INFORMATION PERCOLATION AND CUTOFF FOR THE RANDOM-CLUSTER MODEL

SHIRSHENDU GANGULY AND INSUK SEO

ABSTRACT. We consider the Random-Cluster model on \((\mathbb{Z}/n\mathbb{Z})^d\) with parameters \(p \in (0, 1)\) and \(q \geq 1\). This is a generalization of the standard bond percolation (with open probability \(p\)) which is biased by a factor \(q\) raised to the number of connected components. We study the well known FK-dynamics on this model where the update at an edge depends on the global geometry of the system unlike the Glauber Heat Bath dynamics for spin systems, and prove that for all small enough \(p\) (depending on the dimension) and any \(q > 1\), the FK-dynamics exhibits the cutoff phenomenon at \(\lambda_\infty^{-1} \log n\) with a window size \(O(\log \log n)\), where \(\lambda_\infty\) is the large \(n\) limit of the spectral gap of the process. Our proof extends the Information Percolation framework of Lubetzky and Sly [21] to the Random-Cluster model and also relies on the arguments of Blanca and Sinclair [4] who proved a sharp \(O(\log n)\) mixing time bound for the planar version. A key aspect of our proof is the analysis of the effect of a sequence of dependent (across time) Bernoulli percolations extracted from the graphical construction of the dynamics, on how information propagates.

CONTENTS

1. Introduction and main result 1
2. Idea of the proof and organization of the article 4
3. A priori bounds on mixing time and spectral gap 7
4. Auxiliary percolation models, and disagreement propagation bounds 9
5. Information percolation clusters and time dependent Bernoulli percolations 14
6. Reduction to a product chain 32
7. Proof of main result 38
8. Appendix 41
References 42

1. INTRODUCTION AND MAIN RESULT

The random-cluster (Fortuin-Kasteleyn/FK) model is an extensively studied model in statistical physics, generalizing electrical networks, percolation, and spin systems like the Ising and Potts models, under a single framework. In this work, we study the so called heat-bath Glauber dynamics or FK-dynamics for the model on the \(d\)-dimensional torus. The main result of this paper establishes a sharp convergence to equilibrium for this Markov chain also known as the cutoff phenomenon.

1.1. Random-cluster model (RCM). For \(d \geq 2\), denote by \(\Lambda_n = \mathbb{Z}_n^d\), the \(d\)-dimensional discrete torus and by \(E_n = E(\Lambda_n)\), the set of edges in \(\Lambda_n\).

We will fix the dimension to be \(d\) throughout the entire paper. The random-cluster measure \(\mu^n_{p,q}\) on \((\Lambda_n, E_n)\) with parameters \(p \in (0, 1)\) and \(q > 0\) is a probability measure on the space of subsets of \(E_n\) defined by

\[
\mu^n_{p,q}(S) = \frac{1}{Z^n_{p,q}} p^{|S|} (1 - p)^{|E_n \setminus S|} q^{|c(S)|}; \ S \subset E_n,
\]
where $Z^n_{p,q}$ is the partition function turning $\mu^n_{p,q}$ into a probability measure, and $c(S)$ is the number of connected components of the graph $(\Lambda_n, S)$. Clearly the measure $\mu^n_{p,q}$ can be regarded as a probability measure on $\Omega_n = \{0, 1\}^E_n$, i.e., we will identify $X = (X(e))_{e \in E_n} \in \Omega_n$ with a subset $A$ of $E_n$ where $e \in A$ if and only if $X(e) = 1$. Hence, by slight abusing of notation, we can always regard $X \in \Omega_n$ as a subset of $E_n$. The random-cluster model was introduced by Fortuin and Kasteleyn (see [11, 12]) and unifies the study of various objects in statistical mechanics such as random graphs, spin systems and electrical networks (see [16]). When $q = 1$ this model corresponds to the standard bond percolation but when $q > 1$ (resp., $q < 1$) the probability measure biases subgraphs with more (resp., fewer) connected components. For the special case of integer $q \geq 2$ the random-cluster model is a dual to the classical ferromagnetic $q$-state Potts model, via the so called Edward-Sokal coupling of the models (see, e.g., [10]). However, note that unlike spin systems, the probability that an edge $e$ belongs to $A$ does not depend only on the dispositions of its neighboring edges but on the entire configuration $A$, since connectivity is a global property (see Figure 1.1 for an illustration).

1.2. FK-dynamics (Glauber/Heat-bath dynamics). The FK-dynamics is a reversible Markov process $X_t = \{X_t(e)\}_{e \in E_n}$ on $\Omega_n$ whose invariant measure is given by $\mu^n_{p,q}$. Informally, at rate one, the state of every edge $X(e)$ is resampled conditionally on the state of the remaining edges i.e.,

$$X(e) = \begin{cases} 
1 & \text{w.p. } p \text{ if } e \text{ is not a cut-edge,} \\
1 & \text{w.p. } \frac{p}{p+(1-p)q} \text{ if } e \text{ is a cut-edge,} \\
0 & \text{otherwise.}
\end{cases}$$

where we use the standard terminology cut-edge to denote an edge whose removal increases the number of connected components by one. A more formal treatment appears in Definition 2.1.

Note that unlike Glauber dynamics on spin systems like Ising or Potts models, the FK-dynamics has long range dependencies (see Figure 1.1). The key statistic we will consider is the time taken by the above dynamics to converge to equilibrium.

![Figure 1.1](image_url)

**Figure 1.1.** Illustrating long range dependencies in FK-dynamics. Consider a configuration where the red edges are open while everything else is closed. The probability of the edge $AB$ to be open then depends on whether the edge $CD$ is open or not.
1.3. Mixing of Markov chains and cut-off phenomenon. We review in brief the set up of interest for us from the theory of reversible Markov chains with finite state spaces. For an extensive account of all the details and recent progress in various directions see [17]. For two probability measures $\mu_1$ and $\mu_2$ on $S$ we will be interested in the $L^1$-distance or the so called total variation distance between them to be denoted by $\|\mu_1 - \mu_2\|_{TV}^1$:

$$\sup_{A \subseteq S} (\mu_1(A) - \mu_2(A)) = \frac{1}{2} \sum_{x \in S} |\mu_1(x) - \mu_2(x)| = \frac{1}{2} \sum_{x \in S} |\mu_1(x) - \mu_2(x)| 1 \mu_2(x).$$ (1.1)

For concreteness consider a continuous time reversible Markov chain $Y_t$ with a finite state space $S$ and equilibrium measure $\pi$. We will be primarily interested in the total variation mixing time defined by

$$t_{\text{mix}}(\varepsilon) = \inf \{ t : \sup_{y \in S} \|P_y[Y_t \in ] - \pi\|_{TV} \leq \varepsilon \} ; \varepsilon \in (0, 1).$$

For notational brevity, we will denote by $d(t)$, the worst case total variation distance to stationarity for the FK-dynamics, i.e.,

$$d(t) = d_n(t) := \sup_{x \in \Omega_n} \|P_x[X_t \in ] - \mu_n^n\|_{TV}$$ (1.2)

from now on. Many naturally occurring Markov chains are expected to exhibit a sharp transition in convergence, in the sense that the total variation distance to equilibrium drops from one to zero in a rather short time window. This is formalized by the notion of cutoff formulated by Aldous and Diaconis [1] (see also [5]). Formally a sequence of Markov chain $Y_t^{(1)}, Y_t^{(2)}, \ldots$ with mixing times given by $t_{\text{mix}}^{(1)}(\varepsilon), t_{\text{mix}}^{(2)}(\varepsilon), \ldots$ is said to exhibit the **Cutoff Phenomenon** if for any $\varepsilon \leq 1/2$,

$$\lim_{i \to \infty} \frac{t_{\text{mix}}^{(i)}(\varepsilon)}{t_{\text{mix}}^{(i)}(1 - \varepsilon)} = 1.$$

Moreover cutoff is said to occur with window size $w_i$ if for any $\varepsilon$ one has

$$t_{\text{mix}}^{(i)}(\varepsilon) - t_{\text{mix}}^{(i)}(1 - \varepsilon) = O(\varepsilon w_i),$$

where $w_i = o(t_{\text{mix}}^{(i)}(1/4))$.

1.4. Main result. Given the above definitions, our main result establishes cutoff for the FK-dynamics for a range of sub-critical values of the parameters $p, q$.

**Theorem 1.1.** For any $d \geq 2$, there exists $p_0 = p_0(d) > 0$ such that, for all $p \in (0, p_0)$ and $q > 1$, there exists a constant $\lambda_{\infty} = \lambda_{\infty}(p, q)$ such that the FK-dynamics on $\Omega_n$ exhibits cutoff at $\frac{d}{2\lambda_{\infty}} \log n$ with order $O(\log \log n)$ window size.

Some remarks are in order. Note that the case $q = 1$ is the well known example of random walk on a hypercube where cutoff occurs for all values of $p$. Similarly in the case $d = 1$, one notices that each edge is a cut edge unless the configuration is completely full. Thus the process in this case can also be coupled with a random walk on a hypercube, implying cutoff for all values of $p$ and $q$.

The value of the threshold $p_0$ in the statement above only depends on the dimension through the value of the critical bond percolation probability and does not depend on $q$. We shall assume that $q > 1$ is fixed from now on. Notice that by a duality argument as in [4, Section 7], in the planar case (i.e., $d = 2$) it follows that Theorem 1.1 holds also when $p$ is close enough to 1. We will also elaborate on a description of $\lambda_{\infty}$ in terms of the spectral gap of the Glauber dynamics for the infinite volume RCM in Section 7.2.

---

$^1\|\mu_1 - \mu_2\|_2$ will be used to denote the $L^2$-distance where the 1-norm in (1.1) is replaced by the 2-norm.
1.5. **Background and related work.** There has been much activity over the past two decades in analyzing Glauber dynamics for spin systems in both statistical physics and computer science leading to deep connections between the mixing time and the phase structure of the physical model. In contrast, the Glauber dynamics for the RCM remains less understood. The main reason for this is that connectivity is a global property. Ullrich in a series of important papers [30, 29, 31] established comparison estimates between the FK-dynamics and the well known non-local Swendsen-Wang (SW) dynamics ([28]) using functional analytic arguments. Until recently, all existing bounds on the FK-dynamics were via transferring results for the SW or related dynamics [28] using comparison estimates as above. However these methods typically yield highly sub-optimal bounds and does not provide any insight into the behavior of RCM and furthermore, since it relies on comparison with the Ising/Potts models, this analysis applies only for integer values of $q$.

Recently the authors of [4] established a fast mixing time of order $O(n^2 \log n)$ bound for the discrete time FK-dynamics on RCM in a box of size $n$ in $\mathbb{Z}^2$ with a special class of boundary conditions. The proof works for all $q \geq 1$ and $p \neq p_c(q)$. Furthermore, although not explicitly mentioned, the arguments extend to periodic boundary conditions as well. The key ingredients used were planar duality, tools developed for mixing of spin systems in [24] and most importantly the exponential decay of connectivity below $p_c(q)$ established in the breakthrough work [2]. More recently [3] extends the results to a more general class of boundary conditions with weaker bounds. Among various things, the latter work in particular also shows that boundary conditions can have a drastic effect on the mixing time.

A general conjecture of Peres [26] indicates that one should expect cutoff to occur in the regime of fast mixing for many natural chains as above. In the breakthrough papers, [18, 19], Lubetzky and Sly verified the above conjecture for Glauber dynamics for Ising and Potts models, putting forward a host of new methods using ideas similar to the Propp-Wilson coupling from the past [27] as well as relating $L^1$-mixing to $L^2$-mixing using powerful log-Sobolev inequalities [6]. Subsequently in [21, 22], the results of the above papers were refined by inventing the general Information percolation machinery. Furthermore in very recent work, [25] extended the above framework to prove cutoff results for the non-local SW dynamics for Potts models on the torus in any dimension for suitably high temperatures.

However as indicated above, the FK-dynamics has significant differences with the above described spin models and whether cutoff occurs in the fast mixing regime in this case was left open. The main theorem of this paper answers this question in the affirmative as long as $p$ is small enough and $q > 1$. In the process, we extend the Information Percolation framework to the RCM setting as well. An elaborate description of the various geometric difficulties and how to encounter them is presented in the next section. We end this section by also mentioning the recent work of Lubetzky and Gheissari on proving quasi-polynomial bounds for the mixing time at criticality for FK-dynamics in two dimensions and related bounds for critical spin systems in [13, 14, 15] based on recent breakthroughs in [8, 9].

## 2. Idea of the proof and organization of the article

We first develop a graphical construction (grand coupling of FK-dynamics) which will be quite useful in constructing coupling arguments. We then discuss the key issues that one faces towards proving the main result and what new ideas one needs beyond the existing literature to address them.

### 2.1. Graphical construction/ Monotone coupling

We will define the FK-dynamics formally through the following graphical construction by creating what is now popularly called in the literature as the **Update sequence** (see [18, 25]). For $e \in E_n$, define the sequence of updates as

$$\text{Upd}(e) = \{(t_1, U_1), (t_2, U_2), \ldots\} ,$$ (2.1)
where \( t_1 < t_2 < \cdots \) is a sequence of update times obtained from an independent Poisson process with rate 1 attached at \( e \), and for each \( i, U_i \) is a uniform random variable in \([0, 1]\) independent of all other randomness. The sequence \( \text{Upd}(e) \) is the update sequence corresponding to \( e \). Then, we define the full update sequence as

\[
\text{Upd} = \bigcup_{e \in E_n} \text{Upd}(e) .
\]

Note that \( t_i(e) \neq t_j(e') \) for all \( i, j \in \mathbb{N} \) and \( e, e' \in E_n \) almost surely. It would also be useful to define for \( 0 < t_1 < t_2 \), the update sequence of \( e \) in the time interval \( (t_1, t_2] \) as

\[
\text{Upd}[t_1, t_2](e) = \{ s : (s, U) \in \text{Upd}(e), s \in (t_1, t_2] \} ,
\]

and

\[
\text{Upd}[t_1, t_2] = \bigcup_{e \in E_n} \text{Upd}[t_1, t_2](e) .
\]

For \( X \in \Omega_n \), we say that \( e \in E_n \) is a cut-edge if \( c(X \setminus \{e\}) \neq c(X \cup \{e\}) \). Furthermore, from now on, we shall assume \( q > 1 \) and write

\[
p^* = \frac{p}{q(1-p) + p} < p
\]

for convenience.

We now define a construction of the FK-dynamics suitable for our purposes.

**Definition 2.1** (FK-dynamics). For each \((t, U) \in \text{Upd}(e)\) for some \( e \in E_n \),

1. (a) If \( U < 1 - p + p^* \), we let

\[
X_t(e) = \begin{cases} 
0 & \text{if } U \in [0, 1-p) \\
1 & \text{if } U \in [1-p, 1-p + p^*)
\end{cases}
\]

(b) If \( U \geq 1 - p + p^* \), we let \( X_t(e) = 0 \) if \( e \) is a cut-edge in \((\Lambda_n, X_{t-})\), and \( X_t(e) = 1 \) if \( e \) is not a cut-edge in \((\Lambda_n, X_{t-})\).

2. We set \( X_t(e') = X_{t-}(e') \) for all \( e' \neq e \).

We will denote by \( \mathbb{P}^n_{x_0} = \mathbb{P}^q_{x_0,n} \) the law of the FK-dynamics starting from \( x_0 \in \Omega_n \). Note that the latter is reversible with respect to its invariant measure \( \mu_{p,q}^n \). Naturally the update sequence allows a grand coupling of \((X_t)\) started from all possible configurations \( x_0 \). A well known fact is the monotonicity of FK-dynamics i.e., if \((X_t)\) and \((Y_t)\) are two copies of the Markov chain started from \( x_0 \) and \( y_0 \) with \( x_0 \leq y_0 \) in the usual partial order on \( \Omega_n \) then under the grand coupling for all later times \( t \) one has \( X_t \leq Y_t \). Thus often this coupling is called the monotone coupling and the corresponding law is denoted by \( \mathbb{P}_{x_0,y_0} \). Note that another perhaps more canonical way to define the dynamics would be to first check if \( e \) is a cut-edge (resp. not) and then accordingly set it to 0 or 1 depending on whether \( U < 1 - p^* \) or not (resp. \( U < 1 - p \)). However the above alternative formulation has the nice property that if \( U < 1 - p + p^* \), we do not need to check whether \( e \) is a cut-edge or not, and the randomness at \( e \) only depends on \( U \), not the entire configuration of \( X_t \). This will be used throughout the paper in various coupling arguments.

**2.2. The key ideas of the proof.** In the work of Lubetzky and Sly [18] on the Ising model, the key idea was to break the dependencies in the Markov chain to reduce the analysis to the study of a product chain of Glauber dynamics on small boxes. The proof then relied on the relation between the \( L^1 \)-mixing time of the product chain to \( L^2 \)-mixing time of the individual coordinates and sharp estimates on the latter obtained via log-Sobolev inequalities. Unfortunately such functional analytic tools are not available for the RCM.

Furthermore, to improve the size of the cutoff window to \( O(1) \), in [22, 21], the powerful machinery of information percolation was invented to bypass the use of log-Sobolev inequalities to estimate the \( L^2 \)-mixing time. The proof however still relied heavily on the local nature of Glauber dynamics.
for spin systems. Very recently in [25] the strategy was extended to the non-local Swendsen-Wang (SW) dynamics for Potts model. The latter work is based on the observation that while in Glauber dynamics, in one step the spin at a vertex can only depend on its immediate neighbors, the state of a vertex in SW by definition depends on all the vertices inside an independent percolation cluster sampled at each time step. Thus in the subcritical regime, since the cluster diameters have exponential tails, one can expect the same approach to go through and indeed this is what is made rigorous in [25].

As indicated before, at a very high level, one of the main contributions of our approach is extending the Information Percolation framework to the setting of FK-dynamics. In SW dynamics for the Potts model, one proceeds by sampling an independent bond percolation on each of the mono-chromatic components (connected component of vertices with the same spins) and then for each connected component of the percolation sampled, a uniformly random spin is assigned. This is done at every time step independently of the past and hence the interaction of the spin at every vertex at every time step in only limited to spins within its percolation cluster.

On the other hand in RCM, in one step the update of an edge can depend on the status of an arbitrarily far located edge (see Figure 1.1). To bypass this, we first run the process for an $O(1)$ burn-in time which allows the process to be dominated by a subcritical Bernoulli percolation.

At this point we try to analyze the information percolation clusters. Very informally (see Section 5 for precise definitions) this approach involves keeping track of the interactions between various edges as they are updated, backwards in time. For e.g.: if an edge $e$ is updated using an element $(t, U) \in \text{Upd}(e)$ one of two things could happen (recall Definition 2.1):

- $U < 1-p+p^*$, in which case the updated value of the edge is a Bernoulli variable independent of the state of the system. In this case we call the edge to become Oblivious.
- However if $U > 1-p+p^*$ one needs to check whether $e$ is a cut-edge or not and in the process interacts (shares information) with several edges.

Formally one considers a space time slab (see Figure 5.1) and evolves backward in time by branching out to all possible edges an update shares information with, or gets killed in case of an oblivious update. The key usefulness of this approach as exploited in [18, 19, 22, 21, 25] is that branching out to all possible edges an update shares information with, or gets killed in case of an oblivious update. The key usefulness of this approach as exploited in [18, 19, 22, 21, 25] is that branching out to all possible edges an update shares information with, or gets killed in case of an oblivious update.
which a priori depends on the entire time interval \([0, \tau_i]\). However this is the point at which we use the smallness of \(p\) crucially, which creates an environment which is subcritical and hence the connected component can be bounded by the connected component of \(\Xi_i \cup \Xi_{i-1}\), i.e., instead of the entire interval \([0, \tau_{i+1}]\) we can get by just using the information on \([\tau_{i-1}, \tau_{i+1}]\).

Given the above, the situation is similar to the definition of the SW dynamics considered in [25], except that the percolation sampled at every discrete time step is now 1-dependent across time. This creates the need for a refined and delicate analysis of the information percolation clusters to yield \(L^2\)-mixing bounds. This is stated as Theorem 5.1 and Proposition 5.3. The proof of the latter is the core of this work. The above approach adopted in the paper of extracting dependent percolation models that can be analyzed could be of independent interest and useful in other general contexts in bounding how passage of information occurs in such dynamical settings.

Assuming these results, the arguments used to show cutoff is quite similar to the ones already appearing in [25] based on the methods in [18]. An additional ingredient needed to prove Theorem 5.1 from Proposition 5.3 is that the spectral gap of the FK-dynamics is positive uniformly in the system size. This is in fact a consequence of the a priori mixing time bounds obtained in [4]. In SW the lower bound on the spectral gap follows by path coupling by establishing a one step contraction which unfortunately is absent in our setting. The constant \(\lambda_\infty\) appearing in Theorem 1.1 is the limit of the spectral gaps of the dynamics on finite boxes as in the case of the spin models. Furthermore at the end of the article we include a sketch of the proof of the fact that the limiting constant \(\lambda_\infty\) in Theorem 1.1 is in fact the spectral gap of the infinite volume FK-dynamics although we do not pursue proving it rigorously in this paper (see Section 7.2).

Finally, we mention that for the Ising model, [21] exploited monotonicity of the system, to prove an \(O(1)\) bound on the cutoff window without resorting to the methods of [18]. Such sharp bounds are missing in [25] which deals with the general Potts model. However the RCM is monotone and whether this can be used to prove a similar improvement of Theorem 1.1 is not pursued in this paper and is left for further research.

2.3. Organization of the article. We prove and collect results about a priori bounds on the mixing time and the spectral gap in Section 3 to be used throughout the rest of the article. As mentioned above we need to define several auxiliary percolation models based on the update sequence. This is done in Section 4. Section 5 is the core of this work and the main contribution in this paper which bounds the \(L^2\)-mixing time by defining suitable information percolation clusters. This section is rather long and has several new constructions and delicate geometric arguments. However assuming the main result of this section, the proof of Theorem 1.1 is quite similar to the arguments appearing in [18, 21, 25]. The reader not familiar with the latter papers can choose to first assume the results of Section 5 to see how they are used in the subsequent sections to then come back to the proofs of Section 5.

The proof of the main result Theorem 1.1 spans Section 6 where certain modifications of arguments of [18] and Section 7 where the final proof appears. The outstanding proofs of some of the stated claims are collected in the Appendix (Section 8).

Acknowledgements. The authors thank Antonio Blanca, Fabio Martinelli and Alistair Sinclair for several useful discussions. IS was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2018R1C1B6006896) and (No. 2017R1A5A1015626).

3. A priori bounds on mixing time and spectral gap

We start by recalling the following standard result.

**Proposition 3.1.** Let \((Z_t)\) be a discrete time ergodic reversible Markov chain on a finite state space \(S\) with the equilibrium measure \(\pi\), let \(Q_z\) be the law of Markov chain \((Z_t)\) starting from \(z \in S\), and
let $\gamma$ be the spectral gap of the Markov chain $(Z_t)$. Then,

$$(1 - \gamma)^t \leq 2 \sup_{x \in S} \|Q_t(z_i) - \pi\|_{TV} \leq \frac{1}{\pi_{\text{min}}} (1 - \gamma)^t,$$

where $\pi_{\text{min}} = \min_{x \in S} \pi(x)$.

In [4], a discrete version of FK-dynamics is considered where at every discrete time step, an uniformly chosen edge is updated. Denote by $(\hat{X}_k)_{k \geq 0}$ the discrete FK-dynamics in $\Omega_n$, and by $\hat{p}_{x_0, y_0}$ the law of the monotone coupling (Definition 2.1) of two copies of discrete FK-dynamics $\hat{X}_k$ and $\hat{Y}_k$ starting from two initial conditions $x_0, y_0 \in \Omega_n$ respectively. Moreover, let $\hat{\lambda}(n) = \hat{\lambda}(n, p, q)$ denote the spectral gap of the above process. Furthermore let $\hat{t}_{\text{mix}} = \hat{t}_{\text{mix}}(1/4)$ and $\hat{d}(t)$ be the mixing time and the worst-case distance to stationarity respectively in the sense of (1.2) for the discrete time dynamics. Then, the following sharp mixing time results were either obtained or are consequences of the results in [4]. In the latter, only the two dimensional case was treated but one can easily verify that the arguments extend to general dimensions under exponential decay of connectivity and for our purposes we state the extensions without proof.

**Theorem 3.2.** For any dimension $d$, there exists $p_0 = p_0(d)$ such that for all $q \geq 1$ and $p < p_0$, there exists $C = C(p) > 0$ and $\lambda = \lambda(p) > 0$ such that:

1. For all $x_0, y_0 \in \Omega_n$, $k \leq o(n^{1/(d+2)})$ and $e \in E_n$, it holds that
   $$\hat{p}_{x_0, y_0} \left[ \hat{X}_{kn^d}(e) \neq \hat{Y}_{kn^d}(e) \right] \leq e^{-Ck}.$$

2. The mixing time $\hat{t}_{\text{mix}}$ of discrete process $\hat{X}_k$ is $\Theta(n^d \log n)$.

3. For all $n \in \mathbb{N}$, $\hat{\lambda}(n) \geq \lambda n^{-d}$.

**Remark 3.3.** Indeed, one can take $p_0$ to be the critical Bernoulli bond percolation probability on $\mathbb{Z}^d$. For $d = 2$, thanks to the complete knowledge about exponential decay of connectivity up to the critical point established in [2], the results of Theorem 3.2 were shown to hold for all subcritical $p$, for each $q \geq 1$ in [4].

**Proof.** (1) and (2) appear as [4, (13)], and [4, Theorem 6.1] respectively. Note that (1) proves the upper bound in (2) by taking $k = C \log n$.

The proof of the lower bound of mixing time appears in [4, Theorem 6.1]. Although (3) does not quite appear in [4] it is a consequence of (1). To see this, we will use the well known lower bound of total variation distance in terms of spectral gap recalled in Proposition 3.1 (see [17, Theorems 12.3 and 12.4]). Namely, using the above and union bounding over all elements in $E_n$, we get that $\hat{d}(kn^d)$, the worst-case total variation distance at time $kn^d$ is $e^{-\Omega(k) + d \log n}$, and hence

$$(1 - \hat{\lambda})^{kn^d} \leq e^{-\Omega(k) + d \log n}$$

for all $k \leq o(n^{1/(d+2)})$. Now taking logs we get $-kn^d \hat{\lambda} \leq -\Omega(k) + d \log n$, and therefore for some $C > 0$,

$$\frac{1}{n^d} \left( C - \frac{\log n}{k} \right) \leq \hat{\lambda}.$$

Thus by choosing a large enough $k = o(n^{1/(d+2)})$ the result follows. \hfill \square

However for our purposes, we will need a translation of the result for the continuous time setting. Denote by $\lambda(n) = \lambda(n, p, q)$ the spectral gap of the continuous time FK-dynamics defined in Definition 2.1.

**Corollary 3.4.** For any dimension $d$, there exists $p_0 = p_0(d)$ such that for all $q \geq 1$ and $p < p_0$, there exists $C = C(p) > 0$ and $\lambda = \lambda(p) > 0$ such that:
Lemma 4.2. coupling can couple all of them on the time window \( p \) respectively.

(1) For all \( x_0, y_0 \in \Omega_n \) and \( k \leq o(n^{1/(d+2)}) \), it holds that,
\[
P_{x_0,y_0}[X_t(e) \neq Y_t(e)] \leq e^{-Ct}.
\]

(2) The FK-dynamics in \( \Lambda \) has mixing time of order \( \Omega(\log n) \).

(3) For all \( n \in \mathbb{N} \), it holds that \( \lambda(n) \geq \lambda \).

Proof. All these results are immediate from Theorem 3.2 since the continuous dynamics is \( n^d \) times faster than the discrete counterpart. In particular, to show part (3), see [17, Lemma 20.5, Lemma 20.11]. □

4. Auxiliary percolation models, and disagreement propagation bounds

Given the randomness defined by the update sequence in (2.1), we will need to define several auxiliary percolation models extracted from the graphical construction, which though simple will be useful in various comparison arguments appearing throughout the paper. We will also state useful bounds on speed of propagation of disagreements. We start with the percolation models. Before providing precise definitions, for the reader’s benefit we give short descriptions what each of these models capture. Furthermore, for ease of reference throughout the article, all the definitions are collected in Table 1 at the end of this section.

(1) Standard Percolation dynamics (\( q = 1 \)): Random walk on the hypercube i.e., edges are randomly refreshed at rate one with a Bernoulli(\( p \)) variable independently. This will dominate the FK-dynamics in the regime of our interest.

(2) Update/Non-update percolation: An edge is open if it has not been updated at least once in a given interval of time.

(3) Enlarged percolation: An edge is said to be open if it was open at least once in the Standard Percolation dynamics in a given time interval.

4.1. Standard percolation dynamics (SPD). It will be useful to discretize time as we will see in later applications. Throughout the article we will fix \( \Delta := \Delta(p) = p^{-1/2} \), to be the basic unit of discretization and let \( \tau_i := i\Delta \). (The choice of \( \Delta \) is not special as long as it satisfies the properties discussed in this section.) Also let \( \mathbb{Z}_+ \) be the set of non-negative integers.

Definition 4.1 (SPD associated to the update sequence \( \text{Upd} \)). For each \( i \in \mathbb{Z}_+ \), we construct a SPD \((\mathcal{F}_t^i)_{t \geq \tau_i}\) in \( \Omega_n \) as follows:

(1) \( \mathcal{F}_0^i = E_n \).

(2) For each \( t > \tau_i \) and \( e \in E_n \),
(a) If \( \text{Upd}[\tau_i, t]^i(e) = \emptyset \), we let \( \mathcal{F}_t^i(e) = \mathcal{F}_{\tau_i}^i(e)(= 1) \).
(b) Otherwise, let \( (t^*, U^*) \) be the last update in \( \text{Upd}[\tau_i, t]^i(e) \).
 \begin{align*}
(i) \ & \text{We let } \mathcal{F}_t^i(e) = 1 \text{ if } U^* > 1 - p, \\
(ii) \ & \text{else let } \mathcal{F}_t^i(e) = 0 \text{ if } U^* \leq 1 - p.
\end{align*}

We define the dynamics \((\mathcal{E}_t^i)_{t \geq \tau_i}\) in an identical manner by replacing step (1) with \( \mathcal{E}_0^i = \emptyset \). In other words, \((\mathcal{F}_t^i)\) and \((\mathcal{E}_t^i)\) are the Glauber dynamics of the percolation measure with open probability \( p \) (random walk on the hypercube) in \( \Omega_n \) starting at \( t = \tau_i \) from the full and empty configurations, respectively.

Since \((\mathcal{F}_t^i)\) and \((\mathcal{E}_t^i)\), for \( i \in \mathbb{Z}_+ \), and the FK-dynamics \((X_t)\), share the same update sequence, we can couple all of them on the time window \( [\tau_i, \infty) \) in a natural manner calling this as the canonical coupling. We record some simple but useful lemmas below.

Lemma 4.2. Under the canonical coupling, for all \( i \in \mathbb{Z}_+ \), it holds that
\[
X_t \leq \mathcal{F}_t^i \text{ for all } t \geq \tau_i.
\]
Proof. Denote by $X^\text{full}_t$ the FK-dynamics on $\Omega_n$ with $X_0 = E_n$, the full configuration. Via the monotone coupling, we have $X_t \leq X^\text{full}_t$ for all $t \geq 0$. Now the inclusion $X^\text{full}_t \leq \mathcal{F}_t^i$ for all $t \geq 0$ comes directly from the definitions of FK-dynamics and percolation dynamics. Since we have $\mathcal{F}_t^0 \leq \mathcal{F}_t^i$ for all $t \geq \tau_i$ for all $i \in \mathbb{Z}^+$ under the canonical coupling, we are done. \hfill \Box

For $s \in [0, 1]$, denote by $\text{Perc}_n(s)$ the standard bond percolation on $E_n$ where an edge $e$ is open with probability $s$.

**Lemma 4.3.** For all $i \in \mathbb{Z}^+$ and $t \geq 0$, the law of $\mathcal{F}_{t+\tau_i}^i$ is given by $\text{Perc}_n(e^{-t} + p[1 - e^{-t}])$. Therefore, for all $x_0 \in \Omega_n$, it holds that

$$\mathbb{P}_{x_0}[X_t \in \cdot] \leq \text{Perc}_n(e^{-t} + p[1 - e^{-t}]) .$$

**Proof.** By definition, $\mathcal{F}_{t+\tau_i}^i(e) = 1$ if $\text{Upd}[\tau_i, \tau_i + t](e) = \emptyset$. Otherwise, i.e., if $\text{Upd}[\tau_i, \tau_i + t](e) \neq \emptyset$,

$$\mathcal{F}_{t+\tau_i}^i(e) = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}$$

since the status of $\mathcal{F}_{t+\tau_i}^i(e)$ depends only on the last update for this edge before $t + \tau_i$. Since

$$\mathbb{P}[\text{Upd}[\tau_i, \tau_i + t](e) = \emptyset] = e^{-t} ,$$

it follows that

$$\mathbb{P}[\mathcal{F}_{t+\tau_i}^i(e) = 1] = e^{-t} + p[1 - e^{-t}] .$$

The proof of the first assertion is completed since the status of edges are independent under SPD. The second assertion follows from Lemma 4.2 and choosing $i = 0$. \hfill \Box

As indicated in Section 2, to avoid having to analyze long range correlations that the FK-dynamics allows, we will allow ourselves an $O(1)$ burning time which will be enough by the above domination results for the configuration to look like a sample of a subcritical percolation. This then creates a situation where no connected component is large and hence the interactions between various edges are still rather local. To make this formal, denote by $p_{\text{perc}}(d) \in (0, 1)$ the critical probability of the edge percolation in $\mathbb{Z}^d$.

From now one we will assume that $p \in (0, p_{\text{perc}}(d))$ and further arguments would put additional smallness conditions on $p$. Define

$$p_{\text{init}} = p_{\text{init}}(p) := \frac{1}{2}(p + p_{\text{perc}}(d)) \in (p, p_{\text{perc}}(d)) ,$$

and let $t_{\text{init}} = t_{\text{init}}(p)$ be the solution of the following equation:

$$p(1 - e^{-t_{\text{init}}}) + e^{-t_{\text{init}}} = p_{\text{init}} . \quad (4.1)$$

As the next lemma will show, we can restrict our initial conditions to the class of measures $\nu$ satisfying $\nu \leq \text{Perc}_n(p_{\text{init}})$. More precisely, define

$$\tilde{d}(t) = \sup_{\nu : \nu \leq \text{Perc}_n(p_{\text{init}})} \|\mathbb{P}_\nu[X_t \in \cdot] - \mu^{n}_{p, q}\|_{TV} ,$$

and

$$\hat{t}_{\text{mix}} = \inf \{ t : \tilde{d}(t) < 1/4 \} .$$

**Proposition 4.4.** For all $p < p_{\text{perc}}(d)$ and $t > t_{\text{init}}$, we have

$$\sup_{x_0 \in \Omega_n} \|\mathbb{P}_{x_0}[X_t \in \cdot] - \mu^{n}_{p, q}\|_{TV} \leq \sup_{\nu : \nu \leq \text{Perc}_n(p_{\text{init}})} \|\mathbb{P}_\nu[X_{t-t_{\text{init}}} \in \cdot] - \mu^{n}_{p, q}\|_{TV} . \quad (4.2)$$

Therefore, we have

$$\tilde{t}_{\text{mix}} \leq \hat{t}_{\text{mix}} \leq \hat{t}_{\text{mix}} + t_{\text{init}} . \quad (4.3)$$
Lemma 4.7. Given the above definitions, we have the following comparison results.

This finishes the proof of (1). Part (2) can be readily obtained from the observation that
\[ \tilde{d}(t) \leq d(t) \leq \tilde{d}(t - t_{\text{init}}) \] by (4.2).

4.2. Enlarged percolations. In this section we define the third model indicated at the beginning of the section.

Definition 4.5. We define two sequences of random configurations \((\mathcal{F}_i)_{i \in \mathbb{N}}\) and \((\mathcal{E}_i)_{i \in \mathbb{N}}\) in \(\Omega_n\) as follows:

1. For \(i \in \mathbb{N}\), define \(\mathcal{F}_i \in \Omega_n\) as
   \[ \mathcal{F}_i(e) = 1 \text{ iff } \mathcal{F}_i^{-1}(e) = 1 \text{ for some } t \in [\tau_i, \tau_{i+1}] . \]
   Note that here we consider \(\mathcal{F}_i^{-1}\) instead of \(\mathcal{F}_i\) since otherwise \(\mathcal{F}_i(e)\) would be deterministically 1.
2. For \(i \in \mathbb{Z}_+\), define \(\mathcal{E}_i \in \Omega_n\) as
   \[ \mathcal{E}_i(e) = 1 \text{ iff } \mathcal{E}_i^i(e) = 1 \text{ for some } t \in [\tau_i, \tau_{i+1}] . \]

The following result is a static version of Lemma 4.2.

Lemma 4.6. Under the canonical coupling, for all \(i \in \mathbb{N}\), we have
\[ X_t \leq \mathcal{F}_i \text{ for all } t \in [\tau_i, \tau_{i+1}] . \]

Proof. Since \(X_t \leq \mathcal{F}_i\) for all \(t \in [\tau_i, \tau_{i+1}]\) by Lemma 4.2, the proof is immediate from the definition of \(\mathcal{F}_i\). \(\square\)

Now we investigate the distributions of \(\mathcal{E}_i\) and \(\mathcal{F}_i\). To this end we introduce the non-update percolation \(\mathcal{N}_i \in \Omega_n\), for \(i \in \mathbb{Z}_+\), as the following:
\[ \mathcal{N}_i(e) = \begin{cases} 1 & \text{if } \text{Upd}[\tau_i, \tau_{i+1}](e) = \emptyset , \\ 0 & \text{if } \text{Upd}[\tau_i, \tau_{i+1}](e) \neq \emptyset . \end{cases} \] (4.4)

In order words, \(\mathcal{N}_i(e) = 0\) if and only if there is an update \((t_1, U_1) \in \text{Upd}(e)\) such that \(t_1 \in (\tau_i, \tau_{i+1}]\). Given the above definitions, we have the following comparison results.

Lemma 4.7. The following holds:

1. For all \(i \in \mathbb{Z}_+\), we have \(\mathcal{E}_i \leq \text{Perc}_n(p^{1/2})\).
2. For all \(i \in \mathbb{Z}_+\), we have \(\mathcal{N}_i \leq \text{Perc}_n(p^{1/2})\).
3. For all \(i \in \mathbb{N}\), we have \(\mathcal{F}_i \leq \text{Perc}_n(3p^{1/2})\).

Proof. We start by observing that \(\mathcal{E}_i(e) = 1\) if and only if
\[ \{ U > 1 - p \text{ for some } (t, U) \in \text{Upd}[\tau_i, \tau_{i+1}](e) \} . \] (4.5)

To compute the probability of the latter notice that given the event \(|\text{Upd}[\tau_i, \tau_{i+1}](e)| = k\), the event (4.5) happens with probability \(1 - (1 - p)^k\). Hence, the probability of the event (4.5) can be written as
\[ \sum_{k=0}^{\infty} e^{-\Delta} \frac{\Delta^k}{k!} (1 - (1 - p)^k) = 1 - e^{-p\Delta} \leq p\Delta = p^{1/2} . \]

This finishes the proof of (1). Part (2) can be readily obtained from the observation that
\[ \mathcal{N}_i \sim \text{Perc}_n(e^{-\Delta}) \leq \text{Perc}_n(p^{1/2}) . \]

For part (3), we claim that
\[ \mathcal{F}_i \leq \mathcal{N}_{i-1} \cup \mathcal{E}_{i-1} \cup \mathcal{E}_i . \] (4.6)
CUTOFF FOR RCM

Percolation | Description | Defined in
--- | --- | ---
$\mathcal{E}_i$ | Percolation starting at $\tau_i$ from empty | Def. 4.1
$\mathcal{F}_i$ | Percolation on $[\tau_i, \infty)$, starting at $\tau_{i-1}$ from full | Def. 4.1
$\overline{\mathcal{E}}_i$ | At least one open update in $[\tau_i, \tau_{i+1}]$ | Def. 4.5
$\mathcal{F}_i$ | Open some time in $[\tau_i, \tau_{i+1}]$ starting with full at $\tau_{i-1}$ | Def. 4.5
$\mathcal{N}_i$ | Non-update in $[\tau_i, \tau_{i+1}]$ implies open | (4.4)
$\Xi_i$ | $\overline{\mathcal{E}}_i \cup \mathcal{N}_i$ | (4.7)

Table 1. Different kinds of percolation.

This claim along with parts (1) and (2) will finish the proof. To prove the claim, first suppose that $F_i(e) = 1$ and $N_{i-1}(e) = 0$. Then, $\text{Upd}[\tau_{i-1}, \tau_i](e) \neq \emptyset$ and hence we can take the last update $(t_1, U_1)$ in $\text{Upd}[\tau_{i-1}, \tau_i](e)$. Since $F_i(e) = 1$, at least one update $(t, U)$ in $\{(t_1, U_1)\} \cup \text{Upd}[\tau_i, \tau_{i+1}](e)$ satisfies $U > 1 - p$. It implies either $\overline{E}_{i-1}(e) = 1$ or $\overline{E}_i(e) = 1$. This finishes the proof.

We end this section with a final definition. For $i \in \mathbb{Z}_+$, let

$\Xi_i := \overline{E}_i \cup \mathcal{N}_i \in \Omega_n$. (4.7)

We record a key fact in the next lemma. In short the lemma says that the FK-dynamics across time can be dominated by a sequence of Bernoulli percolations which are one dependent across time. This will be crucially used in the analysis of how information spreads in the FK-dynamics.

**Proposition 4.8.** The following hold:

1. For all $i \in \mathbb{Z}_+$, the distribution of $\Xi_i$ is stochastically dominated by $\text{Perc}_n(2p^{1/2})$.
2. For all $i \in \mathbb{N}$, under the canonical coupling, we have that

$X_t \leq \Xi_{i-1} \cup \Xi_i$ for all $t \in [\tau_i, \tau_{i+1}]$.

**Proof.** The proof is an immediate consequence of Lemma 4.6 and (4.6). □

For purpose of easy reference throughout the article we record all the percolation models defined so far in Table 1.

4.3. Decay of connectivity. We now record some useful exponential decay of connectivity results for a non-equilibrium RCM. It is well-known that for a sub-critical bond percolation or RCM one observes an exponential decay of connectivity, i.e., the probability that two sites $u$ and $v$ belong to same cluster decays exponentially in the graph distance $d(u, v)$ (cf. [2, Theorem 2]). We would need a dynamical version for our purposes and start with some definitions. Note that $\overline{F}_i$ had so far been defined for $i \geq 1$ only. We now define $\overline{F}_0$ as

$\overline{F}_0 = X_0 \cup \overline{E}_0$.

Then, by definition

$X_t \leq \overline{F}_0$ for all $t \in [0, \tau_1]$. (4.8)

**Proposition 4.9.** For all small enough $p$, there exists $\gamma = \gamma(p) > 0$ such that,

$$\sup_{\nu: \nu \leq \text{Perc}_n(p_{init})} \mathbb{P}_\nu \left[ u \overset{\overline{F}_i}{\leftrightarrow} v \right] \leq e^{-\gamma d(u, v)}$$

for all $i \in \mathbb{Z}_+$, $n \in \mathbb{N}$, and $u, v \in \Lambda_n$. 
Proof. By Lemma 4.7, the distribution of $F_{\mathcal{F}}_i$ is dominated by $\text{Perc}_n(3p^{1/2})$ for $i \geq 1$. For $i = 0$, we notice from the definition of $F_{\mathcal{F}}_0$ that the distribution of the latter is dominated by $\text{Perc}_n(p_{\text{init}} + p^{1/2})$.

In conclusion, for all small enough $p$, the distribution of $F_{\mathcal{F}}_i$, $i \geq 0$, is dominated by $\text{Perc}_n(s)$ for some $s < p_{\text{perc}}(d)$ and hence we are done by decay of connectivity for subcritical percolation [7, Theorem 3.7]. \qed

Similar arguments imply the following result which we record for future purposes.

**Lemma 4.10.** Suppose that two disjoint subsets $A$ and $B$ of $E_n$ satisfy $d(A, B) \geq c\log^2 n$ for some $c > 0$. Denote by $\mu_n^+$ the random-cluster measure on $B^c = E_n \setminus B$ under the full boundary condition on $B$. Denote by $X \in \{0, 1\}^{B^c}$ a random-cluster configuration sampled according to $\mu_n^+$, and denote by $\text{Conn}(B; X)$ the set of edges in $B^c$ connected to an edge of $B$ via an open path in $X$. Then, we have that

$$\mu_n^+ [\text{Conn}(B; X) \cap A = \emptyset] \geq 1 - \frac{1}{n^{2d}}.$$  

Proof. By the decay of connectivity established above for $e \in \partial B$, we have

$$\mu_n^+ [\text{Conn}(B; X) \cap A \neq \emptyset] \leq \mu_n^+ \left[ u \leftrightarrow v \text{ for some } u, v \in B^c \text{ such that } d(u, v) \geq c\log^2 n \right] \leq e^{-c\log^2 n \left( \frac{n^d}{2} \right)} \leq \frac{1}{n^{2d}},$$

where the second inequality follows by the union bound. \qed

The final result of this section records a statement about how fast disagreement percolates in FK-dynamics.

### 4.4. Estimates on the propagation of disagreements

We fix a subset $A \subset E = E_n$ this section. Define an enlargement $A^+$ of $A$ as

$$A^+ = \{ e \in E : d(e, A) \leq \log^4 n \}.$$  \hfill (4.9)

where $d(\cdot, \cdot)$ denotes the shortest path metric on the natural graph structure on $E_n$ where the distance between any two adjacent edges is taken to be 1. The main objective in this section is to show that, under monotone coupling, FK-dynamics started from two configurations that agree on $A^+$ and are reasonably sparse, continue to agree on $A$ for all $t \in [0, t_{\text{max}}]$ where

$$t_{\text{max}} = \log^2 n.$$  \hfill (4.10)

Consider two censored dynamics $(Z^+_t)$ (resp. $(Z^-_t)$) as FK-dynamics on $\{0, 1\}^{A^+}$ conditioned on full (resp. empty) configuration on $E_n \setminus A^+$. Let $\text{Perc}_n^{A^+}(p_{\text{init}})$ denote the percolation measure on $A^+$ with open probability $p_{\text{init}}$.

**Lemma 4.11.** Consider two copies of FK-dynamics $(Z^+_t)$ and $(Z^-_t)$ on $\{0, 1\}^{A^+}$ coupled via the monotone coupling. Suppose that the law of the initial condition $Z^+_0$ follows a law $\nu$ on $\{0, 1\}^{A^+}$ satisfying $\nu \preceq \text{Perc}_n^{A^+}(p_{\text{init}})$, and suppose further that $Z^+_0 = Z^-_0$. Then, for all sufficiently small $p$, we have that

$$\mathbb{P} \left[ Z^+_t(A) = Z^-_t(A) \text{ for all } t \in [0, t_{\text{max}}] \right] \geq 1 - \frac{1}{n^{3d}}.$$  \hfill (4.11)

**Remark 4.12.** Note that the probability in (4.11) is with respect to both the FK-dynamics and also the initial measure $\nu$. In other words, this is an annealed probability.

**Remark 4.13.** Even though we considered the two worst boundary conditions, namely, full and empty, a simple monotonicity consideration allows us to conclude that

$$\mathbb{P} \left[ Z^+_t(A) = Z^-_t(A) = Z_t(A) \text{ for all } t \in [0, t_{\text{max}}] \right] \geq 1 - \frac{1}{n^{3d}}.$$
Figure 4.1. Figure illustrating the weak spatial mixing property of the subcritical RCM. Here we consider the equilibrium measures with free (LHS) and wired boundary conditions (RHS). By monotonicity of the equilibrium measures with respect to their boundary conditions, there exists a coupling such that the LHS is dominated by the RHS. However under this coupling by the exponential decay of connectivity the RHS (and hence the LHS) has a closed surface (contour in the planar case) within $O(\log n)$ distance from the boundary and they agree in the interior of the surface in particular on the green region.

where $Z_t$ is one of the following processes on $\{0, 1\}^{A^+}$:

- A censored dynamics on $A^+$ conditioned on any configuration on $E_n \setminus A^+$.
- If $A$, and hence $A^+$, are square boxes, the FK-dynamics on $\{0, 1\}^{A^+}$ with periodic boundary conditions.
- The FK-dynamics projected to $A \subset E_n$, i.e., $X_t(A^+)$.

Remark 4.14. In the above theorem, the size of the ambient space $\Lambda_n$ (which is $n$) is not important. Taking the ambient space to be $\Lambda_m$ which contains $A^+$ suffices. Moreover, we can replace $\log^4 n$ in the statement of lemma with $\log^{4+\delta}$ for any $\delta > 0$ with $t_{\text{max}} = \log^{1+\delta} n$.

The proof follows the arguments in [4, 24] and is postponed to the Appendix (Section 8).

5. Information percolation clusters and time dependent Bernoulli percolations

As emphasized before, this is the section which contains all the new ideas in the paper. The main result is the following bound on $L^2$-mixing. Recall the spectral gap $\lambda(n)$ from Corollary 3.4.

**Theorem 5.1.** For all small enough $p > 0$, there exists $C = C(p) > 0$ such that the following $L^2$-bound holds for all large enough $n$:

$$
\max_{x_0 \in \Omega_n} \|P_{x_0} [X_t \in \cdot] - \mu_{p,q}^n\|_{L^2(\mu_{p,q}^n)} \leq 2 \exp \{-\lambda(n)(t - C \log n)\}
$$

for all $t \geq C \log n$.

**Notation 5.2.** From now on, all the statements are asymptotic in $n$, so that they might hold only when $n$ is large enough. We shall not repeat stating this explicitly.
Recall that the spectral gap governs the rate of decay of $L^2$ norm. More precisely for any $s \leq t$ and any starting state $x_0 \in \Omega_n$ we have (see for e.g. [17, Lemma 20.5]),

$$\left\| \mathbb{P}_{x_0} [X_t \in \cdot] - \mu_{p,q}^n \right\|_{L^2(\mu_{p,q}^n)} \leq e^{-\lambda(n)(t-s)} \left\| \mathbb{P}_{x_0} [X_s \in \cdot] - \mu_{p,q}^n \right\|_{L^2(\mu_{p,q}^n)} .$$

(5.1)

By Corollary 3.4, it suffices to prove the following proposition.

**Proposition 5.3.** For all small enough $p > 0$, there exists $C = C(p) > 0$ such that for $t_* = C \log n$,

$$\max_{x_0 \in \Omega_n} \left\| \mathbb{P}_{x_0} [X_{t_*} \in \cdot] - \mu_{p,q}^n \right\|_{L^2(\mu_{p,q}^n)} \leq 2 .$$

(5.2)

**Notation 5.4.** We shall write $C$ or $c$ for positive constants where different occurrences of $C$ or $c$ may denote different constants.

The proof of Proposition 5.3 is the heart of this work and is rather long, intricate and involves several percolation arguments based on the models introduced in Section 4. As mentioned earlier, using the results of this section as inputs, the arguments of the following sections are quite similar to the ones appearing in [18, 25]. Readers not familiar with these papers, at first read, to get a sense of the overall flow of arguments, could choose to assume Theorem 5.1 and read the following easier sections first, before coming back to this section.

We provide a roadmap for this section for the ease of reading.

- The construction of information percolation is done in Section 5.1 relying on the definitions in Section 4, particularly the percolation models listed in Table 1. At a very high level it amounts to classifying vertices into green, red and blue where the state of the red vertices depend on the initial configuration, the blue vertices are independent Bernoulli variables independent of everything else, whereas the green vertices have a complicated dependency on each other but are still independent of the initial configuration (Theorem 5.7).
- Using the above, the proof of Proposition 5.3 occupies Sections 5.2 and 5.3. The key steps are the following:
  1. To bound the $L^2$-distance it suffices to condition on the green clusters. Then the strategy is to compute the $L^2$-distance of the conditional distribution to a product Bernoulli Measure instead of the equilibrium measure (Lemma 5.15). The Bernoulli measure is exactly the one which describes the law of the blue vertices. Thus this distance would be zero if there does not exist any red cluster.
  2. We then establish the key estimate showing exponential unlikeliness of red vertices with time in Proposition 5.13 which makes the above step sufficient. The proof of this proposition uses a comparison with a subcritical branching process and is presented in Section 5.3. In particular, the proof involves delicate geometric arguments relying on several properties of the auxiliary percolation models defined in Table 1.

### 5.1. Information percolation (IP).

As mentioned before (Section 4.1), we will discretize time using $\tau_i$ and will define IP on the space-time slab $E_n \times [\tau_1, \tau_m]$ for some $m \in \mathbb{N}$. We shall take $m = \Omega(\log n)$ later, but for the moment we think of $m$ as a fixed integer. We also recall the various percolations defined in Table 1.

For $\Xi \in \Omega_n$, and $e = (u, v) \in E_n \cap \Xi$ where $u, v \in \Lambda_n$, denote by Conn$(e; \Xi)$, the connected component of $\Xi$ containing $\{u, v\}$. We will often think of Conn$(e; \Xi)$ also as a set of vertices.

Define $\partial$Conn$(e; \Xi)$ as the edge boundary of Conn$(e; \Xi)$ i.e., as the set of edges in $E_n \setminus$ Conn$(e; \Xi)$ which are adjacent to an edge of Conn$(e; \Xi)$, and define

$$\overline{\text{Conn}}(e; \Xi) = \text{Conn}(e; \Xi) \cup \partial\text{Conn}(e; \Xi) .$$

(5.3)

Note that Conn$(e; \Xi) = \overline{\text{Conn}}(e; \Xi) = \emptyset$ if $\Xi(e) = 0$. Given the above notations, we now define IP for the FK-dynamics. It would be notationally convenient to define $\tau_{i+1/2} = (i + 1/2) \Delta$, for $i \in \mathbb{N}$. Furthermore to distinguish between edges (elements of $E_n$) and connections across time, we
will call the former as ‘space edges’ just edges and the latter as ‘time edges’ (see Figure 5.1 for an illustration).

**Definition 5.5** (Information percolation). The information percolation cluster is defined on the space-time slab \( E_n \times [\tau_1, \tau_m] \) for some fixed \( m \geq 2 \). For an edge \( e \in E_n \), we define the history \( \mathcal{H}_e = (\mathcal{H}_e(t))_{t \in [\tau_1, \tau_m]} \) associated to the edge \( e \) backward in time recursively as follows: Start by setting \( \mathcal{H}_e(\tau_m) = \{e\} \).

1. For each \( t = \tau_{i+1} \) with \( i \in [1, m-1] \) suppose that \( \mathcal{H}_e(\tau_{i+1}) \) is given by a subset of \( E_n \). Then we set \( \mathcal{H}_e(\tau_{i+1}/2) \) be the same as \( \mathcal{H}_e(\tau_{i+1}) \) as well for any \( w \in \mathcal{H}_e(\tau_{i+1}) \) we connect the two edges \((w, \tau_{i+1})\) and \((w, \tau_{i+1/2})\), by a ‘time edge’ in the time direction (see Figure 5.1.)

2. For each \( w \in \mathcal{H}_e(\tau_{i+1/2}) \), we check if it has been updated in the time interval \((\tau_i, \tau_{i+1})\) (recall the various notations from Table 1).
   
   a. If \( \mathcal{N}_i(e) = 1 \), then introduce the ‘space edge’ \((w, \tau_i)\) and connect \((w, \tau_{i+1/2})\) and \((w, \tau_i)\) by a time edge.
   
   b. If \( \mathcal{N}_i(e) = 0 \), we take the last update \((t_0, U_e)\) for \( e \) in \((\tau_i, \tau_{i+1})\).
      
      i. If \( U_e < 1 - p + p^p \), this update is called *oblivious* and we do not take any action on the vertex \((w, \tau_{i+1/2})\).
      
      ii. If \( U_e > 1 - p + p^p \), then \( w \) is open in \( \Xi_i \), and hence is open in \( \Xi_i \) as well. In this case, we include all the edges in \( \text{Conn}(w; \Xi_{i-1} \cup \Xi_i) \) in \( \mathcal{H}_e(\tau_{i+1/2}) \) and \( \mathcal{H}_e(\tau_i) \).
      
      Finally we connect the space edges \((w', \tau_{i+1/2})\) and \((w', \tau_i)\) for all the edges \( w' \) in \( \text{Conn}(w; \Xi_{i-1} \cup \Xi_i) \) using time edges.

3. Steps 1 and 2 above, defines \( \mathcal{H}_e(\tau_i) \) as a subset of \( E_n \). Now return to the first step if \( i \geq 2 \) to use the above construction recursively.

For \( A \subset E_n \), define \( \mathcal{H}_A = (\mathcal{H}_A(t))_{t \in [\tau_1, \tau_m]} \) as \( \mathcal{H}_A = \bigcup_{e \in A} \mathcal{H}_e \). Two histories \( \mathcal{H}_e \) and \( \mathcal{H}_{e'} \) are connected if they share an edge.

Some remarks are in order. First, we emphasize that two histories \( \mathcal{H}_e \) and \( \mathcal{H}_{e'} \) are regarded as two disconnected pieces if they share vertices only. Second, by the construction rule, one can observe that:

\[
\mathcal{H}_e(\tau_{i+1/2}) = \mathcal{H}_e(\tau_{i+1}) \cup \mathcal{H}_e(\tau_i) . \tag{5.4}
\]

Using terminology from existing literature we will often refer to the collection \( \mathcal{H} := \{\mathcal{H}_e\}_{e \in E_n} \) as the history diagram. This induces a new graph structure on \( E_n \). i.e. \( e \) and \( e' \) are connected if \( \mathcal{H}_e \) and \( \mathcal{H}_{e'} \) are connected.

With the above conventions, each connected component of this new graph is called an *information percolation cluster*. We shall simply refer to them as *clusters*. Let them be indexed by the set \( C \).

**Definition 5.6** (IP clusters and their colors). Each cluster \( C \in \mathcal{C} \) is colored red, blue or green according to the following rule:

- Colored **red** if \( \mathcal{H}_C(\tau_1) \neq \emptyset \).
- Colored **blue** if \( \mathcal{H}_C(\tau_1) = \emptyset \) and \( |C| = 1 \).
- Colored **green** if \( \mathcal{H}_C(\tau_1) = \emptyset \) and \( |C| \geq 2 \).

Denote by \( \mathcal{C}_R \), \( \mathcal{C}_B \) and \( \mathcal{C}_G \) the collection of red, blue and green clusters, respectively. Define

\[
E_R = \{e : e \in \mathcal{C}_R\} \tag{5.5}
\]

and define \( E_B \) and \( E_G \) similarly. We use the following simplified notations to denote the history diagrams emanating from the various colored edges:

\[
\mathcal{H}_R := \mathcal{H}_{E_R}, \quad \mathcal{H}_B := \mathcal{H}_{E_B}, \quad \text{and} \quad \mathcal{H}_G := \mathcal{H}_{E_G}.
\]
The various colors indicate the values of the uniform variables for each update: gray $\leftrightarrow \{U < 1 - p\}$, purple $\leftrightarrow \{1 - p \leq U \leq 1 - p + p^*\}$, black $\leftrightarrow \{U > 1 - p + p^*\}$. The purple region denotes $\mathcal{E}_i(e) = 1$, while the yellow region implies that $\mathcal{N}_i(e) = 1$. (Down) In the two graphs, the gray region indicates whether $\Xi_i \cup \Xi_{i-1}(e)$ is 1 or 0 with gray indicating the former. (Down-left) History diagrams for $e_1$, $e_2$, $e_3$, $e_4$. We can assert that $\mathcal{H}_{e_1}$ is red, but not able to say anything about the remaining ones; (Down-right) History diagram for $e_0$ is combined with that of $e_1$. $e_4$ belongs to green cluster although its last update is oblivious. The vertical edges acting as connections across time are referred to as ‘time edges’ in the article.

The following theorem justifies the above definitions. In short, it says that to reconstruct the state of the edges in $\mathcal{H}_A(\tau_{i+1})$, all one needs is the update sequence and the state of the edges $\mathcal{H}_A(\tau_i)$ at time $\tau_i$ provided that $A$ is a cluster.

**Theorem 5.7.** Given a history diagram $\mathcal{H}$, suppose that a set $A \subset E_n$ is a cluster. Then, for each $i \in [1, m - 1]$, the configuration $X_{\tau_{i+1}}(\mathcal{H}_A(\tau_{i+1}))$ is a deterministic function of

$$X_{\tau_i}(\mathcal{H}_A(\tau_i)) \quad \text{and} \quad \bigcup_{e \in \mathcal{H}_A(\tau_{i+1}/2)} \text{Upd}[\tau_i, \tau_{i+1}](e).$$

(5.6)
Consider the following decomposition of dependence on some consequences of the above definitions. We temporarily fix $A \subset E_n$ and suppressing the dependence on $A$, define

$$W_j = \mathcal{H}_A(\tau_j) ; j \in [1, m].$$

(5.7)

Consider the following decomposition of $W_j$

$$W_j = W_j^{\text{NU}} \cup W_j^{\text{Ob}} \cup W_j^{\text{NOb}},$$

(5.8)

where,

$$W_j^{\text{NU}} = \{ e \in W_j : \mathcal{M}_j(e) = 1 \}$$

i.e., the edges that have not been updated in $[\tau_{j-1}, \tau_j]$, 

$$W_j^{\text{Ob}} = \{ e \in W_j : \mathcal{M}_j(e) = 0 \text{ and the last update for } e \text{ in } [\tau_{j-1}, \tau_j] \text{ is oblivious} \},$$

$$W_j^{\text{NOb}} = \{ e \in W_j : \mathcal{M}_j(e) = 0 \text{ and the last update for } e \text{ in } [\tau_{j-1}, \tau_j] \text{ is non-oblivious} \}.$$

For each $j \in [1, m - 1]$, we write

$$C_j = \bigcup_{\tau \in W_j^{\text{NOb}}} \text{Conn}(e; \Xi_{j-1} \cup \Xi_j),$$

(5.9)

Since by (b) in Definition 5.5, one can observe that

$$W_j = C_j \cup W_{j+1}^{\text{NU}},$$

(5.10)

and hence $C_j \subset W_j$. Therefore, by writing

$$N_j = W_j \setminus C_j,$$

(5.11)

we obtain another decomposition of $W_j$ given by

$$W_j = C_j \cup N_j.$$

(5.12)

We record some basic properties of $C_j$ and $N_j$ next.

**Lemma 5.9.** For all $j \in [1, m - 1]$, it holds that

$$W_{j+1}^{\text{NOb}} \subset C_j \text{ and } N_j \subset W_j^{\text{NU}}.$$

*Proof.* For the first inclusion, we note that $e \in W_{j+1}^{\text{NOb}}$ implies that $\overline{\mathcal{M}}_j(e) = 1$ and thus $\Xi_j(e) = 1$. Hence, the definition (5.9) indicates that $e \in C_j$ as well and thus the first inclusion trivially holds. For the latter one, it suffices to recall (5.10) and the definition (5.11) of $N_j$. \qed

For $S \subset E_n$, define $\partial^{-} S$ as the set of edges in $S$ which are adjacent to at least one edge in $S^c$, i.e., $\partial^{-} S = \partial(E \setminus S)$. We record the following simple fact.

**Lemma 5.10.** For all $j \in [1, m - 1]$, all the edges in $\partial^{-} C_j$ are closed in $\Xi_{j-1} \cup \Xi_j$. In particular, there is no open path in $\Xi_{j-1} \cup \Xi_j$ connecting an open edge in $C_j$ and an edge in $N_j$.

*Proof.* The proof is direct from the definition of $C_j$ where we included the closed boundary of $\text{Conn}(e; \Xi_{j-1} \cup \Xi_j)$. \qed

**Lemma 5.11.** For all $j \in [1, m - 1]$, for each $e \in C_j$, and for all $t \in [\tau_j, \tau_{j+1}]$, the process $X_t(e)$ is a deterministic function of

$$X_{\tau_j}(\mathcal{H}_{C_j}(\tau_j)) \text{ and } \bigcup_{e' \in C_j} \text{Upd}[\tau_j, t](e').$$
Proof. Let

$$\mathcal{U}_t = \bigcup_{e \in C_j} \text{Upd}[\tau_j, t](e') ; t \in [\tau_j, \tau_{j+1}] .$$

We fix $e \in C_j$ and $t \in [\tau_j, \tau_{j+1}]$ and denote by $(t_0, U_0)$ the last update for $e$ in $[\tau_j, t]$. If $U_0 < 1 - p + p^*$ then, in view of Definition 2.1, the configuration $X_t(e)$ is 1 if $U_0 < 1 - p$, and 0 if $U_0 \geq 1 - p$. Thus, we can determine $X_t(e)$ solely in terms of $(t_0, U_0) \in \text{Upd}[\tau_j, t](e) \subset \mathcal{U}_t$. Now we consider the case $U_0 > 1 - p + p^*$. In this case, the configuration $X_t(e) = X_{t_0}(e)$ is determined by checking whether $e$ is a cut-edge or not in the configuration $X_{t_0}$. In order to check this, one has to investigate $\text{Conn}(e; X_{t_0}\cup\{e\})$ to determine whether removing $e$ disconnects some component of $X_{t_0}\cup\{e\}$ or not. Note that $e$ is open in $\partial e$ (and hence in $\Xi$) since $U_0 > 1 - p + p^* > 1 - p$. Thus, by Proposition 4.8, we have

$$\text{Conn}(e; X_{t_0}\cup\{e\}) \subset \text{Conn}(e; \Xi_j \cup \Xi_{j+1}) \subset C_j \setminus \partial C_j .$$

Therefore, we can determine $X_t(e)$ in terms of $X_{t_0}(C_j)$ and $(t_0, U_0) \in \mathcal{U}_t$.

If $\mathcal{U}_t = \{(t_0, U_0)\}$, we have $X_{t_0}(C_j) = X_{t_1}(C_j)$, so we can conclude the proof. Otherwise, we take the last update $(t_1, U_1) \in \mathcal{U}_t$ other than $(t_0, U_0)$. Then, we have,

$$X_{t_0}(C_j) = X_{t_1}(C_j) .$$

Since there are finitely many updates in $[\tau_j, \tau_{j+1}]$ almost surely, we can repeat this procedure to finish the proof. An important fact implicitly used above is that in repeating the argument all the edges $\tilde{e}$ that we encounter with an update time $\tilde{t} \in [\tau_j, t_0]$ has the property that the connected component of

$$X_{\tilde{t}}(\tilde{e}) \subset \overline{\text{Conn}(e; \Xi_{j-1} \cup \Xi_j)} \subset C_j ,$$

since the edge boundary of $\text{Conn}(e; \Xi_{j-1} \cup \Xi_j)$ remains closed throughout the interval $[\tau_j, \tau_{j+1}]$. □

The proof of Theorem 5.7 now follows.

Proof of Theorem 5.7. In view of (5.8) and the first inclusion of Lemma 5.9, it suffices to consider the following three cases separately.

- Case 1: $e \in W_{i+1}^{\text{NU}}$. By (2)-(a) of Definition 5.5, we have $X_{\tau_{i+1}}(e) = X_{\tau_i}(e)$ and thus configuration of $X_{\tau_{i+1}}(e)$ is determined by $X_{\tau_i}(W_{i+1}^{\text{NU}})$. Since $W_{i+1}^{\text{NU}} \subset W_i$, the proposition holds for this case.
- Case 2: $e \in W_{i+1}^{\text{Ob}} \setminus C_i$. By (2)-(b)-(i) of Definition 5.5, the configuration $X_{\tau_{i+1}}(e)$ is solely determined by the last update for $e$ in $(\tau_i, \tau_{i+1}]$ and therefore the proposition holds as well.
- Case 3: $e \in W_{i+1} \cap C_i$. This case is immediate from Lemma 5.11.

□

The following corollary is an immediate consequence of the previous theorem.

Corollary 5.12. Given a history diagram $\mathcal{H}$, the following holds.

1. The configurations $X_{\tau_{n}}(E_G)$ and $X_{\tau_{n}}(E_n \setminus E_G)$ are independent.
2. The configuration $X_{\tau_{n}}(E_G)$ is independent of $X_{\tau_{n}}$.
3. Suppose that $e \in E_G$. Then, the distribution of $X_{\tau_{n}}(e)$ is a Bernoulli random variable with parameter $p^*_{1-p+p}$, and is independent of all other randomness.

Proof. Parts (1) and (2) are direct consequences of Theorem 5.7 and the definition of a green cluster. We now consider part (3). For $e \in E_G$, the configuration $X_{\tau_n}(e)$ is determined by the last update $(t, U)$ for $e$ in $[\tau_1, \tau_n]$. Furthermore, since $e \in E_G$, this last update is oblivious and therefore we know that $U < 1 - p + p^*$. Given this condition, we have $X_{\tau_n}(e) = 1$ if $U < 1 - p$ and $X_{\tau_n}(e) = 0$ if $U \in [1 - p, 1 - p + p^*]$ otherwise. This finishes the proof of part (3). □
For each $A \subset E_n$, define
$$\mathcal{H}_A^- = \mathcal{H}_{E_n \setminus A}.$$  
As in [25, 22], it would be crucial to estimate the probability of $A$ being a red cluster or a collection of singleton blue clusters i.e.,
$$\{A \in \mathcal{C}_R \} \cup \{A \subset E_B\}.$$  
Furthermore, technical aspects make it important to estimate the above probabilities conditioned on the history diagram of the complement of $A$. For this conditional probability to be non-zero a necessary condition is that,
$$\mathcal{H}_A^- \cap \{A \times \{t = \tau_{m-1/2}\}\} = \emptyset,$$  
This is because, suppose that $e \in A$ satisfies $(e, \tau_{m-1/2}) \in \mathcal{H}_{e'}$ for some $e' \in E_n \setminus A$. Then, by the definition of the information percolation cluster, the cluster containing $e$ must contain $e'$ as well.

Thus this is a compatibility condition to guarantee that $\{A \in \mathcal{C}_R \} \cup \{A \subset E_B\}$ is a non-empty event which we denote by $\mathcal{H}_A^- \in \mathcal{H}_{\text{com}}(A)$. Given this, we define
$$\mathcal{P}_A = \sup_{\mathcal{H}_A^- \in \mathcal{H}_{\text{com}}(A)} P[\mathcal{H}_R = \emptyset | \mathcal{H}_G \setminus \mathcal{H}_A],$$  
i.e., the maximum probability of $A$ being a red cluster conditioned on a compatible $\mathcal{H}_A^-$. Given the above preparation, the following proposition is the main estimate (similar to [25, Lemma 4.7]) needed. For $A \subset E_n$, we denote by $|\text{Conn}(A)|$, the smallest number of edges in any connected subgraph of $(\Lambda_n, E_n)$ containing $A$.

**Proposition 5.13.** For any $\theta > 0$, we can find two constants $C = C(\theta) > 0$ and $p_0 = p_0(\theta) > 0$ such that, for any $p \in (0, p_0)$, there exists a constant $\alpha = \alpha(p) > 0$ satisfying
$$\mathcal{P}_A \leq Ce^{-(\theta|\text{Conn}(A)|+\alpha m_n)}$$  
for all $A \subset E_n$.

A notable feature of this proposition is the fact that $\alpha$ is independent of $\theta$. In the remaining part of the current section, $\alpha$ always refers to the constant above. The proof of this proposition is postponed to Section 5.3. A corollary of this proposition is the following lemma which lower bounds the probability that there are no red clusters.

**Lemma 5.14.** For all small enough $p$, there exists a constant $C = C(p) > 0$ satisfying
$$\sup_{\mathcal{H}_G} P[\mathcal{H}_R = \emptyset | \mathcal{H}_G] \geq 1 - Cn^2e^{-\alpha m_n}.$$  

**Proof.** By the union bound and the definition of $\mathcal{P}_A$,
$$1 - \mathbb{P}[\mathcal{H}_R = \emptyset | \mathcal{H}_G] \leq \sum_{A \subset E_n, A \neq \emptyset} \mathbb{P}[A \in \mathcal{C}_R | \mathcal{H}_G] \leq \sum_{A \subset E_n, A \neq \emptyset} \mathcal{P}_A.$$  
Now, by Proposition 5.13 and the translation invariance of the periodic lattice,
$$\sum_{A \subset E_n, A \neq \emptyset} \mathcal{P}_A \leq \sum_{e \in E_n} \sum_{A : A \ni e} \mathcal{P}_A \leq Cn^2e^{-\alpha m_n} \sum_{k=1}^{\infty} \sum_{A : A \ni e, |\text{Conn}(A)| = k} e^{-\theta k}.$$  
The proof now follows from the following straightforward claim: For a fixed $e \in E_n$,
$$|\{A \subset E_n : A \ni e, |\text{Conn}(A)| = k\}| \leq (k + 1)(8d^2)^k,$$
(whose proof we leave to the reader); by taking $\theta$ large enough so that $8d^2e^{-\theta} < 1/2$ since,
$$\mathbb{P}[\mathcal{H}_R = \emptyset | \mathcal{H}_G] \geq 1 - Cn^2e^{-\alpha m_n} \sum_{k=1}^{\infty} (k + 1)(8d^2e^{-\theta})^k.$$  

The remainder of the section is now devoted to proving (5.2).
5.2. **Proof of Proposition 5.3.** For $A \subset E_n$, define $\nu_A$ as a Bernoulli percolation measure on $A$ with open probability $\overline{p}$ where

$$\overline{p} = \frac{p^n}{1 - p + p^n}.$$  \hfill (5.18)

In the remaining part of the section, we will simply write $\mu := \mu_{p,q}^n, E := E_n$ and denote by $\mu_A$, $A \subset E$, the projection of $\mu$ on $A$. We first prove the following lemma which shows that the $L^2$ distance to $\mu$ can be controlled by the $L^2$ distance of the measure on the complement of the green clusters to the measure $\nu$.

**Lemma 5.15.** For all small enough $p$, we can find $C = C(p) > 0$ such that for $m \geq C \log n$ we have

$$\|\mathbb{P}_{x_0} [X_t \in \cdot] - \mu\|_{L^2(\mu)} \leq 2 \sup_{\mathcal{H}_G} \|\mathbb{P}_{x_0} [X_t (E \setminus E_G) \in \cdot | \mathcal{H}_G] - \nu_{E \setminus E_G}\|_{L^2(\nu_{E \setminus E_G})} + 1,$$

for all $x_0 \in \Omega_n$.

**Proof.** Consider two copies of FK-dynamics $(X_t)$ and $(Y_t)$ where $X_0 = x_0$ and $Y_0$ is distributed according to $\mu$. We couple them via the monotone coupling introduced in Definition 2.1. Now by Jensen’s inequality (for details see [25, Lemma 4.12]) we obtain

$$\|\mathbb{P}_{x_0} [X_t \in \cdot] - \mu\|_{L^2(\mu)} = \|\mathbb{P}_{x_0} [X_t \in \cdot] - \mathbb{P}_\mu [Y_t \in \cdot]\|_{L^2(\mu)} \leq \int \|\mathbb{P}_{x_0} [X_t \in \cdot | \mathcal{H}_G] - \mathbb{P}_\mu [Y_t \in \cdot | \mathcal{H}_G]\|_{L^2(\mu_{p,q}^n(\cdot | \mathcal{H}_G))} \mathbb{P}(\mathcal{H}_G)$$

$$\leq \sup_{\mathcal{H}_G} \|\mathbb{P}_{x_0} [X_t \in \cdot | \mathcal{H}_G] - \mathbb{P}_\mu [Y_t \in \cdot | \mathcal{H}_G]\|_{L^2(\mu_{p,q}^n(\cdot | \mathcal{H}_G))}. \hfill (5.19)$$

Given $\mathcal{H}_G$, the diagram $\mathcal{H}_{E \setminus E_G}$ is disjoint from $\mathcal{H}_G = \mathcal{H}_{E_G}$, and as we noticed in Corollary 5.12 configurations $X_t(E_G)$ (resp. $Y_t(E_G)$) and $X_t(E \setminus E_G)$ (resp. $Y_t(E \setminus E_G)$) are independent. Moreover, $Y_t(E_G)$ and $X_t(E_G)$ are identical by Theorem 5.7. Thus, the projection onto $E \setminus E_G$ does not change the $L^2$-norm. Combining this observation with (5.19), we obtain

$$\|\mathbb{P}_{x_0} [X_{\tau_m} \in \cdot] - \mu\|_{L^2(\mu)} \leq \sup_{\mathcal{H}_G} \|\mathbb{P}_{x_0} [X_{\tau_m} (E \setminus E_G) \in \cdot | \mathcal{H}_G] - \mathbb{P}_\mu [Y_{\tau_m} (E \setminus E_G) \in \cdot | \mathcal{H}_G]\|_{L^2(\mu_{p,q}^n(\cdot | \mathcal{H}_G))}. \hfill (5.20)$$

Now by Lemma 5.14, for $m \geq C \log n$ where $C = C(p)$ is large enough,

$$\mathbb{P} [\mathcal{H}_\emptyset = \emptyset | \mathcal{H}_G] \geq \frac{1}{2}.$$  \hfill (5.21)

Then, for all $Z \subset \{0, 1\}^{E \setminus E_G}$, we can deduce that,

$$\mathbb{P}_\mu [Y_{\tau_m} (E \setminus E_G) = Z | \mathcal{H}_G] \geq \mathbb{P} [\mathcal{H}_\emptyset = \emptyset | \mathcal{H}_G] \nu_{E \setminus E_G}(Z) \geq \frac{1}{2} \nu_{E \setminus E_G}(Z).$$  \hfill (5.21)

Note that the first inequality follows from the fact that the distribution on $E_\emptyset$ is $\nu_{E_\emptyset}$, and that under $\mathcal{H}_\emptyset = \emptyset$, we have $E \setminus E_\emptyset = E_\emptyset$. We are now able to complete the proof of lemma by combining (5.20), (5.21), and the definition of $L^2$-norm. \hfill $\square$

Thus the task has now been reduced to measuring the $L^2$-distance of certain measures to the product measure $\nu$. The Miller-Peres inequality establishes a simple yet extremely useful bound for such cases. It first appeared in [23] where the product measure was given by independent Ber$(1/2)$s. This was extended later in [21, Lemma 4.3] which is the version we will use.

**Lemma 5.16.** Let $\Omega = \{0, 1\}^S$ for a finite set $S$, and let $\eta$ be a probability measure on the space of subsets of $S$. For each $R \subset S$, suppose that a probability measure $\varphi_R$ on $\{0, 1\}^R$ is given. For $p \in (0, 1/2)$, denote by $\nu_p$ the measure on $\{0, 1\}^S$ given by product of independent Ber$(p)$ variables.
Let $\mu_p$ be a measure on $\Omega$ obtained first by sampling a subset $R$ of $S$ according to $\eta$, and then sampling an element of $\{0,1\}^R$ according to $\varphi_p$, and sampling an element of $\{0,1\}^{S\backslash R}$ according to the restriction of $\nu_p$ on $\{0,1\}^{S\backslash R}$. Then, we have

$$\|\mu_p - \nu_p\|_{L^2(\nu_p)}^2 \leq \mathbb{E}\left[\mathbb{P}^{-|R\cap R'|} \right] - 1,$$

where $R, R' \subset S$ are two independent samples of $\eta$.

In view of Lemmas 5.15 and 5.16, we obtain that

$$\mathbb{P}_{x_0} \left[X_{\tau_m} \in \cdot\right] - \mu \|_{L^2(\mu)} \leq 2 \sup_{\mathcal{H}_G} \mathbb{E}\left[\frac{1}{\mathbb{P}^E_G \cap E_{R'}} \right],$$

provided that $p$ is small enough so that $\mathbb{P} < 1/2$, for all $m > C_1 \log n$ where $C_1$ is the constant in Lemma 5.15 and $E_R$ and $E_{R'}$, are two independent samples of the set $E_R$ of red clusters (see (5.5)) conditioned on $\mathcal{H}_G$. To analyze the right-hand side of (5.22), we recall the following domination results from [20, 21]. Let $\{J_A : A \subset E\}$ be a family of independent indicators such that $\mathbb{P}(J_A = 1) = \mathbb{P}_A$ for all $A \subset E$ and similarly let $\{J_{A,A'} : A, A' \subset E\}$ be a family of independent indicators such that $\mathbb{P}(J_{A,A'} = 1) = \mathbb{P}_A \mathbb{P}_{A'}$ for all $A, A' \subset E$.

**Lemma 5.17** ([20], Lemma 2.3, Corollary 2.4). Then following coupling results hold.

1. The conditional distribution of red clusters given $\mathcal{H}_G$ can be coupled to $J_A$ such that

$$\{A : A \subset C_R \} \subset \{A : J_A = 1\}.$$

2. Similarly, the conditional distribution of $(E_R, E_{R'})$ given $\mathcal{H}_G$ can be coupled such that

$$|E_R \cap E_{R'}| \leq \sum_{A \cap A' \neq \emptyset} |A \cup A'| J_{A,A'}.$$

We are now ready to prove Proposition 5.3.

**Proof of Proposition 5.3.** It suffices to prove that the right-hand side of (5.22) is bounded by $2$ for $m = C \log n$ with large enough $C$. Write $\kappa := \log \frac{1}{\mathbb{P}} > 0$. By part (2) of Lemma 5.17, we have

$$\sup_{\mathcal{H}_G} \mathbb{E}\left[\mathbb{P}^{-|E_R \cap E_{R'}|} \right] \leq \mathbb{E}\left[\exp \left\{ \kappa \sum_{A \cap A' \neq \emptyset} |A \cup A'| J_{A,A'} \right\} \right]$$

$$= \prod_{A \cap A' \neq \emptyset} \mathbb{E}\left[\exp \left\{ \kappa |A \cup A'| J_{A,A'} \right\} \right]$$

$$\leq \prod_{e \in E \cap (A, A') : e \in A, e \in A'} \left[ (e^{\kappa(|A| + |A'|)} - 1) \mathbb{P}_A \mathbb{P}_{A'} + 1 \right]$$

$$\leq \exp \left\{ |E| \left( \sum_{A \in E} e^{\kappa |A|} \mathbb{P}_A \right)^2 \right\},$$

where $e$ in the last line is an arbitrary edge in $E$. The last inequality follows from $x + 1 \leq e^x$ and the translation invariance of the underlying graph. Hence, it suffice to show that

$$\sum_{A \in E} e^{\kappa |A|} \mathbb{P}_A \leq \frac{1}{n^3}$$

for $m = C \log n$ with sufficiently large $C$. To this end, we recall Proposition 5.13 so that

$$\sum_{A \in E} e^{\kappa |A|} \mathbb{P}_A \leq C e^{-\alpha \tau_m} \sum_{A \in E} e^{\kappa |A| - \theta |\text{Conn}(A)|} \leq C e^{-\alpha \tau_m} \sum_{A \in E} e^{(\kappa - \theta) |\text{Conn}(A)|}.$$

Thus, we can proceed as in (5.16) and (5.17) to deduce that the last summation bounded by is bounded by $1$, provided that $\theta$ is large enough. This finishes the proof.  \[\square\]
5.3. Proof of Proposition 5.13: domination by subcritical branching processes. We now prove Proposition 5.13 to complete our discussion on Theorem 5.1. For \( S \subset E \), define \( C^*_R(S) \) to be the collection of red clusters that arises when exposing the joint histories of \( S \) i.e., \( H_S \) only. Similarly define \( C^*_B(S) \) for blue clusters.

**Lemma 5.18.** There exists \( c = c(p) > 0 \) such that, for all \( A \subset E \) we have

\[
P_A \leq e^{-c|\text{Conn}(A)|} \Pr \left[ A \in C^*_R(A) \right].
\]

(5.23)

To prove the above we will first attempt to understand the effect of conditioning on the event \( H_A = X \in \mathcal{H}_{\text{com}}(A) \). We will determine a subset of \( \text{Upd}[0, \tau_m] \) that is enough to determine the event \( \{ H_A = X \} \). We write \( \mathcal{X}_i = \mathcal{X} \cap \{ t = \tau_i \} \) for \( i \in \{ n/2 : n \in \mathbb{Z} \} \). Recall from (5.4) that the event \( \{ H_A = X \} \) is non-empty only when \( \mathcal{X} \) satisfies the consistency condition

\[
\mathcal{X}_{i+1/2} = \mathcal{X}_i \cup \mathcal{X}_{i+1} \text{ for all } i \in [0, m - 1] .
\]

(5.24)

For each \( i \in [1, m - 1] \) and \( e \in E_n \), we define

\[
\mathcal{U}_i(e) = \begin{cases} 
\text{Upd}[\tau_{i-1}, \tau_{i+1}](e) & \text{if } e \in \mathcal{X}_i, \\
\text{Upd}[\tau_i, \tau_{i+1}](e) & \text{if } e \in \mathcal{X}_{i+1} \setminus \mathcal{X}_i, \\
\emptyset & \text{otherwise .}
\end{cases}
\]

(5.25)

Then, we define

\[
\mathcal{U}_i = \bigcup_{e \in E_n} \mathcal{U}_i(e) \quad \text{and} \quad \mathcal{U} = \bigcup_{i=1}^{m-1} \mathcal{U}_i .
\]

Note that \( \mathcal{U} \) depends on \( \mathcal{X} \).

**Lemma 5.19.** The event \( \{ H_A = X \} \) is independent of the update variables not in \( \mathcal{U} \).

**Proof.** Write \( \mathcal{Y}_i = H_A \cap \{ t = \tau_i \} \) and define the event \( \mathcal{E}_i \) by

\[
\mathcal{E}_i = \{ \mathcal{Y}_i = \mathcal{X}_i \} .
\]

If \( \mathcal{X} \) satisfies the condition (5.24), we can write

\[
\{ H_A = \mathcal{X} \} = \bigcap_{i=1}^m \mathcal{E}_i .
\]

We claim that given \( \mathcal{E}_{i+1} \), the event \( \mathcal{E}_i \) depends only on the events in \( \mathcal{U}_i \). Given \( \mathcal{E}_{i+1} \), we decompose \( \mathcal{Y}_{i+1} = \mathcal{X}_{i+1} \) as following (similar to those in Theorem 5.7):

\[
\mathcal{Y}_{i+1}^{\text{NU}} = \{ e \in \mathcal{Y}_{i+1} : \mathcal{N}_i(e) = 1 \} , \\
\mathcal{Y}_{i+1}^{\text{Ob}} = \{ e \in \mathcal{Y}_{i+1} : \mathcal{N}_i(e) = 0 \text{ and the last update for } e \text{ in } [\tau_i, \tau_{i+1}] \text{ is oblivious} \} , \\
\mathcal{Y}_{i+1}^{\text{NOb}} = \{ e \in \mathcal{Y}_{i+1} : \mathcal{N}_i(e) = 1 \text{ and the last update for } e \text{ in } [\tau_i, \tau_{i+1}] \text{ is non-oblivious} \} .
\]

This classification can be carried out if we only know

\[
\bigcup_{e \in \mathcal{Y}_{i+1}} \text{Upd}[\tau_i, \tau_{i+1}](e) \subset \mathcal{U}_i .
\]

Now we suppose that this classification is given. Then, we have

\[
\mathcal{Y}_i = \mathcal{Y}_{i+1}^{\text{NU}} \cup \bigcup_{e \in \mathcal{Y}_{i+1}^{\text{NOb}}} \text{Conn}(e; \Xi_{i-1} \cup \Xi_i) ,
\]
and therefore $X_i = Y_i$ holds if

$$X_i \setminus Y_{i+1}^\text{NU} \subset \bigcup_{e \in Y_{i+1}^\text{NOb}} \text{Conn}(e; \Xi_{i-1} \cup \Xi_i) \subset X_i.$$ 

This event can be determined by knowing $\bigcup_{e \in X_i} U_i(e)$. This finishes the proof. \qed

**Lemma 5.20.** For all $X$ satisfying (5.24), it holds that

$$P\left[ A \in C^*_R(A), \mathcal{H}_A \cap X = \emptyset \right] = P\left[ A \in C^*_R(A), \mathcal{H}_A \cap X = \emptyset \right].$$

**Proof.** We keep the notation from the previous lemma. In view of the previous lemma, it suffices to demonstrate that the event

$$E = \{ A \in C^*_R(A) \} \cap \{ \mathcal{H}_A \cap X = \emptyset \}$$

does not depend on the updates in $U$. Note that this event is the same as saying $\mathcal{H}_A$ reaches $t = \tau_j$ without touching $X$. We prove this by induction (see Figure 5.2 for an illustration). Write $W_i(A) = \mathcal{H}_A \cap \{ t = \tau_i \}$ and for each $i \in [2, m]$, define the event $E_i$ as

$$E_i = \{ W_{i-1}(A) \neq \emptyset \text{ and } W_{i-1}(A) \cap (X_{i-3/2} \cup X_{i-1/2}) = \emptyset \}.$$ 

![Figure 5.2](image-url)

**Figure 5.2.** Illustrating the proofs of Lemmas 5.19 and 5.20. The purple graph is $X$ and $U$ is the set of updates in the purple region. The red graph is the history diagram $\mathcal{H}_A$. At time $t = \tau_k$, the occurrence of the event $E_k$ does not depend on the updates in the purple region. Note that for the latter event to occur $e_3$, $e_4$, $e_5$ cannot hit the purple region and hence the last updates for each of them in $(\tau_{k-1}, \tau_k]$ should be oblivious. This depends on the updates in the red box. For $e_1$ and $e_2$, they can be expanded and one of them must be to ensure that they all together form a red cluster. However this expansion should be confined to $B$. This can be determined by the updates in yellow region and therefore also independent of updates in the purple region.

We suppose that $X$ satisfies $X_m \cap A = \emptyset$ and $X_{m-1/2} \cap A = \emptyset$ since otherwise the event $\{ \mathcal{H}_A \cap X = \emptyset \}$ (and hence $E$) cannot happen. Under this minimal consistency assumption, we can write $E = \bigcap_{i=2}^m E_i$.

We now claim that, for each $k \in [2, m - 1]$, given $\bigcap_{i=k+1}^m E_i$, the event $E_k$ does not depend on updates in $U$. For each $e \in W_k(A)$, we consider two cases:
such that, for any \( \tau_{k-1}, \tau_k \) and \( \Delta \) appeared in the definition of the \( \Delta \) not intersect this implies that condition (5.15) by hypothesis, an event which implies the event in the denominator is the following:

\[
\text{Conn}(e; \Xi_{k-1} \cup \Xi_k) \cap (\mathcal{X}_{k-3/2} \cup \mathcal{X}_{k-1/2}) = \emptyset.
\]

Determined whether this holds or not can be performed by looking only at

\[
\bigcup_{e' \notin \mathcal{X}_{k-3/2} \cup \mathcal{X}_{k-1/2}} \text{Upd}[\tau_{k-2}, \tau_k](e').
\]

These updates are disjoint to \( \mathcal{U} \) since \( e' \notin \mathcal{X}_{k-3/2} \cup \mathcal{X}_{k-1/2} \) implies \( \text{Upd}[\tau_{k-2}, \tau_k](e') \cap \mathcal{U} = \emptyset \) Furthermore, the non-emptiness of \( W_{k-1}(A) \) implies that at least one of the last updates of \( e \in W_k \) is non-oblivious or there is an edge \( e \in W_k \) such that there is no update in \( (\tau_{k-1}, \tau_k) \). By the same reasoning as (1), this is independent of the updates in \( \mathcal{U} \).

Summing up, for the event \( \mathcal{E}_i \) to occur, all the events described above must occur simultaneously and the probability of this is independent of the conditioning on the randomness in \( \mathcal{U} \).

Proof of Lemma 5.18. Given the above preparation the remaining steps of the proof already appears in [19, 25]. Note first that, conditioned on \( \mathcal{H}_A^- = \mathcal{X} \in \mathcal{H}_{\text{com}}(A) \), one has \( \{A \in \mathcal{C}_R \} = \{A \in \mathcal{C}_{R(A)}^* \} \cap \{ \mathcal{H}_A \cap \mathcal{X} = \emptyset \} \) and similarly \( \{A \subset \mathcal{C}_B \} = \{A \subset \mathcal{C}_{B(A)}^* \} \cap \{ \mathcal{H}_A \cap \mathcal{X} = \emptyset \}. \) Therefore, we can deduce

\[
\mathbb{P} \left[ \begin{array}{c}
A \in \mathcal{C}_R \\
\mathcal{H}_A^- = \mathcal{X}, \{A \in \mathcal{C}_R \} \cup \{A \subset \mathcal{C}_B \}
\end{array} \right] = \mathbb{P} \left[ \begin{array}{c}
A \in \mathcal{C}_{R(A)}^*, \mathcal{H}_A \cap \mathcal{X} = \emptyset \\
\mathcal{H}_A^- = \mathcal{X}, \{A \in \mathcal{C}_R \} \cup \{A \subset \mathcal{C}_B \}
\end{array} \right] = \mathbb{P} \left[ \begin{array}{c}
A \in \mathcal{C}_{R(A)}^*, \mathcal{H}_A \cap \mathcal{X} = \emptyset \\
\mathcal{H}_A^- = \mathcal{X}
\end{array} \right] \leq \frac{\mathbb{P} \left[ \begin{array}{c}
A \subset \mathcal{C}_{B(A)}^*, \mathcal{H}_A \cap \mathcal{X} = \emptyset \\
\mathcal{H}_A^- = \mathcal{X}
\end{array} \right]}{\mathbb{P} \left[ \begin{array}{c}
A \subset \mathcal{C}_{R(A)}^* \\
\mathcal{H}_A \cap \mathcal{X} = \emptyset \\
\mathcal{H}_A^- = \mathcal{X}
\end{array} \right]} \quad (\text{by Lemma 5.20})
\]

Now we bound the denominator of (5.26) from below. Since \( \mathcal{X} \) satisfies the compatibility condition (5.15) by hypothesis, an event which implies the event in the denominator is the following: all the edges in \( A \) are updated in the time interval \( [\tau_{m-1/2}, \tau_m] \) with oblivious updates. Note that this implies that \( \mathcal{H}_A \), the history diagram of \( A \), will only intersect \( E \times \{ \tau_{m-1/2}, \tau_m \} \) and hence will not intersect \( \mathcal{X} \). Now the probability of an edge being updated in \( [\tau_{m-1/2}, \tau_m] \) is \( 1 - e^{-\frac{\Delta}{2}} \) where \( \Delta \) appeared in the definition of the \( \tau_i \)'s. Moreover the probability of an update being oblivious is \( 1 - p + p^* \). Putting the above together, we get that the denominator of 5.26 is bounded below by \( e^{-c(p)|A|} \) for some \( c(p) > 0 \). This completes the proof of (5.23).

Proposition 5.21. For any \( \theta > 0 \), we can find two constants \( C = C(\theta) > 0 \) and \( p_0 = p_0(\theta) > 0 \) such that, for any \( p \in (0, p_0) \) there exists a constant \( \alpha = \alpha(p) > 0 \) satisfying

\[
\mathbb{P} \left[ A \in \mathcal{C}_{R(A)}^* \right] \leq Ce^{-\theta(C(\theta)|\text{Conn}(A)| + \alpha \tau_m)} \quad \text{for all } A \subset E.
\]
5.3.1. Domination by sub-critical branching process. To estimate the probability $\mathbb{P}[A \in C_{\mathcal{R}(A)}]$, we fix $A$ and $m$ in the remaining part of the current section. Recall the notation $W_i$ from (5.7). The main idea of the proof is that for sufficiently small $p$, the sequence $W_m, W_{m-1}, \ldots, W_1$ is dominated by a subcritical branching process in a suitable sense that will be explained below. Note that the event \{ $A \in C_{\mathcal{R}(A)}$ \} requires that

1. $\mathcal{H}_e$ for some $e \in A$ starting at time $t = \tau_m$ survives to time $t = \tau_1$.
2. All the history diagrams $\mathcal{H}_e$, $e \in A$, are connected together before arriving at $t = \tau_1$.

Comparing them with sub-critical branching processes will allow us to bound the probabilities of the above events. As Lemma 5.23 and the discussion following that will show, the analysis has to take into account that the dependence across time of the Bernoulli percolation clusters used to define the Information percolation history diagrams prevents a contraction every time step. Nonetheless this is sufficient to yield subcritical behavior once every two steps which is enough for our purposes.

We start with a general lemma. For $r \in (0, p_{\text{perc}}(d))$, let $\omega_r$ be an i.i.d. standard bond percolation configuration on the lattice $(\mathbb{Z}^d, E(\mathbb{Z}^d))$ where each edge is open with probability $r$. Denote by $\text{Conn}(e; \omega_r)$ the closure of the open cluster containing an edge $e$ as in (5.3), and let $m_r$ be the distribution of $\text{Conn}(e; \omega_r)$, i.e.,

$$m_r(k) = \mathbb{P}[|\text{Conn}(e; \omega_r)| = k] ; k \in \mathbb{Z}_+ .$$

It is well-known (see [7, 16]) that there exists a constant $\rho(r) > 0$ such that, for all $e \in E(\mathbb{Z}^d)$,

$$m_r(k) \leq e^{-\rho(r)k} \text{ for all } k \geq 1 .$$

**Lemma 5.22.** Fix a non-empty set $A \subset E$ and consider a random configuration $X \in \Omega_n$ whose distribution is stochastically dominated by $\text{Perc}_n(r)$ for some $r \in (0, p_{\text{perc}}(d))$.

Given $X$, we define

$$A(X) = \bigcup_{e \in A} \text{Conn}(e; X) .$$

Let $(y_i)_{i=1}^\infty$ be a sequence of i.i.d. random variables in $\mathbb{Z}_+$ distributed according to $m_r$. Then, $|A(X)|$ is stochastically dominated by $y_1 + y_2 + \cdots + y_{|A|}$.

**Proof.** Take an arbitrary enumeration $A = \{e_1, e_2, \ldots, e_{|A|}\}$ and define disjoint sets $G_1, G_2, \ldots, G_{|A|}$ as $G_1 = \text{Conn}(e_1; X)$ and

$$G_k = \text{Conn}(e_k; X) \setminus \left( \bigcup_{i=1}^{k-1} \text{Conn}(e_i; X) \right) ; k \in [2, |A|] .$$

Then, the set $A(X)$ can be represented as the disjoint union of $G_1, G_2, \ldots, G_{|A|}$ and thus

$$|A(X)| = \sum_{i=1}^{|A|} |G_i| .$$

We now claim that

$$\sum_{i=1}^k |G_i| \preceq \sum_{i=1}^k y_i \text{ for all } k \in [1, |A|] .$$

Clearly this is true for $k = 1$. To finish the proof by the induction, it suffices to prove that the distribution of $|G_{i+1}|$ given $G_1, \ldots, G_i$, is stochastically dominated by $y_{i+1}$. This follows by the spatial independence of bond percolation. More precisely, given $G_1, \ldots, G_i$, the edge configuration on $(G_1 \cup \cdots \cup G_i)^c$ is a Bernoulli percolation with the same parameter, and thus the distribution of $G_{i+1}$ is dominated by that of $\text{Conn}(e_{i+1}; X)$. By (5.27) and the fact that $X$ is dominated by $\text{Perc}_n(r)$, the size of the latter is dominated by the distribution $m_r(\cdot)$ and the proof is completed. \qed
Recalling (5.7), let
\[ a_i := |W_i| ; i \in [0, m - 1] , \]
and let
\[ p_1 := 2p^{1/2} . \]  
(5.29)
A direct application of Lemma 5.22 is the following bound on \( a_{m-1} \) which along with the fact that \( \mathbb{P}(y_i = 0) = 1 - o_p(1) \) (by (5.28)), shows that the information percolation history diagram exhibits a contraction from \( W_m = A \) to \( W_{m-1} \) similar to a subcritical branching process.

**Lemma 5.23.** Suppose that \( p_1 < \perc(d) \) and let \( (y_i)_{i=1}^{\infty} \) be a sequence of i.i.d. random variables with distribution \( m_{2p_1} \). Then, we have
\[ a_{m-1} \leq y_1 + \cdots + y_{|A|} . \]

**Proof.** We apply Lemma 5.22 with \( X = \Xi_{m-1} \cup \Xi_{m-2} \). By Proposition 4.7 and union bound, the distribution of \( X \) is stochastically dominated by \( \perc_n(2p_1) \). We note that in the construction of the history diagram of an edge \( e \), we have to expand \( e \) to a subset of \( \overline{\text{Conn}}(e; X) \). In particular, when the update is oblivious, this subset is just empty set, and for the non-update situation this subset is \( \{e\} \) which is a subset of \( \overline{\text{Conn}}(e; X) \) since in this case \( e \) is open in \( \mathcal{N}_{m-1} \subseteq \Xi_{m-1} \subseteq X \). Hence, we can conclude that \( W_{m-1} \subset A(X) \). The assertion of the lemma is now immediate from Lemma 5.22. \( \square \)

We now state the main result regarding the domination by branching process.

**Proposition 5.24.** Suppose that \( p \) is small enough so that \( 3p_1 < \perc(d) \). Let \( (y_i)_{i=1}^{\infty} \) be a sequence of i.i.d. random variables with distribution \( m_{3p_1} \) defined in (5.27). For all \( i \in [1, m - 2] \), the distribution of \( a_i \) given \( (\mathcal{F}_{i+2}, W_{i+2}) \) is stochastically dominated by
\[ y_1 + y_2 + \cdots + y_{a_{i+2}} . \]

One might expect that the proof of Proposition 5.24 can be carried out similarly as that of Lemma 5.23. However this does not work, roughly because of the following: Assume that we condition on \( W_{i+1} \) and try to control \( a_i = |W_i| \). Then, \( W_i \) is determined by \( W_{i+1} \), the environment \( \Xi_i \cup \Xi_{i-1} \), and the update sequence in \( [\tau_i, \tau_{i+1}] \). However, by the same reasoning, \( W_{i+1} \) is determined from \( W_{i+2} \), \( \Xi_{i+1} \cup \Xi_i \) and the update sequence in \( [\tau_{i+1}, \tau_{i+2}] \), and thus \( W_{i+1} \) already contains some information on \( \Xi_i \). Therefore, the distribution of \( \Xi_i \) given \( W_i \) is hard to analyze. In particular, in the worst case, if all the edges in \( W_{i+1} \) belong to \( \Xi_i \), one cannot expect a contraction estimate of \( a_i \) in terms of \( a_{i+1} \) described in the previous lemma. However, at this point one notices that \( \Xi_{i-1} \) and \( \Xi_{i-2} \) are still independent of \( W_{i+1} \) and hence one can possibly obtain a bound for \( a_{i-1} \) instead. In other words, if we conditioned on \( W_{i+2} \) and all the relevant information prior to it, the distribution of \( a_i \), instead of \( a_{i+1} \), can be dominated in an appropriate manner.

This is done through the next result whose proof crucially uses the definitions listed in Table 1. Recall the notations \( C_j \) and \( N_j \) from (5.9).

**Proposition 5.25.** For \( i \in [1, m - 2] \), define \( \theta_i \in \Omega_n \) as following:
\[ \theta_i(e) = \begin{cases} \Xi_{i-1} \cup \Xi_i(e) & \text{if } e \in C_{i+2} , \\ \Xi_{i-1} \cup \Xi_i \cup \Xi_{i+1}(e) & \text{if } e \in E \setminus C_{i+2} . \end{cases} \]
Define
\[ Z_i = \bigcup_{e \in W_{i+2}} \overline{\text{Conn}}(e; \theta_i) . \]  
(5.30)
Then, it holds that \( W_i \subset Z_i \).
The proof of this proposition is based on two geometric lemmas (Lemmas 5.26 and 5.27). We refer to Figure 5.3 for the illustration of the proofs of these two lemmas and Proposition 5.25. However before proving the latter we first finish the proof of Proposition 5.24.

Figure 5.3. Figure illustrating the proof of Lemmas 5.26, 5.27, and Proposition 5.25. The crucial fact is that \( e' \in \text{Conn}(e_0; \theta_i) \) (without bar).

For \( i \in [1, m] \), denote by \( \mathcal{F}_i \) the \( \sigma \)-algebra on \( \Omega_n \) generated by update sequence \( \text{Upd}[\tau_i, \tau_m] \) (Hence \( \mathcal{F}_m = \{\emptyset, \Omega_n\} \)). Note that \( W_i \) is determined by \( \text{Upd}[\tau_i, \tau_m] \), and hence \( a_i \) is a random variable measurable with respect to \( \mathcal{F}_i \).

Proof of Proposition 5.24. We consider the following configuration

\[
\Xi_{i+1}(e) = \begin{cases} 
0 & \text{if } e \in C_{i+2}, \\
\Xi_i(e) & \text{if } e \in E \setminus C_{i+2}.
\end{cases}
\]

We first make the following claim.

Claim. Given \( (\mathcal{F}_{i+2}, W_{i+2}) \) the distribution of \( \Xi_{i+1}^o \) is dominated by \( \text{Perc}_n(p_1) \).

Assuming this claim, since \( \theta_i = \Xi_{i-1} \cup \Xi_i \cup \Xi_{i+1}^o \), by Proposition 4.8, it follows that the distribution of \( \theta_i \) given \( (\mathcal{F}_{i+2}, W_{i+2}) \) is stochastically dominated by \( \text{Perc}_n(3p_1) \). Hence, by Lemma 5.22 and the definition (5.30) of \( Z_i \), we can conclude that \( |Z_i| \) is stochastically bounded above by \( y_1 + \cdots + y_{a_{i+2}} \). Thus we are done by Proposition 5.25.

Now we prove the claim. A notable observation is that, if \( e \notin C_{i+2} \), the fact whether \( e \) belongs to \( W_{i+2} \) or not does not affect \( \text{Upd}[\tau_{i+1}, \tau_{i+2}](e) \). Therefore, the distribution of \( \Xi_{i+1}(E \setminus C_{i+2}) \) given \( (\mathcal{F}_{i+2}, W_{i+2}) \) is stochastically bounded by percolation on \( E \setminus C_{i+2} \) with open probability \( p_1 \), by Proposition 4.7 and the definition of \( p_1 \) in (5.29). Since \( \Xi_{i+1}^o(e) = 0 \) for \( e \in C_{i+2} \), the claim holds.

Lemma 5.26. For \( i \in [1, m-1] \), we have that

\[
W_i \subset \bigcup_{e \in W_{i+1}^{\text{NU}}} \overline{\text{Conn}(e; \Xi_{i-1} \cup \Xi_i)}.
\]

Proof. In view of decomposition (5.12) and the first inclusion of Lemma 5.9, it suffices to check that

\[
N_i \subset \bigcup_{e \in W_{i+1}^{\text{NU}}} \overline{\text{Conn}(e; \Xi_{i-1} \cup \Xi_i)}.
\]

If \( e \in W_{i+1}^{\text{NU}} \), we have \( \mathcal{N}_i(e) = 1 \) by the definition of \( \mathcal{N}_i \), and thus \( (\Xi_i \cup \Xi_{i-1})(e) = 1 \) since \( \mathcal{N}_i \leq \Xi_i \). Therefore, we have \( e \in \overline{\text{Conn}(e; \Xi_i \cup \Xi_{i-1})} \). Hence, the right-hand side of (5.32) contains \( W_{i+1}^{\text{NU}} \), and hence contains \( N_i \) by the second inclusion of Lemma 5.9. \( \square \)
Lemma 5.27. For \( i \in [1, m-2] \), define \( \xi_i \in \Omega_n \) as following:

\[
\xi_i(e) = \begin{cases} 
\Xi_i(e) & \text{if } e \in C_{i+2}, \\
(\Xi_i \cup \Xi_{i+1})(e) & \text{if } e \in E \setminus C_{i+2}.
\end{cases}
\]

Then, we have

\[
W_{i+1} \setminus W_{i+2} \subseteq \bigcup_{e \in W_{i+2}} \overline{\text{Comm}}(e; \xi_i). 
\]

Proof. For \( e' \in W_{i+1} \setminus W_{i+2} \), we know from Lemma 5.26 that there exists \( e_0 \in W_{i+2} \) and a path \( e_0, e_1, \cdots, e_k (= e') \) in \( E \) such that \( (\Xi_i \cup \Xi_{i+1})(e_l) = 1 \) for all \( l \in [0, k-1] \). If none of \( e_0, e_1, \cdots, e_k \) belongs to \( C_{i+2} \) then the assertion of lemma is immediate since \( \Xi_i \cup \Xi_{i+1} = \xi_i \) along this path. Otherwise, let

\[
K = \max\{h : e_h \in C_{i+2}\}. 
\]

Since \( e' \notin W_{i+2} \), we have \( K < k \). Then, since \( e_{K+1} \notin C_{i+2} \), we have \( e_K \in \partial C_{i+2} \) and thus \( (\Xi_{i+1} \cup \Xi_{i+2})(e_K) = 0 \) by Lemma 5.10. Since \( (\Xi_i \cup \Xi_{i+1})(e_K) = 1 \), we can conclude that \( \Xi_i(e_K) = 1 \). This implies that \( e' = e_k \in \overline{\text{Comm}}(e_K; \xi_i) \), where \( e_K \in C_{i+2} \subseteq W_{i+2} \). This completes the proof. \( \square \)

Now we are ready to prove Proposition 5.25.

Proof of Proposition 5.25. Fix arbitrary \( e'' \in W_i \). It suffices to verify that \( e'' \in Z_i \). Since \( \Xi_{i-1} \cup \Xi_i \leq \theta_i \), by Lemma 5.26, there exists \( e' \in W_{i+1} \) such that

\[
e'' \in \overline{\text{Comm}}(e'; \Xi_{i-1} \cup \Xi_i) \subseteq \overline{\text{Comm}}(e'; \theta_i). \tag{5.33}
\]

If \( e' \in W_{i+2} \), we can immediately assert that \( e'' \in Z_i \) by the definition of \( Z_i \).

On the other hand, if \( e' \in W_{i+1} \setminus W_{i+2} \), then by Lemma 5.27 and by the fact that \( \xi_i \leq \theta_i \), there exists \( e_0 \in W_{i+2} \) such that

\[
e' \in \overline{\text{Comm}}(e_0; \xi_i) \subseteq \overline{\text{Comm}}(e_0; \theta_i). \tag{5.34}
\]

We remark that (5.33) implies that \( \theta_i(e') = 1 \) since otherwise \( \overline{\text{Comm}}(e'; \theta_i) = \emptyset \). Therefore we can replace \( e' \in \overline{\text{Comm}}(e_0; \theta_i) \) in (5.34) with \( e' \in \overline{\text{Comm}}(e_0; \theta_i) \). Combining this with (5.33) ensures that \( e'' \in \overline{\text{Comm}}(e_0, \theta_i) \subseteq Z_i \). This completes the proof. \( \square \)

5.3.2. Bounds on \( a_i \) based on domination by branching processes. Now we present two consequences of the previous branching process type estimate. These will play a fundamental role in the proof of Proposition 5.21. For a random variable \( y \) in \( \mathbb{Z}_+ \) following the law \( m_{3p_1} \) defined in (5.27), we define \( M = M(p) \) as the solution of

\[
e^{-2M} = \mathbb{E}(y). 
\]

It readily follows that

\[
\lim_{p \to 0} M = \infty.
\]

Lemma 5.28. For \( k \in [1, m] \), select \( \tau \in \{1, 2\} \) so that \( (k - \tau) \mod 2 = 0 \). Then, for some constant \( C = C(p) > 0 \), it holds that

\[
\mathbb{E}[a_\tau | a_k] \leq Ce^{-M_k} a_k. 
\]

Proof. It follows from Lemma 5.24 that, for all \( k \in [3, m] \),

\[
\mathbb{E}[a_{k-2} | a_k, F_k] \leq e^{-2M} a_k. \tag{5.35}
\]

Then, the proof of lemma is completed by the induction. \( \square \)
Lemma 5.29. For all sufficiently small \( p \), there exists \( c_0 = c_0(p) > 0 \) such that, for all \( c \in (0, c_0) \), we have
\[
\mathbb{E} \exp \left\{ c_0 \sum_{i=1}^{m-1} a_i \right\} \leq e^{\ell |A|}.
\]
Furthermore, \( \lim_{p \to 0} c_0(p) = +\infty \).

Proof. By the Cauchy-Schwarz inequality,
\[
\mathbb{E} \left[ \exp \left\{ c \sum_{i=1}^{m-1} a_i \right\} \right]^2 \leq \mathbb{E} \exp \left\{ 2c \sum_{i:2i \in [1,m-1]} a_{2i} \right\} \mathbb{E} \exp \left\{ 2c \sum_{i:2i+1 \in [1,m-1]} a_{2i+1} \right\}.
\] (5.36)

Denote by \( y \) the random variable with distribution \( m_{3p_4} \) defined in (5.27). Note that the following equation on \( c \)
\[
\mathbb{E} e^{(2c+1)y} = e
\] (5.37)
has a positive solution \( c_0(p) \) and we can readily check that \( \lim_{p \to 0} c_0(p) = +\infty \).

Now it suffices to prove that for all \( c \in (0, c_0(p)) \) and for all \( \ell \),
\[
\mathbb{E} \exp \left\{ 2c \sum_{i=1}^{\ell} a_{m-2i} \right\} \leq e^{\ell |A|} \quad \text{and} \quad \mathbb{E} \exp \left\{ 2c \sum_{i=0}^{\ell} a_{m-2i-1} \right\} \leq e^{\ell |A|}.
\] (5.38)

By Lemma 5.24 and (5.37), for all \( i \in [3, m] \), we have
\[
\mathbb{E} \left[ e^{(2c+1)a_{i-2}} \left| a_i, F_i \right. \right] \leq \mathbb{E} \left[ e^{(2c+1)y} \right]^{a_i} \leq e^{a_i}.
\]
Consequently, for all \( \ell \geq 1 \),
\[
\mathbb{E} \left[ e^{a_{m-2\ell}} \exp \left\{ 2c \sum_{i=1}^{\ell} a_{m-2i} \right\} \left| a_{m-2\ell+2}, F_{m-2\ell+2} \right. \right] \leq e^{a_{m-2\ell+2}} \exp \left\{ 2c \sum_{i=1}^{\ell-1} a_{m-2i} \right\}.
\]
Repeating this procedure, we obtain
\[
\mathbb{E} \exp \left\{ 2c \sum_{i=1}^{\ell} a_{m-2i} \right\} \leq \mathbb{E} \left[ e^{a_{2\ell}} \exp \left\{ 2c \sum_{i=1}^{\ell} a_{m-2i} \right\} \right] \leq \mathbb{E} \left[ e^{a_{2\ell+2}} \exp \left\{ 2c \sum_{i=1}^{\ell-1} a_{m-2i} \right\} \right] \leq \cdots \leq \mathbb{E} \left[ e^{(2c+1)a_{m-2}} \right] \leq e^{a_m} = e^{|A|}.
\] (5.39)

This proves the first inequality in (5.38). By a similar argument as above, one can show that
\[
\mathbb{E} \exp \left\{ 2c \sum_{i=0}^{\ell} a_{m-2i-1} \right\} \leq \mathbb{E} e^{(2c+1)a_{m-1}}.
\] (5.40)

By Lemma 5.23 and the fact that \( m_{2p_3} \) is dominated by \( m_{3p_4} \), we have
\[
\mathbb{E} e^{(2c+1)a_{m-1}} \leq \mathbb{E} \left[ e^{(2c+1)y} \right]^{a_m} \leq e^{a_m} \leq e^{|A|}.
\] (5.41)

where the last inequality follows from (5.37). Now, (5.38) is proven by combining (5.39), (5.40), and (5.41). □
5.3.3. **Proof of Proposition 5.21.** Write
\[ \sigma := \begin{cases} \max \{i : a_i = 1\} & \text{if } a_i = 1 \text{ for some } i \in [1, m] \\ 0 & \text{otherwise.} \end{cases} \]

Define two events \( \mathcal{A} \) and \( \mathcal{B} \) as
\[ \mathcal{A} = \{ \text{The history diagram starting from } A \text{ i.e., } \mathcal{H}_A \text{ is reduced to one point in } [\tau_\sigma, \tau_m] \}, \]
\[ \mathcal{B} = \{ \text{The history diagram starting from } W_\sigma \text{ at } t = \tau_\sigma \text{ survives until } t = \tau_1 \}. \]

Then, the event \( \{ A \in C^*_R(A) \} \) is a subset of \( \mathcal{A} \cap \mathcal{B} \) so that we have
\[ \mathbb{P} \left[ A \in C^*_R(A) \right] \leq \mathbb{P} \left[ \mathcal{A} \cap \mathcal{B} \right]. \tag{5.42} \]

If \( \sigma = 0 \), we define \( \mathcal{B} \) as the full event.

**Claim.** Conditioned on \( \sigma \), two events \( \mathcal{A} \) and \( \mathcal{B} \) are independent.

**Proof.** This claim is trivial if \( \sigma = 0 \). Now suppose that \( \sigma \) is given as an element of \([1, m]\). We write \( W_\sigma = \{ e \} \). Then, it suffices prove that the event \( \mathcal{A} \) is independent of \( \text{Upd}[0, \tau_\sigma] \) since \( \mathcal{B} \) depends only on \( \text{Upd}[0, \tau_\sigma] \) conditioned on \( \sigma \). Clearly the behavior of the history diagram starting from \( e \) in \((\tau_{\sigma+1}, \tau_m]\) is independent of \( \text{Upd}[0, \tau_\sigma] \). Hence, it only suffices to check the interval \((\tau_\sigma, \tau_{\sigma+1}]\). The event \( \mathcal{A} \) imposes that all the edges in \( W_{\sigma+1} \setminus \{ e \} \) exhibit the oblivious update in \((\tau_\sigma, \tau_{\sigma+1}]\), while \( e \in W_{\sigma+1} \) survives to \( \tau_\sigma \) without expanding to \( \text{Conn}(e; \varepsilon_{\sigma-1} \cup \varepsilon_\sigma) \). The first event is determined by \( \text{Upd}[\tau_\sigma, \tau_{\sigma+1}] \) and hence is independent of \( \text{Upd}[0, \tau_\sigma] \). The second event occurs only when there is no update at \( e \) in \( \text{Upd}[\tau_\sigma, \tau_{\sigma+1}] \), and hence this event is also independent of \( \text{Upd}[0, \tau_\sigma] \) as well and hence we are done. \( \square \)

By this claim and \( (5.42) \), we deduce that
\[ \mathbb{P} \left[ A \in C^*_R(A) \right] \leq \mathbb{E} \left[ \mathbb{P} \left[ \mathcal{A} \mid \sigma \right] \mathbb{P} \left[ \mathcal{B} \mid \sigma \right] \right]. \tag{5.43} \]

We now estimate \( \mathbb{P} \left[ \mathcal{B} \mid \sigma \right] \). Select \( r \in \{1, 2\} \) so that \((\sigma - r) \mod 2 = 0\). Then, by Lemma 5.28, we have
\[ \mathbb{P} \left[ \mathcal{B} \mid \sigma \right] \leq \mathbb{P} \left[ a_r > 0 \mid a_\sigma = 1 \right] \leq \mathbb{E} \left[ a_r \mid a_\sigma = 1 \right] \leq Ce^{-M_\sigma}. \tag{5.44} \]

By \( (5.43) \) and \( (5.44) \), we have
\[ \mathbb{P} \left[ A \in C^*_R(A) \right] \leq \mathbb{E} \left[ Ce^{-M_\sigma} 1_A \right]. \tag{5.45} \]

If \( |A| = 1 \) so that \( \sigma = m \), this inequality proves the assertion of the proposition. Now we assume that \( |A| \geq 2 \), so that \( \sigma \leq m - 2 \). Note that \( \sigma \) cannot be \( m - 1 \) since \( W_{m-1} \subseteq A \) under \( \mathcal{A} \) since otherwise for \( e \in A \setminus W_{m-1} \), the set \( \mathcal{H}_e \) is a singleton and hence \( \mathcal{H}_e \cap \mathcal{H}_A \setminus \{ e \} = \emptyset \). Thus conditioned on the event \( \{ \sigma = k \} \), the event \( \mathcal{A} \) implies that
\[ \{ a_{k+1} + \cdots + a_{m-1} \geq \text{Conn}(A) - 1 \}, \]
where, recall that \( \text{Conn}(A) \) is the smallest possible number of edges of the connected subgraph of \( (\Lambda_n, E_n) \) containing \( A \). We can neglect \( a_m \) since \( W_m = A \subseteq W_{m-1} \) under \( \mathcal{A} \), and we used the fact that \( a_k = 1 \). Therefore, we can bound the right-hand side of \( (5.45) \) from above by
\[ \sum_{k=1}^{m-2} Ce^{-M_k} \mathbb{E} \left[ 1 \{ a_{k+1} + \cdots + a_{m-1} \geq \text{Conn}(A) - 1 \} 1\{ \sigma = k \} \right] \]
\[ + \mathbb{E} \left[ 1 \{ a_1 + \cdots + a_{m-1} \geq \text{Conn}(A) \} 1\{ \sigma = 0 \} \right]. \]
By applying $1\{\sigma = k\} \leq 1\{a_{k+1}, \cdots, a_{m-1} \geq 2\}$ here, we obtain
\[
\mathbb{E}\left[C e^{-M\sigma} 1_A\right] \leq C \sum_{k=0}^{m+1} H_k.
\] (5.46)

where
\[
H_k = e^{-Mk} \mathbb{E}\left[1\{a_{k+1} + \cdots + a_{m-1} \geq |\text{Conn}(A)| - 1\} \ 1\{a_{k+1}, \cdots, a_{m-1} \geq 2\}\right].
\]

For any $C_1, C_2 > 0$, by the Chebyshev inequality,
\[
H_k \leq e^{-Mk} e^{-C_1(|\text{Conn}(A)| - 1)} e^{-2C_2(m-k)} \mathbb{E}_{\text{Conn}(A)}(C_1 + C_2)(a_{k+1} + \cdots + a_{m-1}).
\]

Now we take $C_1$ and $C_2$ such that $C_1 + C_2 < c_0$ where the constant $c_0$ is the one appeared in Lemma 5.29. Then, by Lemma 5.29 and the fact that $|A| \leq |\text{Conn}(A)|$, we can further obtain
\[
H_k \leq e^{-Mk} e^{-C_1(|\text{Conn}(A)| - 1)} e^{-2C_2(m-k)} e^{|\text{Conn}(A)|}.
\] (5.47)

For given $\theta > 0$, we first take $p$ small enough so that $c_0 > \theta + 2$ and $M > 1$. This is possible since
\[
\lim_{p \to 0} c_0 = \lim_{p \to 0} M = +\infty.
\]

Take $C_1 = \theta + 1$ and $C_2 = 1/2$. With this selection, the bound (5.47) becomes
\[
H_k \leq e^{\theta+1} e^{-(m+\theta|\text{Conn}(A)|)} e^{-(M-1)k}.
\]

Combining this with (5.46) yields
\[
\mathbb{E}\left[C e^{-M\sigma} 1_A\right] \leq C(\theta) e^{-(m+\theta|\text{Conn}(A)|)}.
\]

Thus the statement of the proposition follows by recalling that $\tau_m = m\Delta$. $\square$

6. REDUCTION TO A PRODUCT CHAIN

From now on, we define
\[
r = r(n) = 3 \log^5 n.
\]

Moreover recall $t_{\text{max}}$ from (4.10). Denote by $(X^\dagger_t)_{t \geq 0}$ the FK-dynamics defined on the periodic lattice $\mathbb{Z}_r^d$. Let $\Omega_r = \{0, 1\}^{E_r}$ where $E_r = E(\mathbb{Z}_r^d)$, and denote by $\pi^\dagger := \mu^\dagger_{p,q}$ the random-cluster measure on $\Omega_r = \{0, 1\}^{E_r}$. Let $\Lambda \subset E_r$ be a box of size $2 \log^5 n$. Then, define
\[
\mathbf{d}_t = \max_{x_0 \in \Omega_r} \left\| \mathbb{P}_{x_0} \left[X^\dagger_t(\Lambda) \in \cdot\right] - \pi^\dagger_{\Lambda} \right\|_{L^2(\pi^\dagger_{\Lambda})},
\] (6.1)

where $X^\dagger_t(\Lambda)$ represents the configuration of $X^\dagger_t$ on $\Lambda$, and $\pi^\dagger_{\Lambda}$ stands for the projection of $\pi^\dagger$ onto the set $\Lambda$. The main result of this section is the following theorem.

**Theorem 6.1.** For all sufficiently small $p$, there exists a constant $C_1 = C_1(p)$ such that the following hold.

1. For $s \in [C_1 \log \log n, t_{\text{max}}]$ and $t \in [0, t_{\text{max}}]$, it holds that
\[
\max_{\nu, \nu \leq \text{Per}_n(t_{\text{init}})} \left\| \mathbb{P}_{\nu} [X_{t+s} \in \cdot] - \mu^n_{p,q} \right\|_{TV} \leq \frac{1}{2} \left[ \exp \left( \frac{n^d}{\log^{12d} n} \mathbf{d}_s^2 \right) - 1 \right]^{\frac{1}{2}} + \frac{4}{n^{2d}}.
\]

2. If $t \geq C_1 \log \log n$ and
\[
\lim_{n \to \infty} \left( \frac{n}{\log^{10} n} \right)^d \mathbf{d}_t^2 = +\infty,
\]

then we have
\[
\liminf_{n \to \infty} \max_{\nu, \nu \leq \text{Per}_n(t_{\text{init}})} \left\| \mathbb{P}_{\nu} [X_t \in \cdot] - \mu^n_{p,q} \right\|_{TV} = 1.
\]
Remark 6.2. As the proof will reveal, part (1) of the theorem holds when the underlying lattice is \( \mathbb{Z}^d_m \) with \( m \in [\log^5 n, n] \), by replacing \( \frac{n^d}{\log 123 n} \) on the right-hand side by \( \frac{m^d}{\log 123 n} \).

Henceforth, the constant \( C_1 > 0 \) will always refer to the constant appeared in this theorem. The proof of this theorem will be presented in the remaining part of the current section. We shall assume that \( p \) is small enough so that all the results established in Sections 4 (including 4.4) and 5 are valid. As indicated in Section 2, following the strategy in [18] we will reduce the chain to an approximate product chain.

We start by giving a short roadmap of what the various subsections achieve.

- The first part (Section 6.1) constructs the so called Barrier dynamics where the FK-dynamics on \((\mathbb{Z}/n\mathbb{Z})^d\) gets compared to FK-dynamics on a disjoint collection of \((\mathbb{Z}/r\mathbb{Z})^d\) where \( r = \log 5 n \).
- We define the notion of Update support in Section 6.2. Informally, the goal is the following, to bound total variation distance \( d(t + s) \) where \( t = t_{\text{mix}} = O(\log n) \) and \( s = \log \log(n) \) we try to couple two configurations at time \( t + s \). At this point the key observation is that irrespective of the configuration at time \( t \) all but a sparse set of small boxes couple at time \( t + s \). The remainder is called the ‘Update support’ for reasons which will be clear later and hence the remaining task is to ensure that the time interval \([0, t]\) is sufficient to couple FK-dynamics starting from two arbitrary configurations to couple on the ‘Update support’.
- We prove Theorem 6.1 in Section 6.3.

6.1. Coupling with barrier-dynamics. Divide \( E_n \) into disjoint squares of size \( \log^5 n \) as follows. Let us write \( K = n/\log^5 n \) and assume that \( K \) and \( \log^5 n \) are integers for the simplification of notation. Define
\[
V_n = \{0, \log^5 n, 2 \log^5 n, \ldots, (K - 1) \log^5 n\}^d \subset \Lambda_n.
\]
For each \( v \in V_n \), we define an edge box \( B_v \) by
\[
B_v = \{(u, u + e_j) : u \in v + [0, \log^5 n - 1]^d \text{ and } j \in [1, d]\}, \quad (6.2)
\]
where \( e_j \) represents the \( j \)th standard normal vector in \( \mathbb{R}^d \). One can think of \( B_v \) as a box of size \( \log^5 n \) with some boundary edges are removed. Note that \((B_v)_{v \in V_n}\) is a decomposition of \( E_n \). Furthermore, we mention that all the boxes below of various sizes, are edge boxes and hence for brevity we will refer to them as boxes.

![Figure 6.1](image.png)

**Figure 6.1.** Figure illustrating the maps \( \mathcal{P}_v \).

Then, for each \( v \in V_n \), consider the expanded box \( B^+_v \subset \Lambda_n \) of \( B_v \) in the sense of (4.9). Then, \( B^+_v \) is a box of size \( \log^5 n + 2 \log^4 n \) which is concentric with \( B_v \). Let \( C^+_v \) be another square lattice
of size $\log^5 n + 2 \log^4 n$ and define a natural identification map $\mathcal{P}_v : B_v^+ \to C_v^+$. Define $C_v = \mathcal{P}_v(B_v)$ so that $(C_v, C_v^+)$ is a copy of $(B_v, B_v^+)$ (see Figure 6.1). We define
\[ \hat{E}_n = \bigcup_{v \in V_n} C_v^+ \] and $\hat{\Omega}_n = \{0, 1\}^{\hat{E}_n}$.

Note that the last union is a disjoint union.

**Definition 6.3** (Barrier-dynamics). For each $v \in V_n$, the barrier-dynamics is a FK-dynamics $X^v_t$ on $C_v^+$ coupled with $X_t$ by sharing the same update sequence via the following rules:

1. **(Initial condition)** The initial edge configuration on $C_v^+$ is identical to that of $B_v^+$ through $\mathcal{P}_v$. In other words, $X^v_0(\mathcal{P}_v(e)) = X_0(e)$ for all $e \in B_v$.
2. **(Dynamics)** We define the FK-dynamics $(X^v_t)_{t \geq 0}$ on $C_v^+$ with periodic boundary condition by using the update sequence of $B_v^+$. Formally stating, we perform updates for each $e \in C_v^+$ by using the update sequence $\text{Upd}(\mathcal{P}_v^{-1}(e))$ of the edge $\mathcal{P}_v^{-1}(e) \in B_v$.

For $t \geq 0$, we define a random map $\mathcal{G}_t : \Omega_n \to \Omega_n$ such that, for all $X_0 \in \Omega_n$,
\[ [\mathcal{G}_t(X_0)](e) = X^v_t(\mathcal{P}_v(e)) , \]
where $v \in V_n$ is the unique index such that $e \in B_v$. The next lemma now says that the actual dynamics and the barrier dynamics stay coupled for a significant amount of time provided the initial condition is sparse enough (note that for a spin system the latter condition is not needed since each update only depends on its immediate neighbors).

**Lemma 6.4.** Suppose that $p$ is small enough and the law of the initial condition $X_0$ follows the law $\nu$ such that $\nu \leq \text{Perc}_n(p_{\text{init}})$. Then, we have
\[ \mathbb{P}[X_t = \mathcal{G}_t(X_0) \text{ for all } t \in [0, t_{\text{max}}]] \geq 1 - n^{-2d} . \]

**Proof.** By Lemma 4.11 (cf. Remark 4.13), it holds that
\[ \mathbb{P}[X_t(B_t) = X^v_t(C_t) \text{ for all } t \in [0, t_{\text{max}}]] \geq 1 - n^{-3d} . \]
Thus, the conclusion of the lemma follows from the union bound since $|V_n| < n^d$. \qed

### 6.2. Sparsity of update support.

**Definition 6.5** (Update support). For each $s > 0$, denote by $U_s = \text{Upd}[0, s]$ the update sequence between time $[0, s]$. Then, the random map $\mathcal{G}_s$ is completely determined by $U_s$ and hence we can write $\mathcal{G}_s = g_{U_s}$ for some function $g_{U_s} : \Omega_n \to \Omega_n$. The update support of $U_s$ is the minimum subset $\Gamma_{U_s} \subset E_n$ such that $\mathcal{G}_s$ is a function of $X(\Gamma_{U_s})$ for all $X \in \Omega_n$, i.e.,
\[ g_{U_s}(X) = f_{U_s}(X(\Gamma_{U_s})) \]
for some $f_{U_s} : \{0, 1\}^{\Gamma_{U_s}} \to \Omega_n$.

**Lemma 6.6.** [18, Lemma 3.8] Fix $t \geq 0$ and let $U_s$ represent the update sequence for the time interval $[t, t + s]$ for $s \leq t_{\text{max}}$ where $t_{\text{max}}$ was defined in (4.10). Suppose that $p$ is small enough and a probability measure $\nu$ in $\Omega_n$ satisfies $\nu \leq \text{Perc}_n(t_{\text{init}})$. Then, we have
\[ \left\| \mathbb{P}_\nu[X_{t+s} \in \cdot] - \mu_{p, q}^n \right\|_{TV} \leq \int \left\| \mathbb{P}_\nu[X_t(\Gamma_{U_s}) \in \cdot] - \mu_{\Gamma_{U_s}} \right\|_{TV} d\mathbb{P}(U_s) + 2n^{-3d} , \]
where $\mu_{\Gamma_{U_s}}$ represents the projection of $\mu_{p, q}^n$ on $\Gamma_{U_s}$.

**Proof.** The proof in the above reference relies only on the coupling of $X_t$ and $\mathcal{G}_t(X_0)$ for $t \in [0, t_{\text{max}}]$. For our model this has been established in Lemma 6.4 based on the bound on disagreement percolation using the sparse initial conditions. \qed

Now we establish the sparsity of the update support $\Gamma_{U_s}$.
Definition 6.7 (Sparse set). A $S \in \Omega_n$ is called sparse if for some $K \leq n^d(\log n)^{-12d}$, the graph induced by $S$ can be decomposed into disjoint components $A_1, A_2, \ldots, A_K$ such that

1. For all distinct $i, j \in [1, K]$, there is no open path in $S$ connecting $A_i$ and $A_j$.
2. Every $A_i, i \in [1, K]$, has diameter at most $\log^5 n$. In particular, there is a box of size $2\log^5 n$ containing $A_i$.
3. The distance between any distinct $A_i$ and $A_j$ is at least $4\log^4 n$.

We write $\text{Spa}_n$ to denote the set of sparse configurations in $\Omega_n$.

Lemma 6.8. [18, Lemma 3.9] There exists $C_2 = C_2(p) > 0$ such that, for all $s \geq C_2 \log \log n$,

$$\mathbb{P}[\Gamma \in \text{Spa}_n] \geq 1 - n^{-3d}. \quad (6.3)$$

Proof. The only model-dependent part is the proof of the following fact: For $t \geq C_2 \log \log n$ with a large enough $C_2$,

$$\sum_{e \in C_v^+} \mathbb{P}\left[X_t^{v,\text{full}}(e) \neq X_t^{v,\text{empty}}(e)\right] \leq \log^{-10d} n, \quad (6.4)$$

where $(X_t^{v,\text{full}})_{t \geq 0}$ (resp. $(X_t^{v,\text{empty}})_{t \geq 0}$) is the FK-dynamics on periodic lattice $C_v^+$ with full (resp. empty) initial condition. The proof of this fact in our setting follows from Corollary 3.4 which indicates that, for some $C > 0$,

$$\sum_{e \in C_v^+} \mathbb{P}\left[X_t^{v,\text{full}}(e) \neq X_t^{v,\text{empty}}(e)\right] \leq |C_v^+|e^{-C \log \log n}.$$ 

Hence, the bound (6.4) follows if we take $C$ large enough. The remaining part is identical to cited proofs and will not be repeated here. \hfill \square

By Lemmas 6.6 and 6.8, we obtain the following result.

Proposition 6.9. Suppose that $p$ is small enough and $\nu$ is a probability distribution on $\Omega_n$ satisfying $\nu \preceq \text{Perc}_n(p_{\text{init}})$. Then, for all $s \in [C_0 \log \log n, t_{\text{max}}]$ where $C_0$ is the constant appearing in Lemma 6.8, there exists a measure $\mathbb{Q}$ on $\text{Spa}_n$ such that,

$$\|\mathbb{P}_\nu(X_{t+s} \in \cdot) - \mu^n_{p,q} \|_{TV} \leq \int_{\text{Spa}_n} \|\mathbb{P}_\nu(X_t(\Gamma) \in \cdot) - \mu_{\Gamma} \|_{TV} d\mathbb{Q}(\Gamma) + 3n^{-3d}.$$
6.3. **Proof of Theorem 6.1.** Before jumping into the proof we will need some technical preparation. The next few results use coupling arguments to compare the actual chain to a product chain. We start by defining a notion of good sets, and then introduce a generalized version of barrier dynamics.

**Definition 6.10.** A collection of disjoint subsets $A_1, A_2, \cdots, A_K$ of $\Omega_n$ are $m$-good for some $m \in [\log^4 n, (1/2)\log^5 n]$ if, each $A_i$ is contained in a box of size $2\log^5 n$, and the expanded sets $A_i^+, i \in [1, K]$, are disjoint where

$$A_i^+ = \{ e \in E_n : d(e, A_i) \leq m \}.$$

As a consequence of Lemma 6.8, the sets $A_i$ in the update support are $\log^4 n$–good (see Figure 6.2).

Let us take a box of size $r = 3\log^5 n$ containing $A_i^+$ and denote this box by $A_i^\dagger$ (this is possible since $m \leq (1/2)\log^5 n$). Take $K$ copies $L_1, L_2, \cdots, L_K$ of the periodic lattice $\mathbb{Z}_r^d$, and embed each $A_i^\dagger$ to $L_i$ by a identification map $\mathcal{P}_i : A_i^\dagger \rightarrow L_i$.

For $i \in [1, K]$, denote by $(Y_t^{(i)})_{t \geq 0}$, the FK-dynamics on $L_i$, whose update sequence and initial condition are inherited from that of $A_i^\dagger = \mathcal{P}_i^{-1}(L_i)$ of $(X_t(A_i^\dagger))_{t \geq 0}$. Let $\pi^{(i)}$ be the random-cluster measure on $L_i$ so that $\pi^{(i)}$ is the invariant measure of $Y_t^{(i)}$. Define the product spaces:

$$L = \prod_{i=1}^K L_i, \quad \pi = \prod_{i=1}^K \pi^{(i)}, \quad \text{and} \quad Y_t^* = \prod_{i=1}^K Y_t^{(i)},$$

and let

$$\Gamma = \bigcup_{i=1}^K A_i \subset \Omega_n \quad \text{and} \quad \Gamma^* = \bigcup_{i=1}^K \mathcal{P}(A_i) \subset L.$$

By slight abuse of notations, we identify $A_i$ and $\mathcal{P}(A_i)$, for $i \in [1, K]$ or $\Gamma$ and $\Gamma^*$ and simply write $\mathcal{P}(A_i) = A_i$ and $\Gamma^* = \Gamma$. With this identification, we can regard $X_t(A_i)$ and $Y_t^*(A_i)$ or $X_t(\Gamma)$ and $Y_t^*(\Gamma)$ as processes defined on the same space.

We first recall from Remark 4.13, that we can couple $(X_t)$ and $(Y_t^*)$. In the lemmas below where we record various coupling statements, we assume that the collection $A_1, A_2, \cdots, A_K$ of subsets of $\Omega_n$ is $m$-good for some $m \in [\log^4 n, (1/2)\log^5 n]$.

**Lemma 6.11.** *Suppose that $p$ is small enough and the law of the initial condition $X_0$ follows the law $\nu$ such that $\nu \leq \text{Perc}_n(p_{\text{init}})$*. Then, we have

$$\mathbb{P} [X_t(\Gamma) = Y_t^*(\Gamma) \text{ for all } t \in [0, t_{\text{max}})] \geq 1 - \frac{1}{n^{2d}}.$$

**Proof.** Since $K \leq n^d$, It suffices to show that, for all $i \in [1, K]$,

$$\mathbb{P} [X_t(A_i) = Y_t^*(A_i) \text{ for all } t \in [0, t_{\text{max}})] \geq 1 - \frac{1}{n^{3d}}.$$  

This follows directly from Lemma 4.11. \qed

Now we obtain upper and lower bounds for the total-variation distance of $(Y_t^*)$ in the two lemmas below. Combined with the previous coupling result, they yield bounds on the total-variation distance for $(X_t)$.

**Lemma 6.12.** *For all sufficiently small $p$, we have that*

$$\sup_{x_0 \in \Omega_n} \|\mathbb{P}_{x_0}[Y_t^*(\Gamma) \in \cdot] - \pi_{\Gamma}\|_{TV} \leq \frac{1}{2} \left[ e^{Kd^2} - 1 \right]^\frac{1}{2},$$

*where $\pi_{\Gamma}$ represents the projection of $\pi$ onto $\Gamma$.*
By Lemma 6.11, we have

\[ \mathbb{P}_x \{ Y_t^* (\Gamma) \in \cdot \} \text{ means that the starting configuration of} \]
\[ Y_t^* \text{ is inherited from} \ x_0 \in \Omega_n \text{ by the collection map} \prod_{i=1}^K \mathcal{P}^{(i)} : \prod_{i=1}^K A_i^+ \to \prod_{i=1}^K \mathcal{L}_i. \]
We define \( \mathbb{P}_\nu \{ Y_t^* (\Gamma) \in \cdot \} \) for a probability distribution \( \nu \) on \( \Omega_n \) in the same manner.

**Proof.** By the \( L^1 \)-\( L^2 \) inequality we have

\[ \| \mathbb{P}_x \{ Y_t^* (\Gamma) \in \cdot \} - \pi_\Gamma \|_{TV} \leq \frac{1}{2} \| \mathbb{P}_x \|_{L^2(\pi_\Gamma)}. \]  
(6.5)

Denote by \( \pi_{A_i}^{(i)} \) the projection of \( \pi^{(i)} \) onto \( A_i \). Then, since \( \pi_\Gamma = \prod_{i=1}^K \pi_{A_i}^{(i)} \), by the bound of \( L^2 \)-norm for product space (cf. [18, Section 3.2]), we obtain that

\[ \| \mathbb{P}_x \{ Y_t^* (\Gamma) \in \cdot \} - \pi_\Gamma \|_{L^2(\pi_\Gamma)} \leq \left[ \exp \left( \sum_{i=1}^K \| \mathbb{P}_x \{ Y_t^{(i)} (\Gamma) \in \cdot \} - \pi_{A_i}^{(i)} \|_{L^2(\pi_\Gamma)}^2 \right) - 1 \right]^{1/2}. \]  
(6.6)

By the definition of \( d_t \) (see (6.1)) and by the fact that \( A_i \) is a subset of box of size \( 2 \log^5 n \), we can deduce from the definition of \( d_t \) that

\[ \| \mathbb{P}_x \{ Y_t^{(i)} (\Gamma) \in A_i \} - \pi_{A_i}^{(i)} \|_{L^2(\pi_\Gamma)} \leq d_t. \]  
(6.7)

We now conclude using (6.5), (6.6) and (6.7).

Recall that \( \mu_\Gamma \) represents the projection of \( \mu_{p,q}^{(0)} \) to \( \Gamma \). Using spatial mixing properties, we conclude now that \( \mu_\Gamma \) is close to \( \pi_\Gamma \). This follows from Lemma 4.10 which implies that the effect of the boundary condition does not reach beyond the buffer region \( A_i^+ \setminus A_i \) for each \( i \) (see Figure 6.2). Using this we prove that the total-variation distance between \( \mu_\Gamma \) and \( \pi_\Gamma \) is small.

**Lemma 6.14.** It holds that

\[ \| \mu_\Gamma - \pi_\Gamma \|_{TV} \leq \frac{1}{n^{2d}}. \]

**Proof.** We apply Lemma 4.10 with \( A = \Gamma \) and \( B = E_n \setminus \bigcup_{i=1}^K A_i^+ \). Recall the measure \( \mu_{B^c}^{(0)} \) and the configuration \( X \) from Lemma 4.10. The latter implies that, with probability more than \( 1 - n^{-2d} \), there exists a closed surface in \( X(A_i^+ \setminus A_i) \) enclosing \( A_i \) for all \( i \in \{1, K\} \). This implies the Lemma by the domain Markov property of random cluster measure. For details about this argument, see [4, Proof of Claim 4.2].

**Lemma 6.15.** Suppose that \( p \) is small enough, \( t \in [0, t_{\text{max}}] \) and the collection \( A_1, A_2, \cdots, A_K \) of subsets of \( \Omega_n \) is m-good for some \( m \in \{ \log^4 n, 1/2 \log^5 n \} \). Then under the notations of Definition 6.10, we have

\[ d(t) + \frac{2}{n^{2d}} \geq \sup_{x_0 \in \Omega_n} \| \mathbb{P}_x \{ Y_t^* (\Gamma) \in \cdot \} - \pi_\Gamma \|_{TV}, \]
where \( d(t) \) the total-variation distance at time \( t \) was defined in Section 1.3.

**Proof.** Since projection does not increase total-variation norm, we have

\[ d(t) \geq \sup_{x_0 \in \Omega_n} \| \mathbb{P}_x \{ X_t(\Gamma) \in \cdot \} - \mu_\Gamma \|_{TV}. \]
(6.8)

By Lemma 6.11, we have

\[ \sup_{x_0 \in \Omega_n} \| \mathbb{P}_x \{ X_t(\Gamma) \in \cdot \} - \mathbb{P}_x \{ Y_t^* (\Gamma) \in \cdot \} \|_{TV} \leq \frac{1}{n^{2d}}. \]
(6.9)

By combining (6.8), (6.9), and Lemma 6.14, the proof is completed.

We are finally ready to finish the proof of Theorem 6.1.
6.3.1. Proof of part (1): upper bound. In view of Proposition 6.9, it suffices to prove the following proposition.

**Proposition 6.16.** Suppose that \( p \) is sufficiently small, \( \Gamma \in \text{Spa}_n \), and \( t \in [0, t_{\text{max}}] \). Then, we have

\[
\| \mathbb{P}_\nu(X_t(\Gamma) \in \cdot) - \mu_\Gamma \|_{TV} \leq \frac{1}{2} \left[ \exp \left\{ \frac{n^{d/2} |d|}{\log^{12d} n} \right\} - 1 \right]^{1/2} + \frac{2}{n^{2d}}. \tag{6.10}
\]

**Proof.** Denote by \( A_1, A_2, \ldots, A_K \) the connected components of \( \Gamma \) in the sense of Definition 6.7. Then, then \( A_1, A_2, \ldots, A_K \) are \( m \)-good with \( m = \log^4 n \). Now we recall the notations from Definition 6.10 and Lemma 6.12. We bound the total-variation norm at the left-hand side of (6.10) by

\[
\| \mathbb{P}_\nu(X_t(\Gamma) \in \cdot) - \mathbb{P}_\nu(Y_t^*(\Gamma) \in \cdot)\|_{TV} + \| \mathbb{P}_\nu(Y_t^*(\Gamma) \in \cdot) - \pi_\Gamma \|_{TV} + \| \pi_\Gamma - \mu_\Gamma \|_{TV}. \tag{6.11}
\]

We recall Notation 6.13 for the notation \( \mathbb{P}_\nu(Y_t^*(\Gamma) \in \cdot) \). We now bound these three terms separately to complete the proof. For the first term, by Lemma 6.11 we have

\[
\| \mathbb{P}_\nu(X_t(\Gamma) \in \cdot) - \mathbb{P}_\nu(Y_t^*(\Gamma) \in \cdot)\|_{TV} \leq \frac{1}{n^{2d}}. \tag{6.12}
\]

By the Lemma 6.12, and the fact \( K \leq n^{d/\log^{12d} n} \), the second term is bounded by

\[
\| \mathbb{P}_\nu(Y_t^*(\Gamma) \in \cdot) - \pi_\Gamma \|_{TV} \leq \frac{1}{2} \left[ \exp \left\{ \frac{n^{d/2} |d|}{\log^{12d} n} \right\} - 1 \right]^{1/2}. \tag{6.13}
\]

Finally, the last term at (6.11) is at most \( 1/n^{2d} \) by Lemma 6.14. Combining this with (6.11), (6.12), and (6.13), we can finish the proof. \( \square \)

6.3.2. Proof of Part (2): lower bound. Given the above ingredients the proof of the lower bound is almost verbatim from [18, Section 3.3] but nonetheless we include the proof in the appendix for completeness.

In the following section we finish the proof of Theorem 1.1.

7. Proof of main result

We keep the notation \( r = 3 \log^5 n \). The following lemma provides a sharp bound on \( d_t \).

**Lemma 7.1.** [18, Lemma 4.1] For all small enough \( p \), there exists a constant \( C_3 = C_3(p) > 0 \) such that

\[
e^{-\lambda(r)(t+C_3 \log \log n)} - n^{-2d} \leq d_t \leq e^{-\lambda(r)(t-C_3 \log \log n)} \tag{7.1}
\]

for all \( t \in [C_3 \log \log n, t_{\text{max}}] \).

**Proof.** Since

\[
d_t \leq \max_{x_0 \in \Omega} \left\| \mathbb{P}_{x_0} \left[ X_t^\dagger \in \cdot \right] - \pi^\dagger \right\|_{L^2(\pi^\dagger)},
\]

the upper bound part of (7.1) is immediate from Theorem 5.1. We note from this bound that

\[
r^{-d/2} d_t = o(1) \text{ for } t = C \log \log n \tag{7.2}
\]

with sufficiently large \( C \). Here we implicitly used Corollary 3.4. For the lower bound part, we first recall the bound

\[
e^{-\lambda(r)t} \leq 2 \max_{x_0 \in \Omega} \left\| \mathbb{P}_{x_0} \left[ X_t^\dagger \in \cdot \right] - \pi^\dagger \right\|_{TV} \text{ for all } t \geq 0
\]

which is the continuous time version of Proposition 3.1 (see [17, Lemma 20.11]). By Proposition 4.4 (in particular, (4.2)), we have that

\[
e^{-\lambda(r)(t+t_{\text{init}})} \leq 2 \max_{x_0 \in \text{Perc}_r} \left\| \mathbb{P}_{x_0} \left[ X_t^\dagger \in \cdot \right] - \pi^\dagger \right\|_{TV} \text{ for all } t \geq 0.
\]
We now take \( s = C_1 \log \log n \) and \( t \in [C \log \log n, t_{\text{max}}] \), where \( C_1 \) and \( C \) are the constants appeared in Theorem 6.1 and in (7.2), respectively. Then, by the previous inequality and part (1) of Theorem 6.1 for the lattice \( \mathbb{Z}_r^d \) (cf. Remark 6.2), we have that
\[
e^{-\lambda(r)(t+t_{\text{init}}+C_1 \log \log n)} \leq 2 \max_{\nu, \nu \in \text{Perc}, \nu} \left\| P_{\nu} X_{t+s}^\dagger \right\|_{TV} \leq \left\| \exp \left\{ d t^2 \right\} - 1 \right\|_{TV}^2 + 8n^{-2d} \leq 2e^{d/2} d + 8n^{-2d},
\]
where the last inequality follows from (7.2) and the elementary inequality \( e^x - 1 \leq 4x \) for \( x \in [0, 1] \). We can deduce the lower bound from this computation.

Given the above, the proof of Theorem 1.1 involves two steps:

- Prove a version (Proposition 7.2) with \( \lambda_\infty \) replaced by \( \lambda(r) \) where \( r \) was chosen above.
- Show that \( \lambda(r) \) converges to \( \lambda_\infty \) and have bounds on the convergence rate (Proposition 7.3).

Define \( t(n) = \frac{d}{2\lambda(r)} \log n \) and \( w(n) = \log \log n \).

**Proposition 7.2.** For all small enough \( p \), there exist two constants \( c_1 = c_1(p), c_2 = c_2(p) \) such that

\[
\lim_{n \to \infty} \max_{x_0 \in \Omega_n} \left\| P_{\nu} X_{t(n)+c_1 w(n)} \right\|_{TV} = 1, \quad (7.3)
\]
\[
\lim_{n \to \infty} \max_{x_0 \in \Omega_n} \left\| P_{\nu} X_{t(n)+c_2 w(n)} \right\|_{TV} = 0. \quad (7.4)
\]

**Proof.** By Proposition 4.4, it suffices to consider the initial condition \( \nu \) satisfying \( \nu \preceq \text{Perc}(p_{\text{init}}) \).

We recall the constant \( C_3 \) from the statement of Lemma 7.1. First, by the lower bound in Lemma 7.1 and part (3) of Corollary 3.4,
\[
\left( \frac{n}{\log 10 n} \right)^d d^2 (t(n) - c_1 w(n)) \geq \left( \frac{n}{\log 10 n} \right)^d e^{-2\lambda(r)(t(n) - c_1 w(n) + C_3 \log \log n)} - n^{-d} \geq (\log n)^{(C_1 - C_3) - 12d} \cdot n^{-d}.
\]
Therefore, for \( c_1 > C_3 + \frac{11d}{2} \), we have
\[
\lim_{n \to \infty} \left( \frac{n}{\log 10 n} \right)^d d^2 (t(n) - c_1 w(n)) = +\infty,
\]
and thus by part (2) of Theorem 6.1 we obtain (7.3). Now we turn to (7.4). For \( c \in (0, C_3) \), by the upper bound of Lemma 7.1,
\[
\frac{n^d}{\log 12d n} d^2 (t(n) + c w(n)) \leq \frac{n^d}{\log 12d n} e^{-2\lambda(r)(t(n) + c w(n) - C_3 \log \log n)} \leq (\log n)^{(C_2 - c) - 12d}.
\]
By taking \( c \) close enough to \( C_2 \) we obtain
\[
\frac{n^d}{\log 12d n} d^2 (t(n) + c w(n)) \leq \frac{1}{\log 11d n}.
\]
Let \( c_2 = C_1 + c \) where \( C_1 \) is the constant appeared in Theorem 6.1. Then, by part (1) of Theorem 6.1 (note that this is where the sparsity assumption on \( \nu \) is used) and (7.5),
\[
\max_{\nu \leq \text{Perc}(t_{\text{init}})} \left\| P_{\nu} X_{t(n)+c_2 w(n)} \right\|_{TV} \leq \frac{1}{2} \left\| \exp \left\{ \frac{n^d}{\log 12d n} d^2 (t(n) + c w(n)) \right\} - 1 \right\|_{TV}^2 + \frac{4}{n^{2d}} \leq \frac{1}{2} \left\| \exp \left\{ \frac{1}{\log 11d n} \right\} - 1 \right\|_{TV}^2 + \frac{4}{n^{2d}}.
\]
This completes the proof of (7.4).
Notice that since $\lambda(r) = \Theta(1)$ (Corollary 3.4), we have $w(n) \ll t(n)$, and therefore the previous proposition already exhibits a cut-off phenomenon provided that $p$ is small enough. Next we prove that the sequence $(\lambda(r))_{r \geq 1}$ is a convergent sequence.

**Proposition 7.3.** [18, Lemma 4.3] There exists $\hat{\lambda} = \hat{\lambda}(p) > 0$ such that

$$|\lambda(r) - \hat{\lambda}| \leq r^{-1/4+o(1)}.$$

**Proof.** We only provide the modified choice of parameters needed for our purpose. A careful reading of the proof shows that entire arguments presented above are still in force if we replace $r = 3 \log^5 n$ with $r = \log^{4+\delta}$ for any $\delta$. Of course the constants that we obtained above must be modified to depend on $\delta$, and the time $t_{\text{max}}$ should be defined as $\log^{1+\delta} n$ (cf. Remark 4.14). Taking $r_1 = \log^{4+\delta}$ and $r_2 \in [r_1, r_2^2]$ and applying Proposition 7.2 with $r = r_1$ and $r = r_2$, respectively, yields

$$\frac{d}{2\lambda(r_1)} \log n - C w(n) \leq \frac{d}{2\lambda(r_2)} \log n + C w(n)$$

for some constant $C = C(p, \delta)$. Since $\lambda(\cdot)$ is bounded below, we obtain

$$\lambda(r_1) - \lambda(r_2) \leq C \frac{\log \log n}{\log n} \leq r_1^{-1/4+\delta}$$

for all sufficiently large $n$. The rest of the arguments are exactly the same as [18, Lemma 4.3] and are omitted.

7.1. **Proof of Theorem 1.1.** As mentioned before we can combine Propositions 7.2 and 7.3 to deduce Theorem 1.1. Define

$$t^*(n) = \frac{d}{2\lambda_{\infty}} \log n.$$

Thus we need to show that for all small enough $p$, there exist two constants $c_1 = c_1(p), c_2 = c_2(p)$ such that

$$\lim_{n \to \infty} \max_{x_0 \in \Omega_n} |\mathbb{P}_{x_0} [X_{t^*(n)} - c_1 w(n)] - \mu_{p,q}^n|_{\text{TV}} = 1,$$

$$\lim_{n \to \infty} \max_{x_0 \in \Omega_n} |\mathbb{P}_{x_0} [X_{t^*(n) + c_2 w(n)}] - \mu_{p,q}^n|_{\text{TV}} = 0.

The proof is now immediate from

$$|t^*(n) - t(n)| \leq C |\lambda(r) - \lambda_{\infty}| \log n \leq C \log^{-1/4} n.$$

7.2. **Comparison to infinite volume dynamics.** It is quite natural to predict that $\lambda_{\infty}$ is in fact the spectral gap of the infinite volume FK-dynamics with the same parameters $p$ and $q$. Defining the latter is not trivial but this has been carried out in [16, Chapter 8]. For the Ising model a similar result was shown in [18] using the monotonicity of the underlying dynamics as well as Log-Sobolev inequalities. Even though the lack of monotonicity of the Potts model prevented the authors in [19] to prove a similar conclusion, this was settled in [25, Section 6.2] using the Information Percolation machinery which also implies the same for SW dynamics. Furthermore in [25], the authors remark that the argument relies on bounds on disagreement propagation and an infinite version of the exponential $L^2$ mixing rate and hence holds in more generality for spin systems.

Thus in our context of the FK-dynamics to prove a similar result, given the disagreement propagation bounds, stated in Section 4.4, the only remaining step is to establish an analog of Theorem 5.1 for the infinite system by proving an analog of Proposition 5.3 in the same setting. The argument in [25] proceeds by defining Information Percolation clusters for the infinite process. We believe that this can be carried out in our setting as well, by suitable extensions of the arguments.
for finite systems presented in Section 5. However we do not pursue verifying the precise details in the paper.

8. Appendix

We provide the proofs that were omitted from the main article.

Proof of Lemma 4.11. We recall notations from Section 4.1 and in particular \( t_{\text{max}} \) from (4.10) and write \( t_{\text{max}} = L\Delta \) so that

\[
0 = \tau_0 < \tau_1 < \cdots < \tau_L = \log^2 n.
\]

We regard \( (Z_t^-)_{t \geq 0} \) as a Markov chain on \( \Omega_n \) such that the configuration outside of \( A^+ \) is empty. Also consider \( (X_t)_{t \geq 0} \), the FK-dynamics on \( \Omega_n \) starting from an initial condition which agrees with \( (Z_t^-) \) on \( A^+ \). One can observe that under the monotone coupling,

\[
Z_t^- \leq X_t \leq Z_t^+ \text{ for all } t \geq 0.
\]

We recall the enlarged percolation \( \mathcal{F}_i \) from Table 1, and denote by \( \mathcal{E}_i \) the event that there is no open path of length \( \beta \log n \) in \( \mathcal{F}_i \) for some \( \beta = \beta(p) > 0 \). Then, by Proposition 4.9 and the union bound,

\[
\mathbb{P}[^{\cup}_{i=1} \mathcal{E}_i] \leq \left( \frac{|E_n|}{2} \right) \exp \{-\gamma \beta \log n\} < \frac{1}{n^{4d}}
\]

provided that \( \beta \) is large enough. Define \( \mathcal{E} = \cup_{t=1}^L \mathcal{E}_i \). Since \( L = t_{\text{max}}/\Delta = \Omega(\log^2 n) \), the union bound implies that

\[
\mathbb{P}[\mathcal{E}] \leq \frac{1}{n^{4d}} < \frac{1}{n^{3d}}.
\]

We now claim that \( \mathcal{E}^c \) implies that \( Z_t^+(A) = Z_t^-(A) \) for all \( t \in [0, t_{\text{max}}] \). Thus the claim along with (8.2), finishes the proof of the lemma. To prove the claim, define \( A_i, i \in [0, L] \), inductively as \( A_L = A \) and

\[
A_{i-1} = \{ e \in E_n : d(e, A_i) \leq t_{\text{max}} \Delta \},
\]

so that

\[
A^+ = A_0 \supset A_1 \supset \cdots \supset A_L = A.
\]

For all \( i \in [0, L-1] \), we shall prove that \( Z_t^+(A_i) = Z_t^-(A_i) \) implies \( Z_t^+(A_{i+1}) = Z_t^-(A_{i+1}) \) for all \( t \in (\tau_i, \tau_{i+1}] \). Since then the proof of the claim is completed by the induction. Now it suffices to observe that there exists a closed surface of \( \mathcal{F}_i \) in \( A_i \backslash A_{i+1} \) under \( \mathcal{E}_i \) since the set \( \bigcup_{e \in \partial A_i} \text{Conn}(e; \mathcal{F}_i) \) is disjoint to \( A_{i+1} \) as there is no connected path of length \( \Omega(\log^2 n) \) in \( \mathcal{F}_i \) (call this surface as \( V_i \) ).

The proof now follows by noticing that the FK-dynamics for both \( Z_t^+ \) and \( Z_t^- \) agree on the component of \( E_n \backslash V_i \) (say \( \hat{A}_i \)) containing \( A_{i+1} \) and hence on \( A_{i+1} \) throughout \( [\tau_i, \tau_{i+1}] \), since the starting configurations for both the chains agree on \( \hat{A}_i \) by induction and the dynamics has zero boundary condition throughout \( [\tau_i, \tau_{i+1}] \).  

\[ \square \]

8.0.1. Proof of Theorem 6.1, Part (2): lower bound. Recall \( r = 3 \log^5 n \), and let us divide \( Z_n^d \) by \( K = [n/r]^d \) square boxes \( A_1^+, A_2^+, \cdots, A_k^+ \) of size \( r \) as we did in Section 6.1. Then, let \( A_i \) be the box of size \( 2r/3 \) which is concentric with \( A_i^+ \). Then, the collection \( A_1, A_2, \cdots, A_k \) is \( m \)-good with \( m = (1/2) \log^5 n \). We recall the notations from Definition 6.10. By definition (6.1) of \( d_i \), we can find \( x_0^* = x_0^*(t) \in \Omega_r \) satisfying

\[
d_i = \| \mathbb{P}_{x_0^*} [X_i^+(\Lambda) \in \cdot] - \pi^+_\Lambda \|_{L^2(\pi^+_\Lambda)}.
\]

Let \( U_i \) be a configurations
on $B_i$ distributed according to $\pi_{A_i}^{(i)}$, where $\{U_i, 1 \leq i \leq K\}$ is a collection of independent random variables. Define a sequence of i.i.d. random variable $u_i$ as

$$u_i = \frac{\mathbb{P}[Y_t^{(i)}(A_i) = U_i | Y_0^{(i)} = x_0^*]}{\pi_{B_i}^{(i)}(U_i)}; \quad i \in [1, K].$$

The condition $Y_0^{(i)} = x_0^*$ means $X_0(\widehat{A}_i^+) = x_0^*$. By the definition of $u_i$ and $x_0^*$, one can readily check that

$$\mathbb{E}u_i = 1 \quad \text{and} \quad \text{Var}u_i = d_t^2. \quad (8.3)$$

By the $L^\infty$-$L^2$ reduction for reversible Markov chains, we have

$$\|u_i - 1\|_\infty = \left\| \mathbb{P}\left[Y_t^{(i)}(A_i) \in \cdot | Y_0^{(i)} = x_0^*\right] - \pi_{A_i}^{(i)} \right\|_{L^\infty(\pi_{A_i}^{(i)})} \leq \left\| \mathbb{P}\left[Y_{t/2}^{(i)}(A_i) \in \cdot | Y_0^{(i)} = x_0^*\right] - \pi_{A_i}^{(i)} \right\|_{L^2(\pi_{A_i}^{(i)})} \leq d_t^2. \quad (8.4)$$

Hence, by Theorem 5.1 we obtain

$$\|u_i - 1\|_\infty \leq e^{-c\log\log n} \quad (8.4)$$

for some $c > 0$. Then, by (8.3) and (8.4), we have

$$\mathbb{E}|u_i - 1|^3 \leq e^{-c\log\log n} d_t^2 = o(1) d_t^2. \quad (8.5)$$

Given the above inputs, the rest of the proof follows by arguments identical to [18, Section 3.3] and is omitted.

References

[1] David Aldous and Persi Diaconis. Shuffling cards and stopping times. The American Mathematical Monthly, 93(5):333–348, 1986.
[2] Vincent Beffara and Hugo Duminil-Copin. The self-dual point of the two-dimensional random-cluster model is critical for $q > 1$. Probability Theory and Related Fields, 153(3-4):511–542, 2012.
[3] Antonio Blanca, Reza Gheissari, and Eric Vigoda. Random-cluster dynamics in $\mathbb{Z}^2$: rapid mixing with general boundary conditions. arXiv preprint arXiv:1807.08722, 2018.
[4] Antonio Blanca and Alistair Sinclair. Random-cluster dynamics in $\mathbb{Z}^2$. Probability Theory and Related Fields, 168(3-4):821–847, 2017.
[5] Persi Diaconis. The cutoff phenomenon in finite Markov chains. Proceedings of the National Academy of Sciences, 93(4):1659–1664, 1996.
[6] Persi Diaconis and Laurent Saloff-Coste. Logarithmic sobolev inequalities for finite Markov chains. The Annals of Applied Probability, 6(3):695–750, 1996.
[7] Hugo Duminil-Copin. Introduction to Bernoulli percolation. www.ihes.fr/duminil/publi/2017/percolation.pdf, 2018.
[8] Hugo Duminil-Copin, Maxime Gagnebin, Matan Harel, Ioan Manolescu, and Vincent Tassion. Discontinuity of the phase transition for the planar random-cluster and Potts models with $q > 4$. arXiv preprint arXiv:1611.09877, 2016.
[9] Hugo Duminil-Copin, Vladas Sidoravicius, and Vincent Tassion. Continuity of the phase transition for Planar Random-Cluster and Potts models with $1 \leq q \leq 4$. Communications in Mathematical Physics, 349(1):47–107, 2017.
[10] Robert G Edwards and Alan D Sokal. Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm. Physical review D, 38(6):2009, 1988.
[11] Cornelis Marius Fortuin. On the random-cluster model: III. the simple random-cluster model. Physica, 59(4):545–570, 1972.
[12] Cornelius Marius Fortuin and Piet W Kasteleyn. On the random-cluster model: I. introduction and relation to other models. Physica, 57(4):536–564, 1972.
[13] Reza Gheissari and Eyal Lubetzky. Quasi-polynomial mixing of critical 2d random cluster models. arXiv preprint arXiv:1611.01147, 2016.
[14] Reza Gheissari and Eyal Lubetzky. The effect of boundary conditions on mixing of 2d Potts models at discontinuous phase transitions. Electronic Journal of Probability, 23, 2018.
[15] Reza Gheissari and Eyal Lubetzky. Mixing times of critical two-dimensional Potts models. *Communications on Pure and Applied Mathematics*, 71(5):994–1046, 2018.

[16] Geoffrey Grimmett. Percolation, 1989.

[17] David Asher Levin, Yuval Peres, and Elizabeth Lee Wilmer. *Markov chains and mixing times*. American Mathematical Soc., 2009.

[18] Eyal Lubetzky and Allan Sly. Cutoff for the Ising model on the lattice. *Inventiones mathematicae*, 191(3):719–755, 2013.

[19] Eyal Lubetzky and Allan Sly. Cutoff for general spin systems with arbitrary boundary conditions. *Communications on Pure and Applied Mathematics*, 67(6):982–1027, 2014.

[20] Eyal Lubetzky and Allan Sly. An exposition to information percolation for the Ising model. *Annales de la Faculté des Sciences de Toulouse*, 29(4):745–761, 2015.

[21] Eyal Lubetzky and Allan Sly. Information percolation and cutoff for the stochastic Ising model. *Journal of the American Mathematical Society*, 29(3):729–774, 2016.

[22] Eyal Lubetzky and Allan Sly. Universality of cutoff for the Ising model. *The Annals of Probability*, 45(6A):3664–3696, 2017.

[23] Jason Miller and Yuval Peres. Uniformity of the uncovered set of random walk and cutoff for lamplighter chains. *The Annals of Probability*, 40(2):535–577, 2012.

[24] Elchanan Mossel and Allan Sly. Exact thresholds for Ising–Gibbs samplers on general graphs. *The Annals of Probability*, 41(1):294–328, 2013.

[25] Danny Nam and Allan Sly. Cutoff for the Swendsen-Wang dynamics on the lattice. *arXiv preprint arXiv:1805.04227*, 2018.

[26] Y Peres. American institute of mathematics (AIM) research workshop ‘sharp thresholds for mixing times’(Palo Alto, december 2004). *Summary available at http://www.aimath.org/WWN/mixingtimes*, 12, 2004.

[27] James Propp and David Wilson. Coupling from the past: a user’s guide. *Microsurveys in Discrete Probability*, 41:181–192, 1998.

[28] Robert H Swendsen and Jian-Sheng Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Physical review letters*, 58(2):86, 1987.

[29] Mario Ullrich. Rapid mixing of Swendsen-Wang dynamics in two dimensions. *arXiv preprint arXiv:1212.4908*, 2012.

[30] Mario Ullrich. Comparison of Swendsen-Wang and heat-bath dynamics. *Random Structures & Algorithms*, 42(4):520–535, 2013.

[31] Mario Ullrich. Swendsen–wang is faster than single-bond dynamics. *SIAM Journal on Discrete Mathematics*, 28(1):37–48, 2014.

S. Ganguly
Department of Statistics, UC Berkeley, Berkeley, California, CA 94720, USA.
E-mail address: sganguly@berkeley.edu

I. Seo
Department of Mathematical Science and RIMS, Seoul National University, Seoul, South Korea.
E-mail address: insuk.seo@snu.ac.kr