AN EXPOSITION TO INFORMATION PERCOLATION FOR THE ISING MODEL

EYAL LUBETZKY AND ALLAN SLY

ABSTRACT. Information percolation is a new method for analyzing stochastic spin systems through classifying and controlling the clusters of information-flow in the space-time slab. It yielded sharp mixing estimates (cutoff with an $O(1)$-window) for the Ising model on $\mathbb{Z}^d$ up to the critical temperature, as well as results on the effect of initial conditions on mixing. In this expository note we demonstrate the method on lattices (more generally, on any locally-finite transitive graph) at very high temperatures.

1. INTRODUCTION

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

$$\pi(\sigma) = \mathcal{Z}^{-1} e^{\beta \sum_{uv \in E} \sigma(u)\sigma(v)}, \tag{1.1}$$

where $\mathcal{Z}$ is a normalizer (the partition-function) and β is the inverse-temperature, here taken to be non-negative (ferromagnetic). These definitions extend to infinite locally finite graphs (see, e.g., [5,12]). The (continuous-time) heat-bath Glauber dynamics for the Ising model is the Markov chain—reversible w.r.t. the Ising measure π—where each site is associated with a rate-1 Poisson clock, and as the clock at some site u rings, the spin of u is replaced by a sample from the marginal of π given all other spins.

An important notion of measuring the convergence of a Markov chain (X_t) to its stationarity measure π is its total-variation mixing time, denoted $t_{\mathrm{MIX}}(\varepsilon)$ for a precision parameter $0 < \varepsilon < 1$:

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

where here and in what follows $\mathbb{P}_{x_0}$ denotes the probability given $X_0 = x_0$, and the total-variation distance $\|\nu_1 - \nu_2\|_{\mathrm{TV}}$ is defined as $\max_{A \subset \Omega} |\nu_1(A) - \nu_2(A)| = \frac{1}{2} \sum_{\sigma \in \Omega} |\nu_1(\sigma) - \nu_2(\sigma)|$.

The impact of the precision parameter ε in this definition is addressed by the *cutoff phenomenon*—a concept going back to the pioneering works [1,2,4]—roughly saying the choice of any fixed ε does not change the asymptotics of $t_{\mathrm{MIX}}(\varepsilon)$ as the system size goes to infinity. Formally, a family of ergodic finite Markov chains (X_t), indexed by an implicit parameter n, exhibits *cutoff* iff $t_{\mathrm{MIX}}(\varepsilon) = (1 + o(1)) t_{\mathrm{MIX}}(\varepsilon')$ for any fixed $0 < \varepsilon, \varepsilon' < 1$. The *cutoff window* addresses the correction terms: a sequence $w_n = o(t_{\mathrm{MIX}}(1/2))$ is a cutoff window if $t_{\mathrm{MIX}}(\varepsilon) = t_{\mathrm{MIX}}(1 - \varepsilon) + O(w_n)$ for any $0 < \varepsilon < 1$ with an implicit constant that may depend on ε. That is, the Markov chain exhibits a sharp transition in its convergence to equilibrium, whereby its distance drops abruptly (along the cutoff window) from near 1 to near 0.

Establishing cutoff can be highly challenging (see the survey [3]), even for simple random walk with a uniform stationary measure: e.g., it is conjectured that on *every* transitive expander graph the random walk exhibits cutoff, yet there is not a single example of a transitive expander where cutoff was confirmed (even without transitivity, there were no examples of expanders with cutoff before [6,7]). As for Glauber

dynamics for the Ising model, where our understanding of π is far more limited, until recently cutoff was confirmed only in a few cases, with first results on lattices appearing in the works [8, 9].

The methods used to analyze the dynamics for the Ising model on $\mathbb{Z}^d$ in those works had several caveats: the reliance on delicate features such as log-Sobolev inequalities did not cover the full high temperature regime in dimensions $d \geq 3$, and did not give a correct bound on the cutoff window; furthermore, the argument immediately broke on any geometry with exponential ball growth (expanders). Recently, in [10] and its companion paper [11], we introduced a new framework called *information percolation*, which does not have these limitations and can hopefully be used to analyze a wide range of stochastic spin systems. We demonstrated its application in analyzing the stochastic Ising model on $\mathbb{Z}^d$ up to β_c, and to compare the effects of different starting configurations on mixing—e.g., showing that an initial state of i.i.d. spins mixes about twice faster than the all-plus starting state (which is essentially the worst one), while almost every starting state is as bad as the worst one.

Here we demonstrate a simpler application of the method for the Ising model on a fixed-degree transitive graph at very high temperatures, establishing cutoff within an $O(1)$-window around

$$t_{\mathfrak{m}} = \inf \left\{ \, t > 0 \, : \, \mathfrak{m}_t \leq 1/\sqrt{n} \, \right\} , \tag{1.2}$$

where $\mathfrak{m}_t = \mathbb{E} X_t^+(v)$ is the magnetization at the origin at time $t > 0$ (in which X_t^+ denotes the dynamics started from all-plus). That is, $t_{\mathfrak{m}}$ is the time at which the expected sum-of-spins drops to a square-root of the volume, where intuitively it is absorbed within the normal fluctuations of the Ising measure.

Theorem 1. *For any $d \geq 2$ there exists $\beta_0 = \beta_0(d) > 0$ such that the following holds. Let G be a transitive graph on n vertices with degree d. For any fixed $0 < \varepsilon < 1$, continuous-time Glauber dynamics for the Ising model on G at inverse-temperature $0 \leq \beta \leq \beta_0$ satisfies $t_{\mathrm{MIX}}(\varepsilon) = t_{\mathfrak{m}} \pm O_\varepsilon(1)$.*

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

2. Basic setup of information percolation and proof of Theorem 1

In this section we define and classify the information percolation clusters in their most basic form (commenting how this setup may be altered in more delicate situations such as β close to β_c on $\mathbb{Z}^d$), then reduce the proof of Theorem 1 to estimating the probability that a cluster is "red".

2.1. **Red, green and blue information percolation clusters.** The dynamics can be viewed as a deterministic function of X_0 and a random "update sequence" of the form $(J_1, U_1, t_1), (J_2, U_2, t_2), \ldots$, where $0 < t_1 < t_2 < \ldots$ are the update times (the ringing of the Poisson clocks), the J_i's are i.i.d. uniformly chosen sites (whose clocks ring), and the U_i's are i.i.d. uniform variables on $[0, 1]$ (to generate coin tosses). According to this representation, one processes the updates sequentially: set $t_0 = 0$; the configuration X_t for all $t \in [t_{i-1}, t_i)$ $(i \geq 1)$ is obtained by updating the site J_i via the unit variable as follows: letting $\sigma = \sum_{v \sim J_i} X_{t_{i-1}}(v)$ denote the current sum-of-spins at the neighbors of J_i, if

$$U_i < e^{\beta\sigma}/(e^{\beta\sigma} + e^{-\beta\sigma}) = \tfrac{1}{2}\left(1 + \tanh(\beta\sigma)\right) \tag{2.1}$$

then the new spin at J_i is chosen to be plus and otherwise it is set to minus.

Equivalently, one may evaluate this deterministic function backward in time rather than forward: sort the same update sequence $\{(J_i, U_i, t_i)\}$ such that $t_\star > t_1 > t_2 > \ldots$, where $t_\star$ is a designated time

at which we wish to analyze the distribution $X_{t_\star}$ (and argue it is either close or far from equilibrium). To construct $X_{t_\star}$, we again process the updates sequentially, now setting $t_0 = t_\star$ and determining X_t for all $t \in [t_{i+1}, t_i)$ in step i, where the value of the spins at the neighbors of J_i (determining σ and the probability of ± 1 in the update as above) is evaluated recursively via the suffix of the update sequence.

Examining (2.1), this backward simulation of the dynamics can be made more efficient, since even if all the neighbors of the site that is being updated are (say) minus, there is still a positive probability of $\frac{1}{2}(1 - \tanh(\beta d))$ for a plus update: We may therefore decide to first examine U_i, and if $U_i < \theta$ for

$$\theta = \theta_{\beta,d} := 1 - \tanh(\beta d) \tag{2.2}$$

we will set the new spin to plus/minus as a fair coin toss irrespective of σ (namely, to plus iff $U_i < \theta/2$); otherwise, we will recursively compute the spins at the neighbors of J_i, and set the new spin to plus iff

$$0 \leq U_i - \theta < \tfrac{1}{2}(1 + \tanh(\beta\sigma) - \theta) = \tfrac{1}{2}(\tanh(\beta\sigma) + \tanh(\beta d)) \,.$$

(The right-hand is $0, 1 - \theta$ for the extreme $\sigma = \pm d$, while for other values of σ the rule depends on U_i.)

We have arrived at a branching process in the space-time slab: to recover $X_{t_\star}(v)$ we track its lineage backward in time, beginning with the temporal edge in the space-time slab $V \times [0, t_\star]$ between $(v, t_\star)$ and $(v, t_i \vee 0)$, where t_i is the time of the latest update to v. If no such update was encountered and this "branch" has survived to time $t = 0$, it assumes the value of the initial configuration $X_0(v)$. Alternatively, an update at time t_i has two possible consequences: if it features $U_i < \theta$—an *oblivious update*—the branch is terminated as the spin can be recovered via a fair coin toss using U_i; otherwise, we branch out to the neighbors of v, adding spatial edges between (v, t_i) and (u, t_i) for all $u \sim v$, and continue developing the update histories of each of them until the process dies out or reaches time 0. This produces a graph in the space-time slab which we denote by $\mathscr{H}_v$, and let $\mathscr{H}_v(t)$ be its intersection with the slab $V \times \{t\}$ (viewed as a subset of V); further let $\mathscr{H}_A = \bigcup_{v \in A} \mathscr{H}_v$ and set $\mathscr{H}_A(t)$ analogously. The *information percolation* clusters are the connected components of the graph consisting of $\mathscr{H}_V$.

By definition, $X_{t_\star}(A)$ is a deterministic function of the U_i's corresponding to points in $\mathscr{H}_A$ and of $X_0(\mathscr{H}_A(0))$, the initial values at the intersection of $\mathscr{H}_A$ with the slab $V \times \{0\}$; in particular, if $\mathscr{H}_V(0) = \emptyset$ then $X_{t_\star}$ is independent of X_0, and therefore its law is precisely the Ising distribution π (for instance, in that scenario we could have taken $X_0 \sim \pi$ and then $X_{t_\star} \sim \pi$ by the invariance of π). However, we note that waiting for a time $t_\star$ large enough such that $\mathscr{H}_V$ would be guaranteed not to survive to time $t = 0$ with high probability is an overkill; the correct mixing time is the point at which $|\mathscr{H}_V(0)| \asymp \sqrt{n}$, whence the effect of X_0 on $X_{t_\star}$ would be absorbed in the normal fluctuations of π.

Remark. The above defined rule for developing the update histories either terminated a lineage or branched it to its d neighbors. In different applications of the method, it is crucial to appropriately select other rules with the correct marginal of the heat-bath dynamics.

For instance, in [11] we prove results *à la* Theorem 1 on any graph (including, e.g., expanders) and any $\beta < \kappa/d$ for an absolute constant $\kappa > 0$—the correct dependence on d up to the value of κ—by selecting $k = 0, 1, \ldots, d$ with probability p_k and then deciding the new spin via a function of a uniformly chosen k-element subset of the neighbors, with the probabilities p_k (and the corresponding functions to be applied) following from a discrete Fourier expansion of the original Glauber dynamics update rule.

Furthermore, in [10], to analyze $\mathbb{Z}^d$ arbitrarily close to β_c, instead of describing such a rule explicitly one relies on the *existence* of an efficient rule thanks to the exponential decay of spin-spin correlations.

Example. In the Ising model on $\mathbb{Z}_n$, with probability θ we assign a uniform ± 1 spin, and with probability $1 - \theta$ we expose $\sigma \in \{0, \pm 2\}$ and select the consensus spin in case $\sigma = \pm 2$ or a uniform ± 1 spin in case $\sigma = 0$. Hence, an equivalent rule would be to terminate the branch with probability θ and otherwise to select a uniform neighbor and copy its spin, so $\mathscr{H}_v$ is merely a continuous-time simple random walk killed at rate θ, and $\mathscr{H}_V$ consists of n coalescing random walks killed at rate θ. The probability that $\mathscr{H}_v(0) \neq \emptyset$ (survival from time $t_\star$ to time $t = 0$) is then $\exp(-\theta t_\star)$, and we would have $\mathbb{E}|\mathscr{H}_V(0)| \asymp \sqrt{n}$ at $t_\star = (2\theta)^{-1} \log n + O(1)$, which is exactly the cutoff location.

The key to the analysis is to classify the information percolation clusters to three types, where one of these classes will be revealed (conditioned on), and the other two will represent competing factors, which balance exactly at the correct point of mixing. In the basic case, the classification is as follows:
- A cluster $\mathcal{C} = \mathscr{H}_A$ is RED if, given the update sequence, its final state $X_{t_\star}(A)$ is a nontrivial function of the initial configuration X_0 (in particular, it survives to time $t = 0$, i.e., $\mathscr{H}_A(0) \neq \emptyset$).
- A cluster $\mathcal{C} = \mathscr{H}_A$ is BLUE if A is a singleton ($A = \{v\}$ for some $v \in V$) and its history does not survive to time zero (in particular, $X_{t_\star}(A)$ does not depend on X_0).
- Every other cluster $\mathcal{C}$ is GREEN.

Let $V_{\text{RED}} = \{v : \mathscr{H}_v \subset \mathcal{C} \in \text{RED}\}$ be the set of all vertices whose histories belong to red clusters, and let $\mathscr{H}_{\text{RED}} = \mathscr{H}_{V_{\text{RED}}}$ be their collective history (similarly for blue/green). By a slight abuse of notation, we write $A \in \text{RED}$ to denote that $\mathscr{H}_A \in \text{RED}$ (similarly for blue/green), yet notice the distinction between $A \in \text{RED}$ and $A \subset V_{\text{RED}}$ (the former means that $\mathscr{H}_A$ is a full red cluster, rather than covered by ones).

Remark. Various other classifications to red/green/blue can be used so that RED captures the entire dependence on the initial configuration and BLUE forms a product-measure. For instance, in order to study the effect of initial conditions in [11] we let a cluster be red if at least two branches of it survive to time $t = 0$ and *coalesce upon continuing to develop their history* along $t \in (-\infty, 0]$; and in order to carry the analysis of [10] up to β_c, the classification is more delicate, and involves the histories along a burn-in phase near time $t_\star$ which allows one to amplify the subcritical nature of the clusters.

Note that if $\{v\} \in \text{BLUE}$ then, by definition and symmetry, the distribution of $X_{t_\star}(v)$ is uniform ± 1. On the other hand, while the spins of a green cluster $\mathcal{C}$ are also independent of X_0, its spin-set at time $t_\star$ can have a highly nontrivial distributions due to the dependencies between the intersecting update histories. It is these green clusters that embody the complicated structure of the Ising measure.

Example. In the Ising model on $\mathbb{Z}_n$, as explained above, an information percolation cluster corresponds to a maximal collection of random walks that coalesced. A cluster is red if it survived to time 0 (the rule of copying the spin at the location of the walk guarantees a nontrivial dependency on X_0); it is blue if the random walk started at v dies out before coalescing with any other walk and before reaching time $t = 0$; and it is green otherwise. Observe that the sites of a green cluster at time $t_\star$ all have the same spin—a uniform ± 1 spin, independent of X_0—and the probability of u, v belonging to the same green cluster decays exponentially in $|u - v|$ (as the walks become more likely to die out than merge).

As the green clusters demonstrate the features of the Ising measure, it is tempting to analyze them in order to understand $X_{t_\star}$. The approach we will follow does the opposite: we will *condition on the entire set of histories* $\mathscr{H}_{V_{\mathrm{GREEN}}}$ (or $\mathscr{H}_{\mathrm{GREEN}}$ for brevity), and study the remaining (red and blue) clusters.

As we hinted at when saying above that if $\mathscr{H}_V(0) = \emptyset$ then $X_{t_\star} \sim \pi$, the Ising measure π can be perfectly simulated via developing the histories backwards in time until every branch terminates without any special exception at time $t = 0$. (This would be equivalent to taking $t_\star$ larger and larger until $\mathscr{H}_V$ would be guaranteed not to survive, essentially as in the ingenious *Coupling From The Past* method of perfect simulation due to Propp and Wilson [14]). Thus, we can couple $X_{t_\star}$ to π via the same update sequence, in which case the distribution of V_{GREEN}—albeit complicated—is identical (green clusters never reach $t = 0$), allowing us to only consider blue/red clusters (however in a difficult conditional space where percolation clusters are forbidden from touching various parts of the space-time slab).

The second key is to keep the blue clusters—which could have been coupled to π just like the green ones, as they too do not survive to time $t = 0$—in order to water down the effect of the red clusters (which, by themselves, are starkly differently under $X_{t_\star}$ and π). We will show that, conditioned on $\mathscr{H}_{\mathrm{GREEN}}$, the measures π and $X_{t_\star}$ are both very close to the uniform measure (and therefore to each other), i.e., essentially as if there were no red clusters at all and every $v \in V \setminus V_{\mathrm{GREEN}}$ belonged to V_{BLUE}. Indeed, conditioning on the green clusters replaces the Ising measure by a contest between blue and red clusters: at large $t_\star$, the effect of RED is negligible and we get roughly the uniform measure; at small $t_\star$, the effect of RED dominates and X_0 will have a noticeable affect on $X_{t_\star}$; the balancing point $t_{\mathfrak{m}}$ has the RED clusters make an effect on $X_{t_\star}$, but just within the normal fluctuations of π.

Showing that the effect of the red clusters is negligible just beyond the cutoff location will be achieved via a clever lemma of Miller and Peres [13] that bounds the L^2-distance of a measure from the uniform measure in terms of a certain exponential moment. In our setting, this reduces the problem to showing:

$$\mathbb{E}\left[2^{|V_{\mathrm{RED}} \cap V'_{\mathrm{RED}}|} \mid \mathscr{H}_{\mathrm{GREEN}}\right] \approx 1 \quad \text{in probability as } n \to \infty, \tag{2.3}$$

where V_{RED} and V'_{RED} are independent instances of the vertices whose histories are part of red clusters.

Example. Recall the coalescing random walks representation for the information percolation clusters of the Ising model on $\mathbb{Z}_n$, and suppose we wanted to estimate $\mathbb{E}[2^{|V_{\mathrm{RED}} \cap V'_{\mathrm{RED}}|}]$, i.e., the left-hand of (2.3) *without* the complicated conditioning on $\mathscr{H}_{\mathrm{GREEN}}$. Then $v \in V_{\mathrm{RED}}$ iff its random walk survives to time $t = 0$, which has probability $e^{-\theta t_\star}$. By the independence, $\mathbb{P}(v \in V_{\mathrm{RED}} \cap V'_{\mathrm{RED}}) = e^{-2\theta t_\star}$. If the sites were independent (they are not of course, but the intuition is still correct), then $\mathbb{E}[2^{|V_{\mathrm{RED}} \cap V'_{\mathrm{RED}}|}]$ would break into $\prod_v \mathbb{E}[1 + \mathbb{1}_{\{v \in V_{\mathrm{RED}} \cap V'_{\mathrm{RED}}\}}] \leq \exp(ne^{-2\theta t_\star})$, which for $t_\star = (2\theta)^{-1} \log n + C$ is at most $\exp(e^{-2\theta C})$ that approaches 1 as we increase the constant $C > 0$ in $t_\star$. (The actual calculation of the exponential moment given $\mathscr{H}_{\mathrm{GREEN}}$, especially at very low temperatures, requires quite a bit more care.)

The key to obtaining the bound on the exponential moment in (2.3), which is the crux of the proof, is estimating a conditional probability that $A \in \mathrm{RED}$, in which we condition not only on $\mathscr{H}_{\mathrm{GREEN}}$ but on the entire collective histories of every vertex outside of A, and that A itself is either the full intersection of a red cluster with the top slab $V \times \{t_\star\}$ or a collection of blue singletons. Formally, let

$$\Psi_A = \sup_{\mathcal{X}} \mathbb{P}\left(A \in \mathrm{RED} \mid \mathscr{H}_A^- = \mathcal{X}, \{A \in \mathrm{RED}\} \cup \{A \subset V_{\mathrm{BLUE}}\}\right), \tag{2.4}$$

where

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

noting that, towards estimating the probability of $A \in \mathrm{RED}$, the effect of conditioning on $\mathscr{H}_A^-$ amounts to requiring that $\mathscr{H}_A$ must not intersect $\mathscr{H}_A^-$.

Lemma 2.1. *For any $d \ge 2$ and $\lambda > 0$ there exist $\beta_0, C_0 > 0$ such that if $\beta < \beta_0$ then for any $t_\star > 0$, any $A \subset V$ and any large enough n, the conditional probability that $A \in \mathrm{RED}$ at time $t_\star$ satisfies*

$$\Psi_A \le C_0 \, \mathfrak{m}_{t_\star} \, e^{-\lambda \mathfrak{W}(A)} \,,$$

where $\mathfrak{W}(A)$ is the size of the smallest connected subgraph containing A.

For intuition, recall that $A \in \mathrm{RED}$ if the histories $\{\mathscr{H}_v : v \in A\}$ are all connected and $X_{t_\star}(A)$ is a nontrivial function of X_0 (in particular, $\mathscr{H}_A(0) \ne \emptyset$). This is closely related to the probability that for a single $v \in A$ we have that $X_{t_\star}(v)$ is a nontrivial function of X_0 (i.e., $\mathscr{H}_v$ survives to time $t = 0$ and creates a nontrivial dependence on X_0, whence the connected component of $(v, t_\star)$ is a red cluster). Indeed, the probability of the latter event is exactly

$$\mathfrak{m}_{t_\star} = \mathbb{E} X_{t_\star}^+(v) = \mathbb{P}(X_{t_\star}^+(v) \ne X_{t_\star}^-(v)) \,, \tag{2.5}$$

which explains the term $\mathfrak{m}_{t_\star}$ in Lemma 2.1. The extra term $\exp(-\lambda \mathfrak{W}(A))$ is due to the requirement that the histories of A must spatially connect, thus the projection of the cluster on V is a connected subgraph containing A (whose size is at least $\mathfrak{W}(A)$ by definition).

2.2. **Upper bound modulo Lemma 2.1.** Our goal is to show that $d_{\mathrm{TV}}(t_\star) < \varepsilon$ for $t_\star = t_{\mathfrak{m}} + s_\star$ with a suitably large $s_\star > 0$, where $d_{\mathrm{TV}}(t) = \max_{x_0} \|\mathbb{P}_{x_0}(X_t \in \cdot) - \pi\|_{\mathrm{TV}}$. By Jensen's inequality, $\|\psi - \varphi\|_{\mathrm{TV}} \le \mathbb{E}\big[\|\psi(\cdot \mid Z) - \varphi(\cdot \mid Z)\|_{\mathrm{TV}}\big]$ for any two distributions ψ and φ on a finite probability space Ω and random variable Z. Applied with $\mathscr{H}_{\mathrm{GREEN}}$ playing the role of Z, and letting $X_0' \sim \pi$,

$$d_{\mathrm{TV}}(t) \le \max_{x_0} \mathbb{E}\Big[\Big\| \mathbb{P}_{x_0}(X_t \in \cdot \mid \mathscr{H}_{\mathrm{GREEN}}) - \mathbb{P}_{X_0'}(X_t \in \cdot \mid \mathscr{H}_{\mathrm{GREEN}}) \Big\|_{\mathrm{TV}} \Big]$$

$$\le \sup_{\mathscr{H}_{\mathrm{GREEN}}} \max_{x_0} \Big\| \mathbb{P}_{x_0}(X_t \in \cdot \mid \mathscr{H}_{\mathrm{GREEN}}) - \mathbb{P}_{X_0'}(X_t \in \cdot \mid \mathscr{H}_{\mathrm{GREEN}}) \Big\|_{\mathrm{TV}} \,.$$

As explained above, since $X_t(V_{\mathrm{GREEN}})$ is independent of X_0 we can couple it with the chain started at $X_0' \sim \pi$, whence the projection onto $V \setminus V_{\mathrm{GREEN}}$ does not decrease the total-variation distance, and so

$$d_{\mathrm{TV}}(t) \le \sup_{\mathscr{H}_{\mathrm{GREEN}}} \max_{x_0} \Big\| \mathbb{P}_{x_0}(X_t(V \setminus V_{\mathrm{GREEN}}) \in \cdot \mid \mathscr{H}_{\mathrm{GREEN}}) - \mathbb{P}_{X_0'}(X_t(V \setminus V_{\mathrm{GREEN}}) \in \cdot \mid \mathscr{H}_{\mathrm{GREEN}}) \Big\|_{\mathrm{TV}}$$

$$\le 2 \sup_{\mathscr{H}_{\mathrm{GREEN}}} \max_{x_0} \Big\| \mathbb{P}_{x_0}(X_t(V \setminus V_{\mathrm{GREEN}}) \in \cdot \mid \mathscr{H}_{\mathrm{GREEN}}) - \nu_{V \setminus V_{\mathrm{GREEN}}} \Big\|_{\mathrm{TV}} \,, \tag{2.6}$$

where ν_A is the uniform measure on configurations on the sites in A.

We now appeal to the aforementioned lemma of Miller and Peres [13] that shows that, if a measure μ on $\{\pm 1\}^V$ is given by sampling a variable $R \subset V$ and using an arbitrary law for its spins and a product of Bernoulli($\frac{1}{2}$) for $V \setminus R$, then the L^2-distance of μ from the uniform measure is at most $\mathbb{E} 2^{|R \cap R'|} - 1$ for i.i.d. copies R, R'. (See [10, Lemma 4.3] for a generalization of this to a product of general measures that is imperative for the information percolation analysis on $\mathbb{Z}^d$ at β near β_c.)

Lemma 2.2 ([13]). *Let $\Omega = \{\pm 1\}^V$ for a finite set V. For each $R \subset V$, let φ_R be a measure on $\{\pm 1\}^R$. Let ν be the uniform measure on Ω, and let μ be the measure on Ω obtained by sampling a subset $R \subset V$ via some measure $\tilde{\mu}$, generating the spins of R via φ_R, and finally sampling $V \setminus R$ uniformly. Then*

$$\|\mu - \nu\|_{L^2(\nu)}^2 \le \mathbb{E} 2^{|R \cap R'|} - 1 \qquad \text{where } R, R' \text{ are i.i.d. with law } \tilde{\mu}.$$

Applying this lemma to the right-hand side of (2.6), while recalling that any two measures μ and ν on a finite probability space satisfy $\|\mu - \nu\|_{\mathrm{TV}} = \frac{1}{2}\|\mu - \nu\|_{L^1(\nu)} \le \frac{1}{2}\|\mu - \nu\|_{L^2(\nu)}$, we find that

$$d_{\mathrm{TV}}(t_\star) \le \left(\sup_{\mathscr{H}_{\mathrm{GREEN}}} \mathbb{E}\left[2^{|V_{\mathrm{RED}} \cap V_{\mathrm{RED}'}|} \mid \mathscr{H}_{\mathrm{GREEN}} \right] - 1 \right)^{1/2}, \tag{2.7}$$

where V_{RED} and $V_{\mathrm{RED}'}$ are i.i.d. copies of the variable $\bigcup\{v \in V : \mathscr{H}_v \in \mathrm{RED}\}$. We will reduce the quantity $|V_{\mathrm{RED}} \cap V_{\mathrm{RED}'}|$ to one that involves the Ψ_A variables defined in (2.4) (which will thereafter be controlled via Lemma 2.1) using the next lemma, whose proof is deferred to §2.3 below.

Lemma 2.3. *Let $\{Y_A : A \subset V\}$ be a family of independent indicators satisfying $\mathbb{P}(Y_A = 1) = \Psi_A$. The conditional distribution of V_{RED} given $\mathscr{H}_{\mathrm{GREEN}}$ can be coupled in such a way that*

$$\{A : A \in \mathrm{RED}\} \subset \{A : Y_A = 1\}.$$

The following corollary is then straightforward (see §2.3 for its short proof).

Corollary 2.4. *Let $\{Y_{A,A'} : A, A' \subset V\}$ be a family of independent indicators satisfying*

$$\mathbb{P}(Y_{A,A'} = 1) = \Psi_A \Psi_{A'} \quad \text{for any } A, A' \subset V. \tag{2.8}$$

The conditional distribution of $(V_{\mathrm{RED}}, V_{\mathrm{RED}'})$ given $\mathscr{H}_{\mathrm{GREEN}}$ can be coupled to the $Y_{A,A'}$'s such that

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

Relaxing $|A \cup A'|$ into $|A| + |A'|$, we get

$$\sup_{\mathscr{H}_{\mathrm{GREEN}}} \mathbb{E}\left[2^{|V_{\mathrm{RED}} \cap V_{\mathrm{RED}'}|} \mid \mathscr{H}_{\mathrm{GREEN}} \right] \le \mathbb{E}\left[2^{\sum_{A \cap A' \ne \emptyset} (|A| + |A'|) Y_{A,A'}} \right] = \prod_{A \cap A' \ne \emptyset} \mathbb{E}\left[2^{(|A| + |A'|) Y_{A,A'}} \right],$$

with the equality due to the independence of the $Y_{A,A'}$'s. By the definition of these indicators in (2.8),

$$\prod_{A \cap A' \ne \emptyset} \mathbb{E}\left[2^{(|A| + |A'|) Y_{A,A'}} \right] \le \prod_v \prod_{\substack{A, A' \\ v \in A \cap A'}} \left((2^{|A| + |A'|} - 1) \Psi_A \Psi_{A'} + 1 \right) \le e^{n \left(\sum_{A \ni v_0} 2^{|A|} \Psi_A \right)^2}$$

with v_0 being any fixed vertex, and so, revisiting (2.7), we conclude that

$$d_{\mathrm{TV}}(t_\star) \le \left(e^{n \left(\sum_{A \ni v_0} 2^{|A|} \Psi_A \right)^2} - 1 \right)^{1/2} \wedge 1 \le \sqrt{2n} \sum_{A \ni v_0} 2^{|A|} \Psi_A, \tag{2.9}$$

where we used that $e^x - 1 \le 2x$ for $x \in [0, 1]$. Using Lemma 2.1 with $\lambda = \log(4ed)$ we find that

$$\sum_{A \ni v_0} 2^{|A|} \Psi_A \le C_0 \mathfrak{m}_{t_\star} \sum_k \sum_{\substack{A \ni v_0 \\ \mathfrak{W}(A) = k}} 2^k e^{-\lambda k} \le C_0 \mathfrak{m}_{t_\star} \sum_k (2ede^{-\lambda})^k \le 2C_0 \mathfrak{m}_{t_\star},$$

for some $C_0 = C_0(d) > 0$, and going back to (2.9) shows that $d_{\mathrm{TV}}(t_\star) \leq 2\sqrt{2}C_0\,\mathfrak{m}_{t_\star}\sqrt{n}$.

The upper bound is concluded by the submultiplicativity of the magnetization (see [11, Claim 3.3]):

$$e^{-t}\mathfrak{m}_{t_0} \leq \mathfrak{m}_{t_0+t} \leq e^{-(1-\beta d)t}\mathfrak{m}_{t_0} \qquad \text{for any } t_0, t \geq 0\,, \tag{2.10}$$

since $\mathfrak{m}_{t_\mathfrak{m}} = 1/\sqrt{n}$ (recall (1.2)) and so taking $t_\star = t_\mathfrak{m} + C$ for a suitable $C(\varepsilon) > 0$ yields $d_{\mathrm{TV}}(t_\star) < \varepsilon$. ∎

2.3. **Proof of Lemma 2.3.** We claim that given $\mathscr{H}_{\mathrm{GREEN}}$, if one arbitrarily orders all distinct subsets $A \subset V \setminus V_{\mathrm{GREEN}}$ as $\{A_i\}_{i \geq 1}$ and then successively exposes whether $\{A_i \in \mathrm{RED}\}$, denoting the associated filtration by $\mathcal{F}_i$, then $\mathbb{P}(A_i \in \mathrm{RED} \mid \mathcal{F}_{i-1}) \leq \Psi_{A_i}$. To see this, first increase $\mathbb{P}(A_i \in \mathrm{RED} \mid \mathcal{F}_{i-1})$ into $\mathbb{P}(A_i \in \mathrm{RED} \mid \{A_i \in \mathrm{RED}\} \cup \{A_i \subset V_{\mathrm{BLUE}}\}, \mathcal{F}_{i-1})$, and then further condition on a worst-case $\mathscr{H}_{A_i}^-$. The latter subsumes any information in $\mathbb{1}_{\{A_j \in \mathrm{RED}\}}$ for A_j disjoint from A_i; the former means that A_i is either the full intersection of a red cluster with $V \times \{t_\star\}$ or a collection of blue singletons, either way implying that any A_j intersecting A_i must satisfy $A_j \notin \mathrm{RED}$. Altogether, the events $\{A_j \in \mathrm{RED} : j < i\}$ are measurable under the conditioning on $\{A_i \in \mathrm{RED}\} \cup \{A_i \subset V_{\mathrm{BLUE}}\}, \mathscr{H}_{A_i}^-$, and we arrive at Ψ_{A_i} which immediately implies the stochastic domination. ∎

Proof of Corollary 2.4. By Lemma 2.3 we can stochastically dominate $\{A \in \mathrm{RED}\}$ and $\{A' \in \mathrm{RED}'\}$ by two independent sets of indicators $\{Y_A\}$ and $\{Y_A'\}$. Let $\{(A_l, A_l')\}_{l \geq 1}$ denote all pairs of intersecting subsets $(A, A' \subset V \setminus V_{\mathrm{GREEN}}$ with $A \cap A' \neq \emptyset)$ arbitrarily ordered, and associate each pair with a variable R_l initially set to 0. Process these in order: If we have not yet found $R_j = 1$ for some $j < l$ with either $A_j \cap A_l \neq \emptyset$ or $A_j' \cap A_l' \neq \emptyset$, then set $R_l = \mathbb{1}_{\{A_l \in \mathrm{RED}, A_l' \in \mathrm{RED}'\}}$ (otherwise skip this pair, keeping $R_l = 0$).

Noting $\mathbb{P}(R_l = 1 \mid \mathcal{F}_{l-1}) \leq \mathbb{P}(Y_{A_l} = 1, Y_{A_l'}' = 1) = \Psi_{A_l}\Psi_{A_l'}$ (as testing $R_l = 1$ means we received only negative information on $\{Y_{A_l} = 1\}$ and $\{Y_{A_l'}' = 1\}$) gives the coupling $\{l : R_l = 1\} \subset \{l : Y_{A_l, A_l'} = 1\}$.

Completing the proof is the fact that if $v \in V_{\mathrm{RED}} \cap V_{\mathrm{RED}'}$ then there are subsets A_l, A_l' containing it with $Y_{A_l} = Y_{A_l'}' = 1$, in which case either $R_l = 1$ or we will have an earlier $R_j = 1$ for a pair involving $A_j = A_l$ or $A_j' = A_l'$ (nontrivial intersections with A_l or A_l' will not be red), whence $v \in A_j \cup A_j'$. ∎

2.4. **Lower Bound.** Recalling the choice of $t_\mathfrak{m}$ as the point such that $\mathfrak{m}_{t_\mathfrak{m}} = 1/\sqrt{n}$, we let the sum of spins $f(\sigma) = \sum_v \sigma(v)$ be our distinguishing statistic at time $t_\star^- = t_\mathfrak{m} - s_\star$. Putting $Y = f(X_{t_\star^-}^+)$ for the dynamics (X_t^+) started from all-plus, by (2.10) we have

$$\mathbb{E}Y = n\mathfrak{m}_{t_\star^-} \geq e^{(1-\beta d)s_\star}n\mathfrak{m}_{t_\mathfrak{m}} \geq e^{s_\star/2}\sqrt{n} \tag{2.11}$$

(the last inequality using $\beta d \leq \frac{1}{2}$). For the variance estimate we use the fact (see [11, Claim 3.4]) that for some constants $\beta_0 = \beta_0(d) > 0$ and $\gamma = \gamma(d) > 0$, if $\beta < \beta_0$ then

$$\sum_u \mathrm{Cov}(X_t(u), X_t(v)) \leq \gamma \quad \text{for any } X_0, t > 0 \text{ and } v \in V\,. \tag{2.12}$$

From this inequality it follows that $\mathrm{Var}(Y) \leq \gamma n$, and in light of (2.11), $\mathbb{P}(Y \geq \mathbb{E}Y/2) > 1 - \varepsilon/2$ by Chebyshev's inequality provided $s_\star = s_\star(\varepsilon)$ is chosen large enough.

On the other hand, if $X' \sim \pi$ then $\mathbb{E}[f(X')] = 0$ (as $\mathbb{E}[X'(v)] = 0$ for any v), while $\mathrm{Var}(f(X')) \leq \gamma n$ by the decay of correlation of the Ising measure. By Chebyshev's inequality, for any large enough $s_\star = s_\star(\varepsilon)$ we have $\mathbb{P}(f(X') \geq \mathbb{E}Y/2) < \varepsilon/2$. Altogether, $d_{\mathrm{TV}}(t_\star^-) \geq 1 - \varepsilon$, as required. ∎

3. ANALYSIS OF THE INFORMATION PERCOLATION CLUSTERS

The delicate part of the bounding Ψ_A is of course the conditioning: since A is either the interface of a complete red cluster, or a collection of blue singletons, its combined histories must avoid $\mathcal{H}_A^-$. This immediately implies, for instance, that if any branches of $\mathcal{H}_A^-$ should extend to some point $(v, t) \in A \times [0, t_\star)$ in the space-time slab, necessarily the branch of v must receive an update along the interval $(t, t_\star]$ to facilitate avoiding that branch. Our concern will be such a scenario with $t \in (t_\star - 1, t_\star]$, since for values of t extremely close to $t_\star$ this event might be extremely unlikely (potentially having probability smaller than $\exp[-O(|A|)]$, which would we not be able to afford). For a subset $A' \subset A$ and a sequence of times $\{s_u\}_{u \in A'}$ with $s_u \in (t_\star - 1, t_\star]$, define

$$\mathcal{U} = \mathcal{U}(A', \{s_u\}_{u \in A'}) = \bigcap_{u \in A'} \{u \text{ receives an update along } (s_u, t_\star]\}. \tag{3.1}$$

We will reduce the conditioning on $\mathcal{H}_A^-$ to an appropriate event $\mathcal{U}$, and thereafter we would want to control $\mathcal{H}_A$, the collection of all histories from A, both spatially, measured by its branching edges

$$\chi(\mathcal{H}_A) = \# \left\{ \big((u, t), (v, t)\big) \in \mathcal{H}_A \ : \ uv \in E(G) \,, \ t \in (0, t_\star] \right\}$$

(i.e., those edges that correspond to a site branching out to its neighbors via a non-oblivious update), and temporally, as measured by the following length quantity:

$$\mathfrak{L}(\mathcal{H}_A) = \sum_{u \in V} \int_0^{t_\star} \mathbb{1}_{\{(u, t) \in \mathcal{H}_A\}} dt \,. \tag{3.2}$$

The following lemma bounds an exponential moment of $\chi(\mathcal{H}_A)$ and $\mathfrak{L}(\mathcal{H}_A)$ under any conditioning on an event $\mathcal{U}$ as above, and will be central to the proof of Lemma 2.1.

Lemma 3.1. *For any $d \geq 2$, $0 < \eta < 1$ and $\lambda > 0$ there are $\beta_0(d, \eta, \lambda) > 0$ and $\alpha(d, \eta) > 0$ such that the following holds. If $\beta < \beta_0$ then for any subset A,*

$$\sup_{\mathcal{U}} \mathbb{E}\left[\exp\left(\eta \mathfrak{L}(\mathcal{H}_A) + \lambda \chi(\mathcal{H}_A) \right) \big| \mathcal{U} \right] < \exp\left(\alpha |A| \right) ,$$

where the event $\mathcal{U}$ is as defined in (3.1).

3.1. **Proof of Lemma 2.1.** For a given subset $S \subset V$, let RED_S^* denote the red clusters that arise when exposing the joint histories of $\mathcal{H}_S$ (as opposed to all the histories $\mathcal{H}_V$), noting the events $\{A \in \text{RED}\}$ and $\{A \in \text{RED}_A^*\} \cap \{\mathcal{H}_A \cap \mathcal{H}_A^- = \emptyset\}$ are identical (so that A would be the interface of a full red cluster). Similarly define BLUE_S^*, and by the same reasoning $\{A \subset V_{\text{BLUE}}\} = \{A \subset V_{\text{BLUE}_A^*}\} \cap \{\mathcal{H}_A \cap \mathcal{H}_A^- = \emptyset\}$.

Next, given $\mathcal{H}_A^- = \mathcal{X}$, let $s_u = s_u(\mathcal{X}) = \max\{s \ : \ (u, s) \in \mathcal{X}\}$ be the latest most time at which $\mathcal{X}$ contains $u \in A$, and recall from the discussion above that any u with $s_u \leq t_\star$ must receive an update along $(s_u, t_\star]$ in order to avoid $\mathcal{X}$. Thus, writing $A' = \{u \in A \ : \ s_u > t_\star - 1\}$ and defining $\mathcal{U}(A', \{s_u\}_{u \in A'})$ as in (3.1), we find $\mathbb{P}\left(A \in \text{RED} \mid \mathcal{H}_A^- = \mathcal{X}, \{A \in \text{RED}\} \cup \{A \subset V_{\text{BLUE}}\} \right)$ to be equal to

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

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

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

The numerator is at most $\mathbb{P}(A \in \mathrm{RED}_A^* \mid \mathcal{U})$. As for the denominator, it is at least the probability that, in the space conditioned on $\mathcal{U}$, every $u \in A$ gets updated in the interval $(s_u \vee t_\star - 1, t_\star]$ and the last such update (i.e., the first we expose when revealing $\mathscr{H}_u$) is oblivious (implying $A \subset V_{\mathrm{BLUE}_A^*}$)—which is $\theta^{|A|}(1 - 1/e)^{|A \setminus A'|}$. As this is at least $e^{-|A|}$ for small enough β (recall the definition of θ in (2.2)),

$$\Psi_A \le e^{|A|}\mathbb{P}(A \in \mathrm{RED}_A^* \mid \mathcal{U}). \tag{3.3}$$

To estimate $\mathbb{P}(A \in \mathrm{RED}_A^* \mid \mathcal{U})$, let us expose the history of A backwards from $t_\star$ along the first unit interval $(t_\star - 1, t_\star]$ (moving us past the information embedded in $\mathcal{U}$), then further on until reaching time T where $\mathscr{H}_A(T)$ coalesces to a single point or we reach time $T = 0$. For $A \in \mathrm{RED}_A^*$ to occur, in either situation the projections of X_T^+ and X_T^- on the subset $\mathscr{H}_A(T)$ must differ (otherwise $X_{t_\star}(A)$ will not depend nontrivially on X_0). If $T = 0$ this trivially holds, and if $T > 0$ then, as $\mathscr{H}_A(T) = \{(w, T)\}$ for some vertex w and we did not expose any information on the space-time slab $V \times [0, T]$, the probability of this event is exactly $\mathfrak{m}_T$. Furthermore, using (2.10) we have $\mathfrak{m}_T \le e^{t_\star - T}\mathfrak{m}_{t_\star}$, and by definition, along the interval $[T, t_\star - 1]$ there are at least two branches in $\mathscr{H}_A$, so $\mathfrak{L}(\mathscr{H}_A) \ge 2(t_\star - 1 - T)$; thus, $\mathfrak{m}_T \le e^{\frac{1}{2}\mathfrak{L}(\mathscr{H}_A)+1}\mathfrak{m}_{t_\star}$. Also, on the event $A \in \mathrm{RED}_A^*$ the histories $\mathscr{H}_A$ must all join by time T, and thus $\chi(\mathscr{H}_A) \ge \mathfrak{W}(A) - 1$ since $\mathscr{H}_A$ must observe at least that many branching edges for connectivity. Thus,

$$\mathbb{P}(A \in \mathrm{RED}_A^* \mid \mathcal{U}) \le e\,\mathfrak{m}_{t_\star}\mathbb{E}\left[\mathbb{1}_{\{\chi(\mathscr{H}_A) \ge \mathfrak{W}(A)-1\}}e^{\frac{1}{2}\mathfrak{L}(\mathscr{H}_A)} \mid \mathcal{U}\right],$$

which for any $\lambda > 0$ and $\alpha > 0$ is at most

$$e^{1-(\lambda+\alpha+1)(\mathfrak{W}(A)-1)}\mathfrak{m}_{t_\star}\mathbb{E}\left[e^{(\lambda+\alpha+1)\chi(\mathscr{H}_A)+\frac{1}{2}\mathfrak{L}(\mathscr{H}_A)} \mid \mathcal{U}\right].$$

Plugging in α as given by Lemma 3.1 (which we recall does not depend on the pre-factor of $\chi(\mathscr{H}_A)$ in that lemma), the exponential moment above is at most $e^{\alpha|A|}$, and revisiting (3.3) we conclude that

$$\Psi_A \le \mathfrak{m}_{t_\star}e^{1+(1+\alpha)|A|-(\lambda+\alpha+1)(\mathfrak{W}(A)-1)} \le e^{\lambda+\alpha+2}\,\mathfrak{m}_{t_\star}e^{-\lambda\mathfrak{W}(A)}. \qquad \blacksquare$$

3.2. **Exponential decay of cluster sizes: Proof of Lemma 3.1.** We develop the history of a set A backward in time from $t_\star$ by exposing the space-time slab. Let $W_s = |\mathscr{H}_A(t_\star - s)|$ count the number of vertices in the history of A at time $t_\star - s$, let $Y_s = \chi(\mathscr{H}_A \cap V \times [t_\star - s, t_\star])$ be the total number of vertices observed by the history by time $t_\star - s$ and let $Z_s = \sum_{u \in V} \int_{t_\star - s}^{t_\star} \mathbb{1}_{\{(u,t) \in \mathscr{H}_A\}}dt$ be the total length of the history in the time interval $[t_\star - s, t_\star]$ of the space-time slab. Initially we have $(W_0, Y_0, Z_0) = (|A|, 0, 0)$.

Recall that the probability that an update of a vertex v will branch out to its d neighbors is $1 - \theta$ and that with probability θ it is oblivious which observes no neighbours. Thus we can stochastically dominate (W_s, Y_s, Z_s) by a process $(\bar{W}_s, \bar{Y}_s, \bar{Z}_s)$ defined as follows. Initially, $(\bar{W}_0, \bar{Y}_0, \bar{Z}_0) = (|A|, 0, 0)$

and at rate $\theta \bar{W}_s$ we decrease $\bar{W}_s$ by 1 and at rate $(1-\theta)\bar{W}_s$ both $\bar{W}_s$ and $\bar{Y}_s$ increase by d. The length grows as $d\bar{Z}_s = \bar{W}_s ds$. Now consider the process,

$$Q_s = \exp\left(\eta \bar{Z}_s + \lambda \bar{Y}_s + \alpha \bar{W}_s\right).$$

for some $\alpha > -\log(1-\eta)$ which does not depend on λ. We have that

$$\frac{d}{ds}\mathbb{E}[Q_s \mid Q_{s_0}]\Big|_{s=s_0} = \left(\eta + \theta(e^{-\alpha}-1) + (1-\theta)(e^{(\lambda+\alpha)d}-1)\right)\bar{W}_{s_0}Q_{s_0}$$

which is negative provided θ is sufficiently close to 1 (guaranteed by taking β_0 sufficiently small). Then, letting τ be the first time that $\bar{W}_\tau = 0$, optional stopping for the supermartingale Q_s yields

$$\mathbb{E}\exp\left(\eta \bar{Z}_S + \lambda \bar{Y}_S\right) \leq \mathbb{E}Q_0 = \exp(\alpha|A|).$$

By the stochastic domination we have that $\mathbb{E}\left[\exp\left(\eta \mathfrak{L}(\mathscr{H}_A) + \lambda \chi(\mathscr{H}_A)\right)\right] \leq \exp(\alpha|A|)$. Under this coupling the effect of conditioning on $\mathcal{U}$ is simply to expedite updates and hence reduce the length of the process, thus the conditioning can only reduce the expectation. ∎

References

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

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

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

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

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

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

[7] E. Lubetzky and A. Sly, *Explicit expanders with cutoff phenomena*, Electron. J. Probab. **16** (2011), no. 15, 419–435.

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

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

[10] E. Lubetzky and A. Sly, *Information percolation for the Ising model: cutoff in three dimensions up to criticality*, preprint. Available at `arXiv:1401.6065` (2014).

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

[12] 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.

[13] 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.

[14] 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.

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