



On the tractability of sampling from the Potts model at low temperatures via random-cluster dynamics

Antonio Blanca¹ · Reza Gheissari²

Received: 4 August 2023 / Revised: 17 April 2024 / Accepted: 28 May 2024 /
Published online: 8 June 2024

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

Sampling from the q -state ferromagnetic Potts model is a fundamental question in statistical physics, probability theory, and theoretical computer science. On general graphs, this problem may be computationally hard, and this hardness holds at arbitrarily low temperatures. At the same time, in recent years, there has been significant progress showing the existence of low-temperature sampling algorithms in various specific families of graphs. Our aim in this paper is to understand the minimal structural properties of general graphs that enable polynomial-time sampling from the q -state ferromagnetic Potts model at low temperatures. We study this problem from the perspective of random-cluster dynamics. These are non-local Markov chains that have long been believed to converge rapidly to equilibrium at low temperatures in many graphs. However, the hardness of the sampling problem likely indicates that this is not even the case for all bounded degree graphs. Our results demonstrate that a key graph property behind fast or slow convergence time for these dynamics is whether the independent edge-percolation on the graph admits a strongly supercritical phase. By this, we mean that at large $p < 1$, it has a large linear-sized component, and the graph complement of that component is comprised of only small components. Specifically, we prove that such a condition implies fast mixing of the random-cluster Glauber and Swendsen–Wang dynamics on two general families of bounded-degree graphs: (a) graphs of at most stretched-exponential volume growth and (b) locally treelike graphs. In the other direction, we show that, even among graphs in those families, these Markov chains can converge exponentially slowly at arbitrarily low temperatures if the edge-percolation condition does not hold. In the process, we develop new

An extended abstract of this paper appeared in FOCS 2023 [4] under the title “Sampling from the Potts model at low temperatures via Swendsen–Wang dynamics”.

✉ Reza Gheissari
gheissari@northwestern.edu

Antonio Blanca
ablanca@cse.psu.edu

¹ Department of CSE, Pennsylvania State University, University Park, USA

² Department of Mathematics, Northwestern University, Evanston, USA

tools for the analysis of non-local Markov chains, including a framework to bound the speed of disagreement propagation in the presence of long-range correlations, an understanding of spatial mixing properties on trees with random boundary conditions, and an analysis of burn-in phases at low temperatures.

Keywords Swendsen–Wang dynamics · Potts model · Random-cluster model · Low-temperatures · Mixing times · Disagreement percolation

Mathematics Subject Classification 60J10 · 60K35 · 82C20

1 Introduction

The q -state ferromagnetic Potts model is a classical spin system model central to probability theory and with applications in statistical physics, theoretical computer science, and other fields. It is defined on a graph $G = (V(G), E(G))$ as a probability distribution over configurations in $\Omega_p = \{1, \dots, q\}^{V(G)}$, with a parameter $\beta > 0$, corresponding to the inverse temperature in physical applications, controlling the strength of the interaction between the edges of G . Formally, the probability of each configuration $\sigma \in \Omega_p$ is given by:

$$\mu_{G,\beta,q}(\sigma) = \frac{1}{Z_{G,\beta,q}^p} \exp\left(\beta \sum_{\{u,v\} \in E(G)} \mathbf{1}\{\sigma(u) = \sigma(v)\}\right). \quad (1)$$

The factor $Z_{G,\beta,q}^p$ is a normalization constant and is known as the partition function; $\sigma(u)$ denotes the color or spin value of the configuration σ at vertex u . The classical Ising model corresponds to the $q = 2$ case.

The question of sampling from the ferromagnetic Potts model is an important one and has been extensively studied on a variety of graphs and temperature regimes (i.e., different values of β). In general, it is known that the problem of approximate sampling from (1) for $q \geq 3$ and β large is #BIS-hard, in the sense that there exist graphs, including bounded degree ones, for which the approximate sampling problem is as hard as approximately counting the number of independent sets on bipartite graphs [29, 33]. This latter task is a well-studied computational problem that is believed not to have a polynomial time approximation algorithm. This sharply contrasts with the ferromagnetic Ising case ($q = 2$), where polynomial-time samplers have been known since the 1990s [46, 53]. At the same time, for some families of graphs, notably including $\mathbb{Z}^d$ and expander graphs, the existence of polynomial-time sampling algorithms has recently been shown for the Potts model at low temperatures: see, e.g., [16, 18–20, 32, 40, 41, 43, 45]. This raises the question of what are the underlying graph structures and temperatures that cause tractability or hardness of approximately sampling from (1). We study this from the perspective of widely-used Markov chain-based algorithms.

One fundamental approach to sampling from Gibbs distributions of the form of (1) is via Markov chains whose stationary distribution is exactly $\mu_{G,\beta,q}$. The simplest

such Markov chain is the Glauber dynamics for the Potts model (also known as the Gibbs sampler), which updates the state of a randomly chosen vertex at each step. Its simplicity makes it quite appealing to practitioners, but it is known to take exponentially (in $\text{poly}(|V|)$) many steps to equilibrate at low temperatures (β large). In lieu of this, in order to sample from (1), an oft-used approach is a different family of Markov chains based on the *Edwards–Sokal* coupling of the ferromagnetic Potts model to a graphical model called the random-cluster model [26]: see (3) for its definition. This family of Markov chains includes the extensively studied Swendsen–Wang (SW) dynamics, and its close relative, the Glauber dynamics for the random-cluster model.

These Markov chains make non-local updates, and are often used to bypass the bottlenecks that slow down the convergence of the Potts Glauber dynamics at low temperatures when $q \geq 3$. At the same time, the aforementioned #BIS-hardness of the sampling problem at low temperatures suggests that these Markov chains could not have a polynomial speed of convergence on all graphs. For the sake of completeness, we mention that at high temperatures (small β), these Markov chains converge quickly but have a larger computational overhead per step than the Potts Glauber dynamics [6]; there are also “intermediate” temperature regimes corresponding to first-order phase transitions where these Markov chains are known to converge exponentially slowly to equilibrium [17, 21, 30, 31, 34].

In this paper, we systematically analyze these Edwards–Sokal based Markov chains on general graphs at low temperatures. In the process, we develop new tools for the analysis of non-local chains and arrive at an explanation, in terms of geometric properties of the graph, that dictate whether these Markov chains converge quickly or slowly at low temperatures (i.e., large, but independent of the system size, values of β).

Let us define the Markov chains of interest. For a unified discussion, it is convenient to reparametrize β by $p = 1 - e^{-\beta}$. Notice that low-temperature settings corresponding to β large correspond to p close to 1. The SW dynamics transitions from a configuration $\sigma_t \in \Omega_p$ to $\sigma_{t+1} \in \Omega_p$ as follows:

1. Independently, for every $e = \{u, v\} \in E(G)$ if $\sigma_t(u) = \sigma_t(v)$ include e in E_t with probability p ;
2. Independently, for every connected component $\mathcal{C}$ in $(V(G), E_t)$, draw a color $c \in \{1, \dots, q\}$ uniformly at random, and set $\sigma_{t+1}(v) = c$ for all $v \in \mathcal{C}$.

It can be checked that the SW dynamics is reversible with respect to $\mu_{G,\beta,q}$ and thus converges to it. In effect, the SW dynamics moves on the larger probability space of Potts model configurations together with random-cluster configurations. The configurations of this model consist of edge subsets, i.e., $\Omega_{\text{RC}} = \{0, 1\}^{E(G)}$, and the SW dynamics can be interpreted as alternating steps of sampling a random-cluster configuration E_t conditionally on the Potts configuration σ_t , then sampling the Potts configuration σ_{t+1} conditionally on E_t .

A closely related Markov chain is the Glauber dynamics that moves in the space of random-cluster configurations; for brevity, we call this Markov chain the *FK dynamics* since the random-cluster model is also known as the FK model. Here, given an edge subset $E_t \in \Omega_{\text{RC}}$, we generate E_{t+1} by:

1. Pick an edge $e \in E(G)$ uniformly at random;
2. Set $E_{t+1} = E_t \cup \{e\}$ with probability:

$$\begin{cases} p & \text{if } e \text{ is not a cut-edge in } E_t; \\ \hat{p} := \frac{p}{p+(1-p)q} & \text{if } e \text{ is a cut-edge in } E_t; \end{cases} \quad (2)$$

and $E_{t+1} = E_t \setminus \{e\}$ otherwise.

A cut-edge is an edge whose state affects the number of connected components of the configuration. It can be checked that the FK dynamics converges to the random-cluster distribution (3). After convergence, one may produce a sample from the corresponding q -state Potts measure (the one with β so that $p = 1 - e^{-\beta}$) with little overhead by independently assigning states uniformly amongst $\{1, \dots, q\}$ to each connected component of the random-cluster configuration, as in step (2) of the SW dynamics above. As such, the FK dynamics provides an alternative Markov chain that can be used to sample from (1).

To formalize convergence rates of these Markov chains, recall that the mixing time of a Markov chain is the number of steps required to reach a distribution close to the stationary distribution (in total variation distance), assuming the worst possible starting state: see (4) for the formal definition. It is known that the mixing times of the SW and FK dynamics can differ only up to a $O(|E(G)|)$ factor (see [56]).

Both the SW and FK dynamics are conjectured to overcome some of the key difficulties associated with sampling from the Potts distribution quickly at low temperatures. They are, therefore, quite popular, but their non-locality makes the rigorous analysis of their mixing times significantly more challenging than their Potts Glauber dynamics counterparts. In recent years, significant progress has been made in establishing optimal mixing time bounds for the SW and FK dynamics in high-temperature regimes where the corresponding Potts Glauber dynamics is also known to be fast mixing; see, e.g., [6–8, 32]. These works have resulted in optimal (or nearly optimal) mixing time bounds for SW and FK dynamics that hold under various correlation decay conditions (e.g., strong spatial mixing, tree uniqueness, Dobrushin uniqueness, spectral independence, etc.). In particular, $p \lesssim 1/\Delta$ implies fast mixing for all graphs of maximum degree Δ . By contrast, in the low-temperature setting, where correlations do not decay, the Potts Glauber dynamics converges slowly, and alternative efficient sampling algorithms are most needed, there is no generic criterion guaranteeing that the SW and FK dynamics mix quickly.

In fact, rigorous bounds for the mixing time of the SW and FK dynamics at low temperatures are rare and can be summarized as follows. In the Ising case of $q = 2$, [37] showed that these Markov chains mix in $O(n^{10})$ time on all n -vertex graphs and all p . On the complete graph, [13, 15, 28] establish nearly-optimal mixing time bounds throughout the low-temperature regime. On more sophisticated geometries, progress has been limited to the special case of the integer lattice $\mathbb{Z}^d$. In particular, in [14, 55], fast mixing was shown in the low-temperature regime on subsets of $\mathbb{Z}^2$ via planar duality to high-temperatures (see also [50] for sharper bounds for the SW dynamics at low temperatures in the Ising case). Recently [32] showed fast mixing

at low temperatures in cubes in $\mathbb{Z}^d$. For general graphs, the only low-temperature criterion known to ensure fast mixing is $p > 1 - O(1/|E(G)|)$ [42].

This leaves a wealth of questions to explore on general families of graphs, notably including the mixing times of the SW and FK dynamics at values of p close to 1, but importantly, independent of the graph size. We consider this question for two broad families of bounded-degree graphs: graphs of at most stretched-exponential volume growth, and locally treelike graphs (which allow for exponential volume growth). We show that for all such graphs, fast mixing of the SW and FK dynamics at low enough temperatures is ensured if the independent edge-percolation process on the graph, where an edge-set $\tilde{\omega} \subset E(G)$ is obtained by keeping each edge with probability $\tilde{p}$, independently, has a *strongly supercritical phase* (i.e., for $\tilde{p}$ close to 1, all large connected sets in G intersect the giant component of $\tilde{\omega}$; see Definition 2 and 3 for precise definitions). To illustrate the necessity of this condition, for any arbitrarily large $p < 1$, we construct explicit graphs—both ones of polynomial volume growth, and ones that are locally treelike—on which the edge-percolation is not in a strongly supercritical phase, and, in turn, the SW and FK dynamics mix slowly.

The class of graphs that have strongly supercritical phases for their edge-percolation is an area of extensive study, and it is closely connected to whether the graph has isoperimetric dimension strictly larger than 1. The key takeaway from our results is thus a purely geometric mechanism underlying fast or slow mixing of the SW and FK dynamics at large $p < 1$ on two large families of bounded degree graphs.

1.1 Graphs of at most stretched-exponential growth

The first general class of graphs for which we establish fast mixing of the SW and FK dynamics at low temperatures under the percolation condition are bounded-degree graphs that have at most stretched-exponential volume growth. Let us introduce some notation: in what follows, we think of $G = (V(G), E(G))$ as a connected graph on n vertices, with maximum degree $\Delta \geq 3$, and we fix any $q \geq 2$. For a vertex $v \in V(G)$, let $B_R(v) = \{w : d_G(v, w) \leq R\}$ be the set of vertices at graph distance at most R from v . For a subset $A \subset V(G)$, let $\partial_e A \subset E(G)$ denote the edge boundary of A , i.e., the set of edges in $E(G)$ with exactly one endpoint in A .

Definition 1 The graph G has η -stretched-exponential volume growth if $|B_R(v)| \leq e^{R^\eta}$ for all $v \in V(G)$ and all R sufficiently large (i.e., $R \geq R_0$ for some R_0 independent of n ; for convenience, take $R_0 = 1/\eta^1$).

Natural graph families with at most stretched-exponential volume growth include bounded-degree lattices in $\mathbb{R}^d$; e.g., finite subsets of $\mathbb{Z}^d$, the triangular and hexagonal lattices, etc., and Cayley graphs of polynomial growth groups. This notion is closely related to a quantitative version of amenability.

We show that the SW and FK dynamics on these graphs are rapidly mixing when the independent edge-percolation process on the underlying graph G has a “strong supercritical phase” which we define next. For $\tilde{p} \in (0, 1)$, let $\tilde{\pi}_G = \bigotimes_{e \in E(G)} \text{Bernoulli}(\tilde{p})$

¹ At the level of quantification of our bounds, this choice does not affect our main statements; allowing for general R_0 would simply add R_0 to the set of constants on which the bounds depend.

denote the independent edge-percolation distribution for G . We note that if $\tilde{\omega}$ is drawn from $\tilde{\pi}_G$ (i.e., $\tilde{\omega} \sim \tilde{\pi}_G$), then we can think of $\tilde{\omega}$ as, both, a vector in $\{0, 1\}^{E(G)}$ or as a subset of edges of $E(G)$. For $B \subset E(G)$, let $\tilde{\omega}(B) \in \{0, 1\}^B$ denote the state of the edges from B in $\tilde{\omega}(B)$.

Definition 2 Let $\tilde{\omega} \sim \tilde{\pi}_G$. We say that G has a *strong supercritical phase* (with parameters $\delta, \tilde{p}$) if there exists $\tilde{p} < 1$ and $\delta > 0$ such that for every $v \in V(G)$, the probability that there exists a connected set $A \ni v$ having $\ell \leq |A| \leq n/2$ with $\tilde{\omega}(\partial_e A) \equiv 0$ is at most $\exp(-\ell^{\delta/(1+\delta)})$, for all ℓ sufficiently large (again, meaning $\ell \geq \ell_0$ for some ℓ_0 independent of n , for instance for convenience $\ell_0 = 1/\delta$).

Roughly the definition says that the probability that there exists a set $A \subset V(G)$ that is connected in G , contains v , and has size at least ℓ , but does not intersect the largest component in $\tilde{\omega}$, is stretched-exponentially small in ℓ (with the exponent governed by the parameter $\delta > 0$, which as we will comment on shortly is related to the isoperimetric dimension of the underlying graph). Notice that the existence of a $\tilde{p}$ in Definition 2 implies it for all $\tilde{p}' > \tilde{p}$ by monotonicity of the strong supercritical property in $\tilde{\omega}$.

Theorem 1 *There exists $\eta_0(\delta) > 0$ and $p_0(\Delta, q, \delta, \tilde{p}) < 1$, such that for every graph G with a strong supercritical phase (with parameters $\delta, \tilde{p}$) and η -stretched-exponential volume growth for some $\eta \leq \eta_0$:*

1. *The mixing time of SW dynamics on G is $O(n^2 \log n)$ for every $p \geq p_0$.*
2. *The mixing time of the FK dynamics on G is $O(n \log n)$ for every $p \geq p_0$.*

It is natural to wonder what families of graphs have a strong supercritical phase. The nature of the supercritical phase for edge-percolation on a graph is known to be closely related to the geometric, namely isoperimetric, properties of the graph: see e.g., [1]. One would expect that general graph families with *isoperimetric dimension* at least $1 + \delta$ (meaning that $|\partial_e A| \geq |A|^{\delta/(1+\delta)}$ for all subsets $A \subset V(G)$ with $|A| \leq n/2$) have a strong supercritical phase in the sense of Definition 2. Often at sufficiently large $\tilde{p}$, a strong supercritical phase can be proven using perturbative Peierls-type arguments; by such means, subsets of $\mathbb{Z}^d$ and other lattices (e.g., hexagonal and triangular) that are uniformly at least $(1 + \delta)$ -dimensional, and planar graphs with a bounded-degree planar dual serve as concrete examples of graphs that have a strong supercritical phase. More generally, the structure of the supercritical phase in vertex-transitive graphs of polynomial growth is the subject of deep study (see e.g., [22, 25, 44] which tackle the harder problem of understanding the supercritical phase down to a sharp threshold). Natural graphs of super-polynomial but at most stretched-exponential growth are rarer, but one such family are the well-known construction of Grigorchuk groups; in these, the precise η in the stretched-exponential growth, and the nature of the graphs' supercritical phase are subjects of active research: see e.g., [35].

Our proof of Theorem 1 relies on a novel framework for controlling the rate at which discrepancies spread between two coupled low-temperature FK dynamics chains that agree inside, say, a ball of radius R around a vertex, but that may disagree outside it. This is sometimes called disagreement percolation, and we use it, after a *burn-in period* for the chain (a short period of time after which we can ensure that the certain

“typical” properties of random-cluster configurations are achieved, even though the chain has not equilibrated), to perform space-time recursions to derive our mixing time bounds. To the best of our knowledge, this is the first time disagreement percolation has been analyzed in a low-temperature setting for non-local Markov chains like the SW or FK dynamics, where the giant component could hypothetically spread disagreements instantaneously (except in the special case of $\mathbb{Z}^2$ where low and high temperatures are dual to one another). We say more about the obstacles to proving Theorem 1 using existing tools and the technical novelties in our low-temperature disagreement percolation framework in Sect. 1.4.

Since [57], bounds on the rate of disagreement percolation have been a tool used to prove a variety of other results for spin systems, including bounds on uniqueness thresholds, equivalences of spatial and temporal mixing [24], and tight lower bounds for the mixing time of the Glauber dynamics [39]. We do not explore these directions here, but our low-temperature disagreement percolation, which is self-contained to Sect. 3, opens up those same arguments for the low-temperature random-cluster model and its dynamics. For example, extending the lower bounds of [39] would show that the $O(n \log n)$ in (2) in Theorem 1 is tight; by contrast, the resulting lower bound for SW dynamics would be $\Omega(\log n)$.

1.2 Locally treelike graphs

We consider next the SW and FK dynamics on locally treelike graphs. In this setting, we establish fast mixing on graphs that have a strong supercritical phase with “ $\delta = \infty$ ” in Definition 2. That is to say that we assume true exponential tails on the boundaries of non-giant components, with a rate that goes to ∞ as $\tilde{p} \uparrow 1$, to compete with the exponential volume growth.

Definition 3 We say that G has an *exponentially strong supercritical phase* if there exists $\tilde{p}_0 < 1$ such that for every $\tilde{p} > \tilde{p}_0$ and every $v \in V(G)$, the probability that there exists a connected set $A \ni v$ having $\ell \leq |A| \leq n/2$ with $\tilde{\omega}(\partial_e A) \equiv 0$ is at most $\exp(-c_{\tilde{p}} \ell)$ for some $c_{\tilde{p}}$ going to ∞ as $\tilde{p}$ to 1.

While the notion of an exponentially strong supercritical phase is a property of independent edge-percolation on the graph, a simple geometric criterion of expansion, for instance, ensures that this property holds. In particular, if G is an α -edge-expander graph, in the sense that for all $A \subset V(G)$ such that $|A| \leq n/2$, we have $|\partial_e A| \geq \alpha|A|$, then the assumption holds for a $c_{\tilde{p}}(\alpha, \Delta) > 0$.

In this regime where exponential volume growth is permitted, we restrict to locally treelike graphs.

Definition 4 We say a graph G is (K, L) -locally treelike if for every $v \in V(G)$, the removal of at most K edges from $E(B_L(v))$ induces a tree on $B_L(v)$.

Our main result for locally treelike graphs is the following near-optimal fast mixing bound.

Theorem 2 Fix any $\varepsilon, \eta > 0$. There exists $p_0(\Delta, q, K, \tilde{p}_0, c_{\tilde{p}}, \eta, \varepsilon) < 1$ such that if G has an exponentially strong supercritical phase (with parameter $\tilde{p}_0$), minimum degree 3, and is $(K, \eta \log n)$ -locally treelike:

1. The mixing time of SW dynamics on G is $O(n^{2+\varepsilon})$ for every $p \geq p_0$.
2. The mixing time of FK dynamics on G is $O(n^{1+\varepsilon})$ for every $p \geq p_0$.

The most canonical example of a graph that satisfies all the conditions in this theorem is a Δ -regular random graph (i.e., a graph drawn uniformly at random from the set of all Δ -regular graphs on n -vertices). This is a setting that has attracted plenty of attention (see, e.g., [9, 10, 21, 27, 40]), and Theorem 2 provides fast mixing bounds for the SW and FK dynamics on these graphs at low temperatures. (Note that the bounds will hold with probability $1 - o(1)$ over the choice of the random graph.)

Unlike the sub-exponential growth setting, alternative sampling algorithms were known to exist for the Potts model on expander graphs at low temperatures using the cluster expansion and polymer dynamics (see, e.g., [5, 19, 20, 45]). Still, to our knowledge, ours is the first proof of sub-exponential mixing times for the SW and FK dynamics at low temperatures (even just for random graphs).

Regarding the proof techniques, on graphs of exponential growth, the low-temperature disagreement percolation framework used to establish Theorem 1 breaks down. Even in ideal situations like the Ising model, the optimal recursion obtained from the disagreement percolation framework would not yield rapid mixing on graphs of exponential growth. We, therefore, resort to a vastly different approach, where we utilize a burn-in phase, the censoring technique of [52], and new spatial mixing results for the random-cluster model on trees amongst a (random) class of sufficiently wired boundary conditions. The latter bound applies in settings where spatial mixing between the wired and free boundary conditions does *not* hold.

Remark 1 The bounds of Theorems 1 and 2 are stated for any integer $q \geq 2$ so that statements apply both to the SW and FK dynamics. The random-cluster model also makes sense for non-integer $q \geq 1$ and our fast mixing results for the FK dynamics apply in this level of generality. In fact, the random-cluster model is defined for $q > 0$ but has very different features (negative vs. positive correlations) when $q \in (0, 1)$. It was shown in [2] that the FK dynamics mixes in polynomial time on all graphs when $q \in (0, 1)$.

1.3 Slow mixing in worst-case graphs

We complement our fast mixing result by establishing the existence of graphs for which, even at arbitrarily low temperatures, the SW and FK dynamics slow down exponentially. This is already suggested, though not guaranteed, by the #BIS-hardness of the sampling problem at low temperatures, and our constructions will illuminate the relationship between the notion of a strong supercritical phase for the underlying edge-percolation and the slow mixing of the dynamics.

Theorem 3 Fix any $q \geq 3$ and any $p_0 < 1$. There exists $p \in (p_0, 1)$ and a sequence of graphs G_n on n vertices and maximum degree Δ such that the mixing time of the SW and FK dynamics on G_n is $\exp(\Omega(n))$.

The constructions for Theorem 3 are simple and explicit. In particular, any family of graphs H_n that have slow mixing at *some* parameter value $p_s \in (0, 1)$ —typically the location of its order/disorder phase transition—can be used as a gadget to construct augmented graphs G_n (depending on p_s and p_0) with many of the same properties as H_n (in terms of degree, rate of volume growth, etc.), and a comparable number of edges, for which the SW and FK dynamics are slowly mixing at some $p \in (p_0, 1)$. The graph augmentation leverages the *series law* of the random-cluster model to repeatedly split the edges of H_n , effectively inducing the behavior at p_s in H_n to occur in G_n at $p \in (p_0, 1)$. Using the slow mixing of SW and FK dynamics at the critical point on random regular graphs from [21] as the gadget, Theorem 3 holds even if we impose that the graph is locally treelike and has exponential volume growth. Using the slow mixing at the critical point on $(\mathbb{Z}/n\mathbb{Z})^2$ from [30], a variant of this theorem also holds for graphs of polynomial growth, but the lower bound there is of the form $\exp(\Omega(\sqrt{n}))$: see Theorem 25.

Remark 2 Let us comment on the relationship of Theorems 3 to 1 and 2, given the slow mixing constructions can either have stretched-exponential growth or be locally treelike. Even if H_n has a strongly supercritical phase for its edge-percolation, when we perform the graph augmentation with the series law, the $\tilde{p}$ for which the edge-percolation on G_n has a strongly supercritical phase is pushed closer to 1. In geometric language, this is because the isoperimetric dimension is decreasing to 1, or the edge expansion is decreasing to 0, as the edges are split in series. In turn, this makes the p_0 in Theorems 1 and 2 (above which we can prove fast mixing) larger than the p for which Theorem 3 gives slow mixing.

1.4 Proof ideas

We now discuss our proof ideas for the fast mixing results, which are the more technically involved. (Our bounds on the SW dynamics follow from the bounds on the FK dynamics by [56], so we focus on the FK dynamics.) We begin by describing some of the issues one runs into when applying standard proof approaches to general families of graphs at large p .

1.4.1 Difficulty with classical arguments

The first tool one might try is path coupling, arguing that the number of discrepancies between two configurations that differ on one edge contracts in expectation. The non-locality of the FK dynamics, however, and the presence of $\Omega(\log n)$ sized components at equilibrium means that a single discrepancy at an edge e can cause discrepancies at some $\Omega(\log n)$ many nearby edges, whereas the discrepancy only decreases if the edge e is selected to be updated. A smarter path coupling was used in [23, 42] to deduce fast mixing for the SW dynamics at high enough (but constant) temperatures, but in the low-temperature regime, their argument for fast mixing requires $p > 1 - O(1/|E(G)|)$.

Many of the early fast mixing bounds on, say, the high-temperature Ising model, are based on space-time recursions, i.e., arguments that compare the distance to stationarity across balls of time-dependent radii. When translated to the FK dynamics, this type

of argument runs into the problem that the updates in a small portion of the graph (say, a small ball around a vertex) could depend on the configuration in the entire remainder rather than a local neighborhood. The one exception to this is the approach of [49] for the torus in $\mathbb{Z}^d$, which gave an implication from the weak-spatial mixing (WSM) condition to fast mixing of the Glauber dynamics. This implication was seen to generalize to the FK dynamics in [38] (see also [32] where finite boxes with boundary conditions were allowed). WSM is known to hold for the random-cluster model at large p on $\mathbb{Z}^d$ (see e.g., [36]); however, this is a delicate property whose proofs are very geometry specific. It is the case, for example, that on locally treelike graphs like the random regular graph, WSM fails at arbitrarily large p .

At high temperatures (small p), some of the difficulties with non-locality can be handled using the fact that in the random-cluster model, all connected components are small, and information is only propagated through these connected components. For instance, such an argument was used in [14] to implement the disagreement percolation space-time recursion for the high-temperature regime on $\mathbb{Z}^2$. (In $\mathbb{Z}^2$, the high and low-temperature regimes are dual to one another, so the same argument could be performed using the dual model at low temperatures; that would be similar to the work of [50] on the Ising SW dynamics.)

1.4.2 Low-temperature disagreement percolation bounds

On graphs where the low-temperature random-cluster model does not have a natural high-temperature dual model, however, even at equilibrium, the non-locality of the dynamics is hypothetically not confined since the giant component percolates through the entire graph. The starting point for many of our observations is that a (well-connected) giant component does not create non-local dependencies on its own. In particular, if two configurations that agree at distance R away from an edge e induce different marginals on e , it must be the case that in one of the two configurations, either e is incident to a *non-giant* component of size at least R , or it disconnects a portion of the giant of size at least R from the 2-connected core of the giant. Whereas the giant component percolates throughout the whole graph, we show that under the assumption of a strong supercritical phase (Definition 2), these non-giant, or non-2-connected core connections have (stretched) exponential tails in FK dynamics configurations after an $O(n)$ burn-in period. (In the first $O(n)$ steps, disagreements can spread arbitrarily quickly.)

There are various other delicate points in implementing this argument, both combinatorial and probabilistic in nature, that we describe in greater detail in Sects. 3 and 4. These include having to carefully approach various counting arguments and union bounds due to the non-localities and possible stretched exponential volume growth: see Remark 3 and the proof strategy described in Sect. 4.2. In all, we are able to obtain a space-time recursion on the probability of a disagreement at an edge after time t ; the large value of p is used as a crude initial bound on this probability, which the recursion boosts into exponential decay, leading to the optimal $O(n \log n)$ mixing time bound of Theorem 1.

1.4.3 Mixing after a burn-in phase on locally treelike graphs

When the volume growth is exponentially fast, the bounds and resulting space-time recursions from disagreement percolation break. Our approach here is therefore closer in inspiration to high-temperature arguments from [10, 11] (also [51] in the Ising setting). Those papers localized the dynamics to the treelike balls $B_R(v)$ of the underlying graph using the censoring technique of [52] and used the high-temperature uniqueness on trees to reason that if two censored dynamics chains mix in $B_R(v)$ with their respective boundary conditions, then they are coupled at the root of the ball with high probability. The mixing time on the local balls was relatively simple to deduce since the trees would have nearly free boundary conditions, which induce product chains.

In our low-temperature setting, the key intuition is that after a burn-in period, the boundary conditions induced on balls of radius $\eta \log n$ are “sufficiently wired”, i.e., that they have one (random) linear-sized wired component, and only $O(1)$ many other $O(1)$ -sized components. Given this, to get Theorem 2, we show that FK dynamics on trees with such boundary conditions mix in polynomial time, and that although there is no WSM between the free and wired boundary conditions on trees, two sufficiently wired boundary conditions do induce similar marginals on the root. This latter step requires a careful revealing scheme to prove spatial mixing on trees with randomly wired boundary conditions; that is the content of Sect. 5.1.

1.4.4 On the strong supercritical phase condition

We conclude by remarking on whether the notions of strong supercritical phase in Definitions 2 and 3 could be relaxed. One attempt at such a relaxation would be to only ask that the independent $\hat{p}$ -edge percolation have a unique giant component (of arbitrarily small linear size) and exponential, or stretched-exponential, tails on all its non-giant components. For technical reasons related to the fact that this is a non-monotone criterion, our proofs do not go through with this weaker notion of supercriticality. Nonetheless, we expect that such a condition could be sufficient for the fast mixing of the FK and SW dynamics.

2 Notation and preliminaries

In this section, we outline our global notation and describe some preliminaries on the random-cluster model and the FK dynamics. Throughout the paper, n will be assumed to be sufficiently large. We also use C to denote a generic constant $C > 0$, not depending on n , which may vary from line to line. Our underlying graph will be $G = (V(G), E(G))$ and will have n vertices and maximum degree Δ .

For an edge-subset $A \subset E(G)$, we write $V(A)$ for the set of vertices contained in edges in A , though we sometimes abuse notation and write $v \in A$ for $v \in V(A)$. Its vertex boundary ∂A is the set of vertices in A with neighbors in $A^c = E(G) \setminus A$. Its (outer) edge-boundary $\partial_e A$ is the set of edges in $E(G) \setminus A$ that have one end-point in $V(A)$ and one endpoint in $V(G) \setminus V(A)$. We use $C_v(A)$ to denote the connected component of v in the subgraph $(V(G), A)$. An edge e is a *cut-edge* in A if there is

a vertex v for which $\mathcal{C}_v(A \cup \{e\}) \neq \mathcal{C}_v(A \setminus \{e\})$. We use $\mathcal{C}_1(A)$ to denote the largest component in ω (chosen arbitrarily if two have the same size).

2.1 The random-cluster model

The random-cluster model with parameters $p \in (0, 1)$ and $q > 0$ is a probability distribution over edge-subsets $\omega \subseteq E(G)$, equivalently identified with $\omega \in \{0, 1\}^{E(G)}$, given by

$$\pi_{G,p,q}(\omega) \propto p^{|\omega|} (1-p)^{|E(G)|-|\omega|} q^{k(\omega)}, \quad (3)$$

where $k(\omega)$ denotes the number of connected components in $(V(G), \omega)$. When clear from context, we drop p and q and sometimes G from the notation. The random-cluster model satisfies the following *domain Markov property*: for $A \subset E(G)$, conditional on $\omega(E(G) \setminus A)$, the distribution of $\omega(A)$ is a random-cluster model with the same parameters, with *boundary conditions* on ∂A induced by $\omega(E(G) \setminus A)$. Random-cluster boundary conditions are defined in generality as follows.

Definition 5 Given a graph G and a vertex subset ∂B , a boundary condition ξ on ∂B is a partition of ∂B . The random-cluster model on G with boundary conditions ξ , denoted π_G^ξ is defined as in (3), except that all components intersecting the same element of ξ are identified (“wired”) when counting the number of components $k(\omega)$.

Certain important boundary conditions are the *wired* one, denoted $\mathbf{1}$, where all vertices of ∂A are in the same component of ξ ; the *free*, denoted $\mathbf{0}$, where all vertices of ∂A are in distinct components of ξ , and if A is a subgraph of G , those *induced* by $\omega(E(G) \setminus E(A))$, meaning vertices of ∂A are in the same component of ξ if and only if they are connected through $\omega(E(G) \setminus E(A))$. In this paper, we restrict our attention to the case of $q \geq 1$ where the model exhibits positive correlations. As a consequence, given two boundary conditions ξ and ξ' on G , where $\xi \geq \xi'$ (meaning ξ is a coarser partition than ξ'), we have $\pi_G^\xi \succeq \pi_G^{\xi'}$.

2.2 Mixing times

For a Markov chain $(X_k)_k$ on a finite state space Ω with transition matrix P , reversible with respect to a distribution μ , its *mixing time* is defined as:

$$t_{\text{MIX}} = t_{\text{MIX}}(1/4), \quad \text{where} \quad t_{\text{MIX}}(\varepsilon) = \min\{k : \max_{x_0 \in \Omega} \|\mathbb{P}(X_k^{x_0} \in \cdot) - \mu\|_{\text{TV}} \leq \varepsilon\}, \quad (4)$$

where $X_k^{x_0}$ indicates that X_k is initialized from x_0 , and where $\|\cdot\|_{\text{TV}}$ denotes total-variation distance. The total-variation distance to μ satisfies a sub-multiplicativity property, whereby $t_{\text{MIX}}(\varepsilon) \leq t_{\text{MIX}} \cdot \log_2(2/\varepsilon)$.

2.3 FK dynamics

Recall the definition of (discrete-time) FK dynamics from the introduction. It will be preferable in our proofs to work with the continuous-time FK dynamics $(X_t)_{t \geq 0}$. In this variant, the edges of $E(G)$ are assigned rate-1 Poisson clocks, and if the clock at an edge e rings at time t , we make an update according to (2). It is a standard fact that the mixing time of the discrete-time chain is comparable, up to constants, with $|E(G)|$ times the mixing time of the continuous-time process. In particular, it suffices to show an $O(\log n)$ bound for Theorem 1 and an n^ε bound for Theorem 2 for the continuous-time FK dynamics. FK dynamics updates with boundary conditions ξ are like (2), except that the cut-edge status of e is determined taking into account the wirings of the components of ξ .

The FK dynamics is monotone, meaning that if $x_0 \geq y_0$ (under the natural partial order on subsets) then $X_t^{x_0} \geq X_t^{y_0}$ for all $t \geq 0$, where $\geq$ denotes stochastic domination. I.e., there exists a *grand coupling* of all the Markov chains $\{(X_t^{x_0})_t\}_{x_0 \in \Omega_{\text{RC}}}$ (generated by independent Poisson clocks and independent sequences of i.i.d. $\text{Unif}[0, 1]$ random variables at every edge) such that $X_t^{x_0} \geq X_t^{y_0}$ for all $x_0 \geq y_0$ and $t \geq 0$.

A further monotonicity property we use is with respect to the independent edge-percolation (the random-cluster model with $q = 1$). Recall $\hat{p}$ from (2); it is standard fact that $\pi_{G,p,q} \geq \pi_{G,\hat{p},1}$ (see (3.23) in [36]).

3 Low-temperature disagreement percolation

In this section, we develop the FK dynamics disagreement percolation framework that works at sufficiently low temperatures, in particular in the presence of a giant component. In reality, this new disagreement percolation bound works simultaneously at high and low temperatures and localizes the spread of disagreements even in the presence of a large giant component, so long as all other components (and portions of the giant dangling off of its 2-connected core) are small. Moreover, it can work on graphs that have volume growth up to a stretched exponential, which requires new ideas: see Remark 3.

In this section, G can be an arbitrary graph of maximum degree Δ . We fix an arbitrary $o \in V(G)$ and $R > 0$, and let $B_R = B_R(o)$. The dependencies on o will be kept implicit. The fundamental building blocks of our disagreement set will be *finite-connectivity* clusters; these will disentangle the non-locality of the giant component, which percolates at low temperature, from the edges through which disagreements arise.

Definition 6 Define the *finite (or non-giant) component* of a vertex v in a random-cluster configuration ω as $\mathcal{C}_v(\omega \setminus E(\mathcal{C}_1(\omega)))$, and denote it by $\mathcal{C}_v^{\neq 1}(\omega)$.

Since the FK dynamics updates at edge $e = \{u, v\}$ are oblivious to the state of e in the configuration, and only care about the connectivity of u and v in $\omega \setminus e = \omega \setminus \{e\}$, we consider the finite component of a vertex with respect to the configuration $\omega \setminus e$ rather than ω itself. Specifically, we often consider $\mathcal{C}_v^{\neq 1}(\omega \setminus e)$ for an edge e that is incident to $\mathcal{C}_v(\omega)$.

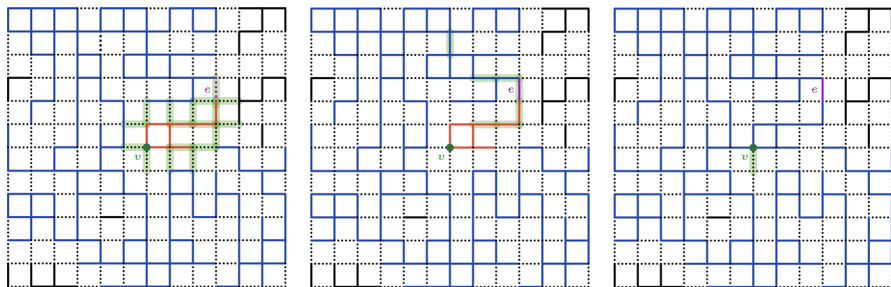


Fig. 1 For a vertex v (dark green) and edge e (purple), the sets $\mathcal{C}_v^{\neq 1}(\omega \setminus e)$ (edges in red) and $\text{CE}_v^{\neq 1}(\omega)$ (edges highlighted in green) are shown in three different cases. Left: v is not part of the giant (blue edges) in ω (blue and black edges). Middle: v is part of the giant component but not of its 2-connected core. Right: v in the 2-connected core of the giant

Definition 7 Let $\text{CE}_v^{\neq 1}(\omega)$ be the set of cut-edges in ω that are in B_R and incident to $\mathcal{C}_v^{\neq 1}(\omega \setminus e)$; i.e.,

$$\text{CE}_v^{\neq 1}(\omega) := \{e \in E(B_R) : e \in \text{Cutedge}(\omega), e \sim \mathcal{C}_v^{\neq 1}(\omega \setminus e)\},$$

where $\text{Cutedge}(\omega)$ denotes the set of cut-edges in ω . Here we are using the notation $e \sim H$ for a subset $H \subset E$ if $V(e) \cap V(H) \neq \emptyset$.

We refer to Fig. 1 for some illustrative depictions of such sets in $\mathbb{Z}^2$ (this is easiest for visualization, but it is key that our definitions do not rely on properties of $\mathbb{Z}^2$ like its dual graph, and thus work on general graphs). We also refer to the proof of Proposition 4 which yields additional insight into these constructions.

Suppose (X_t) and (Y_t) are two instances of the FK dynamics on G coupled via the grand coupling introduced in Sect. 2.3; suppose also that $X_0(E(B_R)) = Y_0(E(B_R))$.

Definition 8 Iteratively construct what we call the *disagreement set* as follows. Let $(t_i)_{i \geq 1}$ be the times of the clock rings in $E(B_R)$, let $t_0 = 0$, and let $I_i = [t_{i-1}, t_i)$. Then

1. Initialize $\mathcal{D}_t = E(G) \setminus E(B_R)$ for $t \in I_1$.
2. Suppose e_i is the edge whose clock rings at time t_i . If e_i is in $E(B_R) \setminus \mathcal{D}_{t_i^-}$ and e_i is in $\text{CE}_v^{\neq 1}(Z)$ for some $v \in \partial \mathcal{D}_{t_i^-}$ and $Z \in \{X_{t_i^-}, Y_{t_i^-}\}$, let

$$\mathcal{D}_t = \mathcal{D}_{t_i^-} \cup \{e_i\} \quad \text{for all } t \in I_{i+1};$$

else, let $\mathcal{D}_t = \mathcal{D}_{t_i^-}$ for all $t \in I_{i+1}$. (We use the standard notation A_{t^-} to denote $\lim_{s \uparrow t} A_s$.)

The role of $\mathcal{D}_t$ is that it confines the set of edges on which a disagreement can possibly exist at time t .

Proposition 4 For all $t \geq 0$, we have $X_t(e) = Y_t(e)$ for all $e \notin \mathcal{D}_t$.

Proof We prove the claim inductively. It holds for all $t \in I_1$ since we assumed $X_t(e) = Y_t(e)$ for $e \in E(B_R)$, and no clock rings occur in the interval I_1 . Supposing it holds for I_i , the only way it can not hold for $t \in I_{i+1}$ is if the disagreement arises at the edge $e_i = \{u_i, v_i\}$ when the clock rings at time t_i . If $e_i \in \mathcal{D}_{t_i}^-$, then since $\mathcal{D}_{t_i}^- \subset \mathcal{D}_t$, we have the claim. Per (2), if $e_i \notin \mathcal{D}_{t_i}^-$, in order for a disagreement to arise at e_i , it must be the case that e_i is a cut-edge in one of $X_{t_i}^-, Y_{t_i}^-$ but not in the other; that is, that $e_i \in \text{Cutedge}(Z)$ for a $Z \in \{X_{t_i}^-, Y_{t_i}^-\}$, but $e_i \notin \text{Cutedge}(\hat{Z})$ for $\hat{Z} = \{X_{t_i}^-, Y_{t_i}^-\} \setminus \{Z\}$. In $\hat{Z} \setminus \{e_i\}$, the two endpoints of e_i must be connected. Since $X_{t_i}^-(\mathcal{D}_{t_i}^c) = Y_{t_i}^-(\mathcal{D}_{t_i}^c)$, this can only happen via a pair of paths from u_i and v_i that reach $\partial \mathcal{D}_{t_i}^-$ in $\mathcal{D}_{t_i}^c$. One of these paths must be in $Z \setminus E(\mathcal{C}_1(Z \setminus \{e_i\}))$, since if both are in $E(\mathcal{C}_1(Z \setminus \{e_i\}))$, then u_i and v_i are in the same connected component of $Z \setminus e_i$, contradicting the claim that e_i was a cut-edge in Z . As such, it must be that $e_i \in \text{CE}_v^{\neq 1}(Z)$ for some $v \in \partial \mathcal{D}_{t_i}^-$. $\square$

We now define an event on the realizations of the coupling $(X_t, Y_t)_{t \geq 0}$ that guarantees that disagreements are unlikely to spread rapidly. This will be a bound on the size of finite connections in $E(B_R)$, as well as a bound on the number of cut-edges in a finite component, which we will later show holds with high probability for the FK dynamics after a burn-in for G satisfying the conditions of Theorem 1: see Proposition 7

Definition 9 A random-cluster configuration ω is in $\mathcal{E}_{\ell, \alpha, \gamma}$ if:

$$(1) \quad \max_{v \in B_R} \max_{e \in E(B_R)} \text{diam}(\mathcal{C}_v^{\neq 1}(\omega \setminus e)) \leq \ell^\alpha, \quad \text{and} \quad (2) \quad \max_{v \in B_R} |\text{CE}_v^{\neq 1}(\omega)| \leq \Delta \ell^\gamma.$$

In the applications in Sect. 4, we will take $\alpha = \gamma$ (depending on the δ for which Definition 2 holds); however, we write this section for general α, γ in case there are better choices in specific situations.

Proposition 5 Suppose $(X_t)_t, (Y_t)_t$ are FK dynamics chains, coupled by the grand coupling, and such that $X_0(E(B_R)) = Y_0(E(B_R))$. There exists $\varepsilon_0(\Delta) > 0$ such that for all $\varepsilon < \varepsilon_0$ and all $t \leq \varepsilon R / \ell^{\alpha+\gamma}$,

$$\mathbb{P}\left(X_t(B_{R/2}) \neq Y_t(B_{R/2}), \bigcap_{s=0}^t \{X_s, Y_s \in \mathcal{E}_{\ell, \alpha, \gamma}\}\right) \leq C |\partial B_R| \exp(-R/(2\ell^\alpha)).$$

Proof By Proposition 4, the probability of the event under consideration is at most that of the event $\{\mathcal{D}_t \cap B_{R/2} \neq \emptyset\} \cap \mathcal{E}_{\ell, t}$ where for ease of notation, $\mathcal{E}_{\ell, t} := \bigcap_{s=0}^t \{X_s, Y_s \in \mathcal{E}_{\ell, \alpha, \gamma}\}$. We construct a witness to that pair of events as follows.

Let f_0 be the edge whose clock rang at time $t_{j_0} := \inf\{s : B_{R/2} \cap \mathcal{D}_s \neq \emptyset\}$ (note that $f_0 \in E(B_{R/2})$). Let w_0 be the vertex in $\partial \mathcal{D}_{t_{j_0}}^-$ for which $f_0 \in \text{CE}_{w_0}^{\neq 1}(Z_0)$ for $Z_0 \in \{X_{t_{j_0}}^-, Y_{t_{j_0}}^-\}$ (if there are multiple choices for w_0 or Z , we choose arbitrarily). Given $(f_j, w_j, Z_j)_{j < i}$, we construct the witness iteratively as follows:

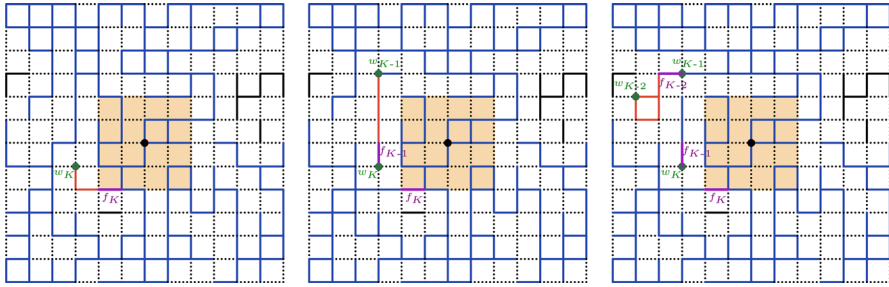


Fig. 2 Three steps of the construction of the witness are shown. The ball $B_{R/2}$ is the highlighted region. For each i , the edges of the finite-connectivity cluster of w_i (green) to f_i (purple) in Z_i (blue and black edges) are depicted in red. Note that the configuration changes from left to right, depicting the evolution of the dynamics (backwards in time)

- Let f_i be the first edge incident to w_{i-1} to be included in $(\mathcal{D}_s)_{s \geq 0}$; i.e., f_i 's clock rang at time

$$t_{j_i} := \inf\{s : w_{i-1} \in \mathcal{D}_s\}$$

- Let w_i be the vertex in $\partial \mathcal{D}_{t_{j_i}}^-$ and $Z_i \in \{X_{t_{j_i}}^-, Y_{t_{j_i}}^-\}$ for which $f_i \in \text{CE}_{w_i}^{\neq 1}(Z_i)$.

(Again, ambiguities are resolved arbitrarily.) Under this construction, the event $X_t(B_{R/2}) \neq Y_t(B_{R/2})$ implies the existence of a witness $(f_i, w_i, Z_i)_{i=0}^K$ for some K such that

1. $f_0 \in E(B_{R/2})$ and $w_K \in \partial B_R$;
2. $f_i \in \text{CE}_{w_i}^{\neq 1}(Z_i)$ for all i ;
3. $w_{i-1} \in f_i$ for all i ;
4. the clock ring at time t_{j_i} is at edge f_i .

Notice that this construction is done *backwards in time*, i.e., $t_{j_0} > t_{j_1} > \dots > t_{j_K}$. See Fig. 2 for a depiction.

We will show that the probability that there exists such a witness and the event $\mathcal{E}_{\ell,t}$ occurs satisfies the claimed bound. On the event $\mathcal{E}_{\ell,t}$, it must be the case that $K \geq R/(2\ell^\alpha)$ since the distances between f_i and f_{i-1} are bounded by ℓ^α . Furthermore, for any witness $(f_i, w_i, Z_i)_{i=0}^K$, there is a projection, which we also call a witness, $(f_i, w_i, L_i)_{i=0}^K$ where the label $L_i \in \{X, Y\}$ indicates whether $Z_i = X_{t_{j_i}}^-$ or $Z_i = Y_{t_{j_i}}^-$.

The total number of clock rings in G in $[0, t]$ has a Poisson($t|E(G)|$) distribution; let M denote this quantity. Note that by standard Poisson concentration, we have $\mathbb{P}(M \geq 4t|E(G)|) \leq \exp(-t|E(G)|)$. Let us work on the event that $M \leq 4t|E(G)|$. Let $\mathcal{T} = \{t_1, \dots, t_M\}$ be the sequence of clock ring times in $|E(G)|$. We start by bounding $\mathbb{P}(X_t(B_{R/2}) \neq Y_t(B_{R/2}), \mathcal{E}_{\ell,t})$ by

$$e^{-t|E(G)|} + \max_{M \leq 4t|E(G)|} \max_{\mathcal{T}} \sum_{K \geq R/(2\ell^\alpha)} 2^K \max_{L \in \{X, Y\}^K} \mathbb{P}(\exists (f_i, w_i)_{i=0}^K : (f_i, w_i, L_i)_{i=0}^K \text{ is witness}, \mathcal{E}_{\ell,t} | \mathcal{T}),$$

where the conditioning on $\mathcal{T}$ indicates conditioning on the clock ring times in $E(G)$ in $[0, t]$ being $\mathcal{T}$ (but importantly not revealing their location yet).

Fix any $M \leq 4t|E(G)|$ and any realization $\mathcal{T}$ and consider the probability on the right. For any subset of times $J = \{j_K, \dots, j_0\}$, given $\mathcal{T}$, we denote by $W_{J,L}$ the event that there exist $(f_i, w_i)_{i=0}^K$ such that the triple $(f_i, w_i, L_i)_{i=0}^K$ is a witness and the clock ring at f_i is at time t_{j_i} for $i = 0, \dots, K$. There being $\binom{M}{K}$ choices of J , for any $L \in \{X, Y\}^K$:

$$\mathbb{P}(\exists (f_i, w_i)_{i=0}^K : (f_i, w_i, L_i)_{i=0}^K \text{ is witness}, \mathcal{E}_{\ell,t} \mid \mathcal{T}) \leq \max_J \binom{M}{K} \mathbb{P}(W_{J,L}, \mathcal{E}_{\ell,t} \mid \mathcal{T}).$$

For $s \geq 0$, let $\mathcal{F}_s$ be the σ -algebra generated by the grand coupling up to time s . We will now sequentially condition on $\mathcal{F}_{t_{j_i}^-}$, and enumerate over the possible choices for the edge f_i (of which there will be at most $\Delta \ell^\gamma$ per item (2) of Definition 9), and then ask that the clock ring at time t_{j_i} be at f_i .

More precisely, if for an edge g , $\mathcal{A}_l^g$ is the event that in the witness $f_l = g$, then we have for every $i < K$,

$$\begin{aligned} & \max_{g_{i+1}, \dots, g_K} \mathbb{P}(W_{J,L}, \mathcal{E}_{\ell,t} \mid \mathcal{T}, \mathcal{F}_{t_{j_i}^-}, (\mathcal{A}_l^{g_l})_{l>i}) \\ & \leq \max_{g_{i+1}, \dots, g_K} \sum_{w_i \in g_{i+1}} \sum_{g_i \in \text{CE}_{w_i}^{\neq 1}(Z_i)} \mathbb{P}(\mathcal{A}_i^{g_i} \mid \mathcal{T}, \mathcal{F}_{t_{j_i}^-}, (\mathcal{A}_l^{g_l})_{l>i}) \\ & \quad \mathbb{P}(W_{J,L}, \mathcal{E}_{\ell,t} \mid \mathcal{T}, \mathcal{F}_{t_{j_i}^-}, (\mathcal{A}_l^{g_l})_{l \geq i}) \\ & \leq \frac{2\Delta \ell^\gamma}{|E(G)|} \cdot \max_{g_i, \dots, g_K} \mathbb{P}(W_{J,L}, \mathcal{E}_{\ell,t} \mid \mathcal{T}, \mathcal{F}_{t_{j_i}^-}, (\mathcal{A}_l^{g_l})_{l \geq i}). \end{aligned} \quad (5)$$

In the above, and in what follows in this proof, when conditioning on a σ -algebra, we mean the inequalities to hold almost surely, i.e., for almost surely every realization of the random variables generating the σ -algebra. Here, the first inequality is a union bound over the potential choices of w_i and then the potential choices of g_i in the witness. (Notice that conditional on $\mathcal{F}_{t_{j_K}^-}$, the configuration Z_K can be read-off from its label L_K .) For the second inequality, we used the definition of $\mathcal{E}_{\ell,t}$ to bound the number of summands in the second line by $2\Delta \ell^\gamma$, and we used the fact that conditionally on $\mathcal{T}$, the locations of the clock rings are independent and uniform on $E(G)$, so given $\mathcal{T}$ and $\mathcal{F}_{t_{j_i}^-}$, the probability that the clock ring at time t_{j_i} is at a fixed edge g_i is $1/|E(G)|$.

We can condition the right-hand side of (5) further on $\mathcal{F}_{t_{j_{i-1}}^-}$ (recalling that $t_{j_i} < t_{j_{i-1}}$) to arrive at the following relation between the probabilities for index i and index $i-1$:

$$\begin{aligned} & \max_{g_{i+1}, \dots, g_K} \mathbb{P}(W_{J,L}, \mathcal{E}_{\ell,t} \mid \mathcal{T}, \mathcal{F}_{t_{j_i}^-}, (\mathcal{A}_l^{g_l})_{l>i}) \leq \frac{2\Delta \ell^\gamma}{|E(G)|} \cdot \max_{g_i, \dots, g_K} \\ & \mathbb{P}(W_{J,L}, \mathcal{E}_{\ell,t} \mid \mathcal{T}, \mathcal{F}_{t_{j_{i-1}}^-}, (\mathcal{A}_l^{g_l})_{l \geq i}). \end{aligned}$$

The same inequality holds for $i = K$ with an extra multiplicative factor of $|\partial B_R|$ for the initial choice of w_K . Iterating this over all i , we arrive at the following bound on $\mathbb{P}(X_t(B_{R/2}) \neq Y_t(B_{R/2}), \mathcal{E}_{\ell,t})$:

$$|\partial B_R| \sum_{K \geq R/(2\ell^\alpha)} 2^K \max_{M \leq 4t|E(G)|} \binom{M}{K} \left(\frac{2\Delta\ell^\gamma}{|E(G)|} \right)^K + e^{-t|E(G)|} \leq |\partial B_R| \sum_{K \geq R/(2\ell^\alpha)} \left(\frac{8\Delta\ell^\gamma e t}{K} \right)^K + e^{-t|E(G)|}.$$

At this stage, we see that if $t \leq \varepsilon R/(16\Delta e\ell^{\alpha+\gamma})$ for $\varepsilon < 1/e$, then this is at most

$$|\partial B_R| \varepsilon^{R/(2\ell^\alpha)} \sum_{j \geq 0} \varepsilon^j + e^{-t|E(G)|} \leq C|\partial B_R| e^{-R/(2\ell^\alpha)} + e^{-t|E(G)|}.$$

The term $e^{-t|E(G)|}$ is absorbed since we have $R \leq \text{diam}(G) \leq E(G)$ trivially. $\square$

Remark 3 Beyond the low-temperature construction of the disagreement region, we point out a subtlety in the above that may have gone unnoticed. In disagreement percolation bounds for high-temperature FK dynamics (e.g., in [14]), the typical analog of $\mathcal{E}_{\ell,\alpha,\gamma}$ is simply that the largest cluster in B_R has volume at most ℓ . When counting the number of possible witnesses, one takes $|B_\ell(w_i)|$ as a worst-case bound for the number of locations of the next disagreement along the chain in the witness. If the volume growth is stretched exponential, however, this does not work. The careful conditioning in (5) was essential to only count those edges $\text{CE}_{w_i}^{\neq 1}$ that could be vulnerable to be the next edge in the witness, rather than the entire volume of a ball, keeping the count to a polynomial even in the presence of exponential volume growth.

4 Fast mixing of FK dynamics on graphs of sub-exponential growth

With the bound on the speed of information propagation at low temperatures from the previous section on hand, we proceed to establish Theorem 1. The event $\mathcal{E}_{\ell,\alpha,\gamma}$ from Definition 9 was crucial to controlling the speed of disagreement propagation, and our first aim (Sect. 4.2) is to establish that after an $O(1)$ time burn-in period, $\mathcal{E}_{\ell,\alpha,\gamma}$ holds for a further $O(1)$ amount of time. Then in Sect. 4.3, we will build a space-time recursion to establish the desired mixing time bound.

4.1 Dominating edge-percolation after a burn-in period

We start with a simple estimate showing that after an $O(1)$ (continuous-time) burn-in period, the FK dynamics started from any initialization stochastically dominates the edge-percolation at a parameter arbitrarily close to $\hat{p} = \frac{p}{q(1-p)+p}$. This will be crucial to many of our arguments throughout the paper.

Lemma 6 Fix p, q, δ . There exists $T_0(\delta)$ such that X_t^0 stochastically dominates $\pi_{\hat{p}-\delta,1}$ for all $t \geq T_0$. This also holds conditioned on $\mathcal{T}_{[T_0, \infty)}$ (the σ -algebra generated by the clock rings from time T_0 on).

Proof Consider any edge e . Uniformly over all the possible randomness (Poisson clocks and uniform random variables) on edges of $E(G) \setminus \{e\}$, as well as all clock rings at e after time T_0 (a to be determined constant depending only on δ), on the event that the clock at e has rung by time t , its distribution stochastically dominates $\text{Ber}(\hat{p})$ per (2). The result follows if we let T_0 be large enough that the probability that the clock at e has not rung by time T_0 is less than δ (this is independent of the clock rings at e after T_0). $\square$

We consistently use the notation $\tilde{\omega}$ for independent edge-percolation processes on $E(G)$ with parameter $\tilde{p}$, i.e., $\tilde{\omega} \sim \pi_{G, \tilde{p}, 1}$. For ease of notation, we simply write $\tilde{\pi}_G$ for the law of $\tilde{\omega}$ on G .

4.2 Burnt-in FK dynamics are in $\mathcal{E}_\ell$

Our next aim is to show that burnt-in FK dynamics configurations are in $\mathcal{E}_{\ell, \alpha, \gamma}$ with high probability. Recall the main assumption on our underlying graphs for Theorem 1, that the independent percolation on them has a strongly supercritical phase: Definition 2.

Recall the event $\mathcal{E}_{\ell, \alpha, \gamma}$ from Definition 9 that governed the size of regions through which disagreements could possibly spread. In what follows, we fix $\delta > 0$ given to us by Definition 2, and let

$$\mathcal{E}_\ell := \mathcal{E}_{\ell, (1+\delta)/\delta, (1+\delta)/\delta}, \quad \text{i.e., } \alpha = \gamma = (1 + \delta)/\delta. \quad (6)$$

For ease of notation, we use $\alpha = (1 + \delta)/\delta$ in the below. We continue to imagine a fixed vertex $o \in V(G)$, and fixed R large, but independent of the graph, and let $B_R = B_R(o)$; events and sets from the previous section are all defined with respect to this ball. Finally, ℓ, R can be assumed to be sufficiently large (depending on η, δ). Our main result in this subsection is the following.

Proposition 7 Suppose G satisfies Definition 2 and has η -stretched exponential growth for η less than some $\eta_0(\delta)$. There exists $T_0(\delta, \eta, q)$ such that for every initial configuration ω_0 and every $t \geq T_0$,

$$\mathbb{P}\left(\bigcup_{s=t}^{2t} \{X_s^{\omega_0} \notin \mathcal{E}_\ell\}\right) \leq Cte^{R^\eta} e^{-\ell}.$$

Proof strategy. Establishing Proposition 7 is quite a bit more involved than it would be in a high-temperature setting (i.e., for small p). This is due both to the delicate graph-theoretic aspects of the event $\mathcal{E}_\ell$, its non-monotonicity, and its non-locality. In particular, the last one means that in the time interval $[t, 2t]$ the number of edge updates that could hypothetically cause the FK dynamics to leave $\mathcal{E}_\ell$ are $t|E(G)|$, rather than $t|B_R|$; a naive union bound over this number would fail. We outline our strategy as follows:

1. In Definition 10 below, we define a proxy event $\mathcal{G}_\ell$ which is monotone, ensures that $\mathcal{E}_\ell$ occurs, and is more “local” than $\mathcal{E}_\ell$. The relationship to $\mathcal{E}_\ell$ under minimal assumptions on G is in Lemma 8.
2. Using Definition 2, we will show in Lemma 10 that $\mathcal{G}_\ell$ holds with high probability for independent edge-percolation on G with a large enough parameter $\tilde{p}$. Since $\mathcal{G}_\ell$ is monotone, we can translate this bound to the FK dynamics after an $O(1)$ burn-in time per Lemma 6.
3. We then perform a careful “union bound” over the update times between $s \in [t, 2t]$. This could be a problem since there are order $|E(G)|$ many updates in this time interval, whereas the probability of $\mathcal{G}_\ell^c$ is only exponentially small in the local quantity ℓ . Importantly, though, we use that the event $\mathcal{G}_\ell$ is “localized” to reason that far away edge updates are unlikely to induce a change in $\mathcal{G}_\ell^c$, in a summable manner. This argument is executed via Lemma 9 in the proof of Proposition 7.

Let us begin by defining the proxy event $\mathcal{G}_\ell$ and its variant $\tilde{\mathcal{G}}_m$ for $m \geq \ell$.

Definition 10 Define the event $\mathcal{G}_\ell$ as the event that there does not exist a connected set A intersecting B_R having $\ell^\alpha \leq |A| \leq n/2$, and an edge $e \in \partial_e A$ such that $\omega(\partial_e A \setminus \{e\}) \equiv 0$.

Define the event $\tilde{\mathcal{G}}_m$ as the event that there does not exist a connected set A intersecting B_R , of size $m^\alpha \leq |A| \leq n/2$, and a pair of edges $e_1, e_2 \in \partial_e A$ such that $\omega(\partial_e A \setminus \{e_1, e_2\}) \equiv 0$.

Notice that, unlike $\mathcal{E}_\ell^c$, the event $\mathcal{G}_\ell^c$ is a monotone decreasing event, since it is the union (over A, e) of decreasing events. Similarly, the event $\tilde{\mathcal{G}}_m^c$ is a decreasing event.

The following graph theoretic lemma demonstrates that $\mathcal{G}_\ell$ controls $\mathcal{E}_\ell$. It is important here to relate the number of cut-edges in the carefully constructed set of vulnerable edges in the disagreement percolation $\text{CE}_v^{\neq 1}$ to an easier quantity: the volume of a set of size smaller than $n/2$ with closed boundary.

Lemma 8 *The event $\mathcal{E}_\ell^c$ is a subset of the event $\mathcal{G}_\ell^c$.*

Proof On the complement of item (1) in Definition 9, there exists $v \in B_R$ and $e \in E(B_R)$ such that $\text{diam}(\mathcal{C}_v^{\neq 1}(\omega \setminus e)) \geq \ell^\alpha$. Letting $A = \mathcal{C}_v^{\neq 1}(\omega \setminus e)$, we notice that $|A| \geq \text{diam}(A) \geq \ell^\alpha$. At the same time, $|A| \leq n/2$ since if $|A| \geq n/2$ then $\mathcal{C}_v(\omega \setminus e) = \mathcal{C}_1(\omega \setminus e)$ and $\mathcal{C}_v^{\neq 1}(\omega \setminus e)$ would be the trivial $\{v\}$. Finally A intersects B_R since it contains $v \in B_R$, and under ω , all of $\partial_e A \setminus \{e\}$ must be closed since A is a connected component of $\omega \setminus e$.

We now show that the complement of item (2) in Definition 9 also implies $\mathcal{G}_\ell^c$. The essence of the argument is that $\text{CE}_v^{\neq 1}$ lower bounds the size of $\mathcal{C}_v^{\neq 1}(\omega \setminus e)$ for some e , but a little care must be taken due to the definition of $\text{CE}_v^{\neq 1}$. We begin by constructing a tree from the set of all $e \in \text{Cutedge}(\omega)$ that are incident to $\mathcal{C}_v(\omega \setminus e)$; note this is a larger set than those $e \in \text{Cutedge}(\omega)$ that have $e \sim \mathcal{C}_v^{\neq 1}(\omega \setminus e)$, but we will subsequently restrict to this latter set. Suppose $\mathbf{e}$ is the set of all edges e in $\text{Cutedge}(\omega)$ such that $e \sim \mathcal{C}_v(\omega \setminus e)$. We iteratively associate a tree $T_v = T_v(\omega)$ to $\{v\} \cup \mathbf{e}$ in the following natural way.

1. Identify the root of the tree with the vertex v ;

2. All cut-edges in $\mathbf{e}$ are associated to descendants of the root.
3. For a vertex w of the tree (associated to a cut-edge e_w in $\mathbf{e}$), a cut-edge $f \in \mathbf{e}$ is associated to a descendant of w if and only if it is disconnected from v by e_w in ω .

This process uniquely determines the tree since the children of $w_{i,j}$ are those descendants of $w_{i,j}$ that are not descendants of any of $w_{i,j}$'s other descendants. The fact that all of these are cut-edges also ensures that no cycles arise in the construction. Notice that the leaves of T_v are exactly $\partial_e C_v(\omega)$.

We next claim that the edges in

$$\mathbf{e}' = \{e \in \mathbf{e} : |C_v(\omega \setminus e)| \leq n/2\},$$

are a sub-tree of T_v . By definition, an edge e can only be in $\mathbf{e}$ but not in $\mathbf{e}'$ if the component $|C_v(\omega \setminus e)| > n/2$. If this occurs for an edge e associated to vertex w in the tree, any edge f associated to a descendant of w will also have $|C_v(\omega \setminus f)| > n/2$ and not be in $\mathbf{e}'$ since the difference in the component structures of $\omega \setminus e$ and $\omega \setminus f$ is that the latter has a larger C_v and one other component is correspondingly smaller. Therefore, the event that an edge is in $\mathbf{e}$ but not in $\mathbf{e}'$ is a decreasing event on the tree T_v . As such the restriction of T_v to $\{v\} \cup \mathbf{e}'$ is itself a tree, which we can call T'_v .

Select an arbitrary vertex in $\partial T'_v$, i.e., its descendants are all in T_v but not in T'_v , and call its corresponding cut-edge $e_\star$; also define $\omega_\star = \omega \setminus e_\star$. Note that the tree $T_v(\omega_\star)$ is exactly $T_v(\omega) \setminus S_\star$ where $S_\star$ are all descendants of $e_\star$. If we let $A = C_v(\omega_\star)$, then evidently $\omega(\partial_e A \setminus e_\star) \equiv 0$ since $\omega_\star(\partial_e A) \equiv 0$. Also, $|A| \leq n/2$ since otherwise $e_\star$ would not belong to T'_v . All edges of $\mathbf{e}'$ belong to $T_v(\omega_\star)$ so they are all incident to A ; therefore

$$\Delta|A| \geq |\mathbf{e}'| \geq |\mathbf{CE}_v^{\neq 1}| \geq \Delta \ell^\alpha,$$

using the fact that $\mathbf{e}' \supset \mathbf{CE}_v^{\neq 1}$. Lastly, A contains $v \in B_R$ since $v \in T_v(\omega_\star)$. Thus, A violates $\mathcal{G}_\ell$. $\square$

The following lemma relates the vulnerability of a configuration to leaving $\mathcal{G}_\ell$ by means of an edge-update at distance m from B_R to the event $\tilde{\mathcal{G}}_m^c$, allowing us to control the probability that far away updates (of which there are many in order-one continuous time) induce the dynamics to leave $\mathcal{G}_\ell$.

Lemma 9 *For any $m \geq \ell$, in order for an edge $e \notin B_{R+m^\alpha}$ to be pivotal to $\omega \in \mathcal{G}_\ell$, i.e., for $\omega \oplus \{e\} \in \mathcal{G}_\ell^c$ while $\omega \in \mathcal{G}_\ell$, it must be that $\omega \in \tilde{\mathcal{G}}_m^c$.*

Proof Suppose $\omega \in \mathcal{G}_\ell$, $e \notin E(B_{R+m^\alpha})$ such that $e \notin \omega$ and $\omega \cup \{e\} \in \mathcal{G}_\ell^c$ or $e \in \omega$ and $\omega \setminus \{e\} \in \mathcal{G}_\ell^c$.

Since $\mathcal{G}_\ell$ is an increasing event, the first case is not possible. In the second case, the removal of an edge $e \in \omega$ outside B_{R+m^α} causes a component A to become part of $\mathcal{G}_\ell^c$. Call A the set in $\omega \setminus e$ that violates $\mathcal{G}_\ell$. Then the edge e must be in $\partial_e A$, meaning that the set A is a set of size at most $n/2$, intersecting B_R and with one other edge $e_1 \in \partial_e A$ such that $\omega(\partial_e A \setminus \{e, e_1\}) \equiv 0$. Since the distance of e to B_R is at least m^α , it must be that $|A| \geq \text{diam}(A) \geq m^\alpha$. $\square$

We now turn to the probabilistic estimates. The following lemma utilizes Definition 2 to bound the probability of $\mathcal{G}_\ell^c$ (as well as of $\bar{\mathcal{G}}_m^c$) under the independent edge-percolation measure $\tilde{\pi} = \pi_{G, \tilde{p}, 1}$.

Lemma 10 *If G satisfies Definition 2 and has η -stretched exponential volume growth for $\eta < \eta_0(\delta)$, for all $\tilde{p}$ sufficiently large,*

$$\tilde{\pi}(\tilde{\omega} \notin \mathcal{G}_\ell) \leq C|B_R| \exp(-\ell),$$

for some $C(\tilde{p}, \eta, \delta)$. Similarly,

$$\tilde{\pi}(\tilde{\omega} \notin \bar{\mathcal{G}}_m) \leq C|B_R| \exp(-m).$$

Proof The lemma is almost a union bound together with Definition 2, the only distinction being that we allow one or two of the edges in $\partial_e A$ to be open. Fix a vertex v , and for every configuration $\tilde{\omega}$ in $\mathcal{G}_\ell^c$ by means of a set $A = A_v(\tilde{\omega})$ containing v , such that $\tilde{\omega}(\partial_e A \setminus e) \equiv 0$, let $\phi_e(\tilde{\omega}) = \omega \setminus e$. Evidently,

$$\frac{\tilde{\pi}(\tilde{\omega})}{\tilde{\pi}(\phi_e(\tilde{\omega}))} \leq \frac{\tilde{p}}{1 - \tilde{p}}.$$

For every $\tilde{\omega}$, the configuration $\phi_e(\tilde{\omega})$ has a set A intersecting v such that $\ell^\alpha \leq |A| \leq n/2$ and such that $(\phi_e(\tilde{\omega}))(\partial_e A) \equiv 0$, the probability of which is governed by Definition 2. Furthermore, if this A has size exactly r , the set of all pre-images of a single $\tilde{\omega}$ under the map ϕ_e is bounded by the set of all e such that $d(v, e) \leq r$, which is at most e^{r^η} per the η -stretched exponential growth assumption. Putting this all together, we get

$$\tilde{\pi}(\tilde{\omega} \notin \mathcal{G}_\ell) \leq \sum_{v \in B_R} \sum_{r=\ell^\alpha}^{n/2} \sum_{\tilde{\omega} \in \mathcal{G}_\ell^c: |A_v(\tilde{\omega})|=r} \tilde{\pi}(\tilde{\omega}) \leq \frac{\tilde{p}}{1 - \tilde{p}} \sum_{v \in B_R} \sum_{r=\ell^\alpha}^{n/2} e^{r^\eta} e^{-r^{\delta/(1+\delta)}}. \quad (7)$$

As long as η is smaller than $\alpha = (1 + \delta)/\delta$, the above quantity is at most some constant (depending on $\eta, \delta, \tilde{p}$) times $|B_R|e^{-\ell}$.

The argument for $\bar{\mathcal{G}}_m$ is essentially identical, with the only differences being that in (7), the pre-factor $\tilde{p}/(1 - \tilde{p})$ is squared, and the number of choices of *two* edges that could be closed contributes a factor of e^{2r^η} instead of e^{r^η} . $\square$

The last lemma we need towards proving Proposition 7 is one for bounding the number of clock rings in B_{R+m^α} . This follows from a standard Poisson tail bound together with a union bound.

Lemma 11 *For a set A , let $N_A^{[t, 2t]}$ be the number of clock rings in A in the time interval $[t, 2t]$. We have*

$$\mathbb{P}\left(\bigcup_{m \geq 1} \{N_{B_{R+m^\alpha}}^{[t, 2t]} \geq 4t|B_{R+m^\alpha}|\}\right) \leq \exp(-t|B_R|).$$

Proof The number of clock rings in a set A in an interval of length $t > 0$ is distributed as a Poisson with rate $|A|t$. Therefore,

$$\mathbb{P}(N_A^{[t, 2t]} \geq 4t|A|) \leq \exp(-2t|A|).$$

By a union bound, we then get

$$\mathbb{P}\left(\bigcup_m \{N_{B_{R+m^\alpha}}^{[t, 2t]} \geq 4t|B_{R+m^\alpha}|\}\right) \leq \sum_m e^{-2t|B_{R+m^\alpha}|}.$$

Using that $|B_{R+m^\alpha}|$ is at least $|B_R| + m^\alpha$ by the fact that m only ranges until expanding the radius doesn't add any vertices, and using that $t > 0$, this sums out to give $Ce^{-2t|B_R|}$, whence we absorb the constant C by changing the 2 in the exponent. $\square$

Proof of Proposition 7 By Lemma 8 it suffices to bound the probability of $\bigcup_{s=t}^{2t} \{X_s \notin \mathcal{G}_\ell\}$. Condition on the clock rings between times t and $2t$; this set of clock rings generates a σ -algebra we denote by $\mathcal{T}_{[t, 2t]}$. Let

$$E_{[t, 2t]} := \bigcap_{m \geq \ell} \{N_{B_{R+m^\alpha}}^{[t, 2t]} \leq 4t|B_{R+m^\alpha}|\},$$

measurable with respect to $\mathcal{T}_{[t, 2t]}$. Lemma 11 showed that $\mathbb{P}(E_{[t, 2t]}^c) \leq \exp(-t|B_R|)$. We can then write

$$\mathbb{P}\left(\bigcup_{s=t}^{2t} \{X_s \notin \mathcal{G}_\ell\}\right) \leq \max_{(e_i, s_i)_{i \in E_{[t, 2t]}}} \mathbb{P}\left(\bigcup_i \{X_{s_i} \notin \mathcal{G}_\ell\} \mid (e_i, s_i)_i\right) + e^{-t|B_R|},$$

where $(e_i, s_i)_i$ denotes the sequence of pairs of edges and corresponding clock rings between times t and $2t$. For ease of notation, let $s_0 = t$ and let $\mathcal{G}_{\ell, s}^c$ be the event $\{X_s \notin \mathcal{G}_\ell\}$. We now write the union above as

$$\bigcup_{i \geq 1} \mathcal{G}_{\ell, s_i}^c \subset \mathcal{G}_{\ell, s_0}^c \cup \bigcup_{i \geq 1} (\mathcal{G}_{\ell, s_{i-1}} \cap \mathcal{G}_{\ell, s_i}^c).$$

Furthermore, given the clock ring times and locations $(e_i, s_i)_i$, we can let $I_0 = B_{R+\ell^\alpha}$ and for $m \geq \ell$, let I_m be the set of i 's for which e_i is in $B_{R+(m+1)^\alpha} \setminus B_{R+m^\alpha}$. Then,

$$\bigcup_{i \geq 1} \mathcal{G}_{\ell, s_i}^c \subset \mathcal{G}_{\ell, s_0}^c \cup \bigcup_{i \in I_0} (\mathcal{G}_{\ell, s_{i-1}} \cap \mathcal{G}_{\ell, s_i}^c) \cup \bigcup_{m \geq \ell} \bigcup_{i \in I_m} (\mathcal{G}_{\ell, s_{i-1}} \cap \mathcal{G}_{\ell, s_i}^c).$$

By Lemma 9, for $i \in I_m$ for $m \geq \ell$, we have

$$(\mathcal{G}_{\ell, s_{i-1}} \cap \mathcal{G}_{\ell, s_i}^c) \subset \{X_{s_{i-1}} \notin \bar{\mathcal{G}}_m\}.$$

Using this bound for $i \in I_m$, and the obvious bound $(\mathcal{G}_{\ell, s_{i-1}} \cap \mathcal{G}_{\ell, s_i}^c) \subset \mathcal{G}_{\ell, s_i}^c$ for $i \in I_0$, we obtain

$$\bigcup_{i \geq 1} \mathcal{G}_{\ell, s_i}^c \subset \{\mathcal{G}_{\ell, t}^c\} \cup \bigcup_{i \in I_0} \{\mathcal{G}_{\ell, s_i}^c\} \cup \bigcup_{m \geq \ell} \bigcup_{i \in I_m} \{X_{s_{i-1}} \notin \bar{\mathcal{G}}_m\}.$$

Taking the probability on either side, conditioning on $\mathcal{T}_{[t, 2t]}$ and using a union bound, we get

$$\begin{aligned} \mathbb{P}\left(\bigcup_{s=t}^{2t} \{X_s \notin \mathcal{G}_\ell\}\right) &\leq \max_{(e_i, s_i)_{i \in E_{[t, 2t]}}} \mathbb{P}(X_t \notin \mathcal{G}_\ell \mid (e_i, s_i)_i) \\ &\quad + \max_{(e_i, s_i)_{i \in E_{[t, 2t]}}} \sum_{i \in I_0} \mathbb{P}(X_{s_i} \notin \mathcal{G}_\ell \mid (e_i, s_i)_i) \\ &\quad + \max_{(e_i, s_i)_{i \in E_{[t, 2t]}}} \sum_{m \geq \ell} \sum_{i \in I_m} \mathbb{P}(X_{s_{i-1}} \notin \bar{\mathcal{G}}_m \mid (e_i, s_i)_i) + e^{-t|B_R|}. \end{aligned}$$

By Lemma 6, conditionally on any $(e_i, s_i)_i \in \mathcal{T}_{[t, 2t]}$, the law of $X_{s_i} \geq \tilde{\omega}$ where $\tilde{\omega}$ is drawn from a $\text{Ber}(\tilde{p})$ distribution, so long as $t \geq T_0$ from that lemma. Since the events $\mathcal{G}_\ell^c$ and $\bar{\mathcal{G}}_m^c$ are decreasing events, each of the probabilities above is bounded above by their analogs for $\tilde{\omega}$. Finally, the number of summands $|I_k| \leq 4t|B_{R+m\alpha}|$ since $(e_i, s_i)_i \in E_{[t, 2t]}$. Together, this means

$$\begin{aligned} \mathbb{P}\left(\bigcup_{s=t}^{2t} \{X_s \notin \mathcal{G}_\ell\}\right) &\leq (1 + 4t|B_{R+\ell\alpha}|)\mathbb{P}(\tilde{\omega} \notin \mathcal{G}_\ell) \\ &\quad + \sum_{m \geq \ell} 4t|B_{R+(m+1)\alpha}| \mathbb{P}(\tilde{\omega} \notin \bar{\mathcal{G}}_m) + e^{-t|B_R|}. \end{aligned}$$

By Lemma 10, this is at most

$$(1 + 4t|B_{R+\ell\alpha}|)e^{-\ell} + 4Ct \sum_{m \geq \ell} |B_{R+(m+1)\alpha}| e^{-m} + e^{-t|B_R|}.$$

Using the stretched-exponential volume growth bound $|B_{R+m\alpha}| \leq e^{(R+m\alpha)\eta} \leq e^{R\eta+m\alpha\eta}$, the first two terms above are summable and yield $4te^{R\eta}e^{-\ell}$ as long as η is small enough depending on δ , and ℓ, R are large enough. The additional term $e^{-t|B_R|}$ is absorbed since $t|B_R| \geq R$ for $t \geq 1$. $\square$

4.3 Exponential relaxation to equilibrium after burn-in

Our aim is to now combine the above ingredients to establish that after a burn-in period that keeps our configuration in the set $\mathcal{E}_\ell$ per Proposition 7, the disagreement propagation bounds of Sect. 3 can be implemented to guarantee exponential relaxation to equilibrium as long as p is sufficiently close to 1 to kickstart the spacetime recursion.

Proposition 12 Fix q, Δ, δ . There exists $\eta_0(\delta) > 0$ and $p_0(q, \Delta, \delta, \eta) < 1$ and $C(p, q, \Delta, \delta, \eta)$ such that for every $\eta < \eta_0$ and $p \geq p_0$ we have the following. For any G satisfying Definition 2 and η -stretched-exponential volume growth, the FK dynamics satisfies

$$\max_{e \in E(G)} (\mathbb{P}(X_s^1(e) = 1) - \mathbb{P}(X_s^0(e) = 1)) \leq e^{-s/C}, \quad \text{for all } s \leq (\log n)^2.$$

Proof Abusing notation slightly, let $(X_s)_{s \geq 0} = (X_s^0)_{s \geq 0}$ and $(Y_s)_{s \geq 0} = (X_s^1)_{s \geq 0}$. Define

$$\rho(t) := \max_{e \in E} \mathbb{P}(X_t(e) \neq Y_t(e)),$$

under the grand coupling (whence the probability is exactly the difference of the probabilities of e taking value 1). Recall the definition of $\mathcal{E}_\ell = \mathcal{E}_{\ell, \alpha, \alpha}$ for $\alpha = (1 + \delta)/\delta$ from Definition 9 and (6). Our first aim is to establish the following recurrence relation,

$$\rho(2t) \leq e^{\sqrt{t}} \rho(t)^2 + e^{-t}. \quad (8)$$

for all $t \geq T_0$ for a large enough T_0 . Toward this aim, let

$$A_{t, R, e} = \{X_t(B_R(e)) \neq Y_t(B_R(e))\}.$$

Then, for any fixed $e \in E(G)$, we have

$$\mathbb{P}(X_{2t}(e) \neq Y_{2t}(e)) \leq \mathbb{P}(X_{2t}(e) \neq Y_{2t}(e) \mid A_{t, R, e}^c) \mathbb{P}(A_{t, R, e}^c) \quad (9)$$

$$+ \mathbb{P}\left(X_{2t}(e) \neq Y_{2t}(e), A_{t, R, e}, \bigcap_{s=t}^{2t} \{X_s, Y_s \in \mathcal{E}_\ell\}\right) \quad (10)$$

$$+ \mathbb{P}\left(\bigcup_{s=t}^{2t} \{X_s \notin \mathcal{E}_\ell\}\right) + \mathbb{P}\left(\bigcup_{s=t}^{2t} \{Y_s \notin \mathcal{E}_\ell\}\right). \quad (11)$$

First notice that

$$\mathbb{P}(X_{2t}(e) \neq Y_{2t}(e) \mid A_{t, R, e}^c) \leq \rho(t),$$

since we can condition on X_t, Y_t , w.r.t. which $A_{t, R, e}$ is measurable, and use the property of the grand coupling that

$$\max_{\omega_0, \omega'_0} \mathbb{P}(X_t^{\omega_0}(e) \neq X_t^{\omega'_0}(e)) \leq \mathbb{P}(X_t^0(e) \neq X_t^1(e)).$$

At the same time, by a union bound over $e \in E(B_R)$, we can bound $\mathbb{P}(A_{t, R, e}^c) \leq |E(B_R)|\rho(t)$. These give the bound on (9) of

$$\mathbb{P}(X_{2t}(e) \neq Y_{2t}(e) \mid A_{t,R,e}^c) \mathbb{P}(A_{t,R,e}^c) \leq |E(B_R)| \rho(t)^2.$$

The quantity in (10) is controlled by Proposition 5, whence as long as $t \leq \varepsilon R / \ell^{2\alpha}$,

$$\mathbb{P}\left(X_{2t}(e) \neq Y_{2t}(e), A_{t,R,e}, \bigcap_{s=t}^{2t} \{X_s, Y_s \in \mathcal{E}_\ell\}\right) \leq C |\partial B_R| \exp(-R/(2\ell^\alpha)).$$

Finally, we control each of the terms in (11) by Proposition 7 giving

$$\mathbb{P}\left(\bigcup_{s=t}^{2t} \{X_s \notin \mathcal{E}_\ell\}\right) + \mathbb{P}\left(\bigcup_{s=t}^{2t} \{Y_s \notin \mathcal{E}_\ell\}\right) \leq C t e^{R^\eta} e^{-\ell}.$$

as long as $t \geq T_0(\delta, \eta, q)$ and $\eta < \eta_0(\delta)$. Putting the above together, and using the bounds on $|E(B_R)|$ and $|\partial B_R|$ from the fact that G has η -stretched-exponential volume growth, for all $T_0(\delta, \eta, q) \leq t \leq \varepsilon R / \ell^{2\alpha}$,

$$\rho(2t) \leq e^{R^\eta} \rho(t)^2 + C e^{R^\eta} e^{-R/(2\ell^\alpha)} + C t e^{R^\eta} e^{-\ell}.$$

If we make the choices

$$\ell = 2R^{2\eta}, \quad \text{and} \quad t = \ell/2 = R^{2\eta}.$$

we find that as long as $\eta < \eta_0(\delta)$ and t, ℓ, R are sufficiently large (as a function of $\delta, \eta, \varepsilon$), we maintain $t \leq \varepsilon R / \ell^{2\alpha}$ and we can absorb the pre-factors above to obtain the claimed (8). That recurrence will hold for all $t \geq T_0(q, \delta, \eta)$ and as an upper bound, for all $t \leq (\log n)^2$ since our arguments are all valid as long as $R \leq \text{diam}(G)$ which for G of η -subexponential volume growth is for all $R \leq (\log n)^{1/\eta}$, which translates to $t \leq (\log n)^2$.

It remains to deduce the exponential decay on $\rho(t)$ from (8); consider the function

$$\phi(t) = e^{\sqrt{t}} (\rho(t) + e^{-t/2})^{1/2}.$$

Then by (8), and the fact that $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$,

$$\phi(2t) \leq e^{\sqrt{2}\sqrt{t}} (e^{\sqrt{t}} \rho(t)^2 + e^{-t} + e^{-t})^{1/2} \leq e^{(\sqrt{2}+.5)\sqrt{t}} \rho(t) + \sqrt{2} e^{\sqrt{2}t} e^{-t/2}.$$

Since $2 \geq \sqrt{2} + .5$ and $\sqrt{2} e^{\sqrt{2}t} \leq e^{2\sqrt{t}}$ for all $t \geq 1$, this is at most $\phi(t)^2$. Therefore, for any $t_0 \geq 1$, we have

$$\phi(2^k t_0) \leq \phi(t_0)^{2^k},$$

whence if $\phi(t_0) < 1/e$, then for $r = 2^k$, we have $\phi(r t_0) \leq e^{-r}$. From there, using the definition of $\phi(t)$ in terms of $\rho(t)$, we see that $\rho(r t_0) \leq e^{-2r}$, whence $\rho(t) \leq e^{-2t/t_0}$

for $t = 2^k t_0$. The fact that $\rho(t)$ is monotone decreasing in time implies the bound $\rho(t) \leq e^{-t/t_0}$ for all $t \geq t_0$.

The last step is to show that $\phi(t_0) < 1/e$ for some t_0 larger than $\max\{6, T_0(q, \Delta, \delta, \eta)\}$. Towards this purpose, notice that by the update rule (2),

$$\rho(s) \leq (1 - \hat{p}) + \mathbb{P}(\text{Pois}(s) = 0) = (1 - \hat{p}) + e^{-s}.$$

There exists s_0 independent of everything else such that $e^{\sqrt{s}}(2e^{-s} + e^{-s/2})^{1/2}$ is less than $1/e$ for all $s > s_0$ because $e^{\sqrt{t}}(2e^{-t} + e^{-t/2})^{1/2}$ is at most $3e^{-t/4+\sqrt{t}}$, say. Let $p_0(q)$ be large enough that $(1 - \hat{p}) \leq e^{-s_0}$ for all $p \geq p_0$. Then for all $p \geq p_0$ and $t_0 \geq s_0$, $\rho(t_0) \leq 2e^{-t_0}$ and we obtain the claimed $\phi(t_0) < 1/e$. $\square$

Proof of Theorem 1 Under the monotone grand coupling, we have for every initial state ω_0 ,

$$\max_{\omega_0} \|\mathbb{P}(X_t^{\omega_0} \in \cdot) - \pi\|_{\text{TV}} \leq \sum_{e \in E(G)} \mathbb{P}(X_t^{\omega_0}(e) \neq X_t^\pi(e)) \leq \sum_{e \in E(G)} \mathbb{P}(X_t^1(e) \neq X_t^0(e)).$$

In turn, by monotonicity, the right-hand side is at most

$$\max_{\omega_0} \|\mathbb{P}(X_t^{\omega_0} \in \cdot) - \pi\|_{\text{TV}} \leq \sum_{e \in E(G)} (\mathbb{P}(X_t^1(e) = 1) - \mathbb{P}(X_t^0(e) = 1)).$$

Let $t = C_1 \log n$ for C_1 a large enough constant (depending on q, Δ, δ, η). Then by Proposition 12, each term in the right-hand side is bounded by n^{-4} for large enough n . Since there are at most $\Delta n \leq n^2$ many summands, the sum above is $o(1)$, implying mixing in $O(\log n)$ time. This gives $O(n \log n)$ mixing time for the discrete-time FK dynamics as described in the preliminaries.

The result for the SW dynamics follows from the comparison result of [56]. $\square$

5 Spatial and temporal mixing on trees with r -wired boundary

Our next goal in the paper is to establish Theorem 2 concerning FK dynamics on tree-like expanders. Recursive mixing time arguments based on disagreement percolation, like those in the previous section, are known to fail on graphs with exponential volume growth. At the same time, localizing the dynamics can be difficult in treelike graphs because there is no weak spatial mixing when p is close to 1; see the discussion at the beginning of Sect. 5.1 for more details. In this section, we define a class of boundary conditions that are sufficiently “wired” to support a notion of weak spatial mixing with respect to the wired boundary conditions. This class of boundary conditions captures the boundary conditions induced by the FK dynamics configuration on a treelike ball centered at a vertex of an expander graph after a short burn-in.

This section focuses on general rooted trees $\mathcal{T}_h = (V(\mathcal{T}_h), E(\mathcal{T}_h))$ having depth h , minimum internal degree 3 and maximum degree Δ . For any $m \leq h$, $\mathcal{T}_m$ will denote the tree given by truncating $\mathcal{T}_h$ at depth m . The boundary $\partial \mathcal{T}_m$ is the set of vertices of

$\mathcal{T}_m$ at depth m . For any vertex $w \in V(\mathcal{T}_h)$, we use $\mathcal{T}_{h,w}$ to denote the sub-tree of $\mathcal{T}_h$ rooted at w , with boundary $\partial\mathcal{T}_{h,w} = \partial\mathcal{T}_h \cap \mathcal{T}_{h,w}$.

Definition 11 A boundary condition ξ of $\mathcal{T}_h$ is *single-component* if the boundary partition corresponding to ξ has at most one non-singleton element; we call this its *wired component*.

Definition 12 A distribution P over boundary conditions ξ on $\mathcal{T}_h$ is *r-wired* if it is supported on single-component boundary conditions, and the distribution of the wired component of ξ stochastically dominates the distribution over random subsets $A \subset \partial\mathcal{T}_h$ in which each vertex of $\partial\mathcal{T}_h$ is included in A with probability r independently (the partial order being the natural one on vertex subsets).

The following shows that except with *double-exponentially* small probability, the random-cluster model on $\mathcal{T}_h$ with *r-wired* boundary conditions satisfies weak spatial mixing with respect to the all-wired boundary condition (the TV distance between the two decays exponentially in the distance from the boundary).

Lemma 13 Let P be *r-wired* and let $\xi \sim P$. Then, with P -probability $1 - e^{-c_{p,r}(1.1)^h}$, we have

$$\|\pi_{\mathcal{T}_h}^{\xi}(\omega(\mathcal{T}_{h/2}) \in \cdot) - \pi_{\mathcal{T}_h}^1(\omega(\mathcal{T}_{h/2}) \in \cdot)\|_{\text{TV}} \leq e^{-c_{p,r}h},$$

for some $c_{p,r} = c_{p,r}(q, \Delta)$, which is positive as long as $p > p_0(q, \Delta)$ and $r > r_0(q, \Delta)$ for suitable $p_0(q, \Delta), r_0(q, \Delta) \in (0, 1)$, and which goes to ∞ as $p, r \rightarrow 1$.

Remark 4 The double exponential concentration under P in Lemma 13 is not strictly needed for the results in this paper. A single-exponential concentration under P would suffice and would be easier to establish by averaging over $\xi \sim P$ and then applying Markov's inequality. However, we include this stronger form since it provides insight into the good qualities of *r-wired* boundary conditions and could be used to get an improved mixing time bound for the FK dynamics on trees with *r-wired* boundary. Unfortunately, with our current methods in Sect. 6, this improvement would not translate into better bounds for the FK dynamics on treelike graphs, so we do not pursue them.

5.1 Spatial mixing on trees with *r-wired* boundary conditions

Our first aim in this section is to establish Lemma 13. Spatial mixing in the traditional sense, where one takes a maximum over the boundary conditions on the tree, does *not* hold for large p , even if it is very close to 1. This can be seen by considering $\mathcal{T}_h$ with wired vs. free boundary conditions, and noticing that the marginal of any edge in the tree with the free boundary condition is $\text{Ber}(\hat{p})$, whereas the marginal of an edge in the wired tree gives at least constant probability to that edge being distributed as $\text{Ber}(p)$, and otherwise, as $\text{Ber}(\hat{p})$, so the total-variation distance on that edge does not go to zero as $h \rightarrow \infty$. Our solution to this issue is to restrict attention to *r-wired* boundary conditions and establish that, at least among these boundary conditions, the

random-cluster model exhibits spatial mixing on $\mathcal{T}_h$. Considering only such boundary conditions will suffice for us, since these are the boundary conditions that appear after a burn-in period of the FK dynamics on treelike expander graphs. We note that with *fully* wired boundary conditions on trees, [47] showed decay of correlations to establish uniqueness at p close to 1.

The mechanism for coupling random-cluster configurations with r -wired and wired boundary conditions is based on what we call wired separating sets. These will be a set of vertices that are all connected down to the wired component of ξ in $\partial\mathcal{T}_h$ and therefore wired together; as such, they separate the influence of the boundary condition of $\partial\mathcal{T}_h$ from, say, $\partial\mathcal{T}_{h/2}$.

Definition 13 A separating set in $\mathcal{T}_h$ is a set of vertices $S \subset V(\mathcal{T}_h) \setminus V(\mathcal{T}_{h/2})$ such that every path from $\partial\mathcal{T}_{h/2}$ to $\partial\mathcal{T}_h$ must intersect S . A configuration $\omega \subset E(\mathcal{T}_h)$ has a *wired separating set* if there exists a separating set S such that every vertex $v \in S$ is connected in $\omega(\mathcal{T}_{h,v})$ to a vertex $u \in \partial\mathcal{T}_{h,v}$ belonging to the wired component of ξ . Let $\mathcal{S}_{h,\xi}$ be the event that ω has a wired separating set in $\mathcal{T}_h$ with boundary condition ξ .

The following lemma shows how the event of having a wired separating set governs the probability of coupling random-cluster configurations to $\omega \sim \pi_{\mathcal{T}_h}^1$ in $\mathcal{T}_{h/2}$.

Lemma 14 For any single-component boundary condition ξ on $\mathcal{T}_h$,

$$\|\pi_{\mathcal{T}_h}^\xi(\omega(\mathcal{T}_{h/2}) \in \cdot) - \pi_{\mathcal{T}_h}^1(\omega(\mathcal{T}_{h/2}) \in \cdot)\|_{\text{TV}} \leq \pi_{\mathcal{T}_h}^\xi(\mathcal{S}_{h,\xi}^c).$$

Proof We construct a monotone coupling for $\pi_{\mathcal{T}_h}^\xi$ and $\pi_{\mathcal{T}_h}^1$ such that if (ω_ξ, ω_1) is sampled from this coupling, then $\omega_\xi \sim \pi_{\mathcal{T}_h}^\xi$, $\omega_1 \sim \pi_{\mathcal{T}_h}^1$ and

$$\{\omega_\xi \in \mathcal{S}_{h,\xi}\} \implies \{\omega_\xi(\mathcal{T}_{h/2}) = \omega_1(\mathcal{T}_{h/2})\}.$$

Our construction relies on revealing the values of (ω_ξ, ω_1) on an edge set $\mathcal{R}$ under a monotone coupling, where crucially, $\mathcal{R}$ will be designed to be the set of edges in the “lowest” wired separating set (if one exists). We construct ω_ξ and ω_1 as follows (and refer to Fig. 3 for a depiction):

1. Initialize $\mathcal{P}_0$ as the parents of $\partial\mathcal{T}_h$, and $\mathcal{R}_0 = \emptyset$.
2. Starting with $i = 1$, pick a vertex $w_i \in \mathcal{P}_{i-1}$, and sample the configurations $(\omega_\xi(\mathcal{T}_{h,w_i}), \omega_1(\mathcal{T}_{h,w_i}))$ from the monotone coupling between the marginals

$$\pi_{\mathcal{T}_h}^\xi(\omega_\xi(\mathcal{T}_{h,w_i}) \in \cdot \mid \omega_\xi(\mathcal{R}_{i-1})), \quad \text{and} \quad \pi_{\mathcal{T}_h}^1(\omega_1(\mathcal{T}_{h,w_i}) \in \cdot \mid \omega_1(\mathcal{R}_{i-1})).$$

3. Let $\mathcal{R}_i = \mathcal{R}_{i-1} \cup E(\mathcal{T}_{h,w_i})$ and form $\omega_\xi(\mathcal{R}_i)$ and $\omega_1(\mathcal{R}_i)$ by adding the configurations $\omega_\xi(\mathcal{T}_{h,w_i})$ and $\omega_1(\mathcal{T}_{h,w_i})$, respectively.
4. If w_i is connected to the wired component of ξ in $\omega_\xi(\mathcal{T}_{h,w_i})$, then let $\mathcal{P}_i = \mathcal{P}_{i-1} \setminus \mathcal{T}_h(w_i)$; else, let $\mathcal{P}_i = (\mathcal{P}_{i-1} \setminus \mathcal{T}_h(w_i)) \cup \{w'_i\}$ where w'_i is the parent of w_i .

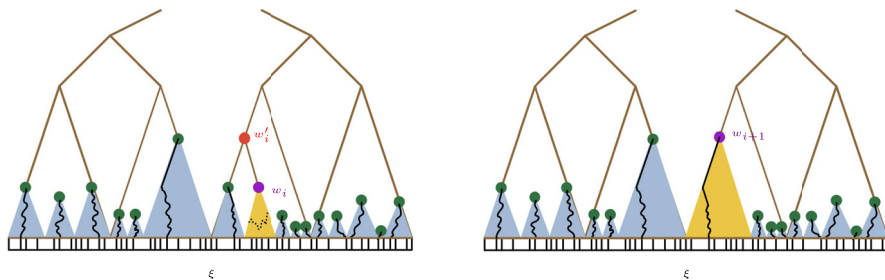


Fig. 3 Two steps of the revealing process used in Lemma 14. Left: the configurations on the sub-tree of $w_i \in \mathcal{P}_{i-1}$ are revealed, and in the configuration with ξ -boundary conditions, w_i is not wired down to ξ . Thus, its parent w'_i is added to $\mathcal{P}_i$. In the next step, this vertex is w_{i+1} , and when the remainder of its sub-tree is revealed to indeed include a wiring to ξ , the vertex is removed from $\mathcal{P}_i$ but its parent is no longer added

While i is such that $\mathcal{P}_{i-1}$ is non-empty, $\mathcal{R}_i \setminus \mathcal{R}_{i-1} \neq \emptyset$. This is because the edges from w_i to its children will be in $\mathcal{R}_i$ but not in $\mathcal{R}_{i-1}$ (if they were in $\mathcal{R}_{i-1}$ then $w_i \in \mathcal{T}_{w_j}$ for some $j < i$ and w_i would have been removed from $\mathcal{P}_j$). Therefore, the revealing process will terminate after a finite number of steps, and we can call $\mathcal{R} = \mathcal{R}_f$ if f is the first index for which $\mathcal{P}_f = \emptyset$.

Now, consider the subset $\partial_{\text{TOP}} \mathcal{R}$ of vertices of $\mathcal{R}$ whose parents are not in $\mathcal{R}$. As long as the process terminates with $\mathcal{R} \cap E(\mathcal{T}_{h/2}) = \emptyset$, the vertices in $\partial_{\text{TOP}} \mathcal{R}$ will be connected to the wired component of ξ in $\omega_\xi(\mathcal{T}_{h,w})$, and so they will form a wired separating set. Conversely, if there exists a wired separating set S in $\omega_\xi(\mathcal{T}_{h,w})$, no parent of any vertex in S will ever be added to $\mathcal{P}$, and therefore, the vertices from $\partial_{\text{TOP}} \mathcal{R}$ will form exactly the lowest wired separating set.

Upon termination of the iterative procedure above, we can then sample $\omega_\xi(\mathcal{T}_h \setminus \mathcal{R}) = \omega_1(\mathcal{T}_h \setminus \mathcal{R})$, since both $\omega_\xi(\mathcal{R})$ and $\omega_1(\mathcal{R})$ induce wired boundary conditions on $E(\mathcal{T}_h) \setminus \mathcal{R}$. Therefore, under this coupling, we always have $\omega_\xi(\mathcal{T}_h \setminus \mathcal{R}) = \omega_1(\mathcal{T}_h \setminus \mathcal{R})$. Since the process terminates at the lowest wired separating set of ω_ξ , on the event $\omega_\xi \in \mathcal{S}_{h,\xi}$, necessarily $\mathcal{T}_h \setminus \mathcal{R}$ contains all of $\mathcal{T}_{h/2}$. As such, using $\mathbb{P}$ to denote the probability under the coupling we just designed,

$$\|\pi_{\mathcal{T}_{h/2}}^\xi - \pi_{\mathcal{T}_{h/2}}^1\|_{\text{TV}} \leq \mathbb{P}(\mathcal{R} \cap \mathcal{T}_{h/2} \neq \emptyset) \leq \mathbb{P}(\omega_\xi \in \mathcal{S}_{h,\xi}^c) = \pi_{\mathcal{T}_h}^\xi(\mathcal{S}_{h,\xi}^c),$$

as claimed. $\square$

5.1.1 Good boundary conditions

Our aim is now to control the probability of $\mathcal{S}_{h,\xi}^c$ under a random r -wired boundary condition $\xi \sim \mathbb{P}$. Recalling the definition of $\mathcal{S}_{h,\xi}$, notice that it is an increasing event in the random-cluster configuration. Since the random-cluster measure with parameters p and q stochastically dominates independent percolation with parameter $\hat{p} = \frac{p}{q(1-p)+p}$ configuration (i.e., the random-cluster measure with parameters $q = 1$ and $\hat{p}$), it will suffice for us to consider the probability of $\mathcal{S}_{h,\xi}$ under the product measure $\tilde{\pi}_{\mathcal{T}_h} = \pi_{\mathcal{T}_h, \tilde{p}, 1}^\xi$ for some $\tilde{p} \leq \hat{p}$ still going to 1 with p (e.g., $\tilde{p} = 2\hat{p} - 1$).

Definition 14 A boundary condition ξ on $\mathcal{T}_h$ is called c -good if

$$\tilde{\pi}_{\mathcal{T}_h}(\tilde{\omega} \in \mathcal{S}_{h,\xi}^c) \leq \exp(-ch).$$

Notice that if ξ is c -good, then any single-component boundary condition $\xi' \geq \xi$ will also be c -good, since any $\tilde{\omega} \in \mathcal{S}_{h,\xi}$ will also be in $\mathcal{S}_{h,\xi'}$. Therefore, the event $\{\xi \text{ is } c\text{-good}\}$ is itself increasing in the partial order on subsets of $\partial\mathcal{T}_h$. In particular, if P is r -wired, then

$$P(\xi \text{ is } c\text{-good}) \geq P_{\text{Ber}(r)}(\xi \text{ is } c\text{-good})$$

where $P_{\text{Ber}(r)}$ is the distribution over boundary conditions on $\partial\mathcal{T}_h$ where the wired component contains each vertex independently with probability r . Given this, the following lemma implies Lemma 13.

Lemma 15 Suppose $\xi \sim P_{\text{Ber}(r)}$. There exists $c = c(p, q, r)$ going to ∞ as $p, r \rightarrow 1$, such that

$$P(\xi \text{ is not } c\text{-good}) \leq \exp(-c(1.1)^h).$$

Proof For some $\gamma = \gamma(p, q, r) > 0$ to be chosen later, going to 1 as $p, r \rightarrow 1$, consider the event $\mathcal{E}_\gamma$ that ξ belongs to the set of boundary conditions on $\mathcal{T}_h$ that satisfy the following property:

(P1) for each downward path from $\partial\mathcal{T}_{h/2}$ to $\partial\mathcal{T}_h$: $v_0, v_1, \dots, v_{h/2}$, for each $i = 1, \dots, h/4$, if we draw a configuration $\tilde{\omega}(\mathcal{T}_{v_i} \setminus \mathcal{T}_{v_{i+1}})$ from $\tilde{\pi}_{\mathcal{T}_{v_i} \setminus \mathcal{T}_{v_{i+1}}}$, the probability that the component of v_i in $\tilde{\omega}(\mathcal{T}_{v_i} \setminus \mathcal{T}_{v_{i+1}})$ intersects the wired component of $\xi \cap \partial(\mathcal{T}_{v_i} \setminus \mathcal{T}_{v_{i+1}})$ is at least γ .

We first show that $\xi \in \mathcal{E}_\gamma$ implies ξ is c -good for a suitable $c > 0$. Fix any $\xi \in \mathcal{E}_\gamma$. In order for $\mathcal{S}_{h,\xi}^c$ to occur, there must exist a path from $\partial\mathcal{T}_{h/2}$ to $\partial\mathcal{T}_h$ such that no v_i is connected to ξ through $\tilde{\omega}(\mathcal{T}_{v_i})$. For any fixed path, that probability is upper bounded by

$$\tilde{\pi}_{\mathcal{T}_h} \left(\bigcap_{i=1}^{h/4} \{ \tilde{\omega}(\mathcal{T}_{v_i} \setminus \mathcal{T}_{v_{i+1}}) \cap \xi = \emptyset \} \right) = \prod_{i=1}^{h/4} \tilde{\pi}_{\mathcal{T}_h}(\tilde{\omega}(\mathcal{T}_{v_i} \setminus \mathcal{T}_{v_{i+1}}) \cap \xi = \emptyset) \leq (1 - \gamma)^{h/4}.$$

Here we have abused notation slightly to identify ξ with the subset of $\partial\mathcal{T}_h$ that is its wired component. (The change from the intersection to the product comes from the fact that $\tilde{\pi}_{\mathcal{T}_h}$ is a product measure and that the sets $\mathcal{T}_{v_i} \setminus \mathcal{T}_{v_{i+1}}$ are disjoint for different i .) A union bound over the Δ^h many paths implies that the probability that $\omega \notin \mathcal{S}_{h,\xi}$ is at most $\Delta^{h/2}(1 - \gamma)^{h/4}$ which is at most $\exp(-ch)$ for some $c > 0$ going to ∞ as $\gamma \rightarrow 1$, which happens as $p, r \rightarrow 1$.

It now suffices to find such a $\gamma > 0$, and a $\hat{c} = \hat{c}(p, q, r) > 0$ such that

$$P_{\text{Ber}(r)}(\xi \notin \mathcal{E}_\gamma) \leq \exp(-\hat{c}(1.1)^h). \quad (12)$$

Fix a downward path $v_0, \dots, v_{h/2}$ from $\partial\mathcal{T}_{h/2}$ to $\partial\mathcal{T}_h$ and an $i \in \{1, \dots, h/4\}$ (these will subsequently be union bounded over), and consider the probability that ξ is such that (P1) above holds for that path and that i .

Define the event $\mathcal{F}^i$ that the connected component of v_i in $\tilde{\omega}(\mathcal{T}_{v_i} \setminus \mathcal{T}_{v_{i+1}})$ intersects $\partial(\mathcal{T}_{v_i} \setminus \mathcal{T}_{v_{i+1}})$ in at least $(2\tilde{p} - \varepsilon)^{h(v_i)-1}$ many sites, where $h(v_i)$ denotes the height of $\mathcal{T}_{v_i}$. Note that $h(v_i) = i - h/2 \geq h/4$, and ε will be chosen later. For intuition, the component of v_i in $\tilde{\omega}(\mathcal{T}_{v_i} \setminus \mathcal{T}_{v_{i+1}})$ is a branching process which (besides its first level) has $2\tilde{p}$ expected number of children since we assumed that in $\mathcal{T}_h$ all internal vertices have degree at least 3. By a standard branching process argument (see Fact 16 below from which this follows after asking that the first level have an open edge with probability at least $\tilde{p}$), we know there exists δ going to 0 as $\tilde{p} \rightarrow 1$ (equivalently as $p \rightarrow 1$) such that

$$\tilde{\pi}_{\mathcal{T}_h}(\tilde{\omega}(\mathcal{T}_{v_i} \setminus \mathcal{T}_{v_{i+1}}) \notin \mathcal{F}^i) \leq \delta.$$

By the independence of $\tilde{\omega}(\mathcal{T}_{v_i} \setminus \mathcal{T}_{v_{i+1}})$ from ξ , we have

$$\mathbb{P}_{\text{Ber}(r)} \otimes \tilde{\pi}_{\mathcal{T}_h} \left(\tilde{\omega}(\mathcal{T}_{v_i} \setminus \mathcal{T}_{v_{i+1}}) \cap \xi = \emptyset, \tilde{\omega}(\mathcal{T}_{v_i} \setminus \mathcal{T}_{v_{i+1}}) \in \mathcal{F}^i \right) \leq (1-r)^{(2\tilde{p}-\varepsilon)^{h(v_i)-1}}.$$

The left-hand side above is exactly the expected value over $\xi \sim \mathbb{P}_{\text{Ber}(r)}$ of the $\tilde{\pi}_{\mathcal{T}_h}$ probability of an event depending on ξ . Thus by Markov's inequality and the fact that $h(v_i) \geq h/4$,

$$\begin{aligned} \mathbb{P}_{\text{Ber}(r)} \left(\xi : \tilde{\pi}_{\mathcal{T}_h}(\tilde{\omega}(\mathcal{T}_{v_i} \setminus \mathcal{T}_{v_{i+1}}) \cap \xi = \emptyset, \tilde{\omega}(\mathcal{T}_{v_i} \setminus \mathcal{T}_{v_{i+1}}) \in \mathcal{F}^i) > (1-r)^{(2\tilde{p}-\varepsilon)^{h/4}/2} \right) \\ \leq \frac{\mathbb{E}_{\xi \sim \text{Ber}(r)} \left[\tilde{\pi}_{\mathcal{T}_h}(\tilde{\omega}(\mathcal{T}_{v_i} \setminus \mathcal{T}_{v_{i+1}}) \cap \xi = \emptyset, \tilde{\omega}(\mathcal{T}_{v_i} \setminus \mathcal{T}_{v_{i+1}}) \in \mathcal{F}^i) \right]}{(1-r)^{(2\tilde{p}-\varepsilon)^{h/4}/2}} \\ = \frac{\mathbb{P}_{\text{Ber}(r)} \otimes \tilde{\pi}_{\mathcal{T}_h} \left(\tilde{\omega}(\mathcal{T}_{v_i} \setminus \mathcal{T}_{v_{i+1}}) \cap \xi = \emptyset, \tilde{\omega}(\mathcal{T}_{v_i} \setminus \mathcal{T}_{v_{i+1}}) \in \mathcal{F}^i \right)}{(1-r)^{(2\tilde{p}-\varepsilon)^{h/4}/2}} \\ \leq \exp((2\tilde{p} - \varepsilon)^{h/4} \log(1-r)/2). \end{aligned}$$

At the same time, for any fixed ξ ,

$$\begin{aligned} \tilde{\pi}_{\mathcal{T}_h}(\tilde{\omega}(\mathcal{T}_{v_i} \setminus \mathcal{T}_{v_{i+1}}) \cap \xi = \emptyset) &\leq \tilde{\pi}_{\mathcal{T}_h}(\tilde{\omega}(\mathcal{T}_{v_i} \setminus \mathcal{T}_{v_{i+1}}) \cap \xi = \emptyset, \tilde{\omega}(\mathcal{T}_{v_i} \setminus \mathcal{T}_{v_{i+1}}) \in \mathcal{F}^i) \\ &\quad + \tilde{\pi}_{\mathcal{T}_h}(\tilde{\omega}(\mathcal{T}_{v_i} \setminus \mathcal{T}_{v_{i+1}}) \notin \mathcal{F}^i). \end{aligned}$$

If the first of these terms is at most $(1-r)^{(2\tilde{p}-\varepsilon_p)^{h/4}/2}$ and the second is at most δ , we set $\gamma = 1 - \delta - (1-r)^{(2\tilde{p}-\varepsilon)^{h/4}/2}$. Then, γ goes to 1 as $p \rightarrow 1$ and by the above, $\tilde{\pi}_{\mathcal{T}_h}(\tilde{\omega}(\mathcal{T}_{v_i} \setminus \mathcal{T}_{v_{i+1}}) \cap \xi \neq \emptyset) \geq \gamma$. Thus,

$$\mathbb{P}_{\text{Ber}(r)}(\xi : \tilde{\pi}_{\mathcal{T}_h}(\tilde{\omega}(\mathcal{T}_{v_i} \setminus \mathcal{T}_{v_{i+1}}) \cap \xi = \emptyset) > 1 - \gamma) \leq \exp((2\tilde{p} - \varepsilon)^{h/4} \log(1-r)/2).$$

We now upgrade this into the probability that ξ is in $\mathcal{E}_\gamma$ by a union bound over all $\Delta^{h/2}$ many paths in $\mathcal{T}_h$ and the $h/4$ many possible i 's; both these terms are absorbed by the double exponential above. As long as ε is sufficiently small, and p is sufficiently large, $(2\tilde{p} - \varepsilon)^{1/4}$ is greater than 1.1, and the prefactor evidently blows up as $r \rightarrow 1$ as claimed. $\square$

For completeness, we have included the following simple branching process concentration estimate.

Fact 16 *In a branching process with progeny distribution stochastically dominating $\text{Bin}(2, \tilde{p})$, let Z_k be the population size at level k . Then for every $\varepsilon > 0$, there exists $\delta > 0$ going to 0 as $p \rightarrow 1$ such that*

$$\mathbb{P}\left(\bigcap_{k \geq 1} \{Z_k \geq (2\tilde{p} - \varepsilon)^k\}\right) \geq 1 - \delta.$$

Proof Since the event in question is an increasing event, it suffices to show the above for the branching process with progeny distribution exactly $\text{Bin}(2, \tilde{p})$. If we let $\mathcal{A}_k$ be the event $\{Z_k \geq (2\tilde{p} - \varepsilon)^k\}$, then we can write

$$\mathbb{P}\left(\bigcup_{k \geq 1} \mathcal{A}_k^c\right) \leq \mathbb{P}(\mathcal{A}_1^c) + \sum_{k \geq 2} \mathbb{P}\left(\mathcal{A}_k^c, \bigcap_{j < k} \mathcal{A}_j\right) \leq \mathbb{P}(\mathcal{A}_1^c) + \sum_{k \geq 2} \mathbb{P}\left(\mathcal{A}_k^c \mid \bigcap_{j < k} \mathcal{A}_j\right).$$

Since $Z_1 = 1$, the probability of $\mathcal{A}_1$ is 1. For any $k \geq 2$, since Z_k is Markov, it suffices to condition on $Z_{k-1} : Z_{k-1} \in \mathcal{A}_{k-1}$; given Z_{k-1} , the distribution of $\mathcal{A}_{k-1}$ is

$$\sum_{i=1}^{Z_{k-1}} X_i \quad \text{where } X_i \text{ are i.i.d. } \text{Bin}(2, \tilde{p}).$$

Thus, $\mathbb{P}(\mathcal{A}_k^c \mid Z_{k-1}, \mathcal{A}_{k-1})$ is at most the probability of a sum of $Z_{k-1} \geq (2\tilde{p} - \varepsilon)^{k-1}$ i.i.d. $\text{Bin}(2, \tilde{p})$ random variables, being at least εZ_{k-1} below its mean. By Hoeffding's inequality (the X_i 's being bounded by 2), this has probability at most $\exp(-\frac{1}{2}\varepsilon^2(2\tilde{p} - \varepsilon)^{k-1})$ for every k . Let K equal $\log_{2\tilde{p}-\varepsilon}(2/(\varepsilon^2\sqrt{1-\tilde{p}}))$. For the first K generations, we can use the simpler union bound over the probability that one of the first $(2\tilde{p} - \varepsilon)^k$ many X_i 's is not equal to 2. Putting these together, we get

$$\mathbb{P}\left(\bigcup_{k \geq 1} \mathcal{A}_k^c\right) \leq \frac{4}{\varepsilon^2}(1 - \tilde{p})^{1/2} + \sum_{k > K} e^{-\frac{1}{\sqrt{1-\tilde{p}}}(k-K)}.$$

For any fixed ε , this is then seen to be at most some δ going to zero as $\tilde{p} \rightarrow 1$, which happens as $p \rightarrow 1$. $\square$

5.2 Mixing time for trees with single-component boundary conditions

In this section, we provide a bound on the mixing time of the FK dynamics on trees with single-component boundary conditions. In view of its application in the following

section for treelike balls of expanders (possibly having a finite number of cycles), it will be helpful for us to recall certain standard definitions that will allow us to relate the convergence rate of various Markov chains.

For a Markov chain on a finite state space Ω with transition matrix P and stationary distribution μ , the Dirichlet form of the chain is defined for any function $f : \Omega \rightarrow \mathbb{R}$ by

$$\mathcal{E}(f, f) := \frac{1}{2} \sum_{\omega, \omega' \in \Omega} \mu(\omega) P(\omega, \omega') (f(\omega) - f(\omega'))^2, \quad (13)$$

and its spectral gap is given by

$$\lambda_P := \min_{f: \text{Var}_\mu[f] \neq 0} \frac{\mathcal{E}(f, f)}{\text{Var}_\mu[f]}, \quad (14)$$

where $\text{Var}_\mu[f] = \mathbb{E}_\mu[f^2] - \mathbb{E}_\mu[f]^2$ with $\mathbb{E}_\mu[f] = \sum_{\omega \in \Omega} \mu(\omega) f(\omega)$.

The inverse of the spectral gap is closely connected to the mixing time of a Markov chain. In particular,

$$(\lambda_P^{-1} - 1) \log\left(\frac{1}{2\varepsilon}\right) \leq t_{\text{MIX}}(\varepsilon) \leq \lambda_P^{-1} \log\left(\frac{1}{\varepsilon \mu_{\min}}\right), \quad (15)$$

where $\mu_{\min} = \min_{\omega \in \Omega} \mu(\omega)$. (We refer the reader to e.g., [48, Chapter 12.2] for more details.)

We establish the following bounds for the mixing time and inverse spectral gap of the FK dynamics on a tree with any single-component boundary condition.

Lemma 17 *Consider any tree $\mathcal{T}_h$ of maximum degree Δ and depth h with single-component boundary condition ξ . There exist constants $a = a(\Delta, q) > 0$ and $C = C(p, q) > 0$ such that the inverse spectral gap of the FK dynamics on $\mathcal{T}_h$ with boundary condition ξ are at most $C \exp(ah)$.*

Proof We use the classical bound on the spectral gap obtained from the cut-width of a graph via the canonical paths method, though a little care is needed for the purpose of handling the random-cluster boundary condition. The edge-cut-width of $\mathcal{T}_h$ (in other words, the cut-width of its line graph) is defined as follows: enumerate the edges of $\mathcal{T}_h$ as $1, \dots, |E(\mathcal{T}_h)|$ and define

$$\text{CW}(\mathcal{T}_h) = \min_{\sigma} \max_i |V(\{e_{\sigma(j)} : j \leq i\}) \cap V(\{e_{\sigma(j)} : j > i\})|$$

where the minimum is over permutations σ on $\{1, \dots, |E(\mathcal{T}_h)|\}$. We claim that there exists a constant $K(\Delta)$ such that uniformly over all trees of degree at most Δ , their edge-cut-width is at most Kh ; this follows e.g., from [3, Lemma 2.1] and the fact that the edge-cut-width is within a factor of Δ of the (vertex) cut-width. Let σ be the permutation that attains this edge-cut-width bound for $\mathcal{T}_h$.

For any two random-cluster configurations $I, F \in \Omega$ on $\mathcal{T}_h$, define the canonical path $\gamma_{I,F}$ as the path of FK dynamics transitions which sequentially processes the edges

of $E(\mathcal{T}_h)$ according to the ordering induced by σ , i.e., $e_{\sigma(1)}, e_{\sigma(2)}, \dots, e_{\sigma(|E(\mathcal{T}_h)|)}$, and whenever there is a discrepancy $I(e_{\sigma(i)}) \neq F(e_{\sigma(i)})$, the transition is the one that changes the state of $e_{\sigma(i)}$ from $I(e_{\sigma(i)})$ to $F(e_{\sigma(i)})$.

For an FK dynamics transition (η, η^i) where $\eta^i := \eta \oplus e_{\sigma(i)}$, construct a bijection from the set of $\{I, F : (\eta, \eta^i) \in \gamma_{I,F}\}$ to Ω by setting

$$\omega_\eta(I, F) = \{I(e_{\sigma(j)}) : j \leq i\} \cup \{F(e_{\sigma(j)}) : j > i\}.$$

This is a bijection because I is recovered via $I = \{\omega_\eta(e_{\sigma(j)}) : j \leq i\} \cup \{\eta(e_{\sigma(j)}) : j > i\}$ and F is analogously recovered via $F = \{\eta(e_{\sigma(j)}) : j \leq i\} \cup \{\omega_\eta(e_{\sigma(j)}) : j > i\}$.

For ease of notation, let $\pi = \pi_{\mathcal{T}_h}^\xi$. The standard canonical paths bound (see [48, Corollary 13.20]) then ensures the inverse gap satisfies

$$\begin{aligned} \lambda_P^{-1} &\leq \max_i \max_\eta \frac{1}{\pi(\eta)P(\eta, \eta^i)} \sum_{I, F: (\eta, \eta^i) \in \gamma_{I,F}} \pi(I)\pi(F)|\gamma_{I,F}| \\ &\leq |E(\mathcal{T}_h)| \max_{\eta, i} \max_{I, F: (\eta, \eta^i) \in \gamma_{I,F}} \frac{1}{P(\eta, \eta^i)} \frac{\pi(I)\pi(F)}{\pi(\eta)\pi(\omega_\eta(I, F))}. \end{aligned}$$

The probability $P(\eta, \eta^i)$ is at least the probability of picking the edge $e_{\sigma(i)}$ to update ($1/|E(\mathcal{T}_h)|$) times the minimal probability of flipping an edge, which is some $C = C(p, q) > 0$. The ratio of probabilities is bounded by noticing that the number of edges present in the multisets $\{I, F\}$ and $\{\eta, \omega_\eta(I, F)\}$ are the same, leaving only the factor of q to contribute. Without the boundary conditions on $\mathcal{T}_h$, we claim that

$$|k(I) + k(F) - k(\eta) - k(\omega_\eta(I, F))| \leq 2|V(e_{\sigma(j):j \leq i}) \cap V(e_{\sigma(j):j > i})|, \quad (16)$$

where we recall that $k(\omega)$ is the number of connected components in the subgraph $(V(G), \omega)$ (3). To see (16), note that the only component counts that can differ between $k(I) + k(F)$ and $k(\eta) + k(\omega_\eta(I, F))$ are from components that intersect the vertex boundary between $\{e_{\sigma(j)} : j \leq i\}$ and $\{e_{\sigma(j)} : j > i\}$.

The addition of the boundary conditions can only change the bound on the left-hand side of (16) additively by at most 2. This is because up to a change of the number of components by at most 1, we can split the boundary condition ξ into two parts, one being its part that intersects vertices of $\{e_{\sigma(j)} : j \leq i\}$ and one being its part that intersects vertices of $\{e_{\sigma(j)} : j > i\}$. With this modification, the same reasoning as in the no boundary condition case holds, that the component counts only differ through components that hit the vertex boundary, and this number of components is evidently bounded above by the size of the vertex boundary. Altogether, we get that the spectral gap satisfies

$$\lambda_P^{-1} \leq C|E(\mathcal{T}_h)|^2 \exp(2(\text{CW}(\mathcal{T}_h) + 1) \log q) \leq C \Delta^h e^{2Kh \log q},$$

which implies the claimed bound up to a change of constants. $\square$

6 Mixing time on locally treelike graphs

We now use the understanding from Sect. 5 on the random-cluster model on trees with r -wired boundary conditions to control their mixing time on treelike graphs having exponentially strong supercritical phases for the edge-percolation: namely, to prove Theorem 2. As mentioned in the introduction, the prototypical example to have in mind in this section is a treelike expander like the random Δ -regular graph.

The proof strategy of Theorem 2 consists of the following three steps, which will organize the section:

1. An $O(1)$ burn-in period to obtain (nearly) r -wired boundary conditions on the locally treelike $O(\log n)$ -radius balls of the graph;
2. Censoring of the dynamics after burn-in to localize to a treelike ball (and an estimate on how long the mixing time will be on the local ball with (nearly) r -wired boundary conditions);
3. A spatial mixing property with (nearly) r -wired boundary conditions to couple the two censored copies after they have each respectively reached equilibrium.

6.1 Burn-in to induce (r, L) -wired boundary conditions on treelike balls

In this subsection, we demonstrate that FK dynamics after an $O(1)$ burn-in period will be such that the boundary conditions it induces on $\frac{\eta}{2} \log n$ sized balls are (almost) r -wired. This will be essential to the application of the spatial mixing results of the previous section. Since the property of being r -wired is a monotone increasing property on the distribution over single-component boundary conditions, it will essentially suffice to establish it for the $\text{Ber}(\tilde{p})$ edge percolation $\tilde{\omega}$, and use Lemma 6. (In reality, there are some added complications by the possible $O(1)$ many extra wirings outside the single-component, as a bound on those extra wirings is no longer an increasing event.)

In what follows, suppose G is a graph of minimum degree 3. Let $B_h = B_h(o)$ be the ball of radius h about a fixed vertex o and suppose it is K -treelike, meaning the removal of at most K edges from $E(B_h)$ leaves a tree. By taking the breadth-first exploration of B_h , any vertex w can be assigned a height via $h - d(w, o)$ and *children* which are all vertices adjacent to w having smaller height. Let the *descendant graph* of w , denoted D_w (in analogy with $T_{h,w}$), be the set of all descendants of w , together with their descendants, etc.

Definition 15 A distribution P over boundary conditions on a K -treelike ball B_h is called (r, L) -wired if it is generated as follows:

- Some arbitrary L vertices in ∂B_h are chosen and an arbitrary wiring is placed on them;
- On the remainder, a subset stochastically dominating the product $\text{Ber}(r)$ subset is wired together into one large component.

The main result of this subsection is that the boundary conditions induced on a treelike ball by the FK dynamics after an $O(1)$ burn-in time are (p, L) -wired. This will follow from the following.

Lemma 18 Suppose G has an exponentially strong supercritical phase per Definition 3, and for $h = \eta \log n$, the graph G is (K, h) -treelike. For every r , there exists $\tilde{p}(\eta, r, c_{\tilde{p}}, \tilde{p}_0)$ such that if $\omega \succeq \tilde{\omega}$, then for every $o \in V(G)$, the distribution over boundary conditions induced by $\omega(E(G) \setminus E(B_{h/2}))$ on $\partial B_{h/2}$ is within TV-distance n^{-5} of a (r, K) -wired boundary condition.

The complications for establishing the above lemma are that the class of (r, L) -wired boundary conditions are not a monotone family, and that the wirings of boundary vertices of B_h are dictated by events of connectivity to a giant which are not independent even under $\tilde{\omega}$. We develop an auxiliary set of events on the random-cluster configuration which provide the necessary monotonicity and independence. For this, we note that all cycles in B_h have either 1 or 2 vertices of minimal height; for a vertex $w \in \partial B_{h/2}$, we say the descendant graph D_w in B_h is a simple subtree if it does not contain any vertices of minimal height of a cycle of B_h .

- Define E_1 as the event that every component of ω of size at least $\frac{\eta}{2} \log n$ coincides (i.e., ω has at most one component of size greater than $\frac{\eta}{2} \log n$).
- For vertices $w \in \partial B_{h/2}$ whose descendant graph D_w in B_h is a simple subtree, let E_w be the event that the configuration $\omega(D_w)$ has size greater than $\frac{\eta}{2} \log n$. (For other w , let E_w be vacuous.)

On the event E_1 , the subset of $w \in \partial B_{h/2}$ such that E_w holds will all be wired together through the giant, and the set of additional wirings of boundary vertices must be confined to those w for which its descendant graph in B_h is not a simple subtree, which is deterministically at most $2K$ since G is (K, h) -treelike. It therefore suffices for us to establish that for all $\tilde{p}$ sufficiently large, if $\omega \succeq \tilde{\omega}$,

$$\mathbb{P}(\omega \notin E_1) \leq n^{-5} \quad \text{and} \quad \{w : \omega \in E_w\} \geq \bigotimes_w \text{Ber}(\tilde{p}),$$

whence on the event E_1 , the boundary conditions induced by $\omega(E(G) \setminus E(B_{h/2}))$ on $\partial B_{h/2}$ would be (r, K) -wired. In particular, Lemma 18 is an immediate consequence of the following two lemmas.

Lemma 19 Suppose G satisfies Definition 3 with some $c_{\tilde{p}}, \tilde{p}_0$. For every $\eta > 0$, there exists $\tilde{p}'_0$ such that if $\tilde{p} \geq \tilde{p}'_0$, and $\omega \succeq \tilde{\omega}$, then

$$\mathbb{P}(\omega \notin E_1) \leq n^{-5}.$$

Proof In order for $\omega \notin E_1$, there must exist a component of size between $\frac{\eta}{2} \log_{\Delta} n$ and $n/2$ (two components both of size at least $n/2$ evidently coincide). Thus, by a union bound the probability of $\omega \notin E_1$ is bounded by the probability that there exists some connected set A of size between $\frac{\eta}{2} \log_{\Delta} n$ and $n/2$ with $\omega(\partial_e A) \equiv 0$; this being a decreasing event, it suffices to upper bound its probability under $\tilde{\omega}$, whence we can apply Definition 3 with $\ell = \frac{\eta}{2} \log_{\Delta} n$ to get

$$\tilde{\pi}(\tilde{\omega} \notin E_1) \leq ne^{-c_{\tilde{p}} \eta \log n}.$$

Since $c_{\tilde{p}}$ goes to ∞ as $\tilde{p} \rightarrow 1$, as long as $\tilde{p}$ is larger than some $\tilde{p}_0(\eta)$ the right-hand side will be at most n^{-5} . $\square$

Lemma 20 *Suppose G is (K, h) -treelike. For every $r < 1$, there exists $\tilde{p}$ such that if $\omega \succeq \tilde{\omega}$ then the distribution of $\{w : \omega \in E_w\}$ stochastically dominates a $\text{Ber}(r)$ subset of $\partial B_{h/2}$.*

Proof Since the E_w are increasing events, $\{w : \omega \in E_w\}$ stochastically dominates $\{w : \tilde{\omega} \in E_w\}$, it suffices to establish the above for $\tilde{\omega}$. Since the descendant graphs D_w are disjoint for the $w \in \partial B_{h/2}$ for which D_w are subtrees (and hence E_w is not vacuous), the events E_w are independent under $\tilde{\omega}$. It remains to argue that $\mathbb{P}(\tilde{\omega} \in E_w) \geq r$ as long as $\tilde{p}$ is sufficiently large. Since the descendant graphs D_w contain full binary trees as subgraphs, the probability of $\tilde{\omega} \in E_w$ is larger than the probability that the branching process with offspring distribution $\text{Bin}(2, \tilde{p})$ survives $\frac{\eta}{2} \log n$ generations, which happens with probability going to 1 as $\tilde{p} \rightarrow 1$: see e.g., Fact 16. In particular, for every r , the probability of $\tilde{\omega} \in E_w$ is greater than r as long as $\tilde{p}$ is large enough (depending only on r). $\square$

6.2 Spatial mixing on treelike balls with (r, L) -wired boundary

We now describe how minor adjustments to Sect. 5.1 lead to spatial mixing on treelike graphs with boundary conditions that are (r, L) -wired.

Lemma 21 *Suppose B_h is K -treelike, and suppose $\mathbb{P}$ is (r, L) -wired. Then except with $\mathbb{P}$ -probability $e^{-c_r(1.1)^{h/(2K)}}$ (with $c_r > 0$ for r large), ξ is such that*

$$\|\pi_{B_h}^{\xi}(\omega(B_{h/2}) \in \cdot) - \pi_{B_h}^1(\omega(B_{h/2}) \in \cdot)\|_{\text{TV}} \leq C e^{-c_p(h/(2K) - K - L)}.$$

for some constant c_p going to ∞ as $p \rightarrow 1$.

Proof Since B_h is K -treelike, there is an edge-set H of size at most K such that $B_h \setminus H$ is a tree. Following the breadth-first search of B_h , there is a stretch of at least $h/(2K)$ consecutive levels between depth $h/2$ and h such that the restriction of B_h to those levels is a forest and every tree in that forest contains a full binary tree as a subgraph (using the minimum degree 3 condition). There exists $m \geq h/2 + h/(2K)$ such that this forest is $B_m \setminus B_{m-h/(2K)}$.

Let $\tilde{S}_{h,\xi}$ be the event that there is a wired separating surface in every one of the constituent trees of height $h/(2K)$ in the configuration $\omega(B_h \setminus H)$. (Notice that since the boundary conditions are at depth h , this depends on the full configuration, not just the restriction to the stretch of $h/(2K)$ heights; moreover by definition of wired separating surface, if $\omega(B_h \setminus H)$ has a wired separating surface then so does $\omega(B_h)$.)

The first claim is that under this definition, there is a minor modification of the revealing procedure of Lemma 14 such that the probability of $\omega_{\xi} \in \tilde{S}_{h,\xi}^c$ upper bounds the total-variation distance. This is done by first revealing the entire configurations ω_{ξ} and ω_1 on $(B_h \setminus H) \setminus B_{m+h/(2K)}$ under the monotone coupling. In this manner, the revealed part of ω_{ξ} induces some single-component boundary conditions

$\tilde{\xi}$ on $B_{m+h/(2K)}$. We can then apply Lemma 14 to each of the constituent trees of $B_m \setminus B_{m-h/(2K)}$ and it follows that on the event that they all have wired separating surfaces in ω_{ξ} , then ω_{ξ} is coupled to ω_1 above those wired separating surfaces and in particular on all of B_m . As such, we have the analogous

$$\|\pi_{B_h}^{\xi}(\omega(B_{h/2}) \in \cdot) - \pi_{B_h}^1(\omega(B_{h/2}) \in \cdot)\|_{\text{TV}} \leq \pi_{B_h}^{\xi}(\omega(B_h \setminus H) \in \tilde{S}_{h,\xi}^c).$$

To control the probability of $\tilde{S}_{h,\xi}^c$, we follow the reasoning of Lemma 15, with the modifications being minimal. The only difference that arises is that when considering the probability of the event in item (2) in that proof, the sub-tree from a vertex $v_i \in B_{m+h/(2K)} \setminus B_m$ in $B_h \setminus H$ no longer necessarily contains a full binary tree: for up to K many vertices, it could be the pruning of a binary tree with up to K many subtrees of depth at least $h/2K$ deleted from it. The easiest thing to do is to simply disregard these vertices which leads to a change from h to $h - K$ in the concentration quality. Similarly, at most L of the subtrees contain at their boundary one of the L vertices where the (r, L) boundary conditions are arbitrarily rewired, and we can disregard these L vertices as well. Nothing else will be affected in the proof. $\square$

6.3 Local mixing time on treelike balls with (r, L) -wired boundary

We now show that the $\eta \log n$ -radius balls in the treelike expander, with the boundary conditions induced on them by the remainder of the FK dynamics configuration, have a polynomial mixing time. By Lemma 18, after a burn-in, these will look like treelike balls with boundary conditions that have an $O(1)$ number of distinct components, one of them being macroscopic, and the rest all being $O(1)$ sized. At this point we can appeal to comparison estimates for Markov chains together with the bound on the inverse spectral gap on trees with single-component mixing times from Lemma 17.

Remark 5 We do not use any further information on the boundary conditions (like the r -wired property or the randomness), only the fact that there is at most 1 large component and $O(1)$ many vertices in the union of all other components. In theory, it is likely that we could use the r -wired property to get a significantly better bound on the mixing time of a tree. However, the extra $O(1)$ wirings and the $O(1)$ many non-tree edges force us to at some point perform a comparison through a spectral gap, and getting back a mixing time bound from this will anyways end up costing a factor of the volume per (15).

To perform comparisons, we formalize a notion of distance between boundary conditions. Two “similar” random-cluster boundary conditions (in terms of the wiring they induce) have similar effects on the underlying random-cluster distribution and on the behavior of the corresponding FK dynamics. In turn, the Dirichlet form, and spectral gaps of their corresponding dynamics should be “close” to one another. We compile a few definitions and results that formalize this idea.

Definition 16 For two boundary conditions $\phi \leq \phi'$, define $D(\phi, \phi') := k(\phi) - k(\phi')$ where $k(\phi)$ is the number of components in ϕ . For two partitions ϕ, ϕ' that are not

comparable, let ϕ'' be the smallest partition such that $\phi'' \geq \phi$ and $\phi'' \geq \phi'$ and set $D(\phi, \phi') = k(\phi) - k(\phi'') + k(\phi') - k(\phi'')$.

The following lemma is then straightforward from the definition of the random-cluster measure (3).

Lemma 22 (E.g., Lemma 2.2 from [12]) *Let G be arbitrary, $p \in (0, 1)$ and $q > 0$. Let ϕ and ϕ' be any two partitions of $V(G)$. Then, for all random-cluster configurations $\omega \in \{0, 1\}^E$, we have*

$$q^{-2D(\phi, \phi')} \pi_G^{\phi'}(\omega) \leq \pi_G^\phi(\omega) \leq q^{2D(\phi, \phi')} \pi_G^{\phi'}(\omega).$$

The following corollary is a standard comparison of spectral gaps, and follows from Lemma 22, the definition of the transition matrix of the FK dynamics, and Theorem 4.1.1 in [54].

Corollary 23 *Let $G = (V, E)$ be an arbitrary graph, $p \in (0, 1)$ and $q > 0$. Consider the FK dynamics on G with boundary conditions ϕ and ϕ' , and let λ, λ' denote their respective spectral gaps. Then,*

$$q^{-5D(\phi, \phi')} \lambda' \leq \lambda \leq q^{5D(\phi, \phi')} \lambda'.$$

Using the above, we are able to deduce the following bound.

Lemma 24 *Consider a K -treelike ball B_h with boundary conditions ξ that have one component of arbitrary size together with at most L many additional boundary wirings. There exists $a(\Delta, q)$ such that the inverse spectral gap on B_h^ξ is at most $C_{p,q} \exp(a(h + K + L))$.*

Proof Consider the modification of B_h^ξ where all endpoints of the set H are wired up to one another via a boundary condition, and denote it by $\widetilde{B}_h^\xi$. By Corollary 23, their spectral gaps are within a factor of $q^{10|H|}$ of one another. The FK dynamics on $\widetilde{B}_h^\xi$ are a product of the FK dynamics on $B_h \setminus H$ with boundary conditions ξ and the wirings of the edges of H (call this $(\widetilde{B}_h \setminus H)^\xi$), and $|H|$ independent FK dynamics chains on single edges with wired boundary. By tensorization of the spectral gap (see e.g., [54]) the spectral gap of the FK dynamics on B_h is then the minimum of the gap on those individual edges, and the gap of the FK dynamics on $(\widetilde{B}_h \setminus H)^\xi$.

The spectral gaps of the individual edges are clearly some constant depending on p , so it suffices to bound the spectral gap of FK dynamics on $(\widetilde{B}_h \setminus H)^\xi$. For this purpose, notice that with a cost of $q^{5(|H|+L)}$ we can perform a further boundary modification and remove the wirings of the edges in H as well as those L additional wirings in ξ to end up with the tree $B_h \setminus H$ with a single-component boundary condition. At this point, we can bound the spectral gap of this resulting single-component tree of depth h using Lemma 17. Putting together the costs from the various comparisons we obtain the desired. $\square$

6.4 Mixing times on locally treelike graphs

In this section, we combine the above parts to establish the near-linear mixing time bound of Theorem 2 for FK dynamics on locally treelike graphs, as long as they have an exponentially strong supercritical phase for edge-percolation.

Proof of Theorem 2 Recall that we use $X_t^{\omega_0}$ to denote the FK dynamics at time t from initialization ω_0 . By monotonicity, under the grand coupling, we have

$$\begin{aligned} \max_{\omega, \omega'} \mathbb{P}(X_t^\omega \neq X_t^{\omega'}) &\leq \sum_{e \in E(G)} \mathbb{P}(X_t^\omega(e) \neq X_t^{\omega'}(e)) \\ &\leq \sum_{e \in E(G)} \mathbb{P}(X_t^0(e) \neq X_t^1(e)) = \sum_{e \in E(G)} \mathbb{E}[X_t^1(e)] - \mathbb{E}[X_t^0(e)]. \end{aligned} \quad (17)$$

Fix an edge $e \in E(G)$ and consider the difference in expectations on the right. Set $\bar{X}_t^0$ to be the censored FK dynamics that agrees with X_t^0 for all times until some T_0 but that then censors (ignores) all updates after time T_0 outside of $B_h = B_h(o)$ for some $o \in e$ and for $h = \frac{\eta}{2} \log n$. Let $\bar{X}_t^1$ be the Markov chain that censors *all* updates of X_t^1 outside of B_h (regardless of t). Then, by the censoring inequality of [52],

$$\mathbb{E}[X_t^1(e)] - \mathbb{E}[X_t^0(e)] \leq \mathbb{E}[\bar{X}_t^1(e)] - \mathbb{E}[\bar{X}_t^0(e)]. \quad (18)$$

By Lemma 6 and Lemma 18, for every r , there exist $T_0(r, q)$ and p_0 large enough that for all $p \geq p_0$, the FK dynamics $X_{T_0}^0, X_{T_0}^1$ induce (r, K) -wired boundary conditions on $\partial B_{h/2}$.

Let $A_\gamma = A_\gamma(B_h)$ be the set of boundary conditions ξ on B_h that are single-component together with at most K additional wirings, and furthermore that are such that the inequality of Lemma 21 holds with constant $c_p = \gamma$. Since $X_{T_0}^0, X_{T_0}^1$ induce (r, K) -wired boundary conditions on $\partial B_{h/2}$, by Lemma 21,

$$\mathbb{P}(X_{T_0}^0(B_{h/2}^c) \notin A_\gamma^c) \leq e^{-c_r(1.1)^{h/2K}}. \quad (19)$$

At the same time, by definition of A_γ , we have

$$\max_{\xi \in A_\gamma} (\pi_{B_h}^1(\omega_e) - \pi_{B_h}^\xi(\omega_e)) \leq C e^{-\gamma h/(3K)}. \quad (20)$$

so long as n is large enough that $h/(2K) - K - K \geq h/(3K)$. Since the event $\{\bar{X}_{T_0}^0(B_h^c) \in A_\gamma\} = \{X_{T_0}^0(B_h^c) \in A_\gamma\}$ is measurable w.r.t. $\mathcal{F}_{T_0}$ (the filtration generated by the grand coupling up to time T_0),

$$\begin{aligned} &\mathbb{E}[\bar{X}_{T_0+S}^1(e)] - \mathbb{E}[\bar{X}_{T_0+S}^0(e)] \\ &\leq \mathbb{P}(\bar{X}_{T_0}^0(B_h^c) \notin A_\gamma) + \mathbb{P}(\bar{X}_{T_0+S}^1(e) \neq \bar{X}_{T_0+S}^0(e) \mid \bar{X}_{T_0}^0(B_h^c) \in A_\gamma) \end{aligned}$$

$$\leq \mathbb{P}(X_{T_0}^{\mathbf{0}}(B_h^c) \notin A_\gamma) + \max_{\xi \in A_\gamma(B_h)} (\mathbb{E}[Y_{S, B_h}^{\mathbf{1}}(e)] - \mathbb{E}[Z_{S, B_h}^{\mathbf{0}}(e)]),$$

where $Y_{s, B_h}^{\mathbf{1}}$ and $Z_{s, B_h}^{\mathbf{0}}$ are Glauber chains on B_h with boundary conditions $\mathbf{1}$ and ξ respectively, initialized from $\mathbf{1}$ and $\mathbf{0}$ respectively. We have used here the definition of censored dynamics and monotonicity. The first term is at most $e^{-c_r(1.1)^{h/2K}}$ by (19).

For the second term, fix any $\xi \in A_\gamma(B_h)$, let $Y_s = Y_{s, B_h}^{\mathbf{1}}$ and $Z_s = Z_{s, B_h}^{\mathbf{0}}$, and write

$$\begin{aligned} \mathbb{E}[Y_S(e)] - \mathbb{E}[Z_S(e)] &= (\mathbb{E}[Y_S^{\mathbf{1}}(e)] - \pi_{B_h}^{\mathbf{1}}(\omega_e)) \\ &\quad + (\pi_{B_h}^{\mathbf{1}}(\omega_e) - \pi_{B_h}^{\xi}(\omega_e)) + (\pi_{B_h}^{\xi}(\omega_e) - \mathbb{E}[Z_S^{\mathbf{0}}(e)]). \end{aligned}$$

The middle term is at most $Ce^{-\gamma h/(3K)}$ by (20). For the first and third, suppose

$$S \geq C_0 \log n \cdot \max_{\xi \in A_\gamma(B_h)} \log \left(\frac{1}{\min_{\omega} \pi_{B_h}^{\xi}(\omega)} \right) \cdot \lambda_{B_h}^{-1},$$

where $\lambda_{B_h}^{-1}$ is the inverse spectral gap of the FK dynamics on B_h with boundary conditions ξ . Then by (15) and sub-multiplicativity of TV-distance to stationarity, for a universal constant C_0 , both the first and third terms will be at most n^{-5} . Combining, we get

$$\mathbb{E}[\bar{X}_{T_0+S}^{\mathbf{1}}(e)] - \mathbb{E}[\bar{X}_{T_0+S}^{\mathbf{0}}(e)] \leq C(e^{-c_r(1.1)^{h/(2K)}} + e^{-\gamma h/(3K)} + n^{-5}). \quad (21)$$

At this point we make the following choices for η , γ , T_0 , S :

1. $h = \frac{\eta}{2} \log n$ for $\eta > 0$ sufficiently small that for every $\xi \in A_\gamma(B_h)$,

$$\log \left(\frac{1}{\min_{\omega} \pi_{B_h}^{\xi}(\omega)} \right) \cdot \lambda_{B_h}^{-1} \leq \Delta^{\frac{\eta}{2} \log n} \lambda_{B_h}^{-1} \log(1-p) \quad \text{is at most } n^{\varepsilon/2} \log(1-p);$$

2. $S = C_1 n^{\varepsilon/2} \log n$ where $C_1 = C_0 \log(1-p)$;
3. γ large enough that $e^{-\gamma h/(2K)}$ is at most n^{-5} ;
4. r large enough that $c_r > 0$
5. T_0 and p large enough that $X_{T_0}^{\mathbf{0}}$ induces (r, K) -wired boundary on $\partial B_{h/2}$.

The existence of such an $\eta(\Delta, q, K)$ follows from Lemma 24 and the fact that a only depends on Δ, q . Furthermore, by taking r large, we can make γ and c_r arbitrarily large, to satisfy items (3)–(4). Finally, Lemma 18 ensures we can take T_0, p_0 large enough that for all $p \geq p_0$, item (5) is satisfied.

With these choices, for n sufficiently large, the right-hand side of (21) is at most n^{-4} . Combining these with (17)–(18), we get that there exists $p_0(c_{\bar{p}}, \tilde{p}_0, \Delta, \eta, q, K)$ such that for all $p \geq p_0$,

$$\max_{\omega} \|\mathbb{P}(X_{T_0+S}^{\omega} \in \cdot) - \pi\|_{\text{TV}} \leq \max_{\omega, \omega'} \mathbb{P}(X_{T_0+S}^{\omega} \neq X_{T_0+S}^{\omega'}) = o(1),$$

which concludes the proof since evidently $T_0 + S = O(n^{\varepsilon/2} \log n) = O(n^{\varepsilon})$. $\square$

7 Graphs with slow mixing at arbitrarily low temperatures

In this section, we establish the following slow mixing result for the FK dynamics. Theorem 3 is the special case where we have restricted to integer q and used the comparison results of [56].

Theorem 25 Fix $\Delta \geq 3$ and $p_0 < 1$.

1. For any $q \geq 3$, there exists $p > p_0$ and a sequence of locally treelike graphs $(G_n)_n$ of maximum degree Δ such that the FK dynamics on G_n have $t_{\text{MIX}} \geq \exp(\Omega(n))$.
2. For any $q > 4$ (possibly non-integer), there exists $p > p_0$ and a sequence $(G_n)_n$ of polynomial volume growth and maximum degree Δ such that the FK dynamics of G_n have $t_{\text{MIX}} \geq \exp(\Omega(\sqrt{n}))$.

The key tool for this proof will be the so-called “series law” for the random-cluster model. That is, splitting any edge e into two edges, and changing the edge probability parameter for e from p_e to approximately $\sqrt{p_e}$ for each of the two new resulting edges preserves the random-cluster measure when one re-identifies the edges and takes the status of e as being open if the two new edges are open. This gives a mechanism for boosting any fixed p into a p' which gets closer and closer to 1, while only increasing the number of edges and vertices in the graph by a constant factor. In this manner, if a graph H has slow mixing for its FK dynamics at some value of p (no matter how small), then its modification G (obtained by multiple applications of the series law) can be made to have slow mixing at p' , for p' that can be arbitrarily close to 1. Moreover, the graph modifications do not distort the maximum degree and volume growth (though they importantly do distort the isoperimetric dimension, and expansion rates, which we recall were fundamental to the presence of a strongly supercritical phase for the edge-percolation on the graph).

Remark 6 Both items (1)–(2) of Theorem 25 should hold for all $q > 2$. The gap for q non-integer in item (1) and $q \in (2, 4]$ in item (2) come from the present lack of proof (to our knowledge) of a slowdown for FK dynamics on bounded degree graphs (satisfying the corresponding graph condition) at those values of q . Such a slowdown is widely expected at the critical points both for the random regular graph and on $(\mathbb{Z}/m\mathbb{Z})^d$ for large d . The values of q for which we can establish our lower bound come from the slowdowns of the random regular graph at integer $q \geq 3$ [21], and the torus $(\mathbb{Z}/\sqrt{n}\mathbb{Z})^2$ at $q > 4$ [30].

Let us precisely recall the series law of the random-cluster model from [36, Theorem 3.89].

Lemma 26 Two edges e, f of a graph $G = (V, E)$ are in series if $e = \{u, v\}$ and $f = \{v, w\}$ and v has no other incident edges. Let

$$\sigma(x, y, q) = \frac{xy}{1 + (q - 1)(1 - x)(1 - y)}. \quad (22)$$

Let $G' = (V \setminus \{v\}, (E \setminus \{e, f\}) \cup \{u, w\})$. For a random-cluster configuration ω on G , define ω' on G' by setting $\omega'(a) = \omega(a)$ for $a \in E \setminus \{e, f\}$ and $\omega'(\{u, w\}) = \omega(e) \cdot \omega(f)$. Then, if ω is sampled from the random-cluster distribution on G with parameters $(p_a)_{a \in E}$ and $q > 0$, ω' is distributed according to the random-cluster distribution on G' with parameters $(p_a)_{a \in E \setminus \{e, f\}}$ and $p_{\{u, w\}} = \sigma(p_e, p_f, q)$.

Proof of Theorem 25 For item (1), let $(H_n)_n$ be a locally treelike sequence of n -vertex graphs of degree at most Δ such that the mixing time of FK dynamics at some fixed value of $p \in (0, 1)$ is $\exp(\Omega(n))$. From Theorem 1.2 of [21], we know that a randomly drawn sequence of Δ -regular graphs satisfies this bound with high probability if $q \geq 3$ is an integer and $p = p_c(q, \Delta)$. For item (2), let $(H_n)_n$ be the torii $(\mathbb{Z}/\sqrt{n}\mathbb{Z})^2$. We then know from Theorem 2 of [30] that at fixed $p = p_c(q) \in (0, 1)$ the mixing time of FK dynamics on H_n is $\exp(\Omega(\sqrt{n}))$ for all real $q > 4$.

Let G_n be the modification of H_n in which every edge of H_n is split into 2^K edges in series, for K determined as follows. Let $\zeta_q(p)$ be the inverse of $\sigma(x, x, q)$ from (22), i.e.,

$$\sigma(\zeta_q(p), \zeta_q(p), q) = p.$$

Such an inverse exists and is increasing for $p \in (0, 1)$ by virtue of the fact that $\sigma(x, x, q)$ is continuously increasing from 0 to 1 as x ranges from $[0, 1]$. Since $x^2/q \leq \sigma(x, x, q) \leq x^2$ it must be the case that $\zeta_q(p)^2/q \leq \sigma(\zeta_q(p), \zeta_q(p), q) \leq \zeta_q(p)^2$. In order for this to be equal to p , it must be that $\sqrt{p} \leq \zeta_q(p) \leq \sqrt{pq}$.

Take K such that

$$\zeta_q^{\circ K}(p) := \underbrace{\zeta_q(\zeta_q(\cdots \zeta_q(p) \cdots))}_K \geq p_0.$$

The inequality $\sqrt{p} \leq \zeta_q(p)$ ensures that for any

$$K \geq \log_2 \left(\frac{\log p}{\log p_0} \right),$$

we have $\zeta_q^{\circ K}(p) \geq p_0$. By Lemma 26, if ω' is a random-cluster configuration on G_n with parameters $p' = \zeta_q^{\circ K}(p)$ and q , and $\omega_e = \prod_{e_i} \omega'_{e_i}$ for every edge $e \in E(H_n)$, where $(e_i)_{i=1}^{2^K}$ are the edges in G_n derived from splitting e , then ω is a sampled from the random-cluster measure on H_n with parameters p and q .

We now claim that the FK dynamics on G_n with parameters p' and q have $\exp(\Omega(n))$ mixing if H_n has $\exp(\Omega(n))$ mixing at parameters (p, q) (the reasoning that it has $\exp(\Omega(\sqrt{n}))$ mixing if H_n has $\exp(\Omega(\sqrt{n}))$ mixing is identical, so we omit it). This is achieved by lifting a bottleneck set from H_n to G_n . In what follows, it is convenient to work with the discrete-time Glauber dynamics (though the same proof would work in the continuous-time setting as well). Let P_H denote the transition matrix of the discrete-time FK dynamics on H_n at parameter p and P_G the transition matrix of the discrete-time FK dynamics on G_n at parameter p' . Similarly, π_{H_n} is at parameter p while π_{G_n} is at parameter p' .

By assumption, the mixing time of FK dynamics on H_n is exponential in n . As such, there exists a subset of configurations $A \subset \{0, 1\}^{E(H_n)}$ of exponentially small conductance. More precisely, there must exist $A \subset \{0, 1\}^{E(H_n)}$ such that $\pi_{H_n}(A) \leq 1/2$ and

$$\Phi_H(A) := \frac{\sum_{x \in A, y \in A^c} \pi_{H_n}(x) P_H(x, y)}{\pi_{H_n}(A)} = \exp(-\Omega(n)); \quad (23)$$

see, e.g., [48, Theorem 13.10]. Let $\partial_{P_H} A = \{x \in A : P_H(x, y) > 0 \text{ for some } y \in A^c\}$. Since for each $x \in \partial_{P_H} A$, each entry of P_{H_n} is at least $\frac{1}{n} \hat{p}$, we see that there exists a constant $c > 0$ such that

$$\frac{\pi_{H_n}(\partial_{P_H} A)}{\pi_{H_n}(A)} = \frac{\sum_{x \in \partial_{P_H} A} \pi_{H_n}(x)}{\pi_{H_n}(A)} \leq n \hat{p}^{-1} \Phi_H(A) \leq \exp(-cn). \quad (24)$$

Let T_A be the subset of $\{0, 1\}^{E(G_n)}$ defined by $T_A = \{\omega' : \omega \in A\}$ where the relationship between ω and ω' is defined per the operation described above. By Lemma 26, $\pi_{G_n}(T_A) = \pi_{H_n}(A)$. Every configuration in ∂T_A must be in $T_{\partial A}$ because if $P_G(\omega', \sigma') > 0$, for $\sigma' \notin T_A$, then σ' projects down to a configuration in A^c , so ω' must project into ∂A . Therefore, by Lemma 26, $\pi_{G_n}(\partial_{P_G} T_A) \leq \pi_{H_n}(\partial_{P_H} A)$. Altogether, it follows from (24) that

$$\frac{\pi_{G_n}(\partial_{P_G} T_A)}{\pi_{G_n}(T_A)} \leq \exp(-cn).$$

Using the facts that for every x , the number of $y \in T_A^c$ for which $P_G(x, y)$ is positive is at most $E(G_n)$,

$$\frac{\sum_{x \in T_A, y \in T_A^c} \pi_{G_n}(x) P_G(x, y)}{\pi_{G_n}(T_A)} \leq |E(G_n)| \frac{\pi_{G_n}(\partial_{P_G} T_A)}{\pi_{G_n}(T_A)},$$

implies that T_A is a set of exponentially small conductance for the FK dynamics on G_n . This then implies the inverse gap and mixing time of FK dynamics on G_n are both exponential in n .

It remains to reason that the number of vertices and edges of G_n are of the same order as the number of vertices and edges of H_n , so that the resulting bounds are indeed exponential in $|V(G_n)|$. Notice that as long as $p = \Omega(1)$ and $1 - p_0 = \Omega(1)$, then $K = O(1)$. As such, G_n will have $|V(G_n)| \leq |V(H_n)| + 2^K |E(H_n)| = O(|V(H_n)|)$, and $|E(G_n)| \leq 2^K |E(H_n)|$. This yields the claimed bound. $\square$

Acknowledgements The authors thank the anonymous referees for their careful reading and comments. The research of AB was supported in part by NSF Grant CCF-2143762. The research of R.G. was supported in part by NSF DMS-2246780.

Author Contributions The authors contributed equally to this work.

Declarations

Conflict of interest The authors have no conflict of interest to declare that are relevant to the content of this article.

References

1. Alon, N., Benjamini, I., Stacey, A.: Percolation on finite graphs and isoperimetric inequalities. *Ann. Probab.* **32**(3), 1727–1745 (2004)
2. Anari, N., Liu, K., Gharan, S.O., Vintzant, C.: Log-concave polynomials II: high-dimensional walks and an FPRAS for counting bases of a matroid. *Ann. Math.* (2023) **(to appear)**
3. Berger, N., Kenyon, C., Mossel, E., Peres, Y.: Glauber dynamics on trees and hyperbolic graphs. *Probab. Theory Relat. Fields* **131**(3), 311–340 (2005). Extended abstract appeared in Proceedings 42nd IEEE Symposium on Foundations of Computer Science (FOCS 2001)
4. Blanca, A., Gheissari, R.: Sampling from the Potts model at low temperatures via Swendsen–Wang dynamics. In: 2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS), pp. 2006–2020, Los Alamitos, CA, USA. IEEE Computer Society (2023)
5. Blanca, A., Cannon, S., Perkins, W.: Fast and perfect sampling of subgraphs and polymer systems. In: Proceedings of APPROX/RANDOM (2022)
6. Blanca, A., Caputo, P., Chen, Z., Parisi, D., Štefankovič, D., Vigoda, E.: On mixing of Markov chains: coupling, spectral independence, and entropy factorization. In: Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 3670–3692 (2022)
7. Blanca, A., Caputo, P., Parisi, D., Sinclair, A., Vigoda, E.: Entropy decay in the Swendsen–Wang dynamics on $\mathbb{Z}^d$. In: Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2021, New York, NY, USA, pp. 1551–1564. Association for Computing Machinery (2021)
8. Blanca, A., Chen, Z., Štefankovič, D., Vigoda, E.: The Swendsen–Wang dynamics on trees. *Rand. Struct. Algorithms* (2022)
9. Blanca, A., Galanis, A., Goldberg, L.A., Štefankovič, D., Vigoda, E., Yang, K.: Sampling in uniqueness from the Potts and random-cluster models on random regular graphs. In: Proceedings of APPROX/RANDOM (2018)
10. Blanca, A., Gheissari, R.: Random-cluster dynamics on random regular graphs in tree uniqueness. *Commun. Math. Phys.* **386**(2), 1243–1287 (2021)
11. Blanca, A., Gheissari, R.: Sampling from Potts on random graphs of unbounded degree via random-cluster dynamics. *Ann. Appl. Probab.* appear (2023). Extended abstract appeared in Proceedings of APPROX/RANDOM 2022
12. Blanca, A., Gheissari, R., Vigoda, E.: Random-cluster dynamics in $\mathbb{Z}^2$: rapid mixing with general boundary conditions. *Ann. Appl. Probab.* **30**(1), 418–459 (2020)
13. Blanca, A., Sinclair, A.: Dynamics for the mean-field random-cluster model. In: Proceedings of APPROX/RANDOM (2015)
14. Blanca, A., Sinclair, A.: Random-cluster dynamics in $\mathbb{Z}^2$. *Probab. Theory Related Fields* **168**, 821–847 (2017). Extended abstract appeared in Proceedings of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2016), pp. 498–513
15. Blanca, A., Sinclair, A., Zhang, X.: The critical mean-field Chayes–Machta dynamics. *Combin. Probab. Comput.* **31**(6), 924–975 (2022)
16. Borgs, C., Chayes, J., Helmuth, T., Perkins, W., Tetali, P.: Efficient sampling and counting algorithms for the Potts model on $\mathbb{Z}^d$ at all temperatures. In: Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020, pp. 738–751, New York, NY, USA. Association for Computing Machinery (2020)
17. Borgs, C., Chayes, J.T., Tetali, P.: Tight bounds for mixing of the Swendsen–Wang algorithm at the Potts transition point. *Probab. Theory Relat. Fields* **152**(3–4), 509–557 (2012)
18. Carlson, C., Davies, E., Kolla, A.: Efficient algorithms for the Potts model on small-set expanders. *arXiv preprint arXiv:2003.01154* (2020)
19. Carlson, C., Davies, E., Fraiman, N., Kolla, A., Potukuchi, A., Yap, C.: Algorithms for the ferromagnetic Potts model on expanders. In: 2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS), pp. 344–355. IEEE (2022)

20. Chen, Z., Galanis, A., Goldberg, L.A., Perkins, W., Stewart, J., Vigoda, E.: Fast algorithms at low temperatures via Markov chains. *Rand. Struct. Algorithms* **58**(2), 294–321 (2021)
21. Coja-Oghlan, A., Galanis, A., Goldberg, L.A., Ravelomanana, J.B., Štefankovič, D., Vigoda, E.: Metastability of the Potts ferromagnet on random regular graphs. *Commun. Math. Phys.* (2023)
22. Contreras, D., Martineau, S., Tassion, V.: Supercritical percolation on graphs of polynomial growth (2021). Preprint available at [arXiv:2107.06326](https://arxiv.org/abs/2107.06326)
23. Cooper, C., Frieze, A.M.: Mixing properties of the Swendsen–Wang process on classes of graphs. *Rand. Struct. Algorithms* **15**, 242–261 (1999)
24. Dyer, M., Sinclair, A., Vigoda, E., Weitz, D.: Mixing in time and space for lattice spin systems: a combinatorial view. *Rand. Struct. Algorithms* **24**(4), 461–479 (2004)
25. Easo, P., Hutchcroft, T.: Supercritical percolation on finite transitive graphs I: uniqueness of the giant component (2021). Preprint available at [arXiv:2112.12778](https://arxiv.org/abs/2112.12778)
26. Edwards, R.G., Sokal, A.D.: Generalization of the Fortuin–Kasteleyn–Swendsen–Wang representation and Monte Carlo algorithm. *Phys. Rev. D* (3) **38**(6), 2009–2012 (1988)
27. Galanis, A., Goldberg, L.A., Smolarova, P.: Sampling from the random cluster model on random regular graphs at all temperatures via Glauber dynamics. Extended abstract to appear in *ph Proceedings of APPROX/RANDOM 2023*
28. Galanis, A., Štefankovič, D., Vigoda, E.: Swendsen–Wang algorithm on the mean-field Potts model. In: *Proceedings of APPROX/RANDOM* (2015)
29. Galanis, A., Štefankovič, D., Vigoda, E., Yang, L.: Ferromagnetic Potts model: refined #BIS-hardness and related results. *SIAM J. Comput.* **45**(6), 2004–2065 (2016)
30. Gheissari, R., Lubetzky, E.: Mixing times of critical two-dimensional Potts models. *Commun. Pure Appl. Math.* **71**(5), 994–1046 (2018)
31. Gheissari, R., Lubetzky, E., Peres, Y.: Exponentially slow mixing in the mean-field Swendsen–Wang dynamics. *Ann. de l’Inst. Henri Poincaré (B)* (2019). Extended abstract appeared in *Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2018)*, pp. 1981–1988
32. Gheissari, R., Sinclair, A.: Spatial mixing and the random-cluster dynamics on lattices. In: *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 4606–4621 (2023)
33. Goldberg, L.A., Jerrum, M.: Approximating the partition function of the ferromagnetic Potts model. *J. ACM* **59**(5) (2012)
34. Gore, V.K., Jerrum, M.R.: The Swendsen–Wang process does not always mix rapidly. In: *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing, STOC ’97*, New York, NY, USA, pp. 674–681. Association for Computing Machinery (1997)
35. Grigorchuk, R., Pak, I.: Groups of intermediate growth: an introduction (2008)
36. Grimmett, G.: The random-cluster model. In: *Probability on Discrete Structures*, Volume 110 of *Encyclopaedia Math. Sci.*, pp. 73–123. Springer, Berlin (2004)
37. Guo, H., Jerrum, M.: Random cluster dynamics for the Ising model is rapidly mixing. In: *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2017*, pp. 1818–1827 (2017)
38. Harel, M., Spinka, Y.: Finitary codings for the random-cluster model and other infinite-range monotone models. *Electron. J. Probab.* **27**, 1–32 (2022)
39. Hayes, T.P., Sinclair, A.: A general lower bound for mixing of single-site dynamics on graphs. In: *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 511–520. IEEE (2005)
40. Helmuth, T., Jenssen, M., Perkins, W.: Finite-size scaling, phase coexistence, and algorithms for the random cluster model on random graphs. *Ann. de l’institut Henri Poincaré (B) Probab. Stat.* (2023) **(to appear)**
41. Helmuth, T., Perkins, W., Regts, G.: Algorithmic Pirogov–Sinai theory. *Probab. Theory Relat. Fields* **176**(3), 851–895 (2020). Extended abstract appeared in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing (STOC 2019)*
42. Huber, M.: Perfect sampling using bounding chains. *Ann. Appl. Probab.* **14**(2), 734–753 (2004)
43. Huijben, J., Patel, V., Regts, G.: Sampling from the low temperature Potts model through a Markov chain on flows. *Rand. Struct. Algorithms* **62**(1), 219–239 (2023)
44. Hutchcroft, T., Tointon, M.: Non-triviality of the phase transition for percolation on finite transitive graphs (2021). Preprint available at [arXiv:2104.05607](https://arxiv.org/abs/2104.05607)
45. Jenssen, M., Keevash, P., Perkins, W.: Algorithms for #BIS-hard problems on expander graphs. *SIAM J. Comput.* **49**(4), 681–710 (2020)

46. Jerrum, M., Sinclair, A.: Polynomial-time approximation algorithms for the Ising model. *SIAM J. Comput.* **22**, 1087–1116 (1993)
47. Jonasson, J.: The random cluster model on a general graph and a phase transition characterization of nonamenability. *Stoch. Process. Their Appl.* **79**(2), 335–354 (1999)
48. Levin, D., Peres, Y.: *Markov Chains and Mixing Times*, 2nd edn. American Mathematical Society, Providence (2017)
49. Martinelli, F., Olivieri, E.: Approach to equilibrium of Glauber dynamics in the one phase region. I. The attractive case. *Commun. Math. Phys.* **161**, 447–486 (1994)
50. Martinelli, F.: Dynamical analysis of low-temperature Monte Carlo cluster algorithms. *J. Stat. Phys.* **66**(5), 1245–1276 (1992)
51. Mossel, E., Sly, A.: Rapid mixing of Gibbs sampling on graphs that are sparse on average. *Rand. Struct. Algorithms* **35**(2), 250–270 (2009)
52. Peres, Y., Winkler, P.: Can extra updates delay mixing? *Commun. Math. Phys.* **323**(3), 1007–1016 (2013)
53. Randall, D., Wilson, D.: Sampling spin configurations of an Ising system. In: *Proceedings of the Tenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '99*, pp. 959–960, USA. Society for Industrial and Applied Mathematics (1999)
54. Saloff-Coste, L.: *Lectures on Finite Markov Chains*, pp. 301–413. Springer, Berlin (1997)
55. Ullrich, M.: Comparison of Swendsen–Wang and heat-bath dynamics. *Random Struct. Algorithms* **42**(4), 520–535 (2013)
56. Ullrich, M.: Swendsen–Wang is faster than single-bond dynamics. *SIAM J. Discrete Math.* **28**(1), 37–48 (2014)
57. van den Berg, J.: A uniqueness condition for Gibbs measures, with application to the 2-dimensional Ising antiferromagnet. *Commun. Math. Phys.* **152**(1), 161–166 (1993)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.