes steps at rate $1 - \theta$ and dies at rate θ; when such walks meet they coalesce and continue together until dying. Each component can only decrease in size over time and all vertices in the component receive the same spin.

In the sequel, the annealed setting (Part 1 of Theorem 2) is studied in §6.1 and §6.2 (upper and lower bounds on mixing, respectively), whereas §6.3 focuses on the quenched setting (Part 2 of Theorem 2).

6.1. **Annealed analysis: upper bound.** We define Y_t to be the process coupled from the past, so the spins at time $t_{\mathfrak{m}}$ are independent on each of the components with equal probability. Now let X_t be the process started from an i.i.d. configuration at time 0. Since the magnetization is simply the probability that the walk has not yet died it follows that $\mathfrak{m}_t = e^{-\theta t}$. We will consider the total variation distance at time $t'_{\mathfrak{m}} = \frac{1}{4\theta} \log n + \frac{1}{\theta} \log \log n - s_\star$ for some large constant $s_\star > 0$.

Theorem 6.1. *With $t'_\star = \frac{1}{4\theta} \log n + \frac{1}{\theta} \log \log n$ we have that $\|\mathbb{P}(X_{t'_\star} \in \cdot) - \pi(\cdot)\|_{\mathrm{TV}} \to 0$.*

Set $t'_{\mathfrak{m}} = t'_\star - s_\star$ where $s_\star > 0$ is a large constant. In order to couple the process with the stationary distribution we consider updates in the range $t \in (-\infty, t'_\star]$ with the block components constructed using the updates in the range $[t'_{\mathfrak{m}}, t'_\star]$ with deferred and undeferred randomness similarly as before.

In this analysis it is necessary to directly compare the annealed distribution with the stationary distribution and for this we use the coupling from the past paradigm and hence consider updates before time 0. We modify the *Intersection* property of the construction of clusters to identify A_i and A_j if $\mathscr{F}_{\mathrm{SUP}}(A_i, t, t_{\mathfrak{m}}) \cap \mathscr{F}_{\mathrm{SUP}}(A_j, t, t_{\mathfrak{m}}) \neq \emptyset$ for some $-\infty < t < t_{\mathfrak{m}}$ (in place of the condition $0 \leq t < t_{\mathfrak{m}}$).

We also redefine the notion of a red cluster to be one containing two vertices whose history reaches time 0 without coalescing, that is,

$$\mathcal{C} \in \mathrm{RED} \quad \text{iff} \quad \left| \bigcup_{v \in A_i \in \mathcal{C}} \mathscr{F}_{\mathrm{SUP}}(v, 0, t_{\mathfrak{m}}) \right| \geq 2 \tag{6.1}$$

(note that the histories of vertices are always of size one). We define blue clusters as before and green clusters as the remaining clusters. We can thus couple X_t and Y_t so that they agree on the blue and green components at time $t_{\mathfrak{m}}$ but possibly not on the red components.

Recalling that $\mathfrak{W}(S)$ is the smallest lattice animal of blocks covering S, we let

$$\mathfrak{W}_2(S) := \min_{S_1, S_2 : S_1 \cup S_2 = S} \mathfrak{W}(S_1) + \mathfrak{W}(S_2).$$

We denote V_ℓ, equivalence classes of components of $\mathcal{C}$ as in §5.4 and this time define $\mathfrak{L}'$ to be the largest ℓ such that $|V_\ell| > 2$. The following claim is a simple extension of Claim 5.4.

Claim 6.2. *For any cluster of components $\mathcal{C}$,*

$$\sum_{\ell=0}^{\mathfrak{L}'} 2^\ell |V_\ell| \geq \mathfrak{W}_2(\cup_{A_i \in \mathcal{C}} A_i) - \sum_{A_i \in \mathcal{C}} \mathfrak{W}(A_i).$$

The proof is essentially the same as Claim 5.4.

Lemma 6.3. *There exists $c(\beta, d), s_0(\beta, d) > 0$ such that, for any $s_\star > s_0$, any large enough n and every $\mathcal{C} \subset \mathcal{A}$, the quantity $\Psi_{\mathcal{C}}$ from (3.1) satisfies*

$$\mathbb{P}\left(\mathcal{C} \in \mathrm{RED} \mid \mathscr{H}_{\mathcal{C}}^-, \mathcal{U} \right) \leq \frac{s_\star^{4d}}{e^{2\theta t'_{\mathfrak{m}}}} e^{3 \sum_i |A_i| - c s_\star \left| \mathfrak{W}_2(\mathcal{C}) - \sum_{A_i \in \mathcal{C}} \mathfrak{W}(A_i) \right|^+}. \tag{6.2}$$

Proof. The proof is similar to Lemma 5.5 and we describe the necessary changes. Throughout the proof we change all instances of $\mathfrak{L}$ to $\mathfrak{L}'$ and $t_{\mathfrak{m}}$ to $t'_{\mathfrak{m}}$. We must incorporate the fact that two histories

independently reaching time 0 are required for red so we denote

$$\breve{J}'_{v,v',\ell} = \left\{ \left| \mathscr{F}_{\mathrm{SUP}}\left(B\left(A_v \cup A_{v'}, \left(\tfrac{2}{5} + \tfrac{1}{100}\right)2^\ell s_\star^2\right), 0, t'_{\mathfrak{m}} - 2^{\ell-1}s_\star\right) \right| \geq 2, \hat{K}^c_{v,\ell}, \hat{K}^c_{v',\ell} \right\}.$$

Since the probability that two separate histories reach time 0 without coalescing is bounded by the square of the probability of a single walk reaching time 0, we have that

$$\mathbb{P}(\breve{J}'_{v_1,v_2,\mathfrak{L}}) \leq e^{-2\theta t'_{\mathfrak{m}}}\prod_{i=1}^2 |B\left(A_{v_i}, \left(\tfrac{2}{5} + \tfrac{1}{100}\right)2^\ell s_\star^2\right)| \leq \exp\left(2^{\mathfrak{L}}s_\star + 2^{-\ell}(|A_{v_1}| + |A_{v_2}|)\right)e^{-2\theta t'_{\mathfrak{m}}}.$$

By essentially the same proof as above we have that $\mathbb{E}|\Gamma| \leq \left(3 \cdot 2^{2\mathfrak{L}}s_\star^4\right)^{2d}\left(\sum_i |A_i|\right)^2$. Also, as we require two histories to reach time 0, we let

$$J'_\Gamma = \left\{ \left| \mathscr{F}_{\mathrm{SUP}}\left(\Gamma, 0, t'_{\mathfrak{m}} - \kappa 2^{\mathfrak{L}}s_\star\right) \right| \geq 2 \right\}$$

and

$$\mathbb{P}\left(J_\Gamma \mid \Gamma\right) \leq \mathfrak{m}^2_{t_{\mathfrak{m}} - \kappa 2^{\mathfrak{L}}s_\star}|\Gamma|^2 \leq e^{-2\theta t'_{\mathfrak{m}}}e^{2\kappa 2^{\mathfrak{L}}s_\star}|\Gamma|^2.$$

With this notation,

$$\mathbb{P}\left(\mathcal{C} \in \mathrm{RED}(\mathcal{C})\right) \leq \mathbb{P}\left(\left(\bigcap_{\ell=0}^{\mathfrak{L}'-1}\bigcap_{v \in V_\ell} R_{v,\ell}\right) \cap \left(\bigcup_{v,v' \in V_{\mathfrak{L}'}} \breve{J}'_{v,v',\mathfrak{L}'}\bigcap_{u \in V_{\mathfrak{L}'}\setminus\{v,v'\}} \hat{R}_{u,\mathfrak{L}'}\right)\right)$$

$$+ \mathbb{P}\left(\left(\bigcap_{\ell=0}^{\mathfrak{L}'-1}\bigcap_{v \in V_\ell} R_{v,\ell}\right) \cap \left(\bigcup_{v \in V_{\mathfrak{L}'}} \hat{K}_{v,\mathfrak{L}'}\bigcap_{u \in V_{\mathfrak{L}'}\setminus\{v\}} \hat{R}_{u,\mathfrak{L}'}\right) \cap J'_\Gamma\right).$$

The result now follows similarly to Lemma 5.5 by substituting our bounds for each of the events. ∎

Since red components under our modified definition are also red components under the previous definition we have by Lemma 5.5,

$$\mathbb{P}\left(\mathcal{C} \in \mathrm{RED} \mid \mathscr{H}^-_\mathcal{C}, \mathcal{U}\right) \leq \frac{s_\star^{4d}}{e^{\theta t'_{\mathfrak{m}}}}e^{3\sum_i |A_i| - cs_\star\left|\mathfrak{W}(\mathcal{C}) - \sum_{A_i \in \mathcal{C}}\mathfrak{W}(A_i)\right|^+},$$

and combining this with (6.2) yields

$$\mathbb{P}\left(\mathcal{C} \in \mathrm{RED} \mid \mathscr{H}^-_\mathcal{C}, \mathcal{U}\right) \leq \frac{s_\star^{4d}}{e^{2\theta t'_{\mathfrak{m}}}}e^{3\sum_i |A_i| - cs_\star\left|\mathfrak{W}_2(\mathcal{C}) - \sum_{A_i \in \mathcal{C}}\mathfrak{W}(A_i)\right|^+ - cs_\star\left|\mathfrak{W}(\mathcal{C}) - \mathfrak{W}_2(\mathcal{C}) - \frac{\theta}{cs_\star}t'_{\mathfrak{m}}\right|^+}. \qquad (6.3)$$

Altogether, this translates into the bound

$$\bar{\Psi}_{\{A_{i_j}\}} \leq \frac{s_\star^{4d}}{e^{2\theta t'_{\mathfrak{m}}}}e^{4\sum_i |A_i| - cs_\star\left|\mathfrak{W}_2(\mathcal{C}) - \sum_{A_i \in \mathcal{C}}\mathfrak{W}(A_i)\right|^+ - cs_\star\left|\mathfrak{W}(\mathcal{C}) - \mathfrak{W}_2(\mathcal{C}) - \frac{\theta}{cs_\star}t'_{\mathfrak{m}}\right|^+}.$$

Having coupled X_t and Y_t as described above where only the red components differ in the two versions of the chain, we can follow the analysis of §4 and in place of equation (4.11) arrive at

$$\frac{2^{2\lambda+6}s_\star^{8d}}{e^{4\theta t'_{\mathfrak{m}}}}\sum_A\left(\sum_{\substack{\mathcal{C}=\{A_{i_j}\}\\A \in \mathcal{C}}}\sum_{\{B_j\}}\exp\left[-\frac{c}{4}\frac{s_\star}{\lambda}\left[\mathfrak{W}(\mathcal{C}) + \sum_j \mathfrak{W}(A_{i_j} \cup B_j) + \left(\mathfrak{W}(\mathcal{C}) - \mathfrak{W}_2(\mathcal{C}) - t'_{\mathfrak{m}}\right)^+\right]\right]\right)^2.$$

$$(6.4)$$

We now count the number of rooted animals $\mathcal{C}$ with $\mathfrak{W}(\mathcal{C}) = \mathcal{S}$ and $\mathfrak{W}_2(\mathcal{C}) = \mathcal{S}'$. We must cover $\mathcal{C}$ with two lattice animals, one containing the root A and the second rooted at A'. Since the distance from A to A' is at most $\mathcal{S}$ there are at most $(2\mathcal{S}+1)^d$ choices for A'. There are $\mathcal{S}'$ ways to choose the sizes of the two animals as some k and $\mathcal{S}' - k$ and then $(2d)^{2k+2(\mathcal{S}'-k)}$ ways of choosing the animals. In total, we have at most $\mathcal{S}'(2\mathcal{S}+1)^d(2d)^{2\mathcal{S}'}$ choices of $\mathcal{C}$. The total number of choices of $\{A_{i_j}\}$ and B_j with $\mathfrak{W}(\mathcal{C}) = \mathcal{S}$, $\mathfrak{W}_2(\mathcal{C}) = \mathcal{S}'$ and $\sum_j \mathfrak{W}(A_{i_j} \cup B_j) = \mathcal{R}$ is therefore $\mathcal{S}'(2\mathcal{S}+1)^d 2^{\mathcal{S}'} 8^{\mathcal{R}}(2d)^{2(\mathcal{S}'+\mathcal{R})}$; thus,

$$\sum_{\substack{\mathcal{C}=\{A_{i_j}\}\,\{B_j\} \\ A\in\mathcal{C}}} \sum \exp\left[-\frac{c}{4}\frac{s_\star}{\lambda}\left[\mathfrak{W}(\mathcal{C}) + \sum_j \mathfrak{W}(A_{i_j} \cup B_j) + \left(\mathfrak{W}(\mathcal{C}) - \mathfrak{W}_2(\mathcal{C}) - t'_{\mathfrak{m}}\right)^+\right]\right]$$

$$\leq \sum_{\substack{\mathcal{S}',\mathcal{R}\geq 1 \\ \mathcal{S}\geq\mathcal{S}'}} e^{-\frac{c}{4}\frac{s_\star}{\lambda}(\mathcal{S}'+\mathcal{R}+(\mathcal{S}-\mathcal{S}'-t'_{\mathfrak{m}})^+)} \mathcal{S}'(2\mathcal{S}+1)^d 2^{\mathcal{S}'} 8^{\mathcal{R}}(2d)^{2(\mathcal{S}'+\mathcal{R})} \leq t'_{\mathfrak{m}} e^{-\frac{c}{5}\frac{s_\star}{\lambda}}$$

provided $s_\star$ is large enough compared to d. Plugging this into (6.4) gives an upper bound on the total variation distance of

$$\frac{2^{2\lambda+6} s_\star^{8d}}{e^{4\theta t'_{\mathfrak{m}}}}|\Lambda|(t'_{\mathfrak{m}})^2 e^{-\frac{2c}{5}\frac{s_\star}{\lambda}}$$

which tends to 0 for $t'_{\mathfrak{m}} = \frac{1}{4\theta}\log n + \frac{1}{\theta}\log\log n - s_\star$, establishing the upper bound on the mixing time.

6.2. **Annealed analysis: lower bound.** We now prove a matching lower bounded on the mixing time from an annealed initial configuration.

Theorem 6.4. *For $t = \frac{1}{4\theta}\log n - \frac{2}{\theta}\log\log n$ we have that $\|\mathbb{P}(X_t \in \cdot) - \pi(\cdot)\|_{\mathrm{TV}} \to 1$.*

It will be convenient to simply omit the deferred and undeferred regions and directly analyze the update support function from time $t = \frac{1}{4\theta}\log n - \frac{2}{\theta}\log\log n$. Taking the coupling above, two vertices v and v' have the same spins in X if their histories merge before time 0 and are conditionally independent otherwise. By contrast, for Y the spins are equal if the histories merge at any time in the past and are conditionally independent otherwise. Defining $\mathcal{A}_{v,v'}$ as the event that the histories of v and v' survive to time 0 and merge at some negative time, we have that

$$\mathbb{P}\left(Y_t(v) = Y_t(v')\right) - \mathbb{P}\left(X_t(v) = X_t(v')\right) = \frac{1}{2}\mathbb{P}\left(\mathcal{A}_{v,v'}\right).$$

Now, as the history of v and v' are both continuous time random walks,

$$\mathbb{P}\left(\mathscr{F}_{\mathrm{SUP}}(v,0,t) = \{v\}, \mathscr{F}_{\mathrm{SUP}}(v',0,t) = \{v'\}\right) \geq c_1 \frac{1}{(t'_{\mathfrak{m}})^2} e^{-2\theta t},$$

since the probability is that two random walks started from neighboring vertices do not intersect until time $t'_{\mathfrak{m}}$ and return to their starting locations is at least $c_1/(t'_{\mathfrak{m}})^2$ and the probability that neither walk dies is at least $e^{-2\theta t}$. They then have a constant probability of merging for some $t_1 < 0$ so

$$\mathbb{P}\left(\mathcal{A}_{v,v'}\right) \geq c_2 \frac{1}{(t'_{\mathfrak{m}})^2} e^{-2\theta t},$$

and if we label the vertices around the cycle as $u_1, \ldots, u_n$ then for large n,

$$\mathbb{E}\left[\sum_{i=1}^{n-1} Y_t(u_i)Y_t(u_{i+1}) - X_t(u_i)X_t(u_{i+1})\right] \geq c_2 \frac{1}{(t'_{\mathfrak{m}})^2} e^{-2\theta t} n \geq \sqrt{n}\log n.$$

By the exponential decay of correlation in the stationary distribution of the 1D Ising model,

$$\mathrm{Var}\left(\sum_{i=1}^{n-1} Y_t(u_i)Y_t(u_{i+1})\right) \leq Cn \,.$$

Since the spins in different clusters are independent and with probability at least $1 - n^{-10}$ there are no clusters whose diameter is greater than $C \log \log n$ for some large $C(\beta)$ we have that when $|i - i'| \geq C \log n$ then

$$\mathrm{Cov}\left(X_t(u_i)X_t(u_{i+1}), X_t(u_{i'})X_t(u_{i'+1})\right) = O(n^{-10})\,,$$

and hence

$$\mathrm{Var}\left(\sum_{i=1}^{n-1} X_t(u_i)X_t(u_{i+1})\right) \leq Cn \log n \,.$$

It follows by Chebyshev's inequality that

$$\mathbb{P}\left(\sum_{i=1}^{n-1} Y_t(u_i)Y_t(u_{i+1}) \geq \mathbb{E}\left[\sum_{i=1}^{n-1} Y_t(u_i)Y_t(u_{i+1})\right] - \frac{1}{2}\sqrt{n}\log n\right) = 1 - o(1)$$

while

$$\mathbb{P}\left(\sum_{i=1}^{n-1} X_t(u_i)X_t(u_{i+1}) < \mathbb{E}\left[\sum_{i=1}^{n-1} Y_t(u_i)Y_t(u_{i+1})\right] - \frac{1}{2}\sqrt{n}\log n\right) = 1 - o(1)\,,$$

and hence $\|X_t - \pi\|_{\mathrm{TV}} \to 1$.

6.3. **Quenched analysis.** Here we show that on the cycle, there is at most a minor, $O(\log \log n)$ improvement when starting from a typical random initial configuration, since almost all configurations bias the magnetizations of most vertices. For a configuration x_0 denote

$$R_t(u, x_0) = \sum_{u' \in V} P_t(u, u')x_0(u')\,, \tag{6.5}$$

where $P_t(u, u')$ is the transition probability of a continuous time walk with jumps at rate $(1 - \theta)$. Observe that

$$\mathbb{E}_{x_0}[X_t(u)] = e^{-\theta t} R_t(u, x_0)\,. \tag{6.6}$$

Proposition 6.5. *Suppose that there exists some sequence $a_n = n^{o(1)}$ such that for any large n,*

$$\frac{1}{n}\sum_{u \in \Lambda} |R_t(u, x_0)| \geq \frac{1}{a_n} \qquad at \qquad t = \frac{1}{2\theta}\log n - \frac{1}{\theta}\log\log n - \frac{1}{\theta}\log a_n \,. \tag{6.7}$$

Then $\|\mathbb{P}_{x_0}(X_t \in \cdot) - \pi(\cdot)\|_{\mathrm{TV}} = 1 - o(1)$ as $n \to \infty$. Furthermore, if x_0 is uniformly chosen over $\{\pm 1\}^{\Lambda}$ then there exists some $C = C(\beta) > 0$ such that (6.7) holds with probability $1 - o(1)$ for $a_n = C \log n$.

Proof. Let X_0 be a uniformly chosen initial configuration. Since $R_t(u, X_0)$ is a sum of independent increments, when t and n are both large, $R_t(u, X_0)$ is approximately $\mathcal{N}(0, \sum_{u' \in V} P_t(u, u')^2)$ by the Central Limit Theorem; in particular, $\mathbb{E}|R_t(u, X_0)| \geq c/t$ for some fixed $c > 0$, and

$$\mathbb{E}\left[\frac{1}{n}\sum_{u \in \Lambda} |R_t(u, X_0)|\right] \geq \frac{c}{\log n}$$

for another fixed $c > 0$ provided t has order $\log n$. From the decay of $P_t(u, u')$ we infer that if $|u - u'| \geq C \log n$ for a large enough $C > 0$ then $\mathrm{Cov}(|R_t(u)|, |R_t(u')|) \leq n^{-10}$, thus implying that

$$\mathrm{Var}\left(\frac{1}{n}\sum_{u\in\Lambda}|R_t(u, X_0)|\right) = O\left(\frac{\log n}{n}\right),$$

and altogether

$$\mathbb{P}\left(\frac{1}{n}\sum_{u\in\Lambda}|R_t(u, X_0)| > \frac{c}{2\log n}\right) \to 1.$$

This establishes (6.7) with probability going to 1 for $a_n = C \log n$ with a suitably chosen $C = C(\theta) > 0$.

By the same decay of correlations, for vertices with $|u - u'| \geq C \log n$ for some large enough $C > 0$ the histories do not merge in the interval $(0, t]$ with probability $1 - O(n^{-10})$. In this case,

$$\mathrm{Cov}_{x_0}\left(\mathrm{sign}(R_t(u, x_0))X_t(u), \mathrm{sign}(R_t(u', x_0))X_t(u')\right) = O(n^{-10}),$$

and again we can deduce that

$$\mathrm{Var}\left(\frac{1}{n}\sum_{u\in\Lambda}\mathrm{sign}(R_t(u, x_0)X_t(u)\right) = O\left(\frac{\log n}{n}\right).$$

Recalling from (6.6) that $\mathbb{E}_{x_0}[\mathrm{sign}(R_t(u, x_0))X_t(u)] = e^{-\theta t}|R_t(u, x_0)|$, Chebyshev's inequality yields

$$\mathbb{P}_{x_0}\left(\frac{1}{n}\sum_{u\in\Lambda}\left(\mathrm{sign}(R_t(u, x_0))X_t(u) - |R_t(u, x_0)|\right) > \frac{1}{2a_n}\right) \leq O\left(\frac{\log n}{n}(e^{\theta t}a_n)^2\right) = O\left(\frac{1}{\log n}\right)$$

by the definition of t. Thus, we conclude that for any x_0 satisfying (6.7),

$$\mathbb{P}_{x_0}\left(\frac{1}{n}\sum_{u\in\Lambda}\mathrm{sign}(R_t(u, x_0))X_t(u) > \frac{1}{2a_n}\right) \to 1. \tag{6.8}$$

By contrast, if Y is chosen independently according to the Ising measure then by the exponential decay of spatial correlations we have that

$$\mathrm{Var}\left(\frac{1}{n}\sum_{u\in\Lambda}\mathrm{sign}(R_t(u, x_0))Y(u)\right) = O(1/n),$$

and since $\mathbb{E}\sum_{u\in\Lambda}\mathrm{sign}(R_t(u))Y(u) = 0$ while $1/a_n \gg 1/\sqrt{n}$ we can infer that

$$\mathbb{P}\left(\frac{1}{n}\sum_{u\in\Lambda}\mathrm{sign}(R_t(u, x_0))Y(u) > \frac{1}{2a_n}\right) \to 0. \tag{6.9}$$

Comparing equations (6.8) and (6.9) implies that

$$\|\mathbb{P}_{x_0}(X_t \in \cdot) - \pi(\cdot)\|_{\mathrm{TV}} = 1 - o(1),$$

completing the result.   ∎

References

[1] M. Aizenman and R. Holley, *Rapid convergence to equilibrium of stochastic Ising models in the Dobrushin Shlosman regime*, (Minneapolis, Minn., 1984), IMA Vol. Math. Appl., vol. 8, Springer, New York, 1987, pp. 1–11.

[2] D. Aldous, *Random walks on finite groups and rapidly mixing Markov chains*, Seminar on probability, XVII, 1983, pp. 243–297.

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

[4] F. Cesi and F. Martinelli, *On the layering transition of an SOS surface interacting with a wall. II. The Glauber dynamics*, Comm. Math. Phys. **177** (1996), no. 1, 173–201.

[5] P. Diaconis, *The cutoff phenomenon in finite Markov chains*, Proc. Nat. Acad. Sci. U.S.A. **93** (1996), no. 4, 1659–1664.

[6] P. Diaconis, R. L. Graham, and J. A. Morrison, *Asymptotic analysis of a random walk on a hypercube with many dimensions*, Random Structures Algorithms **1** (1990), no. 1, 51–72.

[7] P. Diaconis and L. Saloff-Coste, *Comparison techniques for random walk on finite groups*, Ann. Probab. **21** (1993), no. 4, 2131–2156.

[8] P. Diaconis and L. Saloff-Coste, *Comparison theorems for reversible Markov chains*, Ann. Appl. Probab. **3** (1993), no. 3, 696–730.

[9] P. Diaconis and L. Saloff-Coste, *Logarithmic Sobolev inequalities for finite Markov chains*, Ann. Appl. Probab. **6** (1996), no. 3, 695–750.

[10] P. Diaconis and M. Shahshahani, *Generating a random permutation with random transpositions*, Z. Wahrsch. Verw. Gebiete **57** (1981), no. 2, 159–179.

[11] P. Diaconis and M. Shahshahani, *Time to reach stationarity in the Bernoulli-Laplace diffusion model*, SIAM J. Math. Anal. **18** (1987), no. 1, 208–218.

[12] J. Ding, E. Lubetzky, and Y. Peres, *The mixing time evolution of Glauber dynamics for the mean-field Ising model*, Comm. Math. Phys. **289** (2009), no. 2, 725–764.

[13] R. A. Holley, *On the asymptotics of the spin-spin autocorrelation function in stochastic Ising models near the critical temperature*, Spatial stochastic processes, Progr. Probab., vol. 19, Birkhäuser Boston, Boston, MA, 1991, pp. 89–104.

[14] R. A. Holley and D. W. Stroock, *Logarithmic Sobolev inequalities and stochastic Ising models*, J. Statist. Phys. **46** (1987), no. 5-6, 1159–1194.

[15] W. S. Kendall, F. Liang, and J.-S. Wang (eds.), *Markov chain Monte Carlo Innovations and Applications*, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2005.

[16] D. A. Levin, M. Luczak, and Y. Peres, *Glauber dynamics for the Mean-field Ising Model: cut-off, critical power law, and metastability*, Probab. Theory Related Fields **146** (2010), no. 1–2, 223–265.

[17] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov chains and mixing times*, 2008.

[18] T. M. Liggett, *Interacting particle systems*, Classics in Mathematics, Springer-Verlag, Berlin, 2005.

[19] E. Lubetzky and A. Sly, *Cutoff phenomena for random walks on random regular graphs*, Duke Math. J. **153** (2010), no. 3, 475–510.

[20] E. Lubetzky and A. Sly, *Cutoff for the Ising model on the lattice*, Invent. Math. **191** (2013), no. 3, 719-755.

[21] E. Lubetzky and A. Sly, *Cutoff for general spin systems with arbitrary boundary conditions*, Comm. Pure. Appl. Math. **67** (2014), no. 6, 982–1027.

[22] E. Lubetzky and A. Sly, *Universality of cutoff for the Ising model*, preprint. Available at `arXiv:1407.1761` (2014).

[23] F. Martinelli, *Lectures on Glauber dynamics for discrete spin models*, Lectures on probability theory and statistics (Saint-Flour, 1997), Lecture Notes in Math., vol. 1717, Springer, Berlin, 1999, pp. 93–191.

[24] F. Martinelli and E. Olivieri, *Approach to equilibrium of Glauber dynamics in the one phase region. I. The attractive case*, Comm. Math. Phys. **161** (1994), no. 3, 447–486.

[25] F. Martinelli and E. Olivieri, *Approach to equilibrium of Glauber dynamics in the one phase region. II. The general case*, Comm. Math. Phys. **161** (1994), no. 3, 487–514.

[26] F. Martinelli, E. Olivieri, and R. H. Schonmann, *For 2-D lattice spin systems weak mixing implies strong mixing*, Comm. Math. Phys. **165** (1994), no. 1, 33–47.

[27] J. Miller and Y. Peres, *Uniformity of the uncovered set of random walk and cutoff for lamplighter chains*, Ann. Probab. **40** (2012), no. 2, 535–577.

[28] J. G. Propp and D. B. Wilson, *Exact sampling with coupled Markov chains and applications to statistical mechanics*, Random Structures Algorithms **9** (1996), no. 1-2, 223–252.

[29] L. Saloff-Coste, *Lectures on finite Markov chains*, Lectures on probability theory and statistics (Saint-Flour, 1996), Lecture Notes in Math., vol. 1665, Springer, Berlin, 1997, pp. 301–413.

[30] L. Saloff-Coste, *Random walks on finite groups*, Probability on discrete structures, 2004, pp. 263–346.

[31] A. Sokal, *Monte Carlo methods in statistical mechanics: foundations and new algorithms*, Lecture notes, Ecole Polytechnique de Lausanne, 1989.

[32] D. W. Stroock and B. Zegarliński, *The equivalence of the logarithmic Sobolev inequality and the Dobrushin-Shlosman mixing condition*, Comm. Math. Phys. **144** (1992), no. 2, 303–323.

[33] D. W. Stroock and B. Zegarliński, *The logarithmic Sobolev inequality for discrete spin systems on a lattice*, Comm. Math. Phys. **149** (1992), no. 1, 175–193.

Eyal Lubetzky

Courant Institute, New York University, 251 Mercer Street, New York, NY 10012, USA.
  E-mail address: `eyal@courant.nyu.edu`

Allan Sly

Department of Statistics, UC Berkeley, Berkeley, CA 94720, USA.
  E-mail address: `sly@stat.berkeley.edu`