Improved Mixing of Critical Hardcore Model

Zongchen Chen ✉
School of Computer Science, Georgia Institute of Technology, Atlanta, GA, USA

Tianhui Jiang ✉
Zhiyuan College, Shanghai Jiao Tong University, China

Abstract

The hardcore model is one of the most classic and widely studied examples of undirected graphical models. Given a graph G, the hardcore model describes a Gibbs distribution of λ-weighted independent sets of G. In the last two decades, a beautiful computational phase transition has been established at a precise threshold $\lambda_c(\Delta)$ where Δ denotes the maximum degree, where the task of sampling independent sets transitions from polynomial-time solvable to computationally intractable. We study the critical hardcore model where $\lambda = \lambda_c(\Delta)$ and show that the Glauber dynamics, a simple yet popular Markov chain algorithm, mixes in $\tilde{O}(n^{7.44+O(1/\Delta)})$ time on any n-vertex graph of maximum degree $\Delta \geq 3$, significantly improving the previous upper bound $\tilde{O}(n^{12.88+O(1/\Delta)})$ by the recent work [3]. The core property we establish in this work is that the critical hardcore model is $O(\sqrt{n})$-spectrally independent, improving the trivial bound of n and matching the critical behavior of the Ising model. Our proof approach utilizes an online decision-making framework to study a site percolation model on the infinite $(\Delta-1)$-ary tree, which can be interesting by itself.

2012 ACM Subject Classification Theory of computation → Random walks and Markov chains; Mathematics of computing → Markov processes

Keywords and phrases Hardcore model, Phase transition, Glauber dynamics, Spectral independence, Online decision making, Site percolation

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2025.51

Category RANDOM

Related Version *Full Version*: `https://arxiv.org/abs/2505.07515` [8]

1 Introduction

The *hardcore model* is one of the most fundamental undirected graphical models that has been extensively studied in statistical physics, social science, probability theory, combinatorics, and computer science.

Given a graph $G = (V, E)$, we let $\mathcal{I}(G)$ denote the collection of all independent sets of G, where we recall that an independent set is a subset of vertices inducing no edges. The *Gibbs distribution* $\mu_{G,\lambda}$ associated with the hardcore model on G is parameterized by a vertex weight $\lambda > 0$ called the *fugacity*. Each independent set $\sigma \in \mathcal{I}(G)$ receives a probability density given by

$$\mu_{G,\lambda}(\sigma) = \frac{\lambda^{|\sigma|}}{Z_{G,\lambda}},$$

where $Z_{G,\lambda}$ is a normalizing constant call the *partition function* and is defined as

$$Z_{G,\lambda} = \sum_{\sigma \in \mathcal{I}(G)} \lambda^{|\sigma|}.$$

Perhaps the most amazing property of the hardcore model is the phase transition phenomenon associated with it. In fact, the hardcore model was originally proposed by statistical physicists to study and understand the phase transition in systems of hardcore

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2025).
Editors: Alina Ene and Eshan Chattopadhyay; Article No. 51; pp. 51:1–51:22
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

gas particles. Let $\Delta \geq 3$ denote the maximum degree of the underlying graph. The tree-uniqueness threshold $\lambda_c(\Delta) := \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}}$ characterizes the uniqueness of the hardcore Gibbs measure on the infinite Δ-regular tree. Furthermore, it also describes the existence of long-range correlations. Let each vertex be associated with a Bernoulli random variable, called the *spin*, indicating whether the vertex is *occupied* (i.e., included in the independent set) or *unoccupied* (i.e., not included in the independent set). Then, for small fugacity $\lambda \leq \lambda_c(\Delta)$ the configuration at distance ℓ from the root has a vanishing influence on the root as ℓ tends to infinity, while for large fugacity $\lambda > \lambda_c(\Delta)$ the correlation is always bounded away from zero.

In the past two decades, a beautiful computational phase transition has been fully established for the problem of sampling from the hardcore model on graphs of maximum degree Δ, precisely around the uniqueness threshold $\lambda_c(\Delta)$. For $\lambda < \lambda_c(\Delta)$, there exist deterministic approximate counting algorithms for estimating the partition function [28, 2, 21], which in turn gives approximate samplers via standard reduction. Meanwhile, for $\lambda > \lambda_c(\Delta)$, no polynomial-time approximate counting and sampling algorithms exist assuming $\mathsf{RP} \neq \mathsf{NP}$ [23, 24, 14].

While all deterministic approximate counting algorithms run in polynomial time, they suffer from a pretty slow runtime. For example, Weitz's algorithm [28] runs in time $n^{O(\frac{1}{\delta}\log\Delta)}$ where Δ denotes the maximum degree and $\delta \in (0,1)$ the slackness of the fugacity (i.e., $\lambda = (1-\delta)\lambda_c(\Delta)$). In practice, Markov chain Monte Carlo (MCMC) algorithms provide a simpler and significantly faster method for generating random samples from high-dimensional distributions, including the hardcore model studied in this work. Among them, the *Glauber dynamics* (also known as the *Gibbs sampler*) is one of the most important and popular examples. The Glauber dynamics performs a random walk in the space $\mathcal{I}(G)$ of independent sets and, in each step, either stays the same or moves to an adjacent set whose Hamming distance to the current set is 1. More specifically, from the current independent set $\sigma_t \in \mathcal{I}(G)$, the algorithm picks a vertex $v \in V$ uniformly at random and updates its spin: Let $S = \sigma_t \setminus \{v\}$; if $S \cup \{v\} \notin \mathcal{I}(G)$ then set $\sigma_{t+1} = S = \sigma_t$; otherwise, set $\sigma_{t+1} = S \cup \{v\}$ with probability $\lambda/(1+\lambda)$ and, mutually exclusively, set $\sigma_{t+1} = S$ with probability $1/(1+\lambda)$.

Let P_{GD} denote the transition matrix of the Glauber dynamics. From basic Markov chain theories it is easy to show that the Glauber dynamics P_{GD} is irreducible, aperiodic, and reversible with respect to the Gibbs distribution $\mu_{G,\lambda}$, which is the unique stationary distribution (i.e., $\mu_{G,\lambda} P_{\mathrm{GD}} = \mu_{G,\lambda}$). The mixing time of the Glauber dynamics is defined as

$$T_{\mathrm{mix}}(P_{\mathrm{GD}}) = \max_{\sigma_0 \in \mathcal{I}(G)} \min_{t \in \mathbb{Z}_{\geq 0}} \left\{ d_{\mathrm{TV}}\left(P_{\mathrm{GD}}^t(\sigma_0, \cdot),\, \mu_{G,\lambda}\right) \leq \frac{1}{4} \right\},$$

where σ_0 is the initial independent set, $P_{\mathrm{GD}}^t(\sigma_0, \cdot)$ is the distribution of the chain after t steps when starting from σ_0, and $d_{\mathrm{TV}}(\cdot,\cdot)$ denotes the total variation distance.

In the past years, exciting progress has been made in understanding the mixing time of Glauber dynamics for the hardcore model. Anari, Liu, and Oveis Gharan introduced a highly powerful technique known as *spectral independence* [1], leading to significant advancements in this area, including resolutions to major open problems regarding mixing properties. We refer to [18, 25] for a thorough introduction of this technique. In the subcritical regime (i.e., $\lambda < \lambda_c(\Delta)$), the mixing time of the Glauber dynamics was shown to be nearly linear $O(n \log n)$ [1, 9, 5, 7]. Meanwhile, it was long known that in the supercritical regime (i.e., $\lambda > \lambda_c(\Delta)$), the mixing time could be exponentially large $\exp(\Omega(n))$ as witnessed by random Δ-regular bipartite graphs [19].

In a very recent work [3], the mixing property is further investigated at the critical point (i.e., $\lambda = \lambda_c(\Delta)$). For the upper bound, the mixing time of Glauber dynamics is $\tilde{O}(n^{2+4\mathrm{e}+O(1/\Delta)})$ on any n-vertex graph of maximum degree Δ. For the lower bound, there exists an infinite sequence of graphs such that the mixing time is $\Omega(n^{4/3})$, which is, in particular, super-linear.

In this work, we present an improved mixing time upper bound for the Glauber dynamics on the critical hardcore model.

▶ **Theorem 1.** *Let $\alpha \geq 0$ be a constant. For any n-vertex graph $G = (V, E)$ of maximum degree $\Delta \geq 3$, the Glauber dynamics for the hardcore model on G with fugacity $\lambda \leq (1 + \frac{\alpha}{\sqrt{n}})\lambda_c(\Delta)$ satisfies*

$$T_{\mathrm{mix}}(P_{\mathrm{GD}}) = O_\alpha\left(n^{2+2\mathrm{e}+\frac{2\mathrm{e}}{\Delta-2}} \log \Delta\right).$$

Our upper bound scales as $\tilde{O}(n^{7.44+O(1/\Delta)})$, significantly improving the $\tilde{O}(n^{12.88+O(1/\Delta)})$ mixing time established in [3].

Similar to [3], Theorem 1 is established via the spectral independence framework. Our main contribution is to show that the critical hardcore model satisfies spectral independence of order $O(\sqrt{n})$, improving the trivial bound of n used in [3]. We show this new spectral independence result in a novel way by studying an online decision-making problem, which allows us to understand a site percolation model on the infinite tree, from which spectral independence readily follows. We provide an overview of the necessary background and known results on spectral independence, as well as our new contribution and proof approach in Section 2.

2 Proof Overview

2.1 Notations and definitions

Denote the set of non-negative integers by $\mathbb{Z}_{\geq 0}$, and the set of positive integers by $\mathbb{Z}^+$. For any integers $a, b \in \mathbb{Z}$, define $a \wedge b$ by the minimum of a and b, i.e, $a \wedge b := \min\{a, b\}$.

Let $\mathrm{Ber}(p)$ denote the Bernoulli distribution with success probability $p \in [0, 1]$. Let $\mathrm{Bin}(n, p)$ denote the binomial distribution with number of trails $n \in \mathbb{Z}^+$ and success probability $p \in [0, 1]$. Let $d_{\mathrm{TV}}(\cdot, \cdot)$ denote the total variation distance. For any random variables X, Y, let $X \overset{d}{=} Y$ denote that X and Y are equal in distribution.

Let $G = (V, E)$ be a graph. For any $S \subseteq V$, let ∂S denote the set of neighbors of S in G, i.e., $\partial S = \{v \in V \setminus S \mid \exists u \in S, \{u, v\} \in E\}$; and let $G[S]$ denote the subgraph induced in G by S, i.e., the graph with vertex set S and edge set consisting of all edges of G that have both endpoints in S.

Let $T = (V, E)$ be a tree rooted at r. For every vertex $v \in V$, let T_v denote the subtree of T rooted at v that consists of all descendants of v; in particular, $T_r = T$. For any $v \in V$, let $L(v)$ denote the set of children of v in T.

We end this subsection by defining the (t-fold) convolution of distributions on $\mathbb{Z}$.

▶ **Definition 2** ((t-fold) Convolution)**.** *Let μ, ν be two distributions on $\mathbb{Z}$. Define a new distribution $\mu * \nu$ on $\mathbb{Z}$ by*

$$\mu * \nu(k) = \sum_{i=-\infty}^{+\infty} \mu(i)\nu(k - i), \quad \forall k \in \mathbb{Z}.$$

We call $\mu * \nu$ the convolution *of μ and ν. Define μ^{*t} where $t \in \mathbb{Z}^+$ inductively by $\mu^{*1} = \mu$ and $\mu^{*t} = \mu^{*(t-1)} * \mu$ for $t \geq 2$. We call μ^{*t} the* t-fold convolution *of μ with itself.*

2.2 Spectral independence via coupling on trees

The core result of this work is to establish the $O(\sqrt{n})$-spectral independence for the critical hardcore model, from which Theorem 1 readily follows by sophisticated spectral independence techniques that have been developed in a recent line of works.

The following notion of influences is needed to define the meaning of spectral independence.

▶ **Definition 3** (Influence, [1]). *Let μ be a distribution over $\{0,1\}^n$. For any $i,j \in [n]$ such that $\Pr_\mu[\sigma_i = 0] > 0$ and $\Pr_\mu[\sigma_i = 1] > 0$, define the (pairwise) influence from i to j as*

$$\Psi_\mu(i,j) := \Pr_{\sigma\sim\mu}[\sigma_j = 1 \mid \sigma_i = 1] - \Pr_{\sigma\sim\mu}[\sigma_j = 1 \mid \sigma_i = 0].$$

In the setting of the hardcore model, the influences describe the correlation between two vertices, represented as Bernoulli random variables indicating whether the vertices are occupied. Roughly speaking, the influence of one vertex on the other represents the difference of the marginal distribution on the second vertex when flipping the first vertex from occupied to unoccupied.

▶ **Theorem 4** ($O(\sqrt{n})$-Spectral independence of critical hardcore model). *Let $\alpha \geq 0$ be a constant. Consider the hardcore model on an n-vertex graph $G = (V,E)$ of maximum degree $\Delta \geq 3$ with fugacity $\lambda \leq (1 + \frac{\alpha}{\sqrt{n}})\lambda_c(\Delta)$. Then, for any vertex $u \in V$, we have*

$$\sum_{v\in V} \left|\Psi_{\mu_{G,\lambda}}(u,v)\right| \leq C_0\sqrt{n},$$

where $C_0 = C_0(\alpha) > 0$ is a constant depending only on α.

Theorem 4 states that the hardcore model in the regime of interest satisfies ℓ_∞ spectral independence with constant $O(\sqrt{n})$ (see the full version of the paper [8] and also [13]). An analogous result for the Ising model was previously shown in [3].

Many methods have been introduced to establish the spectral independence property for various families of distributions. Here we adopt the *coupling independence* approach introduced in [6] and apply it on a related tree, known as the *self-avoiding walk tree* [28]. The formal definition and construction of this tree are omitted in this paper as we only need its existence, and we refer interested readers to the works [28, 10].

We are interested in coupling two hardcore models on this tree, where in one copy the root is fixed to be occupied while in the other it is fixed to be unoccupied. As we shall see soon in Proposition 5, controlling the number of discrepancies between these two copies under a simple coupling procedure enables us to deduce spectral independence. To formally describe this coupling, we first need a few definitions. Let $T = (V,E)$ be a tree rooted at r of maximum degree at most Δ. Consider the hardcore model on T with fugacity $\lambda > 0$. For each vertex v, let p_v denote the probability that v is occupied in the hardcore model on the subtree T_v, i.e., $p_v := \Pr_{\mu_{T_v,\lambda}}[\sigma_v = 1]$, where we recall that T_v is the subtree of T rooted at v that consists of all descendants of v.

We now describe a natural vertex-by-vertex *greedy coupling* for the hardcore model on T when the spin at root r is flipped.

- Initialization: $X_r = 1$ and $Y_r = 0$;
- While there exists $v \in V$ whose parent u has already been revealed in X and Y:
  - If $X_u = Y_u$, then couple the whole subtree T_v perfectly, i.e., $X_{T_v} = Y_{T_v}$;
  - If $X_u = 1$ and $Y_u = 0$, then set $X_v = 0$ and sample $Y_v \sim \mathrm{Ber}(p_v)$;
  - If $X_u = 0$ and $Y_u = 1$, then sample $X_v \sim \mathrm{Ber}(p_v)$ and set $Y_v = 0$;
- Return (X,Y).

It is straightforward to check that when the greedy coupling ends, X is an independent set distributed as $\Pr_{\mu_{T,\lambda}}[\,\cdot \mid \sigma_r = 1]$ and Y as $\Pr_{\mu_{T,\lambda}}[\,\cdot \mid \sigma_r = 0]$.

▶ **Proposition 5** (Coupling on trees implies spectral independence, [4, Lemma 39], [6, Proposition 4.3])**.** *Consider the hardcore model on an n-vertex graph $G = (V, E)$ of maximum degree $\Delta \geq 3$ with fugacity $\lambda > 0$. For any $u \in V$, there exists a tree $T = T_{\mathrm{SAW}}(G, u)$ rooted at r with maximum degree at most Δ, such that if $(X, Y) \sim \mathcal{C}$ is the greedy coupling on T, then it holds*

$$\sum_{v \in V} \left|\Psi_{\mu_{G,\lambda}}(u, v)\right| \leq \mathop{\mathbb{E}}_{(X,Y)\sim\mathcal{C}} [|X \oplus Y| \wedge n] \,,$$

where $X \oplus Y$ denotes the symmetric difference of two sets X, Y, and recall that $|X \oplus Y| \wedge n :=$ $\min\{|X \oplus Y|, n\}$.

Hence, to establish Theorem 4, it suffices to bound the expected number of disagreements in the greedy coupling for the hardcore model on trees when the spins at the root are distinct.

2.3 Coupling on trees via site percolation

From the greedy coupling procedure above, we observe a natural *site percolation* on the tree T describing the appearance of disagreements. Every vertex v is *open* with probability p_v independently, representing the occurrence of a disagreement at v, that is, $X_v \neq Y_v$; otherwise, the vertex v is *closed*. The root r is always open, i.e., $p_r = 1$, since $X_r \neq Y_r$. Our goal is to bound the size of the *open cluster* containing the root, consisting of all open vertices that are connected to the root via a path of open vertices.

We now introduce some notations for the site percolation model. Let $T = (V, E)$ be a tree rooted at r. For any $v \in V$, let p_v be the probability that v is open, and we call p_v the *occupation probability* of v. For simplicity, we assume $p_r = 1$. Let $P = \{p_v\}_{v \in V}$ be the list of occupation probabilities for all vertices, and call it the *occupation probability list* of the site percolation model. Finally, for the site percolation on T with occupation probability list P, let $N(T, P)$ denote the random variable representing the size of the open cluster containing the root.

From the construction of the greedy coupling and the site percolation above, we see that $|X \oplus Y| \overset{d}{=} N(T, P)$, where $p_v = \Pr_{\mu_{T_v,\lambda}}[\sigma_v = 1]$ for each $v \in V \setminus \{r\}$; see the full version of the paper [8] for a formal statement. Thus, the problem is reduced to studying a site percolation model.

In order to study this site percolation model, we need to know the conditions satisfied by the occupation probabilities $\{p_v\}$. Since these probabilities correspond to the marginal probability of the roots in the respective subtrees, they satisfy a well-known recurrence called the *tree recursion* (see Fact 21). In this work, we present a new inequality satisfied by these marginal probabilities, which is crucial in obtaining our main spectral independence result. In particular, Equation (1) below provides a stronger and simpler contraction property of the tree recursion, which was not known in the literature as far as we are aware. For simplicity, here we consider only the exact critical fugacity $\lambda = \lambda_c(\Delta)$.

▶ **Lemma 6** (Special case of Lemma 20; see also [17])**.** *Let $T = (V, E)$ be a tree rooted at r with maximum degree at most Δ. Consider the critical hardcore model on T with fugacity $\lambda = \lambda_c(\Delta)$. For each vertex $v \in V$, let p_v denote the probability that v is occupied in the critical hardcore model on the subtree T_v rooted at v, i.e., $p_v := \Pr_{\mu_{T_v,\lambda}}[\sigma_v = 1]$. Then, for every non-root vertex $v \in V \setminus \{r\}$, it holds*

$$p_v \sum_{w\in L(v)} p_w \le \frac{1}{\Delta-1}, \tag{1}$$

where we recall that $L(v)$ denotes the set of children of v.

▶ Remark 7. In [17], a potential function approach was applied to study the contraction of the tree recursion for the subcritical hardcore model. Uncovering their result ([17, Lemma 12]), the corresponding condition they established at criticality can be stated as

$$\sqrt{p_v} \sum_{w\in L(v)} \sqrt{p_w} \le 1. \tag{2}$$

Our bound Equation (1) is stronger than Equation (2) since $|L(v)| \le \Delta - 1$. In fact, going through the proof in [17], one is able to recover the stronger inequality Equation (1) though it was not stated explicitly. In this work, we provide a simpler proof of Equation (1) (and hence Equation (2)). To establish the $O(\sqrt{n})$-spectral independence at criticality, we do require the stronger inequality Equation (1), while the weaker Equation (2) is not sufficient.

We are now ready to state our main percolation result, from which spectral independence Theorem 4 readily follows.

▶ **Theorem 8** (Main result for site percolation). *Consider the site percolation model on the infinite d-ary tree $\mathbb{T}_d^{\mathrm{ary}} = (V, E)$ rooted at r with occupation probability list $P = \{p_v\}_{v\in V}$ where $p_r = 1$. Let $N = N(\mathbb{T}_d^{\mathrm{ary}}, P)$ be the size of the open cluster containing the root. Suppose the following hold:*

1. *For every non-root vertex $v \in V \setminus \{r\}$, it holds*

$$p_v \sum_{w\in L(v)} p_w \le \frac{1}{d}; \tag{3}$$

2. *At the root r, it holds*

$$\sum_{w\in L(r)} p_w \le c, \tag{4}$$

where $c > 0$ is a constant.

Then, we have that for any $n \in \mathbb{Z}_{\ge 0}$,

$$\mathbb{E}\left[N \wedge n\right] = O(\sqrt{n}),$$

where the constant in big-O depends only on c.

At this point, we transfer our original problem of bounding the mixing time and proving critical spectral independence into a problem of site percolation on trees. In [3], the same strategy was applied to the critical Ising model, another canonical example of graphical models. There, the occupation probabilities are much simpler; in fact, they can all be set to $p_v = 1/d$, which is a uniform upper bound on the pairwise influence on v from its parent. We remark that the main distinction of [3] and our work lies in the application of Lemma 6 (corresponding to the condition Equation (3)) which substantially extends the uniform setting, $p_v = 1/d$ for all v, appeared in the critical Ising model.

2.4 Site percolation on trees via online decision making

The main part of the paper aims to prove Theorem 8. We introduce a novel approach to study such a site percolation model on the infinite tree, by considering an online decision-making game.

Our strategy is to upper bound $\mathbb{E}[N \wedge n]$ by understanding the worst-case choice of occupation probabilities $\{p_v\}$. We do this by changing the perspective and thinking as an adversary who is allowed to pick the occupation probabilities. Namely, we study an online decision-making problem where a player is allowed to pick the occupation probabilities $\{p_v\}$ every time we need to reveal the status (open or closed) of a few vertices. These probabilities can be arbitrary as long as they satisfy the requirements Equation (3), and the goal of the player is to maximize $\mathbb{E}[N \wedge n]$.

To be more specific, let A_t denote the number of active vertices at time t, where a vertex is said to be *active* if (i) it is open; (ii) there is an open path from it to the root; (iii) the status of its children has not been fully revealed. At the beginning, $A_0 = 1$ since only the root is open and we have not revealed any other vertex. Then, the player picks the occupation probabilities for both the children and the grandchildren of r; these probabilities are required to satisfy Equation (3). By sampling from the corresponding Bernoulli distributions independently, we reveal the number of grandchildren of r, denoted as X_1, that are active (i.e., open and connected to r). With r being deactivated, the number of active vertices becomes $A_1 = A_0 - 1 + X_1 = X_1$. The game is then repeated. Whenever there exists an active vertex v at round t, the player picks occupation probabilities of the children and grandchildren of v satisfying Equation (3), and, after the number X_t of open grandchildren connected to v is revealed, updates $A_t = A_{t-1} - 1 + X_t$. This process stops when there are no active vertices, i.e., when $A_t = 0$.

We remark that we consider only active vertices at even levels because Equation (3) imposes requirements on two adjacent levels.

Suppose the game stops after k rounds. Then, the number of open vertices connected to the root at even levels is precisely k, since every such vertex becomes active at some point and is deactivated at some other time. Therefore, controlling the number of rounds played in the game allows us to bound $\mathbb{E}[N_e \wedge n]$ where N_e is the number of vertices at even levels that are in the open cluster containing the root in the site percolation. (Since $\mathbb{E}[N_e \wedge n] = \sum_{m=1}^{n} \Pr[N_e \geq m]$, in the actual proof we aim to bound $\Pr[N_e \geq m]$ for every m for simplicity. And we can bound $\Pr[N_e \geq m]$ by the maximum probability that the game lasts for at least m rounds.) Finally, combining the upper bound of $\mathbb{E}[N_e \wedge n]$ with Equation (4), it is then not hard to upper bound $\mathbb{E}[N \wedge n]$ as wanted.

Therefore, our goal is to determine the optimal strategy of the player when such an online decision-making game is played. We deal with this in Section 3 where we state and prove our main result, Theorem 19. While a natural guess of the optimal strategy is to set every occupation probability p_v to be $1/d$ so that Equation (3) is satisfied, we show that the optimal strategy of the player should set $p_v p_w = 1/d$ for one child $w \in L(v)$ of v and set $p_{w'} = 0$ for all other children $w' \in L(v)$. We remark that such choices of $\{p_v\}$ are not realizable as marginal probabilities in the hardcore model since they do not satisfy the tree recursion and are way too large (in reality, $p_v = O(\lambda) = O(1/d)$ at criticality); however, they are sufficient to provide meaningful upper bounds on $\mathbb{E}[N \wedge n]$ as we need.

We present the proof of Theorem 8 about site percolation on the infinite tree in Section 4, utilizing Theorem 19. Finally, we deduce spectral independence (Theorem 4) and rapid mixing (Theorem 1) from Theorem 8; see the full version of the paper [8] for details.

Algorithm 1 Online decision-making game.

Input: A family $\mathcal{P}$ of distributions on $\mathbb{Z}_{\geq 0}$, initial number of tokens $a \in \mathbb{Z}_{\geq 0}$
Initialization: $t \leftarrow 0$, $A_0 \leftarrow a$
while $A_t \geq 1$ **do**
  $t \leftarrow t + 1$;
  Player chooses a distribution $\pi_t \in \mathcal{P}$;
  Sample $X_t \sim \pi_t$;
  $A_t \leftarrow A_{t-1} - 1 + X_t$;
end while

3 Online Decision-Making Problem

In this section, we introduce an online decision-making problem that serves as a key tool in proving our main result for site percolation (Theorem 8), and show the optimal strategy as our main result for this section.

First of all, we describe the setup of the online decision-making game in Section 3.1. Then, we define a partial order called second-order stochastic dominance, which plays an important role in finding the optimal strategy, and show some basic properties in Section 3.2. After that, we introduce the Poisson binomial distribution in Section 3.3 and some properties of the random walk hitting time in Section 3.4, which are crucial in the proof of the optimal strategy. Finally, in Section 3.5, we state and prove our main result for the online decision-making game and show the optimal strategy of the game.

3.1 Setup of online decision making

We now describe precisely an online decision-making game under a slightly more general setting.

In the online decision-making game, the player maintains some number of tokens (corresponding to active vertices). Let A_t denote the number of tokens the player owns after round t. At the beginning, the player has $A_0 = a$ tokens. There is a family $\mathcal{P}$ of distributions on non-negative integers $\mathbb{Z}_{\geq 0}$ (corresponding to choices of occupation probabilities). For round t, the player spends one token, assuming $A_{t-1} \geq 1$, and chooses a distribution $\pi_t \in \mathcal{P}$. Then, a sample $X_t \sim \pi_t$ is generated independently, and the player receives X_t tokens as a reward. The number of tokens the player owns then becomes $A_t = A_{t-1} - 1 + X_t$. The game ends when the player uses up all the tokens, i.e., the first time $A_t = 0$. We denote this stopping time by τ. Note that it is possible the game never stops, in which case $\tau = \infty$. The goal of the player is to survive for m rounds for some given integer $m \in \mathbb{Z}_{\geq 0}$. That is, the player wins if and only if $\tau \geq m$. We present the process of the online decision-making game in Algorithm 1.

Define a *strategy* $\mathcal{S}$ for the player by a mapping $\mathcal{S} : \mathbb{Z}^+ \times \mathbb{Z}^+ \to \mathcal{P}$. For any $k, a \in \mathbb{Z}^+$, $\mathcal{S}(k, a)$ is defined by the distribution the player will choose when they needs to survive for k more rounds to win and currently has a tokens. For example, if the winning requirement is m rounds, then at the beginning of round t, the player will choose the distribution $\pi_t = \mathcal{S}(m - t + 1, A_{t-1})$ assuming $A_{t-1} \geq 1$.

For $m, a \in \mathbb{Z}_{\geq 0}$, define the *winning probability* under strategy $\mathcal{S}$ to be

$$\varphi_m^{\mathcal{S}}(a) := \Pr^{\mathcal{S}}\left[\tau \geq m \mid A_0 = a\right].$$

Namely, $\varphi_m^{\mathcal{S}}(a)$ is the probability that the game lasts for at least m rounds, assuming the player has a tokens at the beginning and uses strategy $\mathcal{S}$. We further define

$$\varphi_m^*(a) := \sup_{\mathcal{S}} \varphi_m^{\mathcal{S}}(a).$$

The following lemma establishes a simple recursive formula for the optimal winning probabilities and the existence of an optimal strategy.

▶ **Lemma 9.** *Let $\mathcal{P}$ be a family of distributions on $\mathbb{Z}_{\geq 0}$. Suppose the metric space $(\mathcal{P}, d_{\mathrm{TV}})$ is compact, where we recall d_{TV} is the total variation distance. Then the following holds:*

1. *For all $m, a \in \mathbb{Z}^+$,*

$$\varphi_m^*(a) = \max_{\pi \in \mathcal{P}} \mathop{\mathbb{E}}_{X \sim \pi} \left[\varphi_{m-1}^*(a - 1 + X) \right];$$

2. *There exists a strategy $\mathcal{S}^* : \mathbb{Z}^+ \times \mathbb{Z}^+ \to \mathcal{P}$ such that $\varphi_m^*(a) = \varphi_m^{\mathcal{S}^*}(a)$ holds for any $m, a \in \mathbb{Z}^+$.*

Proof. We note that $\varphi_m^{\mathcal{S}^*}(a)$ is well-defined if $\mathcal{S}^*(k, \cdot)$ is defined for all $1 \leq k \leq m$, and the result of $\mathcal{S}^*(k, \cdot)$ $(k > m)$ does not matter. In our proof, when we write $\varphi_m^{\mathcal{S}^*}(a)$, we only guarantee that for all $1 \leq k \leq m$, $\mathcal{S}^*(k, \cdot)$ is defined.

We verify the recurrence and define the strategy inductively on m.

For the base case $m = 1$, by definition, $\varphi_1^*(a) = \mathbb{1}[a \geq 1]$, and $\varphi_0^*(a) = 1$ for all $a \in \mathbb{Z}_{\geq 0}$. Then Item 1 immediately follows. Furthermore, we can define $\mathcal{S}^*(1, \cdot) := \pi_0$ so that $\varphi_1^*(a) = \varphi_1^{\mathcal{S}^*}(a)$ holds for any $a \in \mathbb{Z}^+$, where $\pi_0 \in \mathcal{P}$ is an arbitrary distribution.

Now suppose $m \geq 2$. Suppose Items 1 and 2 hold for $m - 1$. Let $a \in \mathbb{Z}^+$. Suppose the player chooses $\pi \in \mathcal{P}$ in the first round and obtains $X \sim \pi$ tokens; hence, after the first round, the player has $A_1 = A_0 - 1 + X = a - 1 + X$ tokens. Then to maximize the winning probability, i.e., to maximize the probability that survive for at least $m - 1$ rounds when having $a - 1 + X$ tokens initally, the player should use strategy $\mathcal{S}^*(k, \cdot)$ $(k = 1, \cdots, m - 1)$, and then the winning probability is $\varphi_{m-1}^*(a - 1 + X)$ by the induction hypothesis $\varphi_{m-1}^*(a - 1 + X) = \varphi_{m-1}^{\mathcal{S}^*}(a - 1 + X)$. Therefore, if the player chooses $\pi \in \mathcal{P}$ in the first round, the maximum winning probability for the player is $\mathbb{E}_{X \sim \pi} \left[\varphi_{m-1}^*(a - 1 + X) \right]$. Hence, to obtain the maximum winning probability, we need to choose an optimal π which maximizes $\mathbb{E}_{X \sim \pi} \left[\varphi_{m-1}^*(a - 1 + X) \right]$. By compactness, such optimal π exists, therefore Item 1 for m holds, and we can define $\mathcal{S}^*(m, \cdot)$ by $\mathcal{S}^*(m, a) := \arg\max_{\pi \in \mathcal{P}} \mathbb{E}_{X \sim \pi} \left[\varphi_{m-1}^*(a - 1 + X) \right]$ for all $a \in \mathbb{Z}^+$. Then it is straightforward that $\varphi_m^*(a) = \varphi_m^{\mathcal{S}^*}(a)$ holds for any $a \in \mathbb{Z}^+$. ◀

3.2 Second-order stochastic dominance

Our goal is to find an optimal strategy for the online decision-making game. A first thought that one may consider is to use *first-order stochastic dominance* (also simply called *stochastic dominance*). For any two distributions μ, ν over $\mathbb{Z}_{\geq 0}$, μ is (first-order) stochastically dominated by ν if and only if $\Pr_\mu [X \geq i] \leq \Pr_\nu [Y \geq i]$ for all $i \in \mathbb{Z}_{\geq 0}$. An immediate corollary is that μ is stochastically dominated by ν implies $\mathbb{E}_\mu [X] \leq \mathbb{E}_\nu [Y]$. If there is a largest distribution under stochastic dominance in $\mathcal{P}$, then it can be easily proved that the player can attain the maximum winning probability when always picking the largest distribution. However, the largest distribution under stochastic dominance does not exist in our case, since any two different distributions with the same mean are not comparable under stochastic dominance. In fact, in our application, there are infinitely many distributions attaining the largest mean in $\mathcal{P}$. Therefore, there is no largest distribution under stochastic dominance and we need something else.

It turns out that *second-order stochastic dominance* (see, e.g., [16, 15] for more background and applications) can address the problem of the lack of a largest distribution. However, as a trade-off, it is not always true that the player can achieve the maximum winning probability when always choosing the largest distribution under second-order stochastic dominance. Nonetheless, if the largest distribution satisfies some nice properties (see Theorem 19 for details), this is indeed true.

We now define second-order stochastic dominance.

▶ **Definition 10** (Second-order stochastic dominance). *We define a partial order "$\preceq_{(2)}$" called* second-order stochastic dominance *on the family of distributions on $\mathbb{Z}_{\geq 0}$ with finite expectations. For two distributions μ, ν on $\mathbb{Z}_{\geq 0}$ with finite expectations, $\mu \preceq_{(2)} \nu$ if and only if*

$$\mathop{\mathbb{E}}_{X\sim\mu}[X \wedge i] \leq \mathop{\mathbb{E}}_{Y\sim\nu}[Y \wedge i], \quad \forall i \in \mathbb{Z}^+.$$

The following proposition shows some classical equivalent definitions of second-order stochastic dominance.

▶ **Proposition 11** (Equivalent definitions of second-order stochastic dominance [20, Theorem 8.1.1], see also [22]). *Let μ, ν be distributions on $\mathbb{Z}_{\geq 0}$ with finite expectations. The following definitions are equivalent:*

1. $\mu \preceq_{(2)} \nu$*;*
2. *(Increasing concave order) For any non-decreasing concave function f, it holds:*

$$\mathop{\mathbb{E}}_{X\sim\mu}[f(X)] \leq \mathop{\mathbb{E}}_{X\sim\nu}[f(X)];$$

3. *There exists a coupling $\mathcal{C}$ of μ and ν such that*

$$\mathop{\mathbb{E}}_{(X,Y)\sim\mathcal{C}}[X - Y \mid Y = i] \leq 0, \quad \forall i \in \mathbb{Z}_{\geq 0}.$$

The following two lemmas offer easy ways to find an "upper bound" of a given distribution in the sense of "$\preceq_{(2)}$", and are helpful to us in identifying the largest distribution in $\mathcal{P}$.

▶ **Lemma 12.** *Let μ be a distribution on $\mathbb{Z}_{\geq 0}$ with expectation $\mathbb{E}_\mu[X] \leq \gamma \leq 1$. Then $\mu \preceq_{(2)} \mathrm{Ber}(\gamma)$.*

Proof. For any $i \in \mathbb{Z}^+$, we have

$$\mathop{\mathbb{E}}_{X\sim\mu}[X \wedge i] \leq \mathop{\mathbb{E}}_{X\sim\mu}[X] \leq \gamma = \mathop{\mathbb{E}}_{Y\sim\mathrm{Ber}(\gamma)}[Y \wedge i],$$

which implies $\mu \preceq_{(2)} \mathrm{Ber}(\gamma)$. ◀

▶ **Lemma 13.** *If $\mu_1 \preceq_{(2)} \nu_1$ and $\mu_2 \preceq_{(2)} \nu_2$, then $\mu_1 * \mu_2 \preceq_{(2)} \nu_1 * \nu_2$, where the operator "$*$" represents convolution* (see Definition 2).

Proof. Let X_1, X_2 be independent random variables with distributions μ_1, μ_2 respectively. Let Y_1, Y_2 be independent random variables with distributions ν_1, ν_2 respectively. We also assume X_1, Y_2 are independent and X_2, Y_1 are independent. For $i = 1, 2$, by $\mu_i \preceq_{(2)} \nu_i$ and Proposition 11, there exists a coupling $\mathcal{C}_i$ of (X_i, Y_i), such that

$$\mathop{\mathbb{E}}_{(X_i,Y_i)\sim\mathcal{C}_i}[X_i - Y_i \mid Y_i = j] \leq 0, \quad \forall j \in \mathbb{Z}_{\geq 0}. \tag{5}$$

Let $X = X_1 + X_2, Y = Y_1 + Y_2$. Then $\mu_1 * \mu_2$ is the law of X, and $\nu_1 * \nu_2$ is the law of Y. Let $\mathcal{C}$ be the joint distribution of (X, Y) assuming $(X_i, Y_i) \sim \mathcal{C}_i$ $(i = 1, 2)$. It is clear that $\mathcal{C}$ is a coupling of $\mu_1 * \mu_2$ and $\nu_1 * \nu_2$. For any $k \in \mathbb{Z}_{\geq 0}$,

$$\begin{aligned}\mathop{\mathbb{E}}_{(X,Y)\sim\mathcal{C}}[X - Y|Y = k] &= \sum_{i=1}^{2} \mathop{\mathbb{E}}_{\substack{(X_1,Y_1)\sim\mathcal{C}_1,\\(X_2,Y_2)\sim\mathcal{C}_2}}[X_i - Y_i|Y = k] \\ &= \sum_{i=1}^{2}\sum_{j=0}^{k} \Pr_{\substack{(X_1,Y_1)\sim\mathcal{C}_1,\\(X_2,Y_2)\sim\mathcal{C}_2}}[Y_i = j|Y = k] \mathop{\mathbb{E}}_{(X_i,Y_i)\sim\mathcal{C}_i}[X_i - Y_i|Y_i = j] \\ &\leq 0,\end{aligned}$$

where the last inequality follows from Equation (5). Then Lemma 13 follows from Proposition 11. ◀

3.3 Poisson binomial distribution

As hinted by Item 2 of Proposition 11, in order to apply second-order stochastic dominance, we hope to have some non-decreasing concave functions as the objective/utility function in our decision making. It turns out that when the largest distribution (under "$\preceq_{(2)}$") is a *Poisson binomial distribution* (see, e.g., [26] for a thorough introduction) with expectation at least 1, the maximum winning probability $\varphi_m^*(a)$ is non-decreasing and concave with respect to a.

We first define the Poisson binomial distribution.

▶ **Definition 14** (Poisson binomial distributions (random variables))**.** *We call a random variable X a Poisson binomial random variable if it can be expressed as a finite sum of independent Bernoulli random variables, i.e., $X = \sum_{i=1}^{\ell} X_i$ where $\ell \in \mathbb{Z}^+$, $X_i \sim \mathrm{Ber}(p_i)$ are independent. We call a distribution a Poisson binomial distribution if it is the distribution of a Poisson binomial random variable.*

An immediate property is the following.

▶ **Fact 15.** *If X and Y are two independent Poisson binomial random variables, then $X+Y$ is also a Poisson binomial random variable.*

The crucial property that guarantees the concavity of the maximum winning probability is that a Poisson binomial random variable is unimodal with a mode near the mean of the random variable. We next define unimodality and state the property.

▶ **Definition 16** (Unimodality [12])**.** *Let π be a distribution on $\mathbb{Z}$. Let m be an integer. The distribution π is called* unimodal *about m if*

$$\pi(i) \geq \pi(i-1), \quad \forall i \leq m \quad \text{and} \quad \pi(i) \geq \pi(i+1), \quad \forall i \geq m,$$

and we call m a mode *of π.*

Let Z be a random variable with distribution π. The random variable Z is called unimodal about m if π is unimodal about m.

▶ **Proposition 17** (Darroch's rule for the mode [11])**.** *Let Z be a Poisson binomial random variable. Let q be the expectation of Z. Then there exists $m \in \{\lfloor q \rfloor, \lfloor q \rfloor + 1\}$ if $q \notin \mathbb{Z}$, or $m = q$ if $q \in \mathbb{Z}$, such that Z is unimodal about m. In particular,*

$$\Pr[Z = i] \geq \Pr[Z = i-1], \quad \forall i \leq q \quad \text{and} \quad \Pr[Z = i] \geq \Pr[Z = i+1], \quad \forall i \geq q.$$

3.4 Random walk hitting time

The following proposition of the random walk hitting time will be used in deriving the formula of the maximum winning probability and proving the concavity of the maximum winning probability.

▶ **Proposition 18.** *Let $\{W_t\}_{t=0}^{\infty}$ be a random walk on $\mathbb{Z}$. Specifically, $W_t = \sum_{i=1}^{t} Y_i$, where Y_i are independent and identically distributed integer-valued random variables satisfying $\Pr[Y_1 \geq -1] = 1$ (left-continuous). For any $a \in \mathbb{Z}^+$, define the hitting time $\tau_{-a} := \min\{t \geq 0 \mid W_t = -a\}$, with the convention that $\tau_{-a} := \infty$ if $W_t \neq -a$ for all $t \geq 0$. Then the following holds:*

1. ([27]) *For every $m, a \in \mathbb{Z}^+$,*

$$\Pr[\tau_{-a} = m] = \frac{a}{m} \Pr[W_m = -a];$$

2. *For any $a \in \mathbb{Z}^+$,*

$$\Pr[\tau_{-a} = \infty] = 1 - \Big(1 - \Pr[\tau_{-1} = \infty]\Big)^a \leq a \Pr[\tau_{-1} = \infty].$$

We refer the readers to [27] for the proof of Item 1, and the full version of the paper [8] for the proof of Item 2.

3.5 Determining optimal strategy

In this subsection, we show our main result for the online decision-making game. The following theorem implies that if the largest distribution π^* (under "$\preceq_{(2)}$") of $\mathcal{P}$ exists and is a Poisson binomial distribution with expectation at least 1, then an optimal strategy $\mathcal{S}^*$ for the player is $\mathcal{S}^* \equiv \pi^*$, in other words, the player can achieve the maximum winning probability when always choosing the largest distribution.

▶ **Theorem 19** (Main result for online decision making)**.** *Let $\mathcal{P}$ be a family of distributions on $\mathbb{Z}_{\geq 0}$. Suppose the metric space $(\mathcal{P}, d_{\mathrm{TV}})$ is compact. Suppose there exists a largest distribution in $\mathcal{P}$ under the partial order "$\preceq_{(2)}$", denoted by π^*. Furthermore, suppose π^* is a Poisson binomial distribution with expectation at least 1. Then the following holds:*

1. *(Recurrence) For any $m, a \in \mathbb{Z}^+$,*

$$\varphi_m^*(a) = \mathop{\mathbb{E}}_{X \sim \pi^*}\left[\varphi_{m-1}^*(a - 1 + X)\right];$$

Namely, the player will pick π^* to achieve the maximum winning probability, and the optimal strategy is $\mathcal{S}^*(m, a) = \pi^*$ for all $m, a \in \mathbb{Z}^+$.

2. *(Formula) Let $Z_1, Z_2, \cdots$ be independent and identically distributed random variables with distribution π^*, and let $S_t = \sum_{i=1}^{t}(Z_i - 1)$. For any $a \in \mathbb{Z}^+$, define $\tau_{-a} := \min\{t \geq 0 \mid S_t = -a\}$, with the convention that $\tau_{-a} := \infty$ if $S_t \neq -a$ for all $t \geq 0$. For any $m, a \in \mathbb{Z}^+$, it holds that*

$$\varphi_m^*(a) = \Pr[\tau_{-a} \geq m] = \sum_{t \geq m} \frac{a}{t} \Pr[S_t = -a] + \Pr[\tau_{-a} = \infty];$$

3. *(Concavity) For any $m \in \mathbb{Z}^+$, the function φ_m^* is concave:*

$$2\varphi_m^*(a) \geq \varphi_m^*(a-1) + \varphi_m^*(a+1), \quad \forall a \in \mathbb{Z}^+.$$

We will prove by induction on m. The recurrence follows from the concavity of the maximum winning probability and Proposition 11 about second-order stochastic dominance. Given the recurrence, we obtain the formula using Proposition 18 about the random walk hitting time. Finally, we derive concavity using the formula and the fact that π^* is a Poisson binomial distribution and hence satisfies unimodality with mode near the mean.

Proof. We prove by induction on m.

Base case: $m = 1$. By definition, $\varphi_1^*(a) = \mathbb{1}[a \geq 1]$. It is easy to check that Items 1 and 3 hold for $m = 1$. For Item 2, the first equality holds trivially for $m = 1$, the proof for the second equality for $m = 1$ is the same as that for arbitrary $m \geq 2$, i.e., applying Item 1 of Proposition 18, which we will show in the inductive step below.

Inductive step. For any $m \geq 2$, suppose Items 1–3 hold for $m - 1$, and we aim to prove Items 1–3 hold for m.

1. *Recurrence for m*

By Lemma 9, it holds that

$$\varphi_m^*(a) = \max_{\pi \in \mathcal{P}} \mathop{\mathbb{E}}_{X \sim \pi} \left[\varphi_{m-1}^*(a - 1 + X)\right]. \tag{6}$$

By definition, it is clear that $\varphi_{m-1}^*(a - 1 + X)$ is a non-decreasing function with respect to X. By Concavity for $m - 1$, $\varphi_{m-1}^*(a - 1 + X)$ is a concave function with respect to X. By assumption, $\pi \preceq_{(2)} \pi^*$ holds for any $\pi \in \mathcal{P}$. For any $\pi \in \mathcal{P}$, applying the equivalence between Item 1 and Item 2 of Proposition 11 with $\mu = \pi, \nu = \pi^*$ and $f(X) = \varphi_{m-1}^*(a - 1 + X)$, it holds that

$$\mathop{\mathbb{E}}_{X \sim \pi} \left[\varphi_{m-1}^*(a - 1 + X)\right] \leq \mathop{\mathbb{E}}_{X \sim \pi^*} \left[\varphi_{m-1}^*(a - 1 + X)\right]. \tag{7}$$

Equation (6) and Equation (7) imply Recurrence for m.

2. *Formula for m*

We first explain why $\varphi_m^*(a) = \Pr\left[\tau_{-a} \geq m\right]$ is true intuitively. Recall that $Z_1, Z_2, \cdots$ are independent and identically distributed random variables with distribution π^*, representing the player choosing the largest distribution π^* every time. Then $S_t = \sum_{i=1}^t (Z_i - 1)$ represents the net income of tokens after round t when the player chooses the largest distribution π^* every time. Then, τ_{-a} is the first time that the net income of tokens is $-a$, i.e., the time the game stops if the player initially gets a tokens. Therefore, $\Pr\left[\tau_{-a} \geq m\right]$ is the probability that the game lasts for at least m rounds when the player chooses the largest distribution π^* every time. Then $\varphi_m^*(a) = \Pr\left[\tau_{-a} \geq m\right]$ follows from the fact that maximum winning probability can be obtained from choosing π^* every time, where the fact follows from Recurrence for $1, 2, \cdots, m$.

We next prove $\varphi_m^*(a) = \Pr\left[\tau_{-a} \geq m\right]$ formally from the induction hypothesis. For any $a \in \mathbb{Z}^+$, we have that

$$\begin{aligned}
\varphi_m^*(a) &= \mathop{\mathbb{E}}_{X \sim \pi^*} \left[\varphi_{m-1}^*(a - 1 + X)\right] && \text{(Recurrence for } m\text{)} \\
&= \sum_{i=0}^{\infty} \pi^*(i) \Pr\left[\tau_{-(a-1+i)} \geq m - 1\right] && \text{(Formula for } m - 1\text{)} \\
&= \sum_{i=0}^{\infty} \Pr\left[Z_1 = i\right] \Pr\left[\tau_{-a} \geq m \mid Z_1 = i\right] \\
&= \Pr\left[\tau_{-a} \geq m\right],
\end{aligned}$$

as desired.

For the the second equality of Formula for m, we apply Item 1 of Proposition 18 with $Y_t = Z_t - 1$ and $W_t = \sum_{i=1}^t Y_i = \sum_{i=1}^t (Z_i - 1) = S_t$. Then it follows

$$\begin{aligned}\Pr[\tau_{-a} \geq m] &= \sum_{t=m}^{\infty} \Pr[\tau_{-a} = t] + \Pr[\tau_{-a} = \infty] \\ &= \sum_{t=m}^{\infty} \frac{a}{t} \Pr[S_t = -a] + \Pr[\tau_{-a} = \infty],\end{aligned}$$

as desired.

3. Concavity for m

For any $a \geq 2$, it holds that

$$\begin{aligned}2\varphi_m^*(a) &= \mathop{\mathbb{E}}_{X\sim\pi^*}\left[2\varphi_{m-1}^*(a-1+X)\right] && \text{(Recurrence for } m) \\ &\geq \mathop{\mathbb{E}}_{X\sim\pi^*}\left[\varphi_{m-1}^*(a-2+X) + \varphi_{m-1}^*(a+X)\right] && \text{(Concavity for } m-1) \\ &= \mathop{\mathbb{E}}_{X\sim\pi^*}\left[\varphi_{m-1}^*(a-2+X)\right] + \mathop{\mathbb{E}}_{X\sim\pi^*}\left[\varphi_{m-1}^*(a+X)\right] \\ &= \varphi_m^*(a-1) + \varphi_m^*(a+1). && \text{(Recurrence for } m)\end{aligned}$$

It remains to prove the case $a = 1$, i.e.,

$$2\varphi_m^*(1) \geq \varphi_m^*(2).$$

By Formula for m,

$$\begin{aligned}\varphi_m^*(1) &= \sum_{t=m}^{\infty} \frac{1}{t} \Pr[S_t = -1] + \Pr[\tau_{-1} = \infty] \\ &= \sum_{t=m}^{\infty} \frac{1}{t} \Pr\left[\sum_{i=1}^{t} Z_i = t-1\right] + \Pr[\tau_{-1} = \infty],\end{aligned}$$

$$\varphi_m^*(2) = \sum_{t=m}^{\infty} \frac{2}{t} \Pr\left[\sum_{i=1}^{t} Z_i = t-2\right] + \Pr[\tau_{-2} = \infty].$$

It suffices to prove

$$\Pr\left[\sum_{i=1}^{t} Z_i = t-1\right] \geq \Pr\left[\sum_{i=1}^{t} Z_i = t-2\right], \quad \forall t \in \mathbb{Z}^+, \tag{8}$$

and

$$2\Pr[\tau_{-1} = \infty] \geq \Pr[\tau_{-2} = \infty]. \tag{9}$$

For any $t \in \mathbb{Z}^+$, since π^* is a Poisson binomial distribution, by Fact 15, $\sum_{i=1}^t Z_i$ is a Poisson binomial random variable. By Proposition 17, we have

$$\Pr\left[\sum_{i=1}^{t} Z_i = j\right] \geq \Pr\left[\sum_{i=1}^{t} Z_i = j-1\right]$$

for any $j \leq q$, where $q = \mathbb{E}\left[\sum_{i=1}^t Z_i\right] = t\,\mathbb{E}[Z_1] \geq t$, which implies Equation (8).

Applying Item 2 of Proposition 18 with $a = 2$ yields Equation (9). ◀

4 Site Percolation on Infinite Tree

In this section, we aim to prove the main result for the site percolation model Theorem 8 and show that it can be applied to the critical hardcore model.

In Section 4.1, we prove Lemma 6 about a contraction property of the tree recursion of the hardcore model, which indicates that the main result for the site percolation model Theorem 8 can be applied to the critical hardcore model. In Section 4.2, we prove the main result for the site percolation model Theorem 8 by applying the online decision-making game introduced in Section 3.

4.1 Contraction of tree recursion: Proof of Lemma 6

In this subsection, we prove a general version of Lemma 6. In the general version, we extend the domain of the fugacity λ from at most the critical fugacity to at most $(1+\varepsilon)$ times the critical fugacity. And the original version (Lemma 6) can be obtained simply by letting $\varepsilon = 0$.

▶ **Lemma 20** (General version of Lemma 6). *Let $T = (V, E)$ be a tree rooted at r with maximum degree at most Δ. Let $\varepsilon \geq 0$ be a constant. Consider the hardcore model on T with fugacity $\lambda \leq (1+\varepsilon)\lambda_c(\Delta)$. For each vertex $v \in V$, let p_v denote the probability that v is occupied in the hardcore model with fugacity λ on the subtree T_v rooted at v, i.e., $p_v := \Pr_{\mu_{T_v,\lambda}}[\sigma_v = 1]$. Then, for every non-root vertex $v \in V \setminus \{r\}$, it holds*

$$p_v \sum_{w \in L(v)} p_w \leq \frac{1 + \mathrm{e}\varepsilon}{\Delta - 1}.$$

The following well-known fact gives a natural recurrence of the probability of the root being occupied in the hardcore model of subtrees.

▶ **Fact 21** (Tree recursion, [28]). *For $\{p_v\}$ defined in Lemma 20, and for any $v \in V$, it holds that*

$$\frac{p_v}{1 - p_v} = \lambda \prod_{w \in L(v)} (1 - p_w).$$

Proof of Lemma 20. Fix $v \in V \setminus \{r\}$. Let $d = \Delta - 1$. Then v has at most d children, i.e., $|L(v)| \leq d$.

By tree recursion Fact 21, we have

$$\frac{p_v}{1 - p_v} = \lambda \prod_{w \in L(v)} (1 - p_w) \leq \lambda \left(1 - \frac{1}{d} \sum_{w \in L(v)} p_w\right)^d = \lambda(1 - \bar{p})^d,$$

where $\bar{p} = \frac{1}{d} \sum_{w \in L(v)} p_w$, and the inequality follows from the AM-GM inequality and $|L(v)| \leq d$. Therefore,

$$p_v \sum_{w \in L(v)} p_w \leq \frac{\lambda(1 - \bar{p})^d}{1 + \lambda(1 - \bar{p})^d} \sum_{w \in L(v)} p_w = d\bar{p} \frac{\lambda(1 - \bar{p})^d}{1 + \lambda(1 - \bar{p})^d}.$$

Set

$$f(x) = x \frac{\lambda(1 - x)^d}{1 + \lambda(1 - x)^d}, x \in [0, 1].$$

Since $\bar{p} \in [0,1]$, we have $p_v \sum_{w\in L(v)} p_w \le d \max_{x\in[0,1]} f(x)$. We next bound $\max_{x\in[0,1]} f(x)$. By standard calculus calculation, we have

$$f'(x) = \frac{\lambda(1-x)^{d-1}}{(1+\lambda(1-x)^d)^2}(\lambda(1-x)^{d+1} + (1+d)(1-x) - d).$$

Set $g(x) = \lambda(1-x)^{d+1} + (1+d)(1-x) - d$. Since the first factor of $f'(x)$ is always non-negative, the sign of $f'(x)$ depends on the second factor, i.e., $g(x)$. Clearly, $g(x)$ is decreasing on $[0,1]$, $g(0) > 0$, $g(1) < 0$, by the Intermediate Value Theorem, there exists a unique zero of $g(x)$ on $[0,1]$, denoted by $\hat{x}$. Then,

$$\begin{aligned}\max_{x\in[0,1]} f(x) = f(\hat{x}) &= \hat{x}\left(1 - \frac{1}{1+\lambda(1-\hat{x})^d}\right)\\ &= \hat{x}\left(1 - \frac{1}{1+\frac{d}{1-\hat{x}} - (1+d)}\right) && \text{(by } g(\hat{x}) = 0\text{)}\\ &= \hat{x} + \frac{\hat{x}-1}{d}.\end{aligned}$$

Therefore,

$$p_v \sum_{w\in L(v)} p_w \le d \max_{x\in[0,1]} f(x) = (d+1)\hat{x} - 1.$$

When $\varepsilon = 0$, $\lambda \le \lambda_c(\Delta)$, it holds that

$$g\left(\frac{1}{d}\right) \le \lambda_c(\Delta)\left(1 - \frac{1}{d}\right)^{d+1} + (1+d)\left(1 - \frac{1}{d}\right) - d = 0.$$

By $g(x)$ is decreasing on $[0,1]$, we have $\hat{x} \le \frac{1}{d}$. Then,

$$p_v \sum_{w\in L(v)} p_w \le (d+1)\hat{x} - 1 \le \frac{1}{d},$$

as desired.

We next prove for $\varepsilon > 0$. Consider $f(x) = x\frac{\lambda(1-x)^d}{1+\lambda(1-x)^d}$ as a function with respect to both x and λ. Let $h(\lambda, x) = f(x) = x\frac{\lambda(1-x)^d}{1+\lambda(1-x)^d}$. Then, by the result of the case $\varepsilon = 0$, we have $h(\lambda, x) \le \frac{1}{d^2}$ for all $\lambda \le \lambda_c(\Delta), x \in [0,1]$. For all $\lambda > 0,\ x \in [0,1]$,

$$h'_\lambda(\lambda, x) = \frac{x(1-x)^d}{(1+\lambda(1-x)^d)^2} \le x(1-x)^d \le \frac{1}{d}\left(\frac{(dx + d(1-x))}{d+1}\right)^{d+1} = \frac{1}{d}\left(\frac{d}{d+1}\right)^{d+1},$$

where the last inequality follows from the AM-GM inequality. Then, by the Mean Value Theorem, for any $\lambda \le (1+\varepsilon)\lambda_c(\Delta), x \in [0,1]$, it holds that

$$\begin{aligned}h(\lambda, x) = h(\lambda_c(\Delta), x) + h'_\lambda(\lambda^*, x)(\lambda - \lambda_c(\Delta)) &\le \frac{1}{d^2} + \frac{1}{d}\left(\frac{d}{d+1}\right)^{d+1}\varepsilon\lambda_c(\Delta)\\ &= \frac{1}{d^2}\left(1 + \varepsilon\left(1 + \frac{1}{d^2-1}\right)^{d+1}\right) \le \frac{1}{d^2}\left(1 + \varepsilon \mathrm{e}^{\frac{1}{d-1}}\right) \le \frac{1+\mathrm{e}\varepsilon}{d^2},\end{aligned}$$

where λ^* is a real number between $\lambda_c(\Delta)$ and λ, and the second inequality follows from $1 + x \le \mathrm{e}^x$. Therefore, when $\lambda \le (1+\varepsilon)\lambda_c(\Delta)$, $f(x) \le \frac{1+\mathrm{e}\varepsilon}{d^2}$ holds for any $x \in [0,1]$. Then,

$$p_v \sum_{w\in L(v)} p_w \le d \max_{x\in[0,1]} f(x) \le \frac{1+\mathrm{e}\varepsilon}{d},$$

as desired. ◀

4.2 Site percolation on trees: Proof of Theorem 8

In this subsection, we prove a general version of Theorem 8. In the general version, we extend the upper bound of $p_v \sum_{w \in L(v)} p_w$ from $\frac{1}{d}$ to $\frac{1}{d}(1 + \frac{c_1}{\sqrt{n}})$. We note the latter upper bound can be derived from Lemma 20. The original version (Theorem 8) can be obtained simply by letting $c_1 = 0, c_2 = c$.

▶ **Theorem 22** (General version of Theorem 8)**.** *Consider a site percolation model on the infinite d-ary tree $\mathbb{T}_d^{\text{ary}} = (V, E)$ rooted at r with occupation probability list $P = \{p_v\}_{v \in V}$ where $p_r = 1$. Let $N = N(\mathbb{T}_d^{\text{ary}}, P)$ be the size of the open cluster containing the root. Let $n \in \mathbb{Z}^+$ be a positive integer. Suppose the following conditions hold:*

1. *For every non-root vertex $v \in V \setminus \{r\}$, it holds $p_v \sum_{w \in L(v)} p_w \le \frac{1}{d}\left(1 + \frac{c_1}{\sqrt{n}}\right)$, where $c_1 \ge 0$ is a constant;*
2. *At the root r, it holds*

$$\sum_{w \in L(r)} p_w \le c_2, \tag{10}$$

where $c_2 > 0$ is a constant.

Then, it holds that $\mathbb{E}[N \wedge n] = O(\sqrt{n})$, where the constant in big-O depends only on c_1, c_2.

We first show an upper bound with respect to N_e, the number of vertices at even levels that are in the open cluster containing the root, and it is straightforward to prove the upper bound with respect to N when combining Equation (10).

▶ **Lemma 23.** *Under the setting of Theorem 22, let N_e be the number of vertices at even levels that are in the open cluster containing the root. Then, we have*

$$\mathbb{E}[N_e \wedge n] = O(\sqrt{n}),$$

where the constant in big-O depends only on c_1.

Proof of Theorem 22. Let $r_1, \cdots, r_d$ be d children of r. Recall that N_e is the number of vertices at even levels that are in the open cluster containing the root. For $i = 1, 2, \cdots, d$, let N_i be the number of vertices at odd levels that are both in the open cluster containing the root and T_{r_i}, where we recall T_{r_i} denotes the subtree rooted at r_i. Let $B_i = \mathbb{1}[r_i \text{ is open}]$ $(i = 1, \cdots, d)$. Then,

$$\begin{aligned}
\mathbb{E}[N \wedge n] &= \mathbb{E}\left[\left(N_e + \sum_{i=1}^{d} N_i\right) \wedge n\right] \\
&\le \mathbb{E}\left[N_e \wedge n + \sum_{i=1}^{d} (N_i \wedge n)\right] \\
&= \mathbb{E}[N_e \wedge n] + \sum_{i=1}^{d} \Pr[B_i = 1]\, \mathbb{E}[N_i \wedge n | B_i = 1].
\end{aligned}$$

By Lemma 23, $\mathbb{E}[N_e \wedge n] = O_{c_1}(\sqrt{n})$. Similarly, $\mathbb{E}[N_i \wedge n | B_i = 1] = O_{c_1}(\sqrt{n})$ for any $1 \le i \le d$.

Therefore, there exists a constant $K = K(c_1)$ depending only on c_1, such that

$$\begin{aligned}\mathbb{E}[N \wedge n] &\le \mathbb{E}[N_e \wedge n] + \sum_{i=1}^{d} \Pr[B_i = 1]\, \mathbb{E}[N_i \wedge n | B_i = 1] \\ &\le K\sqrt{n}\left(1 + \sum_{i=1}^{d} \Pr[B_i = 1]\right) \\ &\le K\sqrt{n}(1 + c_2). && \text{(Equation (10))}\end{aligned}$$

This shows $\mathbb{E}[N \wedge n] = O_{c_1,c_2}(\sqrt{n})$ as desired. ◀

We next prove Lemma 23 by considering the process of site percolation described in Section 2.4.

Proof of Lemma 23. Recall that in Section 2.4, we introduced a process of site percolation working in rounds that reveals the status of all children and grandchildren of an active vertex in each round. We note that the process in Section 2.4 is described under an adversary setting. Here, we restate this process more precisely for a fixed site percolation model. For convenience, we call it the *site percolation process.*

Let A_t be the number of active vertices after round t. Let U_t be the set of active vertices after round t. We label all the vertices in V by $1, 2, \cdots$. At first, only the root is open and active, and the status of all other vertices is unrevealed. Therefore, $A_0 = 1$, $U_0 = \{r\}$. For any $t \in \mathbb{Z}^+$, at the beginning of round t, assuming $A_{t-1} \ge 1$, we choose the active vertex with the least label in the current set of active vertices U_{t-1}, denoted by v_t. Then reveal the status of all the children and grandchildren of v_t. To be specific, for each child or grandchild of v_t, denoted by u, sample $B_u \sim \mathrm{Ber}(p_u)$ independently, then let u be open if $B_u = 1$; be closed if $B_u = 0$. Let Q_t be the set of activated grandchildren of v_t. We note that if a grandchild is activated, then it is open; however, the converse does not hold. Let X_t be the number of activated grandchildren of v_t, i.e., $X_t = |Q_t|$. Then $U_t = U_{t-1} \setminus \{v_t\} \cup Q_t$, and $A_t = A_{t-1} - 1 + X_t$. The process stops when there are no active vertices, i.e., the first time $A_t = 0$. If the process stops after k rounds, we have $N_e = k$, because every vertex in the open cluster containing the root at even levels becomes active at some time and will be deactivated later in the process, and the number of rounds equals the number of deactivated vertices. If the process never stops, we have $N_e = \infty$.

Let $\mathcal{V}$ be the set of all possible active vertex sets of the site percolation process. For any $m, a \in \mathbb{Z}_{\ge 0}, S \in \mathcal{V}$ with $a = |S|$, define the *winning probability* $\psi_m(a, S)$ of the size percolation process by the probability that the site percolation process will last for at least m more rounds when currently there are a active vertices and the set of active vertices is S. Then, for any $m \in \mathbb{Z}^+$, we have $\Pr[N_e \ge m] = \psi_m(1, \{r\})$.

Let π_t be the law of X_t. We next find a family of distributions that contains all possible π_t. Let $u_1, \cdots, u_d$ be d children of v_t. Let Z_i be the number of children of u_i that are activated in round t. Then, we have

$$X_t = \sum_{i=1}^{d} Z_i \quad \text{and} \quad \mathbb{E}[Z_i] = p_{u_i} \sum_{w \in L(u_i)} p_w \le \frac{1}{d}\left(1 + \frac{c_1}{\sqrt{n}}\right).$$

Then, $\pi_t = \zeta_1 * \zeta_2 * \cdots * \zeta_d$, where ζ_i is the law of Z_i. Let $\mathcal{D}$ be a family of distributions defined by

$$\mathcal{D} = \left\{ \mu_1 * \cdots * \mu_d \,\middle|\, \mathbb{E}_{\mu_i}[X] \le \frac{1}{d}\left(1 + \frac{c_1}{\sqrt{n}}\right), i = 1 \cdots, d \right\},$$

where $\mu_1, \cdots, \mu_d$ are distributions on $\mathbb{Z}_{\geq 0}$. Then for any $t \in \mathbb{Z}^+$, $\pi_t \in \mathcal{D}$, i.e., $\mathcal{D}$ is a family of distributions that contains all possible π_t.

The site percolation process defined above can be considered as a strategy of the online decision-making game introduced in Section 3 with $\mathcal{P} = \mathcal{D}$. (We note that this strategy is slightly different from the strategy defined in Section 3.1, which we will discuss later.) This inspires us to consider the online decision-making game with $\mathcal{P} = \mathcal{D}$.

We next check $\mathcal{D}$ satisfies the conditions in Theorem 19. It is easy to check that the metric space $(\mathcal{D}, d_{\mathrm{TV}})$ is compact. Let $\gamma = \frac{1}{d}\left(1 + \frac{c_1}{\sqrt{n}}\right)$. We next show $\mathrm{Bin}(d, \gamma)$ is the largest distribution in $\mathcal{D}$ under "$\preceq_{(2)}$". First of all, $\mathrm{Bin}(d, \gamma) = (\mathrm{Ber}(\gamma))^{*d} \in \mathcal{D}$, where we recall that for a distribution μ, μ^{*t} is the t-fold convolution of μ with itself (see Definition 2). For any $\mu \in \mathcal{D}$, we may assume $\mu = \mu_1 * \cdots * \mu_d$, where $\mu_1, \cdots, \mu_d$ are distributions on $\mathbb{Z}_{\geq 0}$ with expectations at most γ. For any $1 \leq i \leq d$, since $\mathbb{E}_{\mu_i}[X] \leq \gamma$, by Lemma 12, we have $\mu_i \preceq_{(2)} \mathrm{Ber}(\gamma)$. By Lemma 13, we have $\mu = \mu_1 * \cdots * \mu_d \preceq_{(2)} \mathrm{Ber}(\gamma) * \cdots * \mathrm{Ber}(\gamma) = (\mathrm{Ber}(\gamma))^{*d} = \mathrm{Bin}(d, \gamma)$. Therefore, $\mathrm{Bin}(d, \gamma)$ is the largest distribution in $\mathcal{D}$ under "$\preceq_{(2)}$". Furthermore, it is clear that $\mathrm{Bin}(d, \gamma)$ is a Poisson binomial distribution with expectation $d\gamma = 1 + \frac{c_1}{\sqrt{n}} \geq 1$.

Therefore, by Theorem 19, for the online decision-making game with $\mathcal{P} = \mathcal{D}$, there is a optimal strategy $\mathcal{S}^*(m, a) \equiv \pi^* := \mathrm{Bin}(d, \gamma)$, and the player can achieve the maximum wining probability by using this strategy, i.e., by choosing $\pi_t = \pi^* := \mathrm{Bin}(d, \gamma)$ for each round $t \in \mathbb{Z}^+$.

As we mentioned above, the site percolation process can be considered as a strategy for a player playing the online decision-making game with $\mathcal{P} = \mathcal{D}$. However, the strategy of the player here is distinct from the strategies defined in Section 3.1, in the sense that the player (the site percolation process) maintains extra storage and uses external randomness (i.e., the evolution of the set of active vertices) beyond provided in the game (the target number of rounds to survive and the number of tokens) to make a decision in each round. However, it is not hard to see intuitively that an optimal strategy should not require any other information and should depend only on m, the target number of rounds to survive, and a, the current number of tokens. Namely, the strategy from the site percolation process is no better than an optimal oblivious strategy as defined in Section 3.1. We formally state this in the following claim, whose proof is similar to Lemma 9.

▷ Claim 24. For any $m \in \mathbb{Z}^+, a \in \mathbb{Z}_{\geq 0}, S \in \mathcal{V}$ with $a = |S|$, it holds that

$$\psi_m(a, S) \leq \varphi_m^*(a),$$

where $\varphi_m^*(a)$ is the maximum winning probability of the online decision-making game with $\mathcal{P} = \mathcal{D}$ when the player needs to survive m more rounds to win and the current number of tokens is a.

Claim 24 implies $\Pr[N_e \geq m] = \psi_m(1, \{r\}) \leq \varphi_m^*(1)$ holds for any $m \in \mathbb{Z}^+$. We next bound $\varphi_m^*(1)$. By Item 2 of Theorem 19, we have for any $m, a \in \mathbb{Z}^+$,

$$\varphi_m^*(a) = \Pr[\tau_{-a} \geq m],$$

where τ_{-a} is defined in the same way as in Item 2 of Theorem 19. To be specific, let $Z_1, Z_2, \cdots$ be independent and identically distributed random variables with distribution $\pi^* = \mathrm{Bin}(d, \gamma)$, and let $S_t = \sum_{i=1}^{t}(Z_i - 1)$, then τ_{-a} is defined by $\tau_{-a} := \min\{t \geq 0 \mid S_t = -a\}$, with the convention that $\tau_{-a} := \infty$ if $S_t \neq -a$ for all $t \geq 0$. Then $\Pr[N_e \geq m] \leq \varphi_m^*(1) = \Pr[\tau_{-1} \geq m]$ holds for any $m \in \mathbb{Z}^+$.

Let $P^* = \{p^*_v\}_{v \in V}$ be an occupation probability list satisfying $p^*_r = 1$, and $p^*_v = \gamma$ for all $v \in V \setminus \{r\}$. It is straightforward to check that $\tau_{-1} \overset{d}{=} N(\mathbb{T}^{\text{ary}}_d, P^*)$, where $N(\mathbb{T}^{\text{ary}}_d, P^*)$ is the random variable representing the size of the open cluster containing the root in the site percolation on $\mathbb{T}^{\text{ary}}_d$ with occupation probability γ for all non-root vertices. For simplicity of notation, we write N^* as shorthand for $N(\mathbb{T}^{\text{ary}}_d, P^*)$. By [3, Lemmas 4.7 and 4.8], we have

$$\Pr[N^* = \infty] = O_{c_1}(n^{-\frac{1}{2}}) \quad \text{and} \quad \Pr[N^* = \ell] = O(\ell^{-\frac{3}{2}}).$$

Then, for any $m \in \mathbb{Z}^+$,

$$\begin{aligned}
\Pr[N_e \geq m] \leq \Pr[\tau_{-1} \geq m] &= \Pr[N^* \geq m] \\
&= \sum_{\ell=m}^{\infty} \Pr[N^* = \ell] + \Pr[N^* = \infty] \\
&\leq K_1 \left(\sum_{\ell=m}^{\infty} \ell^{-\frac{3}{2}} + n^{-\frac{1}{2}} \right) \leq K_2 \left(m^{-\frac{1}{2}} + n^{-\frac{1}{2}} \right),
\end{aligned}$$

where $K_1 = K_1(c_1), K_2 = K_2(c_1)$ are constants depending only on c_1. Therefore,

$$\mathbb{E}[N_e \wedge n] = \sum_{m=1}^{n} \Pr[N_e \geq m] \leq \sum_{m=1}^{n} K_2 \left(m^{-\frac{1}{2}} + n^{-\frac{1}{2}} \right) = O_{c_1}(\sqrt{n}),$$

as desired. ◀

To complete the proof of Lemma 23, we prove Claim 24 by induction.

Proof of Claim 24. We prove by induction on m.

For the base case $m = 1$, by definition, it is easy to check

$$\psi_1(a, S) = \varphi^*_1(a) = \mathbb{1}[a \geq 1]$$

holds for any $a \in \mathbb{Z}_{\geq 0}, S \in \mathcal{V}$ with $a = |S|$.

Now suppose $m \geq 2$ and Claim 24 holds for $m - 1$. We aim to prove Claim 24 holds for m.

For $a = 0$, by definition, $\psi_m(0, \varnothing) = \varphi^*_m(0) = 0$. We now assume $a \geq 1$. For any $S \in \mathcal{V}$ with $|S| = a$, consider that at the beginning of some round of the site percolation process, the current set of activated vertices is S, and the process needs to last for m more rounds to win. Let X be the number of vertices activated in this round, and S' be the set of active vertices after this round. Let $\hat{\pi}$ be the law of X. Then $\hat{\pi} \in \mathcal{D}$. After this round, the site percolation process needs to last for $m - 1$ more rounds to win, and there are $a - 1 + X$ active vertices, and the set of active vertices is S'. Let $\mathcal{L}$ be the law of (X, S'). Then,

$$\begin{aligned}
\psi_m(a, S) &= \mathop{\mathbb{E}}_{(X,S') \sim \mathcal{L}} [\psi_{m-1}(a - 1 + X, S')] \\
&\leq \mathop{\mathbb{E}}_{(X,S') \sim \mathcal{L}} \left[\varphi^*_{m-1}(a - 1 + X)\right] && \text{(induction hypothesis)} \\
&= \mathop{\mathbb{E}}_{X \sim \hat{\pi}} \left[\varphi^*_{m-1}(a - 1 + X)\right] \\
&\leq \sup_{\pi \in \mathcal{D}} \mathop{\mathbb{E}}_{X \sim \pi} \left[\varphi^*_{m-1}(a - 1 + X)\right] && (\hat{\pi} \in \mathcal{D}) \\
&= \varphi^*_m(a). && \text{(Lemma 9)}
\end{aligned}$$

Therefore, Claim 24 holds for m. By the induction principle, Claim 24 holds for all $m \in \mathbb{Z}^+$. ◁

References

1. Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model. In *Proceedings of the Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 1319–1330, 2020. `doi:10.1109/FOCS46700.2020.00125`.
2. Alexander Barvinok. *Combinatorics and complexity of partition functions*, volume 30 of *Algorithms and Combinatorics*. Springer, Cham, 2016. `doi:10.1007/978-3-319-51829-9`.
3. Xiaoyu Chen, Zongchen Chen, Yitong Yin, and Xinyuan Zhang. Rapid mixing at the uniqueness threshold. *arXiv preprint arXiv:2411.03413*, 2024. `doi:10.48550/arXiv.2411.03413`.
4. Xiaoyu Chen and Weiming Feng. Rapid mixing via coupling independence for spin systems with unbounded degree. *arXiv preprint arXiv:2407.04672*, 2024. `doi:10.48550/arXiv.2407.04672`.
5. Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. Optimal mixing for two-state anti-ferromagnetic spin systems. In *Proceedings of the Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 588–599, 2022. `doi:10.1109/FOCS54457.2022.00062`.
6. Xiaoyu Chen and Xinyuan Zhang. A near-linear time sampler for the ising model with external field. In *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 4478–4503, 2023. `doi:10.1137/1.9781611977554.CH170`.
7. Yuansi Chen and Ronen Eldan. Localization schemes: A framework for proving mixing bounds for Markov chains. In *Proceedings of the Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 110–122, 2022.
8. Zongchen Chen and Tianhui Jiang. Improved mixing of critical hardcore model. *Full version*, 2025. `arXiv:2505.07515`.
9. Zongchen Chen, Kuikui Liu, and Eric Vigoda. Optimal mixing of Glauber dynamics: entropy factorization via high-dimensional expansion. In *Proceedings of the Annual ACM Symposium on Theory of Computing (STOC)*, pages 1537–1550, 2021. `doi:10.1145/3406325.3451035`.
10. Zongchen Chen, Kuikui Liu, and Eric Vigoda. Rapid mixing of Glauber dynamics up to uniqueness via contraction. *SIAM J. Comput.*, 52(1):196–237, 2023. `doi:10.1137/20M136685X`.
11. John Newton Darroch. On the distribution of the number of successes in independent trials. *Ann. Math. Stat.*, 35(3):1317–1321, 1964.
12. Sudhakar Dharmadhikari and Kumar Joag-dev. *Unimodality, Convexity, and Applications*. Probability and Mathematical Statistics. Academic Press, San Diego, 1988.
13. Weiming Feng, Heng Guo, Yitong Yin, and Chihao Zhang. Rapid mixing from spectral independence beyond the boolean domain. *ACM Trans. Algorithms*, 18(3):1–32, 2022. `doi:10.1145/3531008`.
14. Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Inapproximability of the partition function for the antiferromagnetic Ising and hard-core models. *Combin. Probab. Comput.*, 25(4):500–559, 2016. `doi:10.1017/S0963548315000401`.
15. Yuanying Guan, Muqiao Huang, and Ruodu Wang. A new characterization of second-order stochastic dominance. *Insur. Math. Econ.*, 119:261–267, 2024.
16. Haim Levy. Stochastic dominance and expected utility: Survey and analysis. *Manag. Sci.*, 38(4):555–593, 1992.
17. Liang Li, Pinyan Lu, and Yitong Yin. Correlation decay up to uniqueness in spin systems. In *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 67–84, 2013. `doi:10.1137/1.9781611973105.5`.
18. Kuikui Liu. *Spectral Independence: A New Tool to Analyze Markov Chains*. PhD thesis, University of Washington, 2023.
19. Elchanan Mossel, Dror Weitz, and Nicholas Wormald. On the hardness of sampling independent sets beyond the tree threshold. *Probab. Theory Related Fields*, 143(3-4):401–439, 2009.
20. Alfred Müller and Dietrich Stoyan. *Comparison of Stochastic Models and Risks*, volume 389 of *Wiley Series in Probability and Statistics*. John Wiley & Sons, 2002.
21. Han Peters and Guus Regts. On a conjecture of Sokal concerning roots of the independence polynomial. *Michigan Math. J.*, 68(1):33–55, 2019.

22 Michael Rothschild and Joseph E. Stiglitz. Increasing risk: I. a definition. *J. Econ. Theory*, 2(3):225–243, 1970.

23 Allan Sly. Computational transition at the uniqueness threshold. In *Proceedings of the Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 287–296, 2010. doi:10.1109/FOCS.2010.34.

24 Allan Sly and Nike Sun. The computational hardness of counting in two-spin models on d-regular graphs. In *Proceedings of the Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 361–369, 2012. doi:10.1109/FOCS.2012.56.

25 Daniel Štefankovič and Eric Vigoda. Lecture notes on spectral independence and bases of a matroid: Local-to-global and trickle-down from a Markov chain perspective. *arXiv preprint*, 2023. arXiv:2307.13826.

26 Wenpin Tang and Fengmin Tang. The poisson binomial distribution – Old & new. *Statist. Sci.*, 38(1):108–119, 2023.

27 Remco van der Hofstad and Michael Keane. An elementary proof of the hitting time theorem. *Amer. Math. Monthly*, 115(8):753–756, 2008. URL: http://www.jstor.org/stable/27642587.

28 Dror Weitz. Counting independent sets up to the tree threshold. In *Proceedings of the Annual ACM Symposium on Theory of Computing (STOC)*, pages 140–149, 2006. doi:10.1145/1132516.1132538.