PROBABILISTIC TECHNIQUES FOR ANALYZING SAMPLING ALGORITHMS AND THE DYNAMICS OF LATTICE MODELS

A Dissertation
Presented to
The Academic Faculty

By

Matthew Fahrbach

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Computer Science

Georgia Institute of Technology

December 2019

PROBABILISTIC TECHNIQUES FOR ANALYZING SAMPLING ALGORITHMS AND THE DYNAMICS OF LATTICE MODELS

Approved by:

Dr. Dana Randall, Advisor
School of Computer Science
Georgia Institute of Technology

Dr. Richard Peng
School of Computer Science
Georgia Institute of Technology

Dr. Eric Vigoda
School of Computer Science
Georgia Institute of Technology

Dr. Mohit Singh
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Prasad Tetali
School of Mathematics
Georgia Institute of Technology

Date Approved: November 6, 2019

ACKNOWLEDGEMENTS

First and foremost, I want to express my deepest gratitude to my advisor, Dana Randall, for her constant guidance and support. I would also like to thank Richard Peng, Mohit Singh, Prasad Tetali, and Eric Vigoda for all they have taught me and for being part of my dissertation committee. I am grateful for my many coauthors and the research collaborations we have had together: Megan Bernstein, Prateek Bhakta, Sara Billey, Shen Chen Xu, Ben Cousins, Robert Cummings, David Durfee, Animesh Fatehpuria, Yu Gao, Gramoz Goranci, Gary L. Miller, Vahab Mirrokni, Richard Peng, Dana Randall, Saurabh Sawlani, Alan Talmage, Junxing Wang, Chi Wang, Tao Xiao, and Morteza Zadimoghaddam. I especially want to thank Morteza Zadimoghaddam for hosting me as a summer intern at Google Research in Zürich, Switzerland. Last, I would not be where I am today without the encouragement and guidance from so many extraordinary and selfless teachers: Sara Billey, Benjamin Braun, James Brennan, Kenneth L. Calvert, Richard Carey, Seth Hunter, Jerzy W. Jaromczyk, Andrew Kemp, Jim Morrow, Richard Peng, and Santosh Vempala.

The most rewarding part of graduate school has been the many conversations, experiences, and friendships along the way. I am fortunate to have spent so much time with those in my cohort: Digvijay Boob, Anna Kirkpatrick, Kevin Lai, Samantha Petti, Samira Samadi, and Saurabh Sawlani. I am also grateful to have known so many of my fellow graduate students: Prateek Bhakta, Greg Bodwin, Travis Burrows, Sarah Cannon, Marcel Celaya, Li Chen, Zongchen Chen, Ben Cousins, David Durfee, Majid Farhadi, Animesh Fatehpuria, Yu Gao, Ezra Goss, Marc Higginson–Rollins, Bhuvesh Kumar, Yacin Nadji, Nathan Opalinski, Robert Pienta, Samira Samadi, Uthaipon Tantipongpipat, Sadra Yazdanbod, Sina Yazdanbod, Youngho Yoo, Peng Zhang, and many others. Life in Atlanta would not have been same without Michelle Hwang, Sebastian Matt, Alison Smith, and Ahna Weeks. Finally, I want to thank my family, especially my parents, for providing me with a lifetime of support and endless encouragement in all of my endeavors. None of this would be possible without them.

TABLE OF CONTENTS

LIST OF FIGURES

SUMMARY

Statistical mechanics bridges the fields of physics and probability theory, providing critical insights into both disciplines. Statistical physics models capture key features of macroscopic phenomena and consist of a set of configurations satisfying various constraints. Markov chain Monte Carlo algorithms are often used to sample from distributions over the exponentially large state space of these models to gain insight about the system and estimate its thermodynamic properties. Similar problems arise throughout machine learning, optimization, and counting complexity. In this dissertation, we present several new techniques based on random walks for analyzing sampling algorithms and the dynamics of various lattice models from statistical physics.

We start by investigating the mixing time of Glauber dynamics for the six-vertex model in its ordered phases. We show that for every Boltzmann weight in the ferroelectric phase, there exist boundary conditions such that local Markov chains require exponential time to converge to equilibrium. This is the first rigorous result about the mixing time of Glauber dynamics for the six-vertex model in the ferroelectric phase. We also analyze the Glauber dynamics with free boundary conditions in the antiferroelectric phase and significantly extend the region for which local Markov chains are known to be slow mixing.

In separate lines of work, we use techniques from the theory of random walks and electrical networks to give nearly tight bounds for the transience class of the Abelian sandpile model, closing an open problem of Babai and Gorodezky. The Abelian sandpile model is the canonical dynamical system used to study the phenomenon of self-organized criticality, and the transience class measures the time needed for the process to reach steady-state behavior. We also explore a new approach for approximately sampling elements with fixed rank from graded posets that relies solely on the mixing time of biased Markov chains. This allows us to bypass the usual obstacle of log-concavity. Last, we take a foray into analytic combinatorics and use the singularity analysis of Dirichlet generating functions to design sampling algorithms for Bose–Einstein condensates.

CHAPTER 1

INTRODUCTION

Randomization has played a fundamental role in the design and analysis of modern algorithms. Many of the current best algorithms for problems in combinatorial optimization, counting complexity, and machine learning leverage random sampling to achieve faster running times without compromising the quality of their solutions. One of the most prevalent classes of sampling algorithms are Markov chain Monte Carlo (MCMC) methods. These algorithms use random walks to explore complex state spaces and generate independent samples from desirable distributions over a state space while visiting only a small fraction of its states. The predominant obstacle when analyzing MCMC algorithms is bounding the number of steps a random walk needs to take so that the process converges close to the target distribution. While there have been several triumphs over the last thirty years, designing provably efficient Markov chains still remains a challenge for countless problems in computer science, discrete mathematics, and statistical physics.

Much of this progress has built on deep insights from statistical physics. One of the most compelling discoveries is the realization that many natural Markov chains undergo phase transitions where their rate of convergence suddenly changes from polynomial-time to exponential-time as some parameter of the system is varied. This phenomenon occurs at critical thresholds governed by thermodynamic properties of the system, similar to many lattice models from statistical physics. Approaching algorithm design from the perspective of statistical physics has repeatedly proven to be a worthwhile endeavor, as demonstrated by the long-standing popularity of the Metropolis–Hastings algorithm for generating random samples [Met+53], simulated annealing in combinatorial optimization [KGV83], and Boltzmann machines in statistical inference [AHS85].

The focus of this dissertation is to introduce several new probabilistic techniques for designing provably efficient sampling algorithms and analyzing the rate at which the dynamics of various

lattice models in statistical physics converge to equilibrium. Our methods are largely inspired by the theory of random walks, leveraging connections to concentration inequalities, electrical networks, and spectral graph theory, and we use these techniques to investigate and resolve open problems about stochastic processes on lattice models. This includes making significant progress towards understanding the conjectured phase transitions of local Markov chains for the *six-vertex model* and the proving nearly tight bounds for the time required for the *Abelian sandpile model* to necessarily reach its steady-state behavior. The approaches in our other works lead to new sampling algorithms for combinatorial structures, where we exploit properties of Markov chains and *Boltzmann* (or *Gibbs*) *sampling*, as well as probabilistic processes motivated by analytic combinatorics. We proceed in the rest of the section by giving a brief overview of MCMC methods and their applications to counting complexity and statistical physics. Then we summarize the main contributions of this dissertation.

1.1 Sampling Algorithms

We start by presenting the Markov chain Monte Carlo method more formally, and then we explore some of its most notable applications. Markov chains provide us with an algorithmic approach for the following fundamental problem. Given an exponentially large, complex set Ω of combinatorial objects and a probability distribution π over Ω, generate a random element from Ω according to the distribution π. This is a core subroutine for many statistical procedures that is immediately useful for examining properties of a typical element drawn from π and constructing unbiased estimators for random variables over Ω. For example, in statistical physics, the state space Ω is often the set of configurations of a finite thermodynamic system and π is a natural probability distribution on Ω in which the probability of a configuration is related to its energy in the system. Although MCMC methods are heavily used across science and engineering, these algorithms often lack rigorous guarantees about the quality of the samples generated. For this reason, the problem of random sampling has received significant attention in theoretical computer science for decades.

One of the subtle but incredibly powerful applications of MCMC algorithms is that they can be used for estimating the cardinality of Ω, or more generally, performing discrete integration. For a given state space Ω and function $w : \Omega \longrightarrow \mathbb{R}_{\geq 0}$, the *discrete integration problem* is to estimate the weighted sum $\sum_{x \in \Omega} w(x)$ with arbitrarily small relative error. Such problems arise routinely in machine learning and statistical physics when estimating the normalizing constant of a probability distribution. In almost all interesting instances, however, exact solutions to these counting and integration problems are #**P**-complete, the counting analog of **NP**-complete decision problems. Therefore, we aim to design efficient *randomized approximation schemes* for these problems instead. Formally, let $f : \{0,1\}^* \longrightarrow \mathbb{Z}_{\geq 0}$ be a function that maps problem instances to the number of their solutions. A *fully polynomial randomized approximation scheme* (FPRAS) for f is a randomized algorithm that takes as input an instance $x \in \{0,1\}^n$ and $\varepsilon > 0$, and outputs a number Y in time that is polynomial in n (the size of the input) and ε^{-1} such that

$$\Pr\left((1-\varepsilon)f(x) \leq Y \leq (1+\varepsilon)f(x)\right) \geq \frac{3}{4}.$$

We can use MCMC methods to give an FPRAS for many such counting problems by performing the following procedure:

1. Partition the state space into two nonempty sets A and $\Omega \setminus A$.
2. Estimate the probability mass p of the set A by generating independent samples.
3. Recursively solve the counting problem on the subset A to get the value Z_A.
4. Output the final estimate $Z = \frac{1}{p} \cdot Z_A$.

This approach was introduced by Jerrum, Valiant, and Vazirani [JVV86] to show that approximate counting and *almost-uniform sampling* (i.e., generating samples that are within ε total variation distance of the uniform distribution) are interreducible for self-reducible problems. Two of the landmark achievements achievements of MCMC methods that heavily build on this technique

are estimating the volume of a convex body [DFK91] and approximating the permanent of a nonnegative matrix [JSV06]. For a more detailed discussion about MCMC methods and their applications to counting complexity, we direct the reader to the surveys [JS96, Jer03].

The results above demonstrate what can be accomplished when we can efficiently generate high-quality samples. Determining how long to simulate a Markov chain so that it produces an approximately unbiased sample from its stationary distribution, however, is often the main challenge when designing MCMC algorithms. The number of steps a Markov chain must take so that its distribution becomes close to stationarity is known as the *mixing time* of the Markov chain. Several techniques for rigorously bounding the mixing time of a Markov chain have been developed over the years, many of which we present in Section 2.1. One of the primary goals of this dissertation is to further advance the techniques used for bounding the mixing time of local Markov chains that sample configurations from weighted lattice path models in statistical physics.

Sampling algorithms for objects that are not necessarily #**P**-complete to count can also greatly benefit from using stochastic processes instead of methods based on dynamic programming. For example, one of the main advantages of using Markov chains to generate random samples is the enormous savings in space complexity compared to the amount typically required for memoization. It is also sometimes beneficial to use *rejection sampling* to generate samples. If there is an efficient algorithm for sampling from a larger state space that (1) contains the target objects and (2) preserves their conditional probabilities, then if the marginal probability of the target set has sufficient mass, we can possibly achieve a faster algorithm by repeatedly drawing samples until we produce one of the target objects. We regularly use rejection sampling in the design and analysis of our algorithms. Furthermore, in addition to MCMC algorithms, we also investigate probabilistic processes inspired by analytic combinatorics in Chapter 6 to design optimized sampling algorithms for highly-structured and decomposable combinatorial objects.

1.2 Connections to Statistical Physics

Markov chain Monte Carlo simulations are used extensively in computational physics for approximating high-dimensional integrals and gaining insight into the nature of mathematical models. While techniques here are often nonrigorous, observations about the behavior of these algorithms in practice have routinely advanced the theory of Markov chains and mixing times. The most notable example is the coexistence of *phase transitions*, where the macroscopic behavior of a model and the mixing time of a corresponding Markov chain suddenly change dramatically as a parameter of the system is varied.

1.2.1 Lattice Models

In statistical physics, lattice models are often the simplest structured models that capture the essential features of nontrivial macroscopic phenomena. They have inspired a wealth of interdisciplinary questions in probability and complexity theory, and despite their initial approachability, they can be immensely difficult to analyze. The most celebrated and deeply studied model in statistical physics is the *Ising model* of ferromagnetism, which models the spontaneous magnetization of ferromagnetic materials as their temperature varies. In an instance of the Ising model there is an underlying graph $G = (V, E)$, which we usually take to be an $n \times n$ region of the square lattice, and each vertex is assigned a *spin* value of either $+1$ or -1, hence the state space is $\Omega = \{+1, -1\}^V$. The probability mass of an Ising configuration $x \in \Omega$ is given by $\pi(x) \propto e^{-\beta H(x)}$, where $\beta = 1/T > 0$ is the inverse temperature and $H(x) = -\sum_{\{i,j\} \in E} x_i x_j$ is the *energy* of the configuration. Thus, at low temperatures the most probable configurations in the state space are those for which many pairs of neighboring vertices take on the same spin value.

A standard approach for sampling configurations from the Ising model is to use an MCMC algorithm called *Glauber dynamics*. This is a local Markov chain where at each time step, the Markov chain selects a vertex $v \in V$ uniformly at random and sets its spin to be $+1$ or -1 with

the appropriate probability conditioned on the spins of its neighbors. Using techniques that we present in Section 2.1, it is easy to show that this process converges to π. A central problem in statistical physics is to evaluate the *partition function* $Z = \sum_{x \in \Omega} e^{-\beta H(x)}$, hence the desire for Glauber dynamics to have polynomially-bounded mixing time. In addition to being a means for sampling, it is reasonable to believe that Glauber dynamics is truly how these idealized systems evolve, further justifying the study of local Markov chains and their mixing times.

One of the primary contributions of this dissertation is our analysis of the mixing time of Glauber dynamics for the *six-vertex model* in Chapter 3. The six-vertex model generalizes *Eulerian orientations* of the square lattice and models the hydrogen-bonding patterns of two-dimensional ice. This model exhibits several physical phase transitions in its parameter space, and our results work towards establishing matching critical thresholds for the mixing time of local Markov chains. In Chapter 4 we investigate the *Abelian sandpile model*, which is the canonical dynamical system for studying a phenomenon called *self-organized criticality*, and we prove nearly tight bounds for number of steps required for this process to necessarily reach its steady-state behavior.

1.2.2 Phase Transitions

One of the main motivations for studying lattice models is that they often undergo phase transitions as parameters of the thermodynamic system are varied. Surprisingly, phase transitions also seem to manifest themselves in the mixing time of local Markov chains that walk along states of the physical system. For example, Onsager [Ons44] showed that the Ising model on the infinite two-dimensional square lattice exhibits a phase transition at the critical inverse-temperature $\beta_c = \frac{1}{2}\log(1 + \sqrt{2})$. An analogous effect has been observed about the mixing time of Glauber dynamics for the Ising model on finite square lattice regions with *free boundary conditions* (i.e., boundary vertices can take either spin). At high temperatures when $\beta < \beta_c$, Glauber dynamics is known to mix in polynomial time [MO94a, MO94b], but at low temperatures when $\beta > \beta_c$, the Markov chain requires exponential time to converge to stationarity [Tho89].

1.3 Summary of Contributions

In this dissertation we introduce several new probabilistic techniques that allow us to prove previously-conjectured convergence rates about the dynamics of lattice models. One of the main themes throughout our analyses is that we leverage properties of random walks to tightly bound quantities that can be interpreted as escape probabilities. Our methods are interdisciplinary by nature, spanning computer science, discrete mathematics and statistical physics, and they have surprising consequences in the design and analysis of MCMC sampling algorithms.

In Chapter 3 we analyze the mixing time of Glauber dynamics for the six-vertex model in its ordered phases. We show that for all Boltzmann weights in the *ferroelectric* phase, there exist boundary conditions for which local Markov chains converge to equilibrium exponentially slowly. This is the first rigorous result about the mixing time of Glauber dynamics for the six-vertex model in the ferroelectric phase. In our analysis, we carefully construct asymmetric cuts in the state space that demonstrate a fundamental connection between *correlated random walks* and intersecting lattice path models. Moreover, this construction suggests an underlying combinatorial interpretation for the phase transition between the ferroelectric and *disordered* phases in terms of the adherence strength of intersecting lattice paths and the momentum parameter of correlated random walks. We remark that one of the major technical contributions here is a new tail inequality for correlated random walks.

Taking a completely different approach, we also analyze Glauber dynamics for the six-vertex model with free boundary conditions in the *antiferroelectric* phase and significantly extend the region for which local Markov chains are known to be slowly mixing. This mixing time result builds on the *topological obstruction* framework of Randall [Ran06b] and relies on a Peierls argument combined with novel properties of *weighted non-backtracking walks*, which we obtain by solving a system of linear recurrence relations. This is based on joint work with Dana Randall that appeared in the 2019 International Conference on Randomization and Computation [FR19].

In Chapter 4 we develop techniques in the theory of random walks and electrical networks that give nearly tight bounds for the *transience class* of the Abelian sandpile model on the two-dimensional grid, closing an open problem of Babai and Gorodezky [BG07, CV12]. The Abelian sandpile model is a discrete process on graphs that is intimately related to the phenomenon of *self-organized criticality*. In this process, vertices receive grains of sand, and once the number of grains exceeds their degree, they topple by sending grains to their neighbors. The transience class of the Abelian sandpile model is the maximum number of grains of sand that can be added to the system before it necessarily reaches its steady-state behavior or, equivalently, a recurrent state. Using a more refined and global analysis of electrical potentials on the $n \times n$ grid, we prove that the transience class of the grid has an upper bound of $O(n^4 \log^4 n)$ and a lower bound of $\Omega(n^4)$. Our methods also naturally generalize to d-dimensional grids to give $O(n^{3d-2} \log^{d+2} n)$ upper bounds and $\Omega(n^{3d-2})$ lower bounds. Our work builds on a reduction of Choure and Vishwanathan [CV12] that bounds the transience class in terms of *vertex potentials* when the underlying graph is viewed as an electrical network. One of the main highlights of our analysis is that we view voltages as escape probabilities and demonstrate a systematic method for decoupling two-dimensional random walks into simple symmetric random walks. This allows us to then extend well-known results for one-dimensional random walks and achieve tight inequalities for the vertex potentials. This chapter is based on joint work with David Durfee, Yu Gao, and Tao Xiao that appeared in the 2018 Annual ACM-SIAM Symposium on Discrete Algorithms [DFGX18].

In Chapter 5 we show that for certain classes of graded posets, biased Markov chains that walk along the edges of *Hasse diagrams* allow us to approximately generate uniform samples with any fixed rank in expected polynomial time. Our arguments do not rely on the typical proofs of *log-concavity*, which are often used to construct a stationary distribution with a specific mode in order to lower bound the probability of outputting an element of the desired rank. Instead, we infer the rejection rate directly from bounds on the mixing time of the Markov chains through a method that we call *balanced bias*. This approach demonstrates an unconventional way of leveraging

the high conductance of a *rapidly mixing* Markov chain. We investigate how our balanced bias technique can be applied to the problem of uniformly sampling integer partitions of n subject to a variety of constraints. In particular, we present the first provably efficient Markov chain algorithm for uniformly sampling *region-restricted* integer partitions of n. Several problem-specific observations allow us to improve our uniform sampling algorithm for integer partitions to run in expected $O(n^{9/4})$ time and to use $O(n^{1/2} \log n)$ space. Lastly, some of the related applications that we explore include sampling permutations with a fixed number of inversions and sampling lozenge tilings on the triangular lattice with a fixed average height. This is based on joint work with Prateek Bhakta, Ben Cousins, and Dana Randall that appeared in the 2017 ACM-SIAM Symposium on Discrete Algorithms [BCFR17].

In Chapter 6 we take a foray into analytic combinatorics and use techniques from Boltzmann sampling to give provably efficient sampling algorithms for a broad class of combinatorial structures known as *weighted partitions* and *selections*. In particular, we focus on a family of weighted integer partitions related to *Bose–Einstein condensation* from statistical physics. Our algorithms are probabilistic interpretations of the ordinary generating functions for these objects, derived from the symbolic method in analytic combinatorics. Using the *Khintchine–Meinardus probabilistic method* to bound the rejection rate of these Boltzmann samplers through the singularity analysis of Dirichlet generating functions, we offer an alternative approach to analyze Boltzmann samplers for objects with multiplicative structure. The main technical contributions in this chapter include developing a new tail inequality for negative binomial distributions and applying intricate bounds from analytic number theory for the Riemann zeta function to analyze the singularities of Dirichlet generating functions. This is based on joint work with Megan Bernstein and Dana Randall that appeared in the 2018 Workshop on Analytic Algorithms and Combinatorics [BFR18].

CHAPTER 2

PRELIMINARIES

We start by presenting background on Markov chains and mixing times. In particular, we show how to design Markov chains that converge to a given stationary distribution using the Metropolis–Hastings algorithm [Met+53], and we introduce several established techniques for analyzing their rates of convergence. We also examine Boltzmann distributions and their role in Markov chain Monte Carlo sampling algorithms. For a more comprehensive review, we direct the reader to the references [Ran06a, LPW17].

2.1 Markov Chains

A discrete-time Markov chain is a stochastic process on a state space Ω such that starting anywhere, the probability of moving to the next state depends solely on the current state and is independent of all previous states. Formally, a sequence of random variables $(X_0, X_1, X_2, \dots)$ is a Markov chain with state space Ω and transition matrix P if for all $x, y \in \Omega$, all $t \geq 0$, and any realizable sequence of states $(x_0, x_1, \dots, x_{t-1}, x)$, we have

$$\Pr(X_{t+1} = y \mid X_0 = x_0, X_1 = x_1, \dots, X_{t-1} = x_{t-1}, X_t = x) = \Pr(X_{t+1} = y \mid X_t = x) = P(x, y). \tag{2.1}$$

By a realizable sequence, we mean $\Pr(X_0 = x_0, X_1 = x_1, \dots, X_{t-1} = x_{t-1}, X_t = x) > 0$, so that these conditional probabilities are well-defined. Equation (2.1) is often called the *Markov* (or *memoryless*) *property* of the stochastic process, and it shows that a finite Markov chain can be completely described by its $|\Omega| \times |\Omega|$ transition matrix P. The x-th row of P is the probability distribution $P(x, \cdot)$. Therefore, P is a right stochastic matrix, i.e., all of its entries are nonnegative and $\sum_{y \in \Omega} P(x, y) = 1$ for all $x \in \Omega$.

There are two important properties of a Markov chain that govern the long-term behavior. A Markov chain is *irreducible* if for any two states $x, y \in \Omega$ there exists a $t \geq 0$ such that $P^t(x, y) > 0$, where $P^t(x, y)$ is the probability of moving from x to y in exactly t steps. Second, we say that a Markov chain is *aperiodic* if for all $x \in \Omega$ we have $\gcd\{t \geq 0 : P^t(x, x) > 0\} = 1$. A Markov chain is *ergodic* if it is irreducible and aperiodic. Ergodicity turns out to be a useful minimum requirement when designing Markov chain Monte Carlo algorithms. If a chain is not aperiodic, we can easily remedy its periodicity by introducing self-loops at each state. Concretely, we can construct a new Markov chain $Q = (I + P)/2$, where I is the $|\Omega| \times |\Omega|$ identity matrix. We call Q the *lazy version* of P because every state has a self-loop probability of at least $1/2$ and the long-term behavior of Q is the same as that of P. It is often more convenient to analyze the lazy version of a Markov chain.

2.1.1 Stationary Distributions

A *stationary distribution* of a Markov chain is a distribution π satisfying $\pi = \pi P$. In other words, we have $\pi(y) = \sum_{x \in \Omega} \pi(x) P(x, y)$ for all $y \in \Omega$. The fundamental theorem of Markov chains states that ergodicity is a necessary and sufficient condition for a Markov chain to converge to a unique stationary distribution regardless of its initial state.

Theorem 2.1.1 (Fundamental Theorem of Markov Chains)**.** *If a finite Markov chain P is ergodic, it has a unique stationary distribution π. Moreover, we have $\lim_{t \to \infty} P^t(x, y) = \pi(y)$, for all $x, y \in \Omega$.*

This result is a consequence of the Perron–Frobenius theorem, which implies that π is the unique left eigenvector of P with eigenvalue 1 and that all other eigenvalues have absolute value strictly less than 1. To reason about the stationary distribution of a Markov chain, we often use the *detailed balance equations*. These conditions state that if a distribution π satisfies

$$\pi(x)P(x, y) = \pi(y)P(y, x),$$

for all $x, y \in \Omega$, then π is a stationary distribution of P. A Markov chain that satisfies the detailed

balance equations is said to be *reversible*.

Next, we present a well-established framework for designing Markov chains that converge to any desired stationary distribution on a state space Ω using only an allowable set of transitions. The Metropolis–Hastings algorithm [Met+53] is a simple but robust idea that tells us how to assign transition probabilities so that the Markov chain converges to the target distribution π. To see how it works, let $N(x)$ denote the *neighborhood* of a state x (i.e., the set of states to which x can transition), and let the *degree* of x be the size of its neighborhood. Starting from any $x_0 \in \Omega$, at each time step the Metropolis–Hastings algorithm chooses a neighbor $y \in N(x_t)$ uniformly at random with probability $1/(2\Delta)$, where Δ is the maximum degree over all states, and sets $x_{t+1} \leftarrow y$ with probability $\min\{1, \pi(y)/\pi(x_t)\}$. The process stays at $x_{t+1} \leftarrow x_t$ with all remaining probability. Using the detailed balance equations, it is easy to verify that if this Markov chain is irreducible, then π must be the stationary distribution. The term $\min\{1, \pi(y)/\pi(x_t)\}$ is known as the *Metropolis filter*, and while the individual probabilities $\pi(y)$ and $\pi(x_t)$ are often intractable computations, we design algorithms using stationary distributions on suitably connected state spaces so that this quotient is simple to calculate.

2.1.2 Mixing Times

The Metropolis–Hastings algorithm allows us to engineer Markov chains that converge to a particular distribution, but it does not provide any guarantees about the rates of convergence. We typically measure the time required for a Markov chain to become close to its stationary distribution in terms of *total variation distance*, which for two probability distributions μ and ν on Ω is the norm

$$\|\mu - \nu\|_{\mathrm{TV}} = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|.$$

The *mixing time* of a Markov chain is the number of steps needed for the total variation distance between the t-step distribution $P^t(x, \cdot)$ and the stationary distribution π to become sufficiently

small, for all possible initial states $x \in \Omega$. Formally, for any $\varepsilon > 0$, the mixing time is defined as

$$\tau(\varepsilon) = \min\left\{ t \geq 0 \,:\, \max_{x \in \Omega} \left\| P^t(x, \cdot) - \pi \right\|_{\mathrm{TV}} \leq \varepsilon \right\}.$$

It is often convenient to assume $\varepsilon = 1/4$, because for any $\varepsilon < 1/2$, we have $\tau(\varepsilon) \leq \tau(1/4)\lceil \log_2(\varepsilon^{-1}) \rceil$. We say that a Markov chain is *rapidly* (or *polynomially*) *mixing* if its mixing time is bounded above by a polynomial in $\log(\varepsilon^{-1})$ and n, where n is the size of a configuration in the state space. Similarly, a Markov chain is said to be *slowly mixing* if its mixing time is bounded below by $\exp(n^c)$, for some constant $c > 0$.

The mixing time of a reversible, ergodic Markov chain can also be analyzed through the spectral representation of its transition matrix. Label the eigenvalues of P in decreasing order as

$$1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_{|\Omega|} > -1,$$

and let $\lambda^* = \max\{|\lambda| \,:\, \lambda \text{ is an eigenvalue of } P \text{ and } \lambda \neq 1\}$. We call the difference $1 - \lambda^*$ the *spectral gap* of P, and we can use it to prove both upper and lower bounds on the mixing time.

Theorem 2.1.2 ([LPW17, Theorem 12.3 and Theorem 12.4])**.** *Let P be a reversible, ergodic Markov chain on the state space Ω with stationary distribution π, and let $\pi^* = \min_{x \in \Omega} \pi(x)$. The mixing time satisfies*

$$\left(\frac{1}{1 - \lambda^*} - 1 \right) \log\left(\frac{1}{2\varepsilon} \right) \leq \tau(\varepsilon) \leq \frac{1}{1 - \lambda^*} \log\left(\frac{1}{\varepsilon \pi^*} \right).$$

We note that the spectrum of a lazy Markov chain is nonnegative, so the spectral gap is $1 - \lambda_2$. We also reiterate that for ergodic Markov chains, the spectral gap $1 - \lambda^* \neq 0$ by the Perron–Frobenius theorem.

2.1.3 Coupling Arguments

Coupling is a powerful technique for bounding mixing times because it reduces a comparison between distributions to a comparison between random variables, which can be considerably simpler. A coupling of a Markov chain with transition matrix P is a stochastic process $(X_t, Y_t)_{t=0}^{\infty}$ on $\Omega \times \Omega$ such that both marginals $(X_t, \cdot)_{t=0}^{\infty}$ and $(\cdot, Y_t)_{t=0}^{\infty}$ are faithful copies of the original Markov chain given the initial states $X_0 = x$ and $Y_0 = y$. Moreover, a coupling can be modified so that once the two chains simultaneously occupy the same state, they agree from that time forward. Formally, we only consider couplings with the property that if $X_t = Y_t$, then $X_{t+1} = Y_{t+1}$. For any initial states $x, y \in \Omega$, let $T_{x,y} = \min\{t \geq 0 : X_t = Y_t \mid X_0 = x, Y_0 = y\}$ be a random variable that denotes the time required for the processes to *coalesce*. The *coupling time* is the worst-case expected coalescence time $T = \max_{x,y \in \Omega} \mathbb{E}[T_{x,y}]$, and it allows us to upper bound the mixing time by $\tau(\varepsilon) \leq \lceil Te \log(\varepsilon^{-1}) \rceil$ (see, e.g., [Ald83]).

Path coupling is a technique introduced by Bubley and Dyer [BD97] that greatly simplifies the construction of couplings. Instead of arguing about the expected coalescence time for *all* pairs of initial states $x, y \in \Omega$, we can use path coupling to restrict our comparison to a small subset of neighboring states $U \subseteq \Omega \times \Omega$ that are close according to some distance metric. To show that a Markov chain is rapidly mixing, we need to choose an appropriate metric φ on Ω and construct a local coupling for the joint process starting from each pair of states $(x, y) \in U$. If we can prove that the expected change in distance of the coupled process is nonincreasing for all neighboring pairs, then we can extend the argument to all pairs of initial states $x, y \in \Omega$ by our choice of φ and the linearity of expectation, and therefore bound the time required for coalescence.

We call a metric $\varphi : \Omega \times \Omega \longrightarrow \mathbb{R}_{\geq 0}$ a *path metric* for the set $U \subseteq \Omega \times \Omega$ if for all $(x, y) \in \Omega \times \Omega$ there exists a shortest path $(z_0, z_1, \ldots, z_r)$ from $x = z_0$ to $y = z_r$ such that $(z_i, z_{i+1}) \in U$ for $0 \leq i < r$ and

$$\varphi(x, y) = \sum_{i=0}^{r-1} \varphi(z_i, z_{i+1}).$$

The following result of Dyer and Greenhill is a convenient version of the path coupling theorem.

Theorem 2.1.3 ([DG98]). *Let φ be a path metric for $U \subseteq \Omega \times \Omega$ that takes values in $\{0\} \cup [1, D]$. Let $\mathcal{M}$ be an ergodic Markov chain and let (X_t, Y_t) be a coupling of $\mathcal{M}$. Suppose there exists a $\beta \leq 1$ such that, for all $(x, y) \in U$, we have*

$$\mathbb{E}\left[\varphi\left(X_{t+1}, Y_{t+1}\right) \mid X_t = x, Y_t = y\right] \leq \beta\varphi\left(x, y\right).$$

1. *If $\beta < 1$, then the mixing time satisfies*

$$\tau(\varepsilon) \leq \frac{\log\left(D\varepsilon^{-1}\right)}{1-\beta}.$$

2. *If $\beta = 1$, i.e., $\mathbb{E}[\varphi(X_{t+1}, Y_{t+1}) - \varphi\left(x, y\right) \mid X_t = x, Y_t = y] \leq 0$, let $\alpha > 0$ be a parameter satisfying $\Pr[\varphi(X_{t+1}, Y_{t+1}) \neq \varphi(x, y) \mid X_t = x, Y_t = y] \geq \alpha$ for all $(x, y) \in U$ such that $x \neq y$. The mixing time of $\mathcal{M}$ satisfies*

$$\tau(\varepsilon) \leq \left\lceil \frac{eD^2}{\alpha} \right\rceil \left\lceil \log\left(\varepsilon^{-1}\right) \right\rceil.$$

Greenberg, Pascoe, and Randall [GPR09] showed how to modify the path coupling theorem to allow for cases where the distance between states is exponentially large and the expected change in distance is at most zero. In particular, they showed that to prove rapid mixing, it suffices for the expected change in the absolute value of the distance to be proportional to the current distance.

Theorem 2.1.4 ([GPR09]). *Let φ be a path metric for $U \subseteq \Omega \times \Omega$ that takes values in $\{0\} \cup [1, D]$. Let $\mathcal{M}$ be an ergodic Markov chain and let (X_t, Y_t) be a coupling of $\mathcal{M}$. If the expected change in distance satisfies $\mathbb{E}[\varphi(X_{t+1}, Y_{t+1}) - \varphi(x, y) \mid X_t = x, Y_t = y] \leq 0$ and there exists $\kappa, \eta \in (0, 1)$ such that $\Pr\left(|\varphi(X_{t+1}, Y_{t+1}) - \varphi(x, y)| \geq \eta\varphi(x, y) \mid X_t = x, Y_t = y\right) \geq \kappa$, for all $(x, y) \in U$, then*

$$\tau(\varepsilon) \leq \left\lceil \frac{e \log^2(D)}{\log^2(1+\eta)\kappa} \right\rceil \left\lceil \log\left(\varepsilon^{-1}\right) \right\rceil.$$

We use this path coupling result for exponential metrics in Section 5.4.3 to bound the mixing time of a biased Markov chain on region-restricted integer partitions.

2.1.4 Conductance and Isoperimetric Inequalities

The mixing time of a Markov chain can also be characterized by its *conductance*, a quantity that measures the presence of bottlenecks in the overall state space. Given a Markov chain on Ω with the transition matrix P and stationary distribution π, the conductance of a nonempty set $S \subseteq \Omega$ is

$$\Phi(S) = \frac{\sum_{x \in S, y \notin S} \pi(x) P(x, y)}{\pi(S)}.$$

It is often useful to view conductance as an escape probability—starting from stationarity and conditioned on being in S, the conductance $\Phi(S)$ is the probability that the Markov chain leaves S in one step. The *global conductance* of a Markov chain is defined as its worst-case escape probability

$$\Phi^* = \min_{S \subseteq \Omega \,:\, 0 < \pi(S) \leq 1/2} \Phi(S).$$

The notion of conductance is closely connected to the Cheeger constant in graph theory, and therefore provides us with lower and upper bounds on the spectral gap of P. In particular, Jerrum and Sinclair [JS89] proved the isoperimetric inequalities $\Phi^* \leq \sqrt{2(1 - \lambda^*)}$ and $1 - \lambda^* \leq 2\Phi^*$. Combining their results with Theorem 2.1.2, we can lower bound the mixing time as follows.

Theorem 2.1.5 ([JS89])**.** *For an ergodic, reversible Markov chain with conductance Φ^*, we have*

$$\tau(\varepsilon) \geq \left(\frac{1}{2\Phi^*} - 1 \right) \log \left(\frac{1}{2\varepsilon} \right).$$

This theorem is useful for showing that a Markov chain is slowly mixing because it only requires the construction of a cut $S \subseteq \Omega$ in the state space whose conductance $\Phi(S)$ is exponentially small. We use this technique for our analysis of the six-vertex model in Chapter 3.

A *Peierls argument* is a mathematically rigorous and intuitive method for proving the existence of such a bottleneck. For any cut $S \subseteq \Omega$, let $\partial S = \{x \in S : \text{there exists a } y \in \overline{S} \text{ with } P(x, y) > 0\}$ denote the boundary of S. If we can construct an injection $f : \partial S \longrightarrow \Omega$ such that $\pi(f(x)) \geq c^n \pi(x)$ for all $x \in \partial S$ and some constant $c > 1$, then if S is a cut satisfying $1/\text{poly}(n) \leq \pi(S) \leq 1/2$, we can show that the conductance of the Markov chain Φ^* is exponentially small. In particular, observe that

$$\Phi^* \leq \Phi(S) = \frac{\sum_{x \in S, y \notin S} \pi(x) P(x, y)}{\pi(S)} \leq \frac{\sum_{x \in \partial S} \pi(x)}{\pi(S)} \leq \frac{\sum_{x \in \partial S} \pi(f(x))}{c^n \pi(S)} \leq \frac{1}{c^n \pi(S)} \leq \frac{\text{poly}(n)}{c^n}.$$

We use the *topological obstruction* framework introduced in [Ran06b] and a Peierls argument to show slow mixing of the six-vertex model in its antiferroelectric phase in Section 3.3.

2.2 Boltzmann Distributions

One of the central objects in statistical physics is the *Boltzmann distribution* (or *Gibbs measure*). This distribution gives the probability of observing a particular state $x \in \Omega$ when considering the state space as a thermodynamic system at equilibrium for a fixed temperature T. The probability mass function takes the form

$$\Pr_\beta(x) = \frac{e^{-\beta H(x)}}{Z_\beta},$$

where $\beta = 1/T$ is the inverse temperature, $H(x)$ is the *Hamiltonian* (or *energy*) of the state x, and Z_β is a normalizing constant called the *partition function*. Computing the partition function is often an intractable problem, but we may be able to rely on Markov chain Monte Carlo methods to approximately sample from Boltzmann distributions.

A convenient property of Boltzmann distributions that can be leveraged when designing uniform sampling algorithms is that states with equal energy appear with equal probability. This provides us with some flexibility when designing Markov chains because we can embed our target set of objects into a larger state space where we might be able to show that a related Markov

chain is rapidly mixing. As long as the desired objects are assigned an equal energy value, any sample generated with this energy will be a uniform sample from the target set. The Hamiltonian of a combinatorial object is typically given by the negative of its size (e.g., the cardinality of a matching in a graph or the sum of parts in an integer partition). We can bias a distribution to favor objects with a given energy value by varying β to change the temperature of the system. Most of the analysis in this situation is then devoted to finding a viable temperature parameter, if one exists, and lower bounding the probability of generating an object from the target set at this temperature. We use this approach extensively in the design of our sampling algorithms for region-restricted integer partitions in Chapter 5 and for Bose–Einstein condensates in Chapter 6.

CHAPTER 3

SLOW MIXING OF GLAUBER DYNAMICS FOR THE SIX-VERTEX MODEL

In this chapter we show that the mixing time of Glauber dynamics for the *six-vertex model* can be exponentially slow in its ordered phases. The six-vertex model is a weighted generalization of the *ice model* (i.e., Eulerian orientations) and the *zero-temperature three-state Potts model* (i.e., proper three-colorings) on the two-dimensional square lattice. We prove for all *ferroelectric* weights that there exist boundary conditions for which local Markov chains are slowly mixing. This is the first rigorous result about the mixing time of Glauber dynamics in the ferroelectric phase. We also analyze Glauber dynamics in the *antiferroelectric* phase subject to free boundary conditions and significantly extend the subregion for which Glauber dynamics is known to be slow mixing.

3.1 Introduction

The six-vertex model was first introduced by Pauling in 1935 [Pau35] to study the thermodynamics of crystalline solids with ferroelectric properties, and has since become one of the most compelling models in statistical mechanics. The prototypical instance of the model is the hydrogen-bonding pattern of two-dimensional ice—when water freezes, each oxygen atom must be surrounded by four hydrogen atoms such that two of the hydrogen atoms bond covalently with the oxygen atom and two are farther away. The state space of the six-vertex model consists of orientations of the edges in a finite region of the two-dimensional square lattice where every internal vertex has two incoming edges and two outgoing edges, also represented as Eulerian orientations of the underlying lattice graph. The model is most often studied on the $n \times n$ square lattice $\Lambda_n \subseteq \mathbb{Z}^2$ with $4n$ additional edges so that each internal vertex has degree 4. There are six possible edge orientations incident to a vertex (see, e.g., Figure 3.1). We assign Boltzmann weights $w_1, w_2, w_3, w_4, w_5, w_6 \in \mathbb{R}_{>0}$

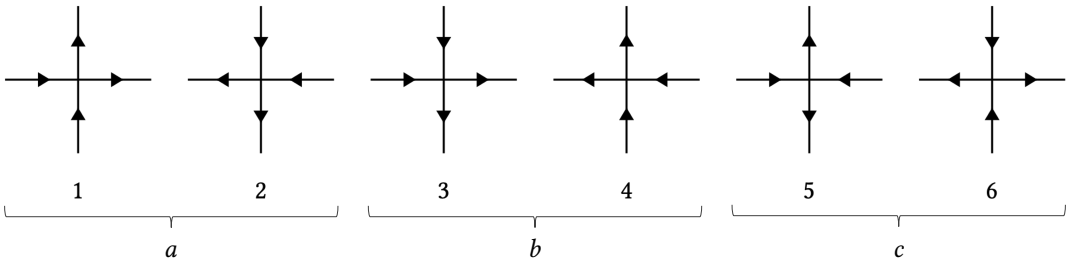


Figure 3.1: The six valid edge orientations for internal vertices in the six-vertex model and their corresponding Boltzmann weights.

to the six vertex types and define the partition function as $Z = \sum_{x \in \Omega} \prod_{i=1}^{6} w_i^{n_i(x)}$, where Ω is the set of Eulerian orientations of Λ_n and $n_i(x)$ is the number of type-i vertices in the configuration x.

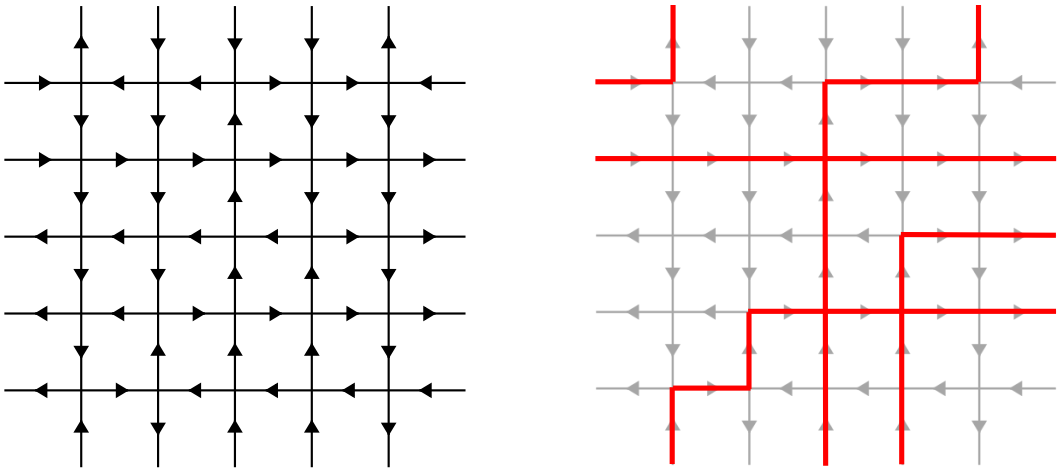
In 1967, Lieb discovered exact solutions to the six-vertex model with periodic boundary conditions (i.e., on the torus) for three different parameter regimes [Lie67a, Lie67b, Lie67c]. In particular, he famously showed that if all six vertex weights are set to $w_i = 1$, the energy per vertex is $\lim_{n \to \infty} Z^{1/n^2} = (4/3)^{3/2} = 1.5396007...$, which is known as "Lieb's square ice constant". His results were immediately generalized to all parameter regimes and to account for external electric fields [Sut67, Yan67]. An equivalence between periodic and free boundary conditions in the limit was established in [BKW73], and since then the primary object of study has been the six-vertex model subject to *domain wall boundary conditions*, where the lower and upper boundary edges point into the square and the left and right boundary edges point outwards [ICK92, KZJ00, BPZ02, BF06, BL09, BL10]. The six-vertex model serves as an important "counterexample" in statistical physics because the *surface free energy* in the thermodynamic limit depends on the boundary conditions. In particular, it is different for periodic and domain wall boundary conditions.

There have been several surprisingly profound connections to combinatorics and probability in this line of work. For example, Zeilberger gave a sophisticated computer-assisted proof of the *alternating sign matrix conjecture* in 1995 [Zei96]. A year later, Kuperberg [Kup96] produced an elegant and significantly shorter proof using analysis of the partition function of the six-vertex model with domain wall boundary conditions. Other connections to combinatorics include the

dimer model on the Aztec diamond and the arctic circle theorem [CEP96, FS06], sampling lozenge tilings [LRS01, Wil04, BCFR17], and counting 3-colorings of lattice graphs [RT00, CR16].

While there has been extraordinary progress in understanding properties of the six-vertex model with periodic or domain wall boundary conditions in mathematical physics, remarkably less is known when the model is subject to arbitrary boundary conditions. Sampling configurations using Markov chain Monte Carlo (MCMC) algorithms has been one of the primary means for discovering mathematical and physical properties of the six-vertex model [AR05, LKV17, LKRV18, KS18]. However, the model is empirically very sensitive to boundary conditions, and numerical studies have often observed slow convergence of local MCMC algorithms under certain parameter settings. For example, according to [LKRV18], "it must be stressed that the Metropolis algorithm might be impractical in the antiferromagnetic phase, where the system may be unable to thermalize." There are very few rigorous results about natural Markov chains and the computational complexity of sampling from the six-vertex model when the Boltzmann distribution is nonuniform, thus motivating our study of Glauber dynamics for the six-vertex model, the most widely used MCMC sampling algorithm, in the *ferroelectric* and *antiferroelectric phases*.

At first glance, the model has six degrees of freedom. However, this conveniently reduces to a two-parameter family because of invariants that relate pairs of vertex types. To see this, it is useful to view the configurations of the six-vertex model as intersecting lattice paths by erasing all of the edges that are directed south or west and keeping the others (see, e.g., Figure 3.2). Using this bijective "routing interpretation," it is simple to see that the number of type-5 and type-6 vertices must be closely correlated. In addition to revealing invariants, the lattice path representation of configurations turns out to be exceptionally useful for analyzing Glauber dynamics. Moreover, the total weight of a configuration should remain unchanged if all the edge directions are reversed in the absence of an external electric field, so we let $w_1 = w_2 = a$, $w_3 = w_4 = b$, and $w_5 = w_6 = c$. This complementary invariance is known as the *zero field assumption*, and it is often convenient to exploit the conservation laws of the model [BL09] to reparameterize the system so that $w_1 = a^2$



(a) (b)

Figure 3.2: Examples of a configuration in the six-vertex model: (a) illustrates the edge orientations and internal Eulerian constraints, and (b) overlays the corresponding routing interpretation in red.

and $w_2 = 1$. This allows us to ignore empty sites and focus solely on weighted lattice paths. Furthermore, since our goal is to sample configurations from the Boltzmann distribution, we can normalize the partition function by a factor of c^{-n^2} and consider the weight $(a/c, b/c, 1)$ instead of the parameter (a, b, c). Collectively, we refer to these properties as the invariance of the Gibbs measure for the six-vertex model.

The single-site *Glauber dynamics* for the six-vertex model is the Markov chain that makes local moves by (1) choosing an internal cell of the lattice uniformly at random and (2) reversing the orientations of the edges that bound the chosen cell if they form a cycle. In the lattice path interpretation, these dynamics correspond to the "mountain-valley" Markov chain that flips corners. Transitions between states are made according to the Metropolis-Hastings acceptance probability so that the Markov chain converges to the desired distribution. For a review on Markov chains, mixing times, and the Metropolis–Hastings algorithm, we refer the reader to Chapter 2.

The phase diagram of the six-vertex model represents distinct thermodynamic properties of the system and is partitioned into three regions: the *disordered* (DO) phase, the *ferroelectric* (FE)

phase, and the *antiferroelectric* (AFE) phase. To establish these regions, we consider the parameter

$$\Delta = \frac{a^2 + b^2 - c^2}{2ab}.$$

The disordered phase is the set of parameters $(a, b, c) \in \mathbb{R}^3_{>0}$ that satisfy $|\Delta| < 1$, and Glauber dynamics is expected to be rapidly mixing in this region because there are no long-range correlations in the system. The ferroelectric phase is defined by $\Delta > 1$, or equivalently when we have $a > b + c$ or $b > a + c$. The antiferroelectric phase is defined by $\Delta < -1$, or equivalently when $a + b < c$.

The phase diagram is symmetric over the positive diagonal, which follows from the fact that a and b are interchangeable under the automorphism that rotates each of the six vertex types by ninety degrees clockwise. This is equivalent to rotating the entire model under the zero field assumption. Therefore, we can assume that mixing results are symmetric over the main diagonal. Combinatorially, we show in Section 3.2 that configurations in the ferroelectric phase can be interpreted as intersecting lattice paths that prefer to adhere to each other. We carefully exploit this property to show that Glauber dynamics slow mixing. In the antiferroelectric phase, configurations prefer vertices of type-c and tend to be closely aligned with one of states with maximum probability that are arrow reversals of each other.

3.1.1 Related Works

Cai, Liu, and Lu [CLL19] recently investigated the six-vertex model for 4-regular graphs and provided strong evidence that the complexity of approximating the partition function agrees with the phase diagram from statistical physics. In particular, they give a *fully randomized approximation scheme* (FPRAS) for all 4-regular graphs in the subregion of the disordered phase defined by the inequalities $a^2 \leq b^2 + c^2$, $b^2 \leq a^2 + c^2$, and $c^2 \leq a^2 + b^2$ (i.e., the blue region in Figure 3.3a). Their algorithm builds on the *winding technique* for Holant problems developed in [McQ13, HLZ16] and requires $O(n^{10})$ time to sample a six-vertex configuration from the Boltzmann distribution, where n

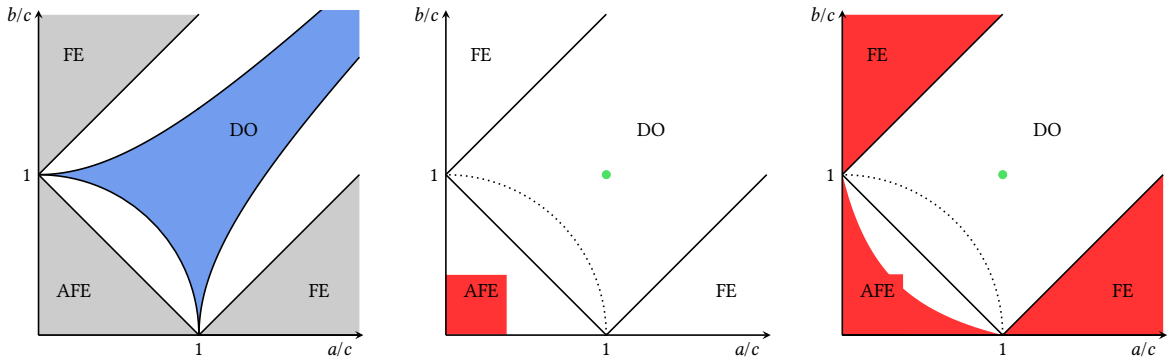


Figure 3.3: Diagram (a) shows the computational complexity of approximating the partition function of the six-vertex model for 4-regular graphs. There exists an FPRAS in the blue subregion of the disordered phase, and there cannot exist an FRPAS in the ferroelectric or antiferroelectric (gray) regions unless **RP** = **NP**. Diagram (b) shows the previously known slow mixing regions of Glauber dynamics in red, and diagram (c) shows the current slow mixing regions. Glauber dynamics is conjectured to be rapidly mixing in all of the disordered phase, but it has only been shown for the uniform distribution indicated by the green point $(1, 1)$.

is the number of vertices in the graph. The Markov chain they use is not Glauber dynamics, but rather a *directed loop algorithm* whose state space is augmented with "near-perfect" configurations that slightly violate the Eulerian orientation constraint. This Markov chain can be understood as gradually reversing a large directed loop in a valid six-vertex configuration, whereas Glauber dynamics is restricted to reversing cycles that form the perimeter of a cell. Cai, Liu, and Lu also showed that an FPRAS for 4-regular graphs cannot exist in the ferroelectric or antiferroelectric regions unless **RP** = **NP** (i.e., the gray regions in Figure 3.3a). Their hardness results use nonplanar 4-regular gadgets to reduce from 3-MIS, the **NP**-hard problem of computing the cardinality of a maximum independent set in a 3-regular graph [GJS74], and therefore so not directly reveal anything about the mixing time of Glauber dynamics for the six-vertex model on regions of $\mathbb{Z}^2$. A dichotomy theorem for the (exact) computability of the partition function of the six-vertex model on 4-regular graphs was also recently proven in [CFX18].

As for the positive results about the mixing time of Glauber dynamics, Luby, Randall, and Sinclair [LRS01] proved rapid mixing of a Markov chain that leads to a fully polynomial almost

uniform sampler for Eulerian orientations on any region of the Cartesian lattice with fixed boundaries (i.e., the unweighted case when $a/c = b/c = 1$). Randall and Tetali [RT00] then used a comparison technique to argue that Glauber dynamics for Eulerian orientations on lattice graphs is rapidly mixing by relating this Markov chain to the Luby-Randall-Sinclair chain. Goldberg, Martin, and Paterson [GMP04] extended their approach to show that Glauber dynamics is rapidly mixing on rectangular lattice regions with free boundary conditions.

Liu [Liu18] gave the first rigorous result showing that Glauber dynamics can be slowly mixing in a subregion of an ordered phase. In particular, Liu showed that local Markov chains subject to free boundary conditions require exponential time to converge to stationarity in the antiferroelectric subregion defined by $\max(a, b) < c/\mu$ (i.e., the red region in Figure 3.3b), where $\mu = 2.6381585...$ is the *connective constant* for self-avoiding walks on the square lattice. We note that the connective constant is defined by the limit $\mu = \lim_{n\to\infty} \gamma_n^{1/n}$, where γ_n is the number of self-avoiding walks of length n on the square lattice. Liu also showed that the directed loop algorithm used in [CLL19] mixes slowly in the same antiferroelectric subregion and for all of the ferroelectric region. This, however, has no bearing on the efficiency of Glauber dynamics in the ferroelectric region. As an aside, we also remark that the partition function is exactly computable for all boundary conditions at the free-fermion point when $\Delta = 0$, or equivalently $a^2 + b^2 = c^2$, via a reduction to domino tilings and a Pfaffian computation [FS06].

3.1.2 Main Results

In this chapter we show that there exist boundary conditions for which Glauber dynamics mixes slowly for the six-vertex model in the ferroelectric and antiferroelectric phases. We start by proving that there are boundary conditions that cause Glauber dynamics to be slow for all Boltzmann weights that lie in the ferroelectric region of the phase diagram, where the mixing time is exponential in the number of vertices in the lattice. This is the first rigorous result for the mixing time of Glauber dynamics in the ferroelectric phase and it gives a complete characterization.

Theorem 3.1.1 (Ferroelectric Phase)**.** *For any* $(a, b, c) \in \mathbb{R}^3_{>0}$ *such that* $a > b + c$ *or* $b > a + c$*, there exist boundary conditions for which Glauber dynamics mixes exponentially slowly on* Λ_n*.*

We note that our approach naturally breaks down at the critical line of the conjectured phase diagram for the mixing time in a way that reveals a trade-off between the energy and entropy of the system. Additionally, our analysis suggests an underlying combinatorial interpretation for the phase transition between the ferroelectric and disordered phases in terms of the adherence strength of intersecting lattice paths and the momentum parameter of correlated random walks.

Our second mixing result builds on the topological obstruction framework developed in [Ran06b] to show that Glauber dynamics with free boundary conditions mixes slowly in most of the antiferroelectric region. Specifically, we generalize the recent antiferroelectric mixing result in [Liu18] with a Peierls argument that uses multivariate generating functions for weighted non-backtracking walks instead of the connectivity constant for (unweighted) self-avoiding walks to better account for the discrepancies in Boltzmann weights.

Theorem 3.1.2 (Antiferroelectric Phase)**.** *For any* $(a, b, c) \in \mathbb{R}^3_{>0}$ *such that* $ac + bc + 3ab < c^2$*, Glauber dynamics mixes exponentially slowly on* Λ_n *with free boundary conditions.*

We illustrate the new regions for which Glauber dynamics can be slowly mixing in Figure 3.3. Observe that our antiferroelectric subregion significantly extends Liu's and pushes towards the conjectured threshold.

3.1.3 Techniques

We take significantly different approaches for our analysis of the ferroelectric and antiferroelectric phases. In the ferroelectric phase, where $a > b + c$ and type-a vertices are preferred to type-b and type-c vertices, we construct boundary conditions that induce polynomially-many paths separated by a critical distance that allows all of the paths to (1) behave independently and (2) simultaneously intersect with their neighbors maximally. (This analysis also covers the case $b > a + c$ by a standard

invariant that shows symmetry in the phase diagram over the line $y = x$.) From here, we analyze the dynamics of a single path in isolation as an escape probability, which eventually allows us to bound the conductance of the Markov chain. The dynamics of a single lattice path is equivalent to that of a *correlated random walk*. In Section 3.4 we present a new tail inequality for correlated random walks that accurately bounds the probability of large deviations from the starting position. We note that decomposing the dynamics of lattice models into one-dimensional random walks has recently been shown to achieve nearly tight bounds for escape probabilities in a different setting [DFGX18].

One of the key technical contributions in this chapter is our analysis of the tail behavior of correlated random walks in Section 3.4. While there is a simple combinatorial expression for the position of a correlated random walk written as a sum of marginals, it is not immediately useful for bounding the displacement from the origin. To achieve an exponentially small tail bound for these walks, we first construct a smooth function that tightly upper bounds the marginals and then optimize this function to analyze the asymptotics of the log of the maximum marginal. Once we obtain an asymptotic equality for the maximum marginal, we can upper bound the deviation of a correlated random walk, and hence the deviation of a lattice path in a configuration. Ultimately, this allows us to show that there exists a balanced cut in the state space that has an exponentially small escape probability, which implies that the Glauber dynamics are slowly mixing.

In the antiferroelectric phase, on the other hand, the weights satisfy $a + b < c$, so type-c vertices are preferred. It follows that there are two (arrow-reversal) symmetric ground states of maximum probability containing only type-c vertices. To move between configurations that agree predominantly with different ground states, the Markov chain must pass through configurations with a large number of type-a or type-b vertices. Using the idea of *fault lines* introduced in [Ran06b], we use *weighted non-backtracking walks* to characterize such configurations and construct a cut set with exponentially small probability mass that separates the ground states.

3.2 Slow Mixing in the Ferroelectric Phase

We start with the ferroelectric phase where $a > b + c$ or $b > a + c$, and we give a conductance-based argument to show that Glauber dynamics can be slowly mixing in the entire ferroelectric region. Specifically, we show that there exist boundary conditions that induce an exponentially small, asymmetric bottleneck in the state space, revealing a natural trade-off between the energy and entropy in the system. Viewing the six-vertex model in the intersecting lattice path interpretation suggests how to plant polynomially-many paths in the grid that can (1) be analyzed independently, while (2) being capable of intersecting maximally. This path independence makes our analysis tractable and allows us to interpret the dynamics of a path as a *correlated random walk*, for which we develop an exponentially small tail bound in Section 3.4. Since conductance governs mixing times (see Section 2.1.4), we show how to relate the expected maximum deviation of a correlated walk to the conductance of the Markov chain and prove slow mixing. In addition to showing slow mixing up to the conjectured threshold, a surprising feature of our argument is that it potentially gives a combinatorial explanation for the phase transition from the ferroelectric to disordered phase. In particular, Lemma 3.2.6 demonstrates how the parameters of the model delicately balance the probability mass of the Markov chain.

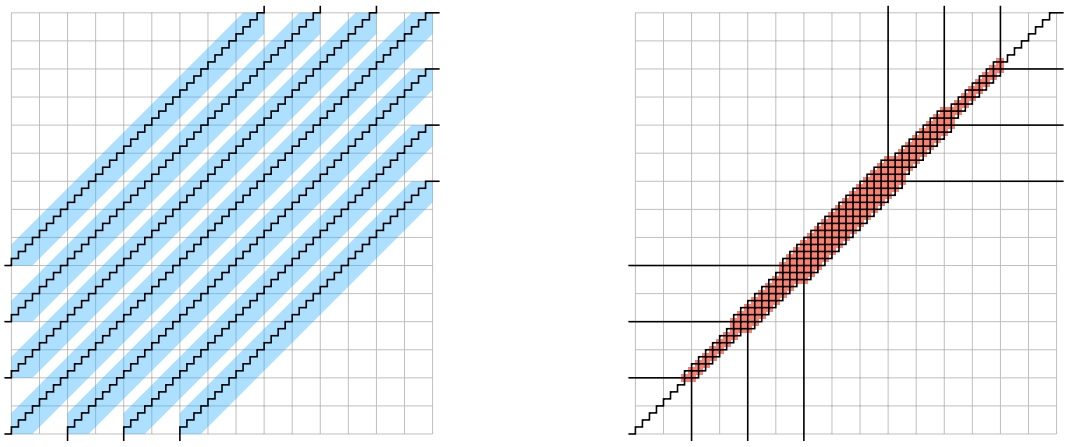
We start by leveraging the invariance of the Gibbs measure and the lattice path interpretation of the six-vertex model to conveniently reparameterize the Boltzmann weights. Recall that for a fixed boundary condition, the invariants of the model [BL09] imply that $a = \sqrt{w_1 w_2}$. Therefore, we set $w_1 = \lambda^2$ and $w_2 = 1$ to ignore empty sites while letting $a = \lambda$. We also set $b = w_3 = w_4 = \mu$ and $c = w_5 = w_6 = 1$ so that the weight of a configuration only comes from straight segments and intersections of neighboring lattice paths.

3.2.1 Constructing the Boundary Conditions and Cut

We begin with a few colloquial definitions for lattice paths that allow us to easily construct the boundary conditions and make arguments about the conductance of the Markov chain. We call a $2n$-step, north-east lattice path γ starting from $(0, 0)$ a *path of length* $2n$, and if the path ends at (n, n) we describe it as *tethered*. If $\gamma = ((0, 0), (x_1, y_1), (x_2, y_2), \dots, (x_{2n}, y_{2n}))$, we define the *deviation* of γ to be $\max_{i=0..2n} \|(x_i, y_i) - (i/2, i/2)\|_1$. Geometrically, path deviation captures the (normalized) maximum perpendicular distance of the path to the line $y = x$. We refer to vertices (x_i, y_i) along the path as *corners* or *straights* depending on whether or not the path turned. If two paths intersect at a vertex we call this site a *cross*. Note that this classifies all vertex types in the six-vertex model.

We consider the following *independent paths boundary condition* for an $n \times n$ six-vertex model for the rest of the section. To construct this boundary condition, we consider its lattice path interpretation. First, place a tethered path γ_0 that enters $(0, 0)$ horizontally and exits (n, n) horizontally. Next, place $2\ell = 2\lfloor n^{1/8} \rfloor$ translated tethered paths of varying length above and below the main diagonal, each separated from its neighbors by distance $d = \lfloor 32n^{3/4} \rfloor$. Specifically, the paths $\gamma_1, \gamma_2, \dots, \gamma_\ell$ below the main diagonal begin at the vertices $(d, 0), (2d, 0), \dots, (\ell d, 0)$ and end at the vertices $(n, n-d), (n, n-2d), \dots, (n, n-\ell d)$, respectively. The paths $\gamma_{-1}, \gamma_{-2}, \dots, \gamma_{-\ell}$ above the main diagonal begin at $(0, d), (0, 2d), \dots, (0, \ell d)$ and end at $(n-d, n), (n-2d, n), \dots, (n-\ell d, n)$. The deviation of a translated tethered path is the deviation of the same path starting at $(0, 0)$. To complete the boundary condition, we force the paths below the main diagonal to enter vertically and exit horizontally. Symmetrically, we force the paths above the main diagonal to enter horizontally and exit vertically. See Figure 3.4a for an illustration of the construction when all paths have small deviation.

Next, we construct an asymmetric cut in the state space induced by this boundary condition in terms of its internal lattice paths. In particular, we analyze a set S of configurations such that



(a) (b)

Figure 3.4: Examples of states with the independent paths boundary condition: (a) is a state in S with the deviation bounds highlighted and (b) is the ground state in the ferroelectric phase.

every path in a configuration has small deviation. Formally, we let

$$S \stackrel{\text{def}}{=} \left\{ x \in \Omega \,:\, \text{the deviation of each path in } x \text{ is less than } 8n^{3/4} \right\}.$$

Observe that by our choice of separation distance $d = \lfloor 32n^{3/4} \rfloor$ and the deviation limit for S, no paths in any configuration of S intersect. It follows that the partition function for S factors into a product of $2\ell + 1$ partition functions, one for each path with bounded deviation. This intuition is useful when analyzing the conductance $\Phi(S)$ as an escape probability from stationarity.

3.2.2 Lattice Paths as Correlated Random Walks

Now we weight the internal paths according to the parameters of the six-vertex model defined in the beginning of Section 3.2. The main result in this subsection is Lemma 3.2.1, which states that random tethered paths are exponentially unlikely to deviate past $\omega(n^{1/2})$, even if drawn from a Boltzmann distribution that favors straights. Start by defining $\Gamma(\mu, n)$ to be the distribution over

tethered paths of length $2n$ with the property that

$$\Pr(\gamma) \propto \mu^{(\#\text{ of straights in } \gamma)}.$$

Lemma 3.2.1. *Let $\mu, \varepsilon > 0$ and $m = o(n)$. For n sufficiently large and $\gamma \sim \Gamma(\mu, n)$, we have*

$$\Pr(\gamma \text{ deviates by at least } 2m) \leq e^{-(1-\varepsilon)\frac{m^2}{\mu n}}.$$

Before giving the proof of Lemma 3.2.1, we first introduce the concept of correlated random walks. Then we present three prerequisite results about correlated random walks and briefly explain their connection to the deviation of biased tethered paths. Our goal here is to show how the supporting lemmas interact prior to the proof of Lemma 3.2.1.

A key idea in our analysis of the ferroelectric phase is the notion of a *correlated random walk*, which generalize a simple symmetric random walk by accounting for momentum. A correlated random walk with momentum parameter $p \in [0, 1]$ starts at the origin and is defined as follows. Let X_1 be a uniform random variable with support $\{-1, 1\}$. For all subsequent steps $i \geq 2$, the direction of the process is correlated with the direction of the previous step and satisfies

$$X_{i+1} = \begin{cases} X_i & \text{with probability } p, \\ -X_i & \text{with probability } 1 - p. \end{cases}$$

We denote the position of the walk at time t by $S_t = \sum_{i=1}^{t} X_i$. It will often be useful to make the change of variables $p = \mu/(1 + \mu)$ when analyzing the six-vertex model, where $\mu > 0$ is the weight of a straight vertex. In many cases this also leads to cleaner expressions. We use the following probability mass function (PMF) for the position of a correlated random walk to develop our new tail inequality (Lemma 3.2.5), which holds for all values of p.

Lemma 3.2.2 ([HF98])**.** *For any $n \geq 1$ and $m \geq 0$, the PMF of a correlated random walk is*

$$\Pr(S_{2n} = 2m) = \begin{cases} \frac{1}{2}p^{2n-1} & \text{if } 2m = 2n, \\ \sum_{k=1}^{n-m} \binom{n+m-1}{k-1}\binom{n-m-1}{k-1}(1-p)^{2k-1}p^{2n-1-2k}\left(\frac{n(1-p)+k(2p-1)}{k}\right) & \text{if } 2m < 2n. \end{cases}$$

Now that we have defined correlated random walks, we proceed by observing that there is a natural measure-preserving bijection between biased tethered paths of length $2n$ and correlated random walks of length $2n$ that return to the origin. To see this, observe that every vertical edge in the tethered path corresponds to a step to the right in the correlated random walk (i.e., $X_i = 1$), and every horizontal edge in the tethered path corresponds to a step to the left in the correlated random walk (i.e., $X_i = -1$). Concretely, for a correlated random walk $(S_0, S_1, \ldots, S_{2n})$ parameterized by $p = \mu/(1+\mu)$, we have

$$\Pr(\gamma \text{ deviates by at least } 2m) = \Pr\left(\max_{i=0..2n}|S_i| \geq 2m \;\middle|\; S_{2n} = 0\right). \tag{3.1}$$

The first prerequisite lemma we present is an asymptotic equality that generalizes the return probability of simple symmetric random walks. This allows us to relax the condition in Equation (3.1) where the correlated random walk must return to the origin, and instead we bound $\Pr(\max_{i=0..2n}|S_i| \geq 2m)$ at the expense of an polynomial factor.

Lemma 3.2.3 ([Gil55])**.** *For any constant $\mu > 0$, the return probability of a correlated random walk is*

$$\Pr(S_{2n} = 0) \sim \frac{1}{\sqrt{\mu\pi n}}.$$

The second result that we need in order to prove Lemma 3.2.1 is that the PMF for correlated random walks is monotone.

Lemma 3.2.4. *For any momentum parameter $p \in (0,1)$ and n sufficiently large, the probability of*

the position of a correlated random walk is monotone. Concretely, for $m \in \{0, 1, \dots, n-1\}$, we have

$$\Pr(S_{2n} = 2m) \geq \Pr(S_{2n} = 2(m+1)).$$

Proof. We consider the cases $m = n-1$ and $m \in \{0, 1, 2, \dots, n-2\}$ separately. Using Lemma 3.2.2, the probability density function for the position of a correlated random walk is

$$\Pr(S_{2n} = 2m) = \begin{cases} \frac{1}{2}p^{2n-1} & \text{if } 2m = 2n, \\ \sum_{k=1}^{n-m} \binom{n+m-1}{k-1}\binom{n-m-1}{k-1}(1-p)^{2k-1}p^{2n-1-2k}\left(\frac{n(1-p)+k(2p-1)}{k}\right) & \text{if } 2m < 2n. \end{cases}$$

If $m = n-1$, then we have the equations

$$\begin{aligned} \Pr(S_{2n} = 2m) &= (1-p)p^{2n-3}(n(1-p) + 2p - 1), \\ \Pr(S_{2n} = 2(m+1)) &= \frac{1}{2}p^{2n-1}. \end{aligned}$$

Therefore, we have $\Pr(S_{2n} = 2m) \geq \Pr(S_{2n} = 2(m+1))$ for all

$$n \geq \frac{1}{1-p} \cdot \left(\frac{p^2}{2(1-p)} + 1 - 2p\right) > 0.$$

Now we assume that $m \in \{0, 1, 2, \dots, n-2\}$. Writing $\Pr(S_{2n} = 2m) - \Pr(S_{2n} = 2(m+1))$ as a difference of sums and matching the corresponding terms, it is instead sufficient to show for all values of $k \in \{1, 2, \dots, n-(m+1)\}$, we have

$$\binom{n+m-1}{k-1}\binom{n-m-1}{k-1} - \binom{n+(m+1)-1}{k-1}\binom{n-(m+1)-1}{k-1} \geq 0.$$

Next, rewrite the binomial coefficients as

$$\binom{n+(m+1)-1}{k-1} = \frac{n+m}{n+m-(k-1)} \cdot \binom{n+m-1}{k-1},$$
$$\binom{n-(m+1)-1}{k-1} = \frac{n-m-k}{n-m-1} \cdot \binom{n-m-1}{k-1}.$$

Therefore, it remains to show that

$$1 - \frac{n+m}{n+m-(k-1)} \cdot \frac{n-m-k}{n-m-1} \geq 0.$$

Since all of the values in $\{n+m, n+m-(k-1), n-m-k, n-m-1\}$ are positive for any choice of m and k, it is equivalent to show that

$$(n+m-(k-1))(n-m-1) \geq (n+m)(n-m-k).$$

Observing that

$$(n+m-(k-1))(n-m-1) - (n+m)(n-m-k) = (2m+1)(k-1) \geq 0$$

completes the proof. □

The third result we need is an upper bound for the position of a correlated random walk. We fully develop this inequality in Section 3.4 by analyzing the asymptotic behavior of the PMF in Lemma 3.2.2. We note that Lemma 3.2.5 shows exactly how the tail behavior of simple symmetric random walks generalizes to correlated random walks as a function of μ.

Lemma 3.2.5. *Let $\mu, \varepsilon > 0$ and $m = o(n)$. For n sufficiently large, a correlated random walk satisfies*

$$\Pr(S_{2n} = 2m) \leq e^{-(1-\varepsilon)\frac{m^2}{\mu n}}.$$

Now that we have established these supporting lemmas, we are prepared to complete the proof of Lemma 3.2.1, which also heavily relies on union bounds and relaxing conditional probabilities.

Proof of Lemma 3.2.1. Using the measure-preserving bijection between tethered paths of length $2n$ and correlated random walks of length $2n$ (Section 3.2.2) along with the definition of conditional probability and Lemma 3.2.3, we have

$$\begin{aligned}\Pr(\gamma \text{ deviates by at least } 2m) &= \Pr\left(\max_{i=0..2n}|S_i| \geq 2m \;\middle|\; S_{2n} = 0\right) \\ &\leq \frac{\Pr(\max_{i=0..2n}|S_i| \geq 2m)}{\Pr(S_{2n} = 0)} \\ &\leq 2\sqrt{\mu\pi n} \cdot \Pr\left(\max_{i=0..2n}|S_i| \geq 2m\right),\end{aligned}$$

where the last inequality uses the definition of asymptotic equality with $\varepsilon = 1/2$. Next, a union bound and the symmetry of correlated random walks imply that

$$\begin{aligned}\Pr\left(\max_{i=0..2n}|S_i| \geq 2m\right) &\leq \Pr\left(\max_{i=0..2n} S_i \geq 2m\right) + \Pr\left(\min_{i=0..2n} S_i \leq -2m\right) \\ &= 2 \cdot \Pr\left(\max_{i=0..2n} S_i \geq 2m\right).\end{aligned}$$

Now we focus on the probability that the maximum position of the walk is at least $2m$. For this event to be true, the walk must reach $2m$ at some time $i \in \{0, 1, 2, \ldots, 2n\}$, so by a union bound,

$$\Pr\left(\max_{i=0..2n} S_i \geq 2m\right) \leq \sum_{i=0}^{2n} \Pr(S_i = 2m) \leq \sum_{i=1}^{n} \Pr(S_{2i} \geq 2m).$$

The second inequality takes into account the parity of the random walk, the fact that if $i = 0$ the walk can only be at position 0, and the relaxed condition that the final position is at least $2m$. Lemma 3.2.4 implies that the distribution is unimodal on its support centered at the origin for sufficiently large n. Moreover, for walks of the same parity with increasing length and a fixed tail

threshold, the probability of the tail is nondecreasing. Combining these two observations, we have

$$\sum_{i=1}^{n} \Pr(S_{2i} \geq 2m) \leq n \cdot \Pr(S_{2n} \geq 2m) \leq n^2 \cdot \Pr(S_{2n} = 2m).$$

Using the chain of previous inequalities and the upper bound for $\Pr(S_{2n} = 2m)$ in Lemma 3.2.5 with the smaller error $\varepsilon/2$, it follows that

$$\Pr(\gamma \text{ deviates by at least } 2m) \leq 4n^2\sqrt{\mu\pi n} \cdot \Pr(S_{2n} = 2m) \leq e^{-(1-\varepsilon)\frac{m^2}{\mu n}},$$

which completes the proof. □

3.2.3 Bounding the Conductance and Mixing Time

Next, we bound the conductance of the Markov chain by viewing $\Phi(S)$ as an escape probability. We start by claiming that $\pi(S) \leq 1/2$ (as required by the definition of conductance) if and only if the parameters are in the ferroelectric phase. Then we use the correspondence between tethered paths and correlated random walks (i.e., Section 3.2.2) to prove that $\Phi(S)$ is exponentially small.

Lemma 3.2.6. *Let $\mu > 0$ and $\lambda > 1 + \mu$ be constants. For n sufficiently large, $\pi(S) \leq 1/2$.*

Proof. We start by upper bounding $\pi(S)$ in terms of the partition function Z. No paths in any state of S deviate by more than $2n^{3/4}$ by the definition of S. Moreover, since adjacent paths are separated by distance $d = \lfloor 32n^{3/4} \rfloor$, no two can intersect (Figure 3.4a). Therefore, it follows that the paths are independent of each other, which is convenient because it allows us to implicitly factor the generating function for configurations in S.

Next, observe that an upper bound for the generating function of any single path is $(1 + \mu)^{2n+1}$. This is true because all paths have length at most $2n$, and we introduce an additional $(1 + \mu)^2$ factor

to account for boundary conditions. Since all the paths are independent and $\ell = \lfloor n^{1/8} \rfloor$, we have

$$\pi(S) \leq \frac{\left((1+\mu)^{2n+1}\right)^{2\ell+1}}{Z} = \frac{(1+\mu)^{4n^{9/8}(1+o(1))}(1+o(1))}{Z}.$$

Now we lower bound the partition function Z of the entire model by considering the weight of the ferroelectric ground state (Figure 3.4b). Recall that we labeled the ℓ paths below the main diagonal path $\gamma_1, \gamma_2, \ldots, \gamma_\ell$ such that γ_ℓ is farthest from the main diagonal. Let $c \leq 10$ be a constant that accounts for subtle misalignments between adjacent paths. It follows that each path γ_k uniquely corresponds to at least $n - (2kd + k + c)$ intersections. Using the last path γ_ℓ as a lower bound for the number of intersections that each path contributes and accounting also for the paths above the main diagonal, it follows that there are at least

$$2\ell(n - (2\ell d + \ell + c)) = 2n^{9/8}(1 - o(1))$$

intersections in the ground state.

Similarly, we bound the number of straights that each path γ_k contributes. Note that we may also need an upper bound for this quantity in order to lower bound the partition function since it is possible that $0 < \mu < 1$. The number of straights in γ_k is $2(kd \pm c)$, and γ_0 has two straights on the boundary. Therefore, the total number of straights in the ground configuration is

$$2\sum_{k=1}^{\ell} 2(kd \pm c) = 64n(1 + o(1)).$$

Since intersections are weighted by λ^2 and straights by μ in our reparameterized model, by considering the ground state and using the previous enumerations, it follows that

$$Z \geq \left(\lambda^2\right)^{2n^{9/8}(1-o(1))} \mu^{64n(1+o(1))}.$$

Combining these inequalities allows us to upper bound the probability mass of the cut $\pi(S)$ by

$$\pi(S) \le \left(\frac{1+\mu}{\lambda}\right)^{4n^{9/8}(1+o(1))} \mu^{-64n(1+o(1))}(1+o(1)).$$

Using the assumption that $\lambda > 1+\mu$, we have $\pi(S) \le 1/2$ for n sufficiently large, as desired. □

Our analysis of the escape probability from S critically relies on the fact that paths in any state $x \in S$ are non-intersecting. Combinatorially, we exploit the factorization of the generating function for states in S as a product of $2\ell+1$ independent path generating functions.

Lemma 3.2.7. *Let $\mu, \varepsilon > 0$ be constants. For n sufficiently large, $\Phi(S) \le e^{-(1-\varepsilon)\mu^{-1}n^{1/2}}$.*

Proof. The conductance $\Phi(S)$ can be understood as the following escape probability. Sample a state $x \in S$ from the stationary distribution π conditioned on $x \in S$, and run the Markov chain from x for one step to get a neighboring state y. The definition of conductance implies that $\Phi(S)$ is the probability that $y \notin S$. Using this interpretation, we can upper bound $\Phi(S)$ by the probability mass of states that are near the boundary of S in the state space, since the process must escape in one step. Therefore, it follows from the independent paths boundary condition and the definition of S that

$$\Phi(S) \le \Pr\Big(\text{there exists a path in } x \text{ deviating by at least } 4n^{3/4} \;\Big|\; x \in S\Big).$$

Next, we use a union bound over the $2\ell+1$ different paths in a configuration and consider the event that a particular path γ_k deviates by at least $4n^{3/4}$. Because all of the paths in S are independent, we only need to consider the behavior of γ_k in isolation. This allows us to rephrase

the conditional event. Relaxing the conditional probability of each term in the sum gives

$$\begin{aligned}\Phi(S) &\le \sum_{k=-\ell}^{\ell} \Pr\big(\gamma_k \text{ deviates by at least } 4n^{3/4} \;\big|\; x \in S\big) \\ &= \sum_{k=-\ell}^{\ell} \Pr\big(\gamma_k \text{ deviates by at least } 4n^{3/4} \;\big|\; \gamma_k \text{ deviates by less than } 8n^{3/4}\big) \\ &\le \sum_{k=-\ell}^{\ell} \frac{\Pr\big(\gamma_k \text{ deviates by at least } 4n^{3/4}\big)}{1 - \Pr\big(\gamma_k \text{ deviates by at least } 8n^{3/4}\big)}.\end{aligned}$$

For large enough n, the length of every path γ_k is in the range $[n, 2n]$ since we eventually have the inequality $n - \ell d \ge n/2$. Therefore, we can apply Lemma 3.2.1 with the error $\varepsilon/2$ to each term and use the universal upper bound

$$\frac{\Pr\big(\gamma_k \text{ deviates by at least } 4n^{3/4}\big)}{1 - \Pr\big(\gamma_k \text{ deviates by at least } 8n^{3/4}\big)} \le \frac{e^{-\left(1-\frac{\varepsilon}{2}\right)\frac{16n^{3/2}}{\mu n}}}{1 - e^{-\left(1-\frac{\varepsilon}{2}\right)\frac{64n^{3/2}}{\mu n}}} \le 2e^{-\left(1-\frac{\varepsilon}{2}\right)\frac{16n^{3/2}}{\mu n}}.$$

It follows from the union bound and previous inequality that the conductance $\Phi(S)$ is bounded by

$$\Phi(S) \le (2\ell + 1) \cdot 2e^{-\left(1-\frac{\varepsilon}{2}\right)\frac{16n^{3/2}}{\mu n}} \le e^{-(1-\varepsilon)\mu^{-1}n^{1/2}},$$

which completes the proof. □

Now that we have constructed a cut in the state space with exponentially small conductance, we can obtain a bound on the mixing time when the probability mass is properly distributed.

Theorem 3.2.8. *Let $\mu, \varepsilon > 0$ and $\lambda > 1 + \mu$. For n sufficiently large, $\tau(1/4) \ge e^{(1-\varepsilon)\mu^{-1}n^{1/2}}$.*

Proof. Since $\pi(S) \le 1/2$ by Lemma 3.2.6, we have $\Phi^* \le \Phi(S)$. The proof follows from Theorem 2.1.5 and the conductance bound in Lemma 3.2.7 with a smaller error $\varepsilon/2$. □

Last, we restate our main theorem and use Theorem 3.2.8 to show that Glauber dynamics for the six-vertex model can be slow mixing for all parameters in the ferroelectric phase.

Theorem 3.1.1 (Ferroelectric Phase)**.** *For any* $(a, b, c) \in \mathbb{R}^3_{>0}$ *such that* $a > b + c$ *or* $b > a + c$*, there exist boundary conditions for which Glauber dynamics mixes exponentially slowly on* Λ_n*.*

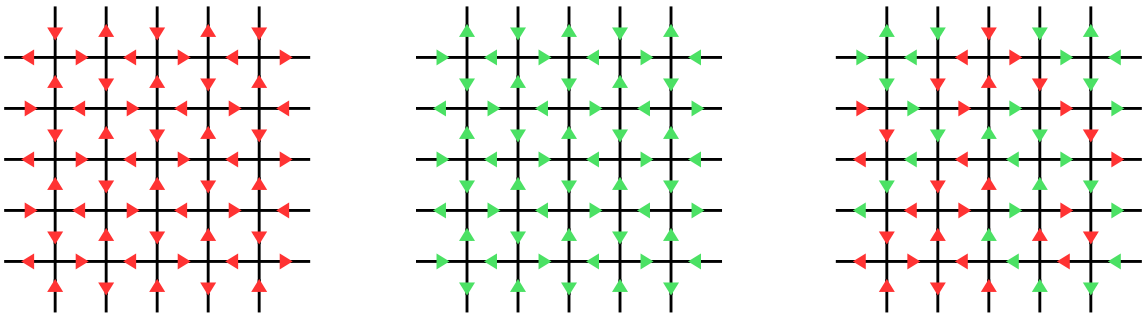
Proof. Without loss of generality, we reparameterized the model so that $a = \lambda$, $b = \mu$, and $c = 1$. Therefore, Glauber dynamics with the independent paths boundary condition is slow mixing if $a > b + c$ by Theorem 3.2.8. Since the rotational invariance of the six-vertex model implies that a and b are interchangeable parameters, this mixing time result also holds in the case $b > a + c$. □

3.3 Slow Mixing in the Antiferroelectric Phase

While Glauber dynamics can be slowly mixing in the ferroelectric phase, we find it is true for substantially different reasons. In the antiferroelectric phase, Boltzmann weights satisfy $a+b < c$, so configurations tend to favor corner (i.e., type-c) vertices. The main insight behind our slow mixing proof is that when c is sufficiently large, the six-vertex model can behave like the low-temperature hardcore model on $\mathbb{Z}^2$ where configurations predominantly agree with one of two ground states. Liu recently formalized this argument in [Liu18] and showed that Glauber dynamics for the six-vertex model with free boundary conditions requires exponential time when $\max(a, b) < \mu c$, where $\mu \leq 2.639$ is the connective constant of self-avoiding walks on the square lattice [GC01]. His proof uses a Peierls argument based on topological obstructions introduced by Randall [Ran06b] in the context of independent sets. In this section, we extend Liu's result to the region depicted in Figure 3.3c by computing a closed-form multivariate generating function that upper bounds the number of self-avoiding walks and better accounts for disparities in their Boltzmann weights induced by the parameters of the six-vertex model.

3.3.1 Topological Obstruction Framework

We start with a recap of the definitions and framework laid out in [Liu18]. There are two ground states in the antiferroelectric phase such that every interior vertex is a corner: x_R (Figure 3.5a)



(a) (b) (c)

Figure 3.5: Edge colorings of (a) the red ground state x_R, (b) the green ground state x_G, and (c) an example configuration with free boundary conditions that does not have a monochromatic cross.

and x_G (Figure 3.5b). These configurations are edge reversals of each other, so for any state $x \in \Omega$ we can color its edges *red* if they are oriented as in x_R or *green* if they are oriented as in x_G. See Figure 3.5c for an example of how a configuration is colored. It follows from case analysis of the six vertex types in Figure 3.1 that the number of red edges incident to any internal vertex is even, and if there are only two red edges then they must be rotationally adjacent to each other. The same property holds for green edges by symmetry. Note that the four edges bounding a cell of the lattice are monochromatic if and only if they are oriented cyclically, and thus reversible by Glauber dynamics. We say that a simple path from a horizontal edge on the left boundary of Λ_n to a horizontal edge on the right boundary is a *red horizontal bridge* if it contains only red edges. We define green horizontal bridges and monochromatic vertical bridges similarly. A configuration has a *red cross* if it contains both a red horizontal bridge and a red vertical bridge. Likewise, we can define a *green cross*. Let $C_R \subseteq \Omega$ be the set of all states with a red cross, and let $C_G \subseteq \Omega$ be the set of all states with a green cross. It follows from Lemma 3.3.1 that $C_R \cap C_G = \emptyset$.

Next, we define the dual lattice L_n to describe configurations in $\Omega \setminus (C_R \cup C_G)$. The vertices of L_n are the centers of the cells in Λ_n, including the cells on the boundary that are partially enclosed, and we connect dual vertices by an edge if their corresponding cells are diagonally adjacent. Note that L_n is a union of two disjoint graphs (Figure 3.6a). For any state $x \in \Omega$ there is a corresponding dual subgraph L_x defined as follows: for each interior vertex v in Λ_n, if v is incident to two red

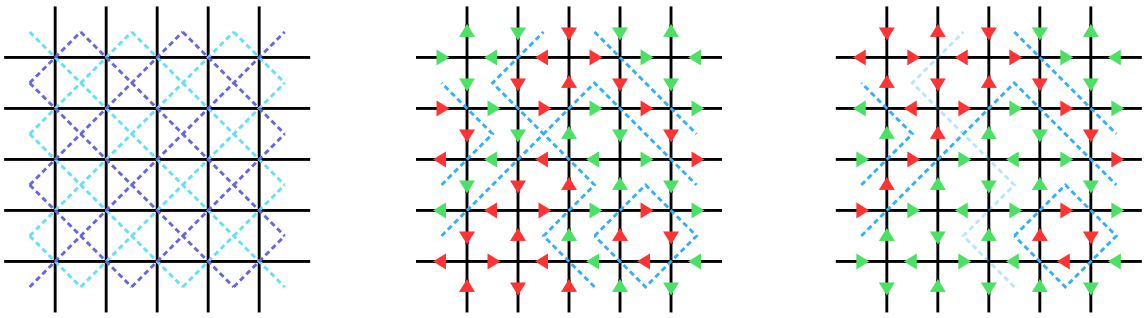


Figure 3.6: Illustrations of (a) the dual lattice L_n as a union of disjoint cyan and purple subgraphs, (b) an example configuration overlaid with its dual graph, and (c) the example under the injective fault line map.

edges and two green edges, then L_x contains the dual edge passing through v that separates the two red edges from the two green edges. This construction is well-defined because the red edges are rotationally adjacent. See Figure 3.6b for an example of a dual configuration. For any $x \in \Omega$, we say that x has a *horizontal fault line* if L_x contains a simple path from a left dual boundary vertex to a right dual boundary vertex. We define horizontal fault lines similarly and let $C_{\text{FL}} \subseteq \Omega$ be the set of all states containing a horizontal or vertical fault line. Fault lines completely separate red and green edges, and hence are topological obstructions that prohibit monochromatic bridges.

Last, we extend the notion of fault lines to *almost fault lines*. We say that $x \in \Omega$ has a horizontal almost fault line if there is a simple path in L_n connecting a left dual boundary vertex to a right dual boundary vertex such that all edges except for one are in L_x. We define vertical almost fault lines similarly and let the set $C_{\text{AFL}} \subseteq \Omega$ denote all states containing an almost fault line. Finally, let $\partial C_{\text{R}} \subseteq \Omega$ denote the set of states not in C_{R} that one move away from C_{R} in the state space according to the Glauber dynamics.

Lemma 3.3.1 ([Liu18]). *We can partition the state space into* $\Omega = C_{\text{R}} \cup C_{\text{FL}} \cup C_{\text{G}}$. *Furthermore, we have* $\partial C_{\text{R}} \subseteq C_{\text{FL}} \cup C_{\text{AFL}}$.

3.3.2 Bounding the Mixing Time with a Peierls Argument

In this subsection we show that $\pi(C_{\mathrm{FL}} \cup C_{\mathrm{AFL}})$ is an exponentially small bottleneck in the state space Ω. The analysis relies on Lemma 3.3.1 and a new multivariate upper bound for weighted self-avoiding walks (Lemma 3.3.2). Our key observation is that when a fault line changes direction, the vertices in its path change from type-a to type-b or vice versa. Therefore, our goal in this subsection is to generalize the trivial 3^{n-1} upper bound for the number of self-avoiding walks by accounting for their changes in direction in aggregate. We achieve this by using generating functions to solve a system of linear recurrence relations.

We start by encoding *non-backtracking walks* that start from the origin and take their first step northward using the characters in $\{\mathsf{S}, \mathsf{L}, \mathsf{R}\}$, representing straight, left, and right steps. For example, the walk SLRSSL corresponds to the sequence $((0,0),(0,1),(-1,1),(-1,2),(-1,3),(-1,4),(-2,4))$. If a fault line is the same shape as SLRSSL up to a rotation about the origin, then there are only two possible sequences of vertex types through which it can pass: $abaaab$ and $babbba$. This follows from the fact that once the first vertex type is determined, only turns in the self-avoiding walk (i.e., the L and R characters) cause the vertex type to switch. We define the weight of a fault line to be the product of the vertex types through which it passes. More generally, we define the weight of a non-backtracking walk that initially passes through a fixed vertex type to be the product of the induced vertex types according to the rule that turns toggle the current type. Formally, we let the function $g_a(\gamma) : \{\mathsf{S}\} \times \{\mathsf{S}, \mathsf{L}, \mathsf{R}\}^{n-1} \longrightarrow \mathbb{R}_{\geq 0}$ denote the weight of a non-backtracking walk γ that starts by crossing a type-a vertex. We define the function $g_b(\gamma)$ similarly and provide the examples $g_a(\mathsf{SLRSSL}) = a^4b^2$ and $g_b(\mathsf{SLRSSL}) = a^2b^4$ for clarity. Last, observe that a sequence of vertex types can have many different walks in its preimage. The non-backtracking walk SRRSSR also maps to $abaaab$ and $babbba$—in fact, there are $2^3 = 8$ such walks in this example since we can interchange L and R characters.

The idea of enumerating the preimages of a binary string corresponding to sequence of vertex

types suggests a recursive approach for computing the sum of weighted non-backtracking walks. This naturally leads to the use of generating functions, so we overload the variables x and y to also denote function arguments. For nonempty binary string $s \in \{0,1\}^n$, let $h(s)$ count the number of pairs of adjacent characters that are not equal and let $|s|$ denote the number of ones in s (e.g., if $s = 010001$ then $h(s) = 3$ and $|s| = 2$). The sum of weighted self-avoiding walks is upper bounded by the sum of weighted non-backtracking walks, so we proceed by analyzing the following function:

$$F_n(x,y) \stackrel{\text{def}}{=} \sum_{\gamma \in \{\mathsf{S}\} \times \{\mathsf{S},\mathsf{L},\mathsf{R}\}^{n-1}} g_x(\gamma) + g_y(\gamma) = \sum_{s \in \{0,1\}^n} 2^{h(s)} x^{|s|} y^{n-|s|}. \tag{3.2}$$

Note that $F_n(1,1) = 2 \cdot 3^{n-1}$ recovers the number of non-backtracking walks that initially cross type-a or type-b vertices.

In the next section, we compute the closed-form solution for $F_n(x,y)$ by diagonalizing a matrix corresponding to the system of recurrence relations, which allows us to accurately quantify the discrepancy between fault lines when the Boltzmann weights a and b differ. For now, we use the following upper bound for $F_n(x,y)$ in our Peierls argument and defer its proof to Section 3.3.3.

Lemma 3.3.2. *Let $F_n(x,y)$ be the generating function for weighted non-backtracking walks defined in Equation* (3.2)*. For any integer $n \geq 1$ and $x, y \in \mathbb{R}_{>0}$, we have*

$$F_n(x,y) \leq 3(x+y)\left(\frac{x+y+\sqrt{x^2+14xy+y^2}}{2}\right)^{n-1}.$$

The first step of our Peierls argument is to upper bound $\pi(C_{\mathrm{FL}} \cup C_{\mathrm{AFL}})$, which then gives us a bound on the conductance and allows us to prove Theorem 3.1.2. We start by defining the subset of antiferroelectric parameters that cause $F_n(a/c, b/c)$ to decrease exponentially fast.

Lemma 3.3.3. *If $(a, b, c) \in \mathbb{R}^3_{>0}$ is antiferroelectric and $3ab + ac + bc < c^2$, then*

$$a + b + \sqrt{a^2 + 14ab + b^2} < 2c.$$

Proof. Let $x = a/c$ and $y = b/c$, and observe that $0 < x < 1$ by the antiferroelectric assumption. It follows from our hypothesis that $y < (1 - x)/(1 + 3x)$. Therefore, we have

$$\begin{aligned} x + y + \sqrt{x^2 + 14xy + y^2} &< x + \frac{1 - x}{1 + 3x} + \sqrt{\frac{x^2(1 + 3x)^2 + 14x(1 - x)(1 + 3x) + (1 - x)^2}{(1 + 3x)^2}} \\ &= \frac{x(1 + 3x) + 1 - x + \sqrt{(3x^2 - 6x - 1)^2}}{1 + 3x} \\ &= \frac{x(1 + 3x) + 1 - x - \left(3x^2 - 6x - 1\right)}{1 + 3x} \\ &= \frac{2(1 + 3x)}{1 + 3x} \\ &= 2, \end{aligned}$$

which completes the proof. □

Lemma 3.3.4. *If $(a, b, c) \in \mathbb{R}^3_{>0}$ is antiferroelectric and $3ab + ac + bc < c^2$, then for Glauber dynamics with free boundary conditions we have*

$$\pi\left(C_{\mathrm{FL}} \cup C_{\mathrm{AFL}}\right) \leq \mathrm{poly}(n)\left(\frac{a + b + \sqrt{a^2 + 14ab + b^2}}{2c}\right)^n.$$

Proof. For any self-avoiding walk γ and dual vertices $s, t \in L_n$ on the boundary, let $\Omega_{\gamma,s,t} \subseteq \Omega$ be the set of states that contain γ as a fault line or an almost fault line such that γ starts at s and ends at t. Without loss of generality, assume that the (almost) fault line is vertical. Reversing the direction of all edges on the left side of γ defines the injective map $f_{\gamma,s,t} : \Omega_{\gamma,s,t} \longrightarrow \Omega \setminus \Omega_{\gamma,s,t}$ such that if γ is a fault line of $x \in \Omega_{\gamma,s,t}$, then the weight of its image $f_{\gamma,s,t}(x)$ is amplified by $c^{|\gamma|}/g_a(\gamma)$ or $c^{|\gamma|}/g_b(\gamma)$. For an example of this injection, see Figure 3.6c. Similarly, if γ is an almost fault line,

decompose γ into subpaths γ_1 and γ_2 separated by a type-c vertex such that γ_1 starts at s and γ_2 ends at t. In this case, the weight of the images of almost fault lines is amplified by a factor of $\min(a,b)/c \cdot c^{|\gamma_1|+|\gamma_2|}/(g_\alpha(\gamma_1)g_\beta(\gamma_2))$ for some $(\alpha,\beta) \in \{a,b\}^2$. Using the fact that $f_{\gamma,s,t}$ is injective and summing over the states containing γ as a fault line and an almost fault line separately gives us

$$\pi\left(\Omega_{\gamma,s,t}\right) \leq \frac{g_a(\gamma)+g_b(\gamma)}{c^{|\gamma|}} + \frac{c}{\min(a,b)} \sum_{\gamma_1+\gamma_2=\gamma} \frac{g_a(\gamma_1)+g_b(\gamma_1)}{c^{|\gamma_1|}} \cdot \frac{g_a(\gamma_2)+g_b(\gamma_2)}{c^{|\gamma_2|}}, \tag{3.3}$$

where the sum is over all $\Theta(|\gamma|)$ decompositions of γ into γ_1 and γ_2.

Equipped with Equation (3.3) and Lemma 3.3.2, we use a union bound over all pairs of terminal vertices (s,t) and fault line lengths ℓ to bound $\pi(C_{\mathrm{FL}} \cup C_{\mathrm{AFL}})$ in terms of the generating function for weighted non-backtracking walks $F_\ell(x,y)$. Since antiferroelectric weights satisfy $3ab+ac+bc < c^2$, it follows from Lemma 3.3.3 that

$$\begin{aligned}
\pi(C_{\mathrm{FL}} \cup C_{\mathrm{AFL}}) &\leq \sum_{(s,t)} \sum_{\ell=n}^{n^2} \left(F_\ell(a/c,b/c) + \frac{c}{\min(a,b)} \sum_{k=0}^{\ell} F_k(a/c,b/c) F_{\ell-k}(a/c,b/c) \right) \\
&\leq \sum_{(s,t)} \sum_{\ell=n}^{n^2} \mathrm{poly}(\ell) \left(\frac{a+b+\sqrt{a^2+14ab+b^2}}{2c} \right)^{\ell} \\
&\leq \mathrm{poly}(n) \left(\frac{a+b+\sqrt{a^2+14ab+b^2}}{2c} \right)^{n}.
\end{aligned}$$

Note that the convolutions in the first inequality generate all *almost* weighted non-backtracking walks. □

Theorem 3.1.2 (Antiferroelectric Phase)**.** *For any $(a,b,c) \in \mathbb{R}^3_{>0}$ such that $ac+bc+3ab < c^2$, Glauber dynamics mixes exponentially slowly on Λ_n with free boundary conditions.*

Proof of Theorem 3.1.2. Let $\Omega_{\mathrm{MIDDLE}} = C_{\mathrm{FL}} \cup C_{\mathrm{AFL}}$, $\Omega_{\mathrm{LEFT}} = C_{\mathrm{R}} \setminus \Omega_{\mathrm{MIDDLE}}$, and $\Omega_{\mathrm{RIGHT}} = C_{\mathrm{G}} \setminus \Omega_{\mathrm{MIDDLE}}$. It follows from Lemma 3.3.1 that $\Omega = \Omega_{\mathrm{LEFT}} \cup \Omega_{\mathrm{MIDDLE}} \cup \Omega_{\mathrm{RIGHT}}$ is a partition with the properties that $\partial\Omega_{\mathrm{LEFT}} \subseteq \Omega_{\mathrm{MIDDLE}}$ and $\pi(\Omega_{\mathrm{LEFT}}) = \pi(\Omega_{\mathrm{RIGHT}})$. Since the partition is symmetric, Lemma 3.3.4 implies that $1/4 \leq \pi(\Omega_{\mathrm{LEFT}}) \leq 1/2$, for n sufficiently large. Therefore, we can upper bound the

conductance by $\Phi^* \le \Phi(\Omega_{\text{LEFT}}) \le 4\pi(\Omega_{\text{MIDDLE}})$. Using Theorem 2.1.5 along with Lemma 3.3.4 and Lemma 3.3.3 gives the desired mixing time bound. □

3.3.3 Weighted Non-Backtracking Walks

In this section we present a closed-form formula for the weighted non-backtracking walks generating function $F_n(x, y)$, and we give the proof of Lemma 3.3.2. We start by decomposing the generating function $F_n(x, y)$ into two sums over disjoint sets of bit strings defined by their final character. Formally, for any $n \ge 1$, let

$$F_{n,0}(x, y) = \sum_{s \in \{0,1\}^{n-1} \times \{0\}} 2^{h(s)} x^{|s|} y^{n-|s|}$$
$$F_{n,1}(x, y) = \sum_{s \in \{0,1\}^{n-1} \times \{1\}} 2^{h(s)} x^{|s|} y^{n-|s|}.$$

First, note that $F_n(x, y) = F_{n,0}(x, y) + F_{n,1}(x, y)$. Second, observe that by recording the final character of the bit strings, we can design a system of linear recurrences to account for the $2^{h(s)}$ term appearing in Equation (3.2), which counts the number of non-backtracking walks that map to a given sequence of vertex types.

Lemma 3.3.5. *For any integer $n \ge 1$ and $x, y \in \mathbb{R}_{>0}$, we have the system of recurrence relations*

$$F_{n+1,0}(x, y) = xF_{n,0}(x, y) + 2xF_{n,1}(x, y)$$
$$F_{n+1,1}(x, y) = 2yF_{n,0}(x, y) + yF_{n,1}(x, y),$$

where the base cases are $F_{1,0}(x, y) = x$ and $F_{1,1}(x, y) = y$.

Proof. This immediately follows from the definitions of the functions $F_{n,0}(x, y)$ and $F_{n,1}(x, y)$. □

Lemma 3.3.6. *For any integer $n \geq 1$ and $x, y \in \mathbb{R}_{>0}$, define the values*

$$\begin{aligned} m &= \sqrt{x^2 + 14xy + y^2} \\ \lambda_1 &= \frac{1}{2}(x + y - m) \\ \lambda_2 &= \frac{1}{2}(x + y + m). \end{aligned}$$

The generating $F_n(x, y)$ can be written in closed-form as

$$F_n(x, y) = \frac{1}{2m}\Big(\big(x^2 + 6xy + y^2 + m(x + y)\big)\lambda_2^{n-1} - \big(x^2 + 6xy + y^2 - m(x + y)\big)\lambda_1^{n-1}\Big).$$

Proof. For brevity, we let $F_{n,0} = F_{n,0}(x, y)$ and $F_{n,1} = F_{n,1}(x, y)$. It follows from Lemma 3.3.5 that

$$\begin{bmatrix} F_{n+1,0} \\ F_{n+1,1} \end{bmatrix} = \begin{bmatrix} x & 2x \\ 2y & y \end{bmatrix} \begin{bmatrix} F_{n,0} \\ F_{n,1} \end{bmatrix}.$$

Next, observe that the recurrence matrix is diagonalizable. In particular, we have

$$A = \begin{bmatrix} x & 2x \\ 2y & y \end{bmatrix} = P\Lambda P^{-1},$$

where

$$P = \frac{1}{4y}\begin{bmatrix} x - y - m & x - y + m \\ 4y & 4y \end{bmatrix} \qquad \Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \qquad P^{-1} = \frac{1}{2m}\begin{bmatrix} -4y & x - y + m \\ 4y & -(x - y - m) \end{bmatrix}.$$

Since the base cases are $F_{n,0} = x$ and $F_{n,1} = y$, it follows that

$$\begin{bmatrix} F_{n,0} \\ F_{n,1} \end{bmatrix} = A^{n-1} \begin{bmatrix} F_{1,0} \\ F_{1,1} \end{bmatrix} = P\Lambda^{n-1}P^{-1} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Using the fact $F_n(x, y) = F_{n,0}(x, y) + F_{n,1}(x, y)$ and simplifying the matrix equation above gives us

$$\begin{aligned} F_n(x, y) &= \frac{1}{8my}\big(y(3x + y + m)(x + 3y + m)\lambda_2^{n-1} - y(3x + y - m)(x + 3y - m)\lambda_1^{n-1}\big) \\ &= \frac{1}{2m}\Big(\big(x^2 + 6xy + y^2 + m(x + y)\big)\lambda_2^{n-1} - \big(x^2 + 6xy + y^2 - m(x + y)\big)\lambda_1^{n-1}\Big), \end{aligned}$$

as desired. □

Lemma 3.3.2. *Let $F_n(x, y)$ be the generating function for weighted non-backtracking walks defined in Equation* (3.2)*. For any integer $n \geq 1$ and $x, y \in \mathbb{R}_{>0}$, we have*

$$F_n(x, y) \leq 3(x + y)\left(\frac{x + y + \sqrt{x^2 + 14xy + y^2}}{2}\right)^{n-1}.$$

Proof. We start by using Lemma 3.3.6 to rewrite the closed-form solution of $F_n(x, y)$ as

$$F_n(x, y) = \frac{1}{2m}\Big(\big(x^2 + 6xy + y^2\big)\big(\lambda_2^{n-1} - \lambda_1^{n-1}\big) + m(x + y)\big(\lambda_2^{n-1} + \lambda_1^{n-1}\big)\Big).$$

Next, we observe that the eigenvalue λ_1 satisfies $\lambda_1 < 0$ and $|\lambda_1| \leq \lambda_2$. Since $(x + y)^2 < m^2$, it follows that $x + y - m = 2\lambda_1 < 0$. Furthermore, we have $2|\lambda_1| \leq |x + y| + |-m| = 2\lambda_2$ by the triangle inequality. Together these two properties imply that

$$\lambda_2^{n-1} - \lambda_1^{n-1} \leq 2\lambda_2^{n-1} \qquad \text{and} \qquad \lambda_2^{n-1} + \lambda_1^{n-1} \leq 2\lambda_2^{n-1}.$$

Therefore, we can upper bound $F_n(x, y)$ by

$$F_n(x, y) \leq \frac{1}{m}\left(x^2 + 6xy + y^2\right)\lambda_2^{n-1} + (x + y)\lambda_2^{n-1}.$$

Since $x^2 + 6xy + y^2 < m^2$, we have the inequalities

$$\frac{1}{m}(x^2 + 6xy + y^2) < \sqrt{x^2 + 6xy + y^2} < \sqrt{(2x + 2y)^2} = 2(x + y).$$

The result follows from the definition of λ_2. □

3.4 Tail Behavior of Correlated Random Walks

In this section we prove Lemma 3.2.5, which gives an exponentially small upper bound for the tail of a correlated random walk as a function of its momentum parameter μ. Our proof builds off of the PMF for the position of a correlated random walk restated below, which is combinatorial in nature and not readily amenable for tail inequalities. Specifically, the probability $\Pr(S_{2n} = 2m)$ is a sum of marginals conditioned on the number of turns that the walk makes [RH81].

Lemma 3.2.2 ([HF98])**.** *For any $n \geq 1$ and $m \geq 0$, the PMF of a correlated random walk is*

$$\Pr(S_{2n} = 2m) = \begin{cases} \frac{1}{2}p^{2n-1} & \text{if } 2m = 2n, \\ \sum_{k=1}^{n-m} \binom{n+m-1}{k-1}\binom{n-m-1}{k-1}(1-p)^{2k-1}p^{2n-1-2k}\left(\frac{n(1-p)+k(2p-1)}{k}\right) & \text{if } 2m < 2n. \end{cases}$$

There are two main ideas in our approach to develop a more useful bound for the position of a correlated random walk $\Pr(S_{2n} = 2m)$. First, we construct a smooth function that upper bounds the marginals as a function of x (a continuation of the number of turns in the walk k), and then we determine its maximum value. Next we show that the log of the maximum value is asymptotically equivalent to $m^2/(\mu n)$ for $m = o(n)$, which gives us desirable bounds for sufficiently large values

of n. We note that our analysis illustrates precisely how correlated random walks generalize simple symmetric random walks and how the momentum parameter μ controls the exponential decay.

3.4.1 Upper Bounding the Marginal Probabilities

We start by using Stirling's approximation to construct a smooth function that upper bounds the marginal terms in the sum of the PMF for correlated random walks. For $x \in (0, n-m)$, let

$$f(x) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } x = 0, \\ \frac{(n+m)^{n+m}}{x^x(n+m-x)^{n+m-x}} \cdot \frac{(n-m)^{n-m}}{x^x(n-m-x)^{n-m-x}} \cdot \mu^{-2x} & \text{if } x \in (0, n-m), \\ \mu^{-2(n-m)} & \text{if } x = n-m. \end{cases} \tag{3.4}$$

It can easily be checked that $f(x)$ is continuous on all of $[0, n-m]$ using the fact that $\lim_{x \to 0} x^x = 1$.

Lemma 3.4.1. *For any integer $m \geq 0$, a correlated random walk satisfies*

$$\Pr(S_{2n} = 2m) \leq \text{poly}(n) \sum_{k=0}^{n-m} \left(\frac{\mu}{1+\mu} \right)^{2n} f(k).$$

Proof. Consider the probability density function for $\Pr(S_{2n} = 2m)$ in Lemma 3.2.2. If $2m = 2n$ the claim is clearly true, so we focus on the other case. We start by bounding the rightmost polynomial term in the sum. For all $n \geq 1$, we have

$$\frac{n(1-p) + k(2p-1)}{k} \leq 2n.$$

Next, we reparameterize the marginals in terms of μ, where $p = \mu/(1+\mu)$, and use a more convenient

upper bound for the binomial coefficients. Observe that

$$\begin{aligned}\Pr(S_{2n} = 2m) &\le 2n \sum_{k=1}^{n-m} \binom{n+m-1}{k-1}\binom{n-m-1}{k-1}\left(\frac{1}{1+\mu}\right)^{2k-1}\left(\frac{\mu}{1+\mu}\right)^{2n-1-2k} \\ &\le \text{poly}(n) \sum_{k=0}^{n-m} \binom{n+m}{k}\binom{n-m}{k}\left(\frac{\mu}{1+\mu}\right)^{2n}\mu^{-2k}.\end{aligned}$$

Stirling's approximation states that for all $n \ge 1$ we have

$$e\left(\frac{n}{e}\right)^n \le n! \le en\left(\frac{n}{e}\right)^n,$$

so we can bound the products of binomial coefficients up to a polynomial factor by

$$\begin{aligned}\binom{n+m}{k}\binom{n-m}{k} &\le \text{poly}(n) \cdot \frac{\left(\frac{n+m}{e}\right)^{n+m}}{\left(\frac{k}{e}\right)^k\left(\frac{n+m-k}{e}\right)^{n+m-k}} \cdot \frac{\left(\frac{n-m}{e}\right)^{n-m}}{\left(\frac{k}{e}\right)^k\left(\frac{n-m-k}{e}\right)^{n-m-k}} \\ &= \text{poly}(n) \cdot \frac{(n+m)^{n+m}}{k^k(n+m-k)^{n+m-k}} \cdot \frac{(n-m)^{n-m}}{k^k(n-m-k)^{n-m-k}}.\end{aligned}$$

The proof follows the definition of $f(x)$ given in Equation (3.4). □

There are polynomially-many marginal terms in the sum of the PMF, so if the maximum term is exponentially small, then the total probability is exponentially small. Since the marginal terms are bounded above by an expression involving $f(x)$, we proceed by maximizing $f(x)$ on its support.

Lemma 3.4.2. *The function $f(x)$ is maximized at the critical point*

$$x^* = \begin{cases} \frac{n^2-m^2}{2n} & \text{if } \mu = 1, \\ \frac{n}{1-\mu^2}\left(1 - \sqrt{\mu^2 + (1-\mu^2)\frac{m^2}{n^2}}\right) & \text{otherwise.} \end{cases}$$

Proof. We start by showing that $f(x)$ is log-concave on $(0, n-m)$, which implies that it is unimodal. It follows that a local maximum of $f(x)$ is a global maximum. Since n and k are fixed as constants

and because the numerator is positive, it is sufficient to show that

$$\begin{aligned} g(x) &= -\log\left(x^x(n+m-x)^{n+m-x} \cdot x^x(n-m-x)^{n-m-x} \cdot \mu^{2x}\right) \\ &= -(2x\log(\mu x) + (n+m-x)\log(n+m-x) + (n-m-x)\log(n-m-x)) \end{aligned}$$

is concave. Observe that the first derivative of $g(x)$ is

$$\begin{aligned} g'(x) &= -2(1+\log(\mu x)) + (1+\log(n+m-x)) + (1+\log(n-m-x)) \\ &= -2\log(\mu x) + \log(n+m-x) + \log(n-m-x), \end{aligned}$$

and the second derivative is

$$g''(x) = -\frac{2}{x} - \frac{1}{n+m-x} - \frac{1}{n-m-x}.$$

Because $g''(x) < 0$ on $(0, n-m)$, the function $f(x)$ is log-concave and hence unimodal.

To identify the critical points of $f(x)$, it suffices to determine where $g'(x) = 0$ since $\log x$ is increasing. Using the previous expression for $g'(x)$, it follows that

$$g'(x) = \log\left[\frac{(n-x)^2 - m^2}{\mu^2 x^2}\right]. \tag{3.5}$$

Therefore, the critical points are the solutions of $(n-x)^2 - m^2 = \mu^2 x^2$, so we have

$$x^* = \begin{cases} \frac{n^2-m^2}{2n} & \text{if } \mu = 1, \\ \frac{n-\sqrt{n^2-(1-\mu^2)(n^2-m^2)}}{1-\mu^2} & \text{otherwise.} \end{cases}$$

It remains and suffices to show that x^* is a local maximum since $f(x)$ is unimodal. Observing that

$$\frac{\partial}{\partial x} \log f(x) = g'(x)$$

and differentiating $f(x) = \exp(\log f(x))$ using the chain rule, the definition of x^* gives

$$\begin{aligned} f''(x^*) &= e^{\log f(x^*)} \left[g''(x^*) + g'(x^*)^2 \right] \\ &= f(x^*) g''(x^*). \end{aligned}$$

We know $f(x^*) > 0$, so $f''(x^*)$ has the same sign as $g''(x^*) < 0$. Therefore, x^* is a local maximum of $f(x)$. Using the continuity of $f(x)$ on $[0, n-m]$ and log-concavity, $f(x^*)$ is a global maximum. □

Remark 3.4.3. It is worth noting that for $m = o(n)$, the asymptotic behavior of the critical point is continuous as a function of $\mu > 0$. In particular, it follows from Lemma 3.4.2 that $x^* \sim n/(1+\mu)$.

3.4.2 Asymptotic Behavior of the Maximum Log Marginal

Now that we have a formula for x^*, and hence an expression for $f(x^*)$, we want to show that

$$\left(\frac{\mu}{1+\mu} \right)^{2n} f(x^*) \leq e^{-n^c},$$

for some constant $c > 0$. Because there are polynomially-many marginals in the sum, this leads to an exponentially small upper bound for $\Pr(S_{2n} = 2m)$. Define the *maximum log marginal* to be

$$h(n) \stackrel{\text{def}}{=} -\log \left[\left(\frac{\mu}{1+\mu} \right)^{2n} f(x^*) \right]. \tag{3.6}$$

Equivalently, we show that $h(n) \geq n^c$ for sufficiently large n using asymptotic equivalences.

Lemma 3.4.4. *The maximum log marginal $h(n)$ can be symmetrically expressed as*

$$h(n) = (n+m)\log\left[\left(\frac{1+\mu}{\mu}\right)\left(1-\frac{x^*}{n+m}\right)\right] + (n-m)\log\left[\left(\frac{1+\mu}{\mu}\right)\left(1-\frac{x^*}{n-m}\right)\right].$$

Proof. Grouping the terms of $h(n)$ by factors of n, m and x^* gives

$$n\log\left[\left(\frac{1+\mu}{\mu}\right)^2\frac{(n-x^*)^2-m^2}{(n+m)(n-m)}\right] + m\log\left[\frac{(n-m)(n+m-x^*)}{(n+m)(n-m-x^*)}\right] + x^*\log\left[\frac{(\mu x^*)^2}{(n-x^*)^2-m^2}\right].$$

Using Equation (3.5), observe that the last term is

$$x^*\log\left[\frac{(\mu x^*)^2}{(n-x^*)^2-m^2}\right] = -x^*g'(x^*) = 0.$$

The proof follows by grouping the terms of the desired expression by factors of n and m. □

The following lemma is the crux of our argument, as it presents an asymptotic equality for the maximum log marginal in the PMF for correlated random walks. We remark that we attempted to bound this quantity directly using Taylor expansions instead of an asymptotic equivalence, and while this seems possible, the expressions are unruly. Our asymptotic equivalence demonstrates that second derivative information is needed, which makes the earlier approach even more unmanageable.

Lemma 3.4.5. *For any $\mu > 0$ and $m = o(n)$, the maximum log marginal satisfies $h(n) \sim m^2/(\mu n)$.*

Proof. The proof is by case analysis for μ. In both cases we analyze $h(n)$ as expressed in Lemma 3.4.4, consider a change of variables, and use L'Hospital's rule twice. In the first case, we assume $\mu = 1$. The value of x^* in Lemma 3.4.2 gives us

$$1-\frac{x^*}{n+m} = \frac{2n(n+m)-\left(n^2-m^2\right)}{2n(n+m)} = \frac{n+m}{2n}$$
$$1-\frac{x^*}{n-m} = \frac{2n(n-m)-\left(n^2-m^2\right)}{2n(n-m)} = \frac{n-m}{2n}.$$

It follows that $h(n)$ can be simplified as

$$\begin{aligned} h(n) &= n \log\left[\left(\frac{1+\mu}{\mu}\right)^2 \left(\frac{n^2 - m^2}{4n^2}\right)\right] + m \log\left(\frac{n+m}{n-m}\right) \\ &= n \log\left(1 - \frac{m^2}{n^2}\right) + m \log\left(1 + \frac{2m}{n-m}\right). \end{aligned}$$

To show $h(n) \sim m^2/n$, by the definition of asymptotic equivalence we need to prove that

$$\lim_{n\to\infty} \frac{n \log\left(1 - \frac{m^2}{n^2}\right) + m \log\left(1 + \frac{2m}{n-m}\right)}{\frac{m^2}{n}} = 1.$$

Make the change of variables $y = m/n$. Since $m = o(n)$, this is equivalent to showing

$$\lim_{y\to 0} \frac{\log\left(1 - y^2\right) + y \log\left(1 + \frac{2y}{1-y}\right)}{y^2} = 1.$$

Using L'Hospital's rule twice with the derivatives

$$\begin{aligned} \frac{\partial}{\partial y}\left[\log\left(1 - y^2\right) + y \log\left(1 + \frac{2y}{1-y}\right)\right] &= \log\left(-\frac{y+1}{y-1}\right) \\ \frac{\partial^2}{\partial y^2}\left[\log\left(1 - y^2\right) + y \log\left(1 + \frac{2y}{1-y}\right)\right] &= \frac{2}{1-y^2}, \end{aligned}$$

it follows that

$$\lim_{y\to 0} \frac{\log\left(1 - y^2\right) + y \log\left(1 + \frac{2y}{1-y}\right)}{y^2} = \lim_{y\to 0} \frac{\log\left(-\frac{y+1}{y-1}\right)}{2y} = \lim_{y\to 0} \frac{\frac{2}{1-y^2}}{2} = 1.$$

This completes the proof for $\mu = 1$.

The case when $\mu \neq 1$ is analogous but messier. Making the same change of variables $y = m/n$,

it is equivalent to show that

$$(1+y)\log\left[\left(\frac{1+\mu}{\mu}\right)\left(1-\frac{1}{1-\mu^2}\cdot\frac{1}{1+y}\cdot\left(1-\sqrt{\mu^2+(1-\mu^2)y^2}\right)\right)\right]$$
$$+(1-y)\log\left[\left(\frac{1+\mu}{\mu}\right)\left(1-\frac{1}{1-\mu^2}\cdot\frac{1}{1-y}\cdot\left(1-\sqrt{\mu^2+(1-\mu^2)y^2}\right)\right)\right]\sim\mu^{-1}y^2, \qquad (3.7)$$

because the value of x^* for $\mu\neq 1$ in Lemma 3.4.2 gives us

$$1-\frac{x^*}{n+m}=1-\frac{1}{n+m}\cdot\frac{n}{1-\mu^2}\cdot\left(1-\sqrt{\mu^2+(1-\mu^2)\frac{m^2}{n^2}}\right).$$

Denoting the left-hand side of Equation (3.7) by $g(y)$, one can verify that the first two derivatives of $g(y)$ are

$$g'(y)=\log\left(\frac{\mu^2-\sqrt{\mu^2-\mu^2y^2+y^2}+\left(\mu^2-1\right)y}{(\mu-1)\mu(y+1)}\right)-\log\left(\frac{-\mu^2+\sqrt{\mu^2-\mu^2y^2+y^2}+\left(\mu^2-1\right)y}{(\mu-1)\mu(y-1)}\right)$$
$$g''(y)=\frac{2}{(1+y)(1-y)\sqrt{y^2-\mu^2(y^2-1)}}.$$

Observing that $g(0)=g'(0)=0$ due to convenient cancellations and using L'Hospital's rule twice,

$$\lim_{y\to 0}\frac{g(y)}{\mu^{-1}y^2}=\lim_{y\to 0}\frac{g'(y)}{2\mu^{-1}y}=\lim_{y\to 0}\frac{2}{(1+y)(1-y)\sqrt{y^2-\mu^2(y^2-1)}}\cdot\frac{\mu}{2}=1.$$

This completes the proof for all cases of μ. □

Lemma 3.2.5. *Let $\mu,\varepsilon>0$ and $m=o(n)$. For n sufficiently large, a correlated random walk satisfies*

$$\Pr(S_{2n}=2m)\leq e^{-(1-\varepsilon)\frac{m^2}{\mu n}}.$$

Proof. For n sufficiently large, the asymptotic equality for $h(n)$ in Lemma 3.4.5 gives us

$$h(n) \geq \left(1 - \frac{\varepsilon}{2}\right)\frac{m^2}{\mu n}.$$

It follows from our construction of $f(x)$ and the definition of the maximum log marginal that

$$\begin{aligned}
\Pr(S_{2n} = 2m) &\leq \text{poly}(n) \cdot \left(\frac{\mu}{1+\mu}\right)^{2n} f(x^*) \\
&= \text{poly}(n) \cdot e^{-h(n)} \\
&\leq \text{poly}(n) \cdot e^{-\left(1-\frac{\varepsilon}{2}\right)\frac{m^2}{\mu n}} \\
&\leq e^{-(1-\varepsilon)\frac{m^2}{\mu n}},
\end{aligned}$$

as desired. □

3.5 Summary and Discussion

We have made significant progress in this chapter towards rigorously establishing the conjectured slow regions of the phase diagram for the six-vertex model. In particular, we prove that there exist boundary conditions for which Glauber dynamics requires exponential convergence time for the entire ferroelectric region and most of the antiferroelectric region. Furthermore, our proofs demonstrate why sharp boundaries exist between the ferroelectric phase and the disordered phase, where Glauber dynamics is believed to transition to polynomial-time convergence. We have not fully characterized the antiferroelectric phase, but our improvement over the best previous bounds in [Liu18] cover a significantly larger part of the region.

Our arguments for the slow mixing of Glauber dynamics completely break down in the disordered phase, as expected, but there has not been any rigorous work showing that in this region of the phase diagram we have fast convergence. The single exception is the unweighted case when we have $a = b = c$, which corresponds to Eulerian orientations of the lattice region. This

was shown to converge in polynomial time for all boundary conditions [RT00, LRS01, GMP04]. The approaches in these works are inherently combinatorial, and it seems that generalizing them to weighted cases will require significantly different ideas. Lastly, we emphasize that our proofs of slow mixing rely on new techniques for analyzing lattice models, which include the exponentially small tail inequality for correlated random walks developed in Section 3.4 and the closed-form generating function for weighted non-backtracking walks derived in Section 3.3.

CHAPTER 4

NEARLY TIGHT BOUNDS FOR SANDPILE TRANSIENCE ON THE GRID

In this chapter we prove nearly tight bounds for the *transience class* of the *Abelian sandpile model* on the two-dimensional grid, closing an open problem of Babai and Gorodevsky [BG07, CV12]. The transience class of the Abelian sandpile model quantifies how long it takes for the system to reach its steady-state behavior, starting from a worst-case configuration. Our main results in this chapter show that the transience class of the Abelian sandpile model on the $n \times n$ grid has an upper bound of $O(n^4 \log^4 n)$ and a lower bound of $\Omega(n^4)$. We remark that the notion of convergence to equilibrium here differs from that of the stationary distribution of ergodic Markov chains.

4.1 Introduction

The Abelian sandpile model is the canonical dynamical system used to study *self-organized criticality*. In their seminal paper, Bak, Tang, and Wiesenfeld [BTW87] proposed the idea of self-organized criticality to explain several ubiquitous patterns in nature typically viewed as complex phenomena, such as catastrophic events occurring without any triggering mechanism, the fractal behavior of mountain landscapes and coastal lines, and the presence of pink noise in electrical networks and stellar luminosity. Since their discovery, self-organized criticality has been observed in an abundance of disparate scientific fields [Bak96, Wat+16], including condensed matter theory [WWAM06], economics [SW94, BPR15], epidemiology [SMM14], evolutionary biology [Phi14], high-energy astrophysics [MTN94, Asc11], materials science [RAM09], neuroscience [LHG07, Bro+16], statistical physics [Man91, Dha06], seismology [SS89], and sociology [KG09]. A stochastic process is a self-organized critical system if it naturally evolves to highly imbalanced critical states where slight local disturbances can completely alter the current state. For example, when pouring

grains of sand onto a table, the pile initially grows in a predictable way, but as it becomes steeper and more unstable, dropping a single grain can spontaneously cause an avalanche that affects the entire pile. Self-organized criticality differs from the critical point of a phase transition in statistical physics, because a self-organizing system does not rely on tuning an external parameter. Instead, it is insensitive to all parameters of the model and simply requires time to reach criticality, which is known as the transient period. Natural events empirically operate at a critical point between order and chaos, thus justifying our study of self-organized criticality.

Dhar [Dha90] developed the Abelian sandpile model on finite directed graphs with a sink vertex to further understand self-organized criticality. The Abelian sandpile model, also known as a chip-firing game [BLS91], on a graph with a sink is defined as follows. In each iteration a grain of sand is added to a non-sink vertex of the graph. While any non-sink vertex v contains at least $\deg(v)$ grains of sand, a grain is transferred from v to each of its neighbors. This is known as a *toppling*. When no vertex can be toppled, the state is *stable* and the iteration ends. The sink absorbs and destroys grains, and the presence of a sink guarantees that every toppling procedure eventually stabilizes. An important property of the Abelian sandpile model is that the order in which vertices topple does not affect the stable state. Therefore, as the process evolves it produces a sequence of stable states. From the theory of Markov chains, we say that a stable state is *recurrent* if it can be revisited; otherwise it is *transient*.

In the self-organized critical state of the Abelian sandpile model on a graph with a sink, transient states have zero probability and recurrent states occur with equal probability [Dha90]. As a result, recurrent configurations model the steady-state behavior of the system. Therefore, the natural algorithmic question to ask about self-organized criticality for the Abelian sandpile model is:

Question 4.1.1. *How long in the worst case does it take for the process to reach its steady-state behavior or, equivalently, a recurrent state?*

Starting with an empty configuration, if the vertex that receives the grain of sand is chosen

uniformly at random in each step, Babai and Gorodezky [BG07] give a simple solution that is polynomial in the number of edges of the graph using a coupon collector argument. In the worst case, however, an adversary can choose where to place the grain of sand in each iteration. Babai and Gorodezky analyze the *transience class* of the model to understand its worst-case behavior, which is defined as the maximum number of grains that can be added to the empty configuration before the configuration necessarily becomes recurrent. An upper bound for the transience class of a model is an upper bound for the time needed to enter self-organized criticality.

4.1.1 Main Results

We give the first nearly tight bounds (up to polylogarithmic factors) for the transience class of the Abelian sandpile model on the $n \times n$ grid with all boundary vertices connected to the sink. This model was first studied in depth by Dhar, Ruelle, Sen, and Verma [DRSV95], and it has since been the most extensively studied Abelian sandpile model due to its role in algebraic graph theory, theoretical computer science, and statistical physics. Babai and Gorodezky [BG07] initially established that the transience class of the grid is polynomially bounded by $O(n^{30})$, which was unexpected because there are graphs akin to the grid with exponential transience classes. Choure and Vishwanathan [CV12] improved the upper bound for the transience class of the grid to $O(n^7)$ and gave a lower bound of $\Omega(n^3)$ by viewing the graph as an electrical network and relating the Abelian sandpile model to random walks on the underlying graph. Moreover, they conjectured that the transience class of the grid is $O(n^4)$, which we answer nearly affirmatively.

Theorem 4.1.2. *The transience class of the Abelian sandpile model on the $n \times n$ grid is $O(n^4 \log^4 n)$.*

Theorem 4.1.3. *The transience class of the Abelian sandpile model on the $n \times n$ grid is $\Omega(n^4)$.*

Our results establish how fast the system reaches its steady-state behavior in the adversarial case, and they corroborate empirical observations about natural processes exhibiting self-organized

criticality. Our analysis directly generalizes to higher-dimensional cubic hypergrids and gives the following result.

Theorem 4.1.4. *For any integer $d \geq 2$, the transience class of the Abelian sandpile model on the d-dimensional cubic hypergrid with n^d vertices is $O(n^{3d-2} \log^{d+2} n)$ and $\Omega(n^{3d-2})$.*

In addition to addressing the main open problem in [BG07] and [CV12], we begin to shed light on Babai and Gorodezky's inquiry about sequences of graphs that exhibit polynomially bounded transience classes. Specifically, for cubic hypergrids (a family of locally finite graphs with high symmetry) we quantify how the transience class grows as a function of the size and local degree of the graph. When viewed through the lens of graph connectivity, such transience class bounds are surprising because grids have low algebraic connectivity, yet we are able to make global structural arguments using only the fact that grids have low maximum effective resistance when viewed as electrical networks. By doing this, we avoid spectral analysis of the grid and evade the main obstacle in Choure and Vishwanathan's analysis. Our techniques suggest that low effective resistance captures a different but similar phenomenon to high conductance and high edge expansion for stochastic processes on graphs. This distinction between the role of a graph's effective resistance and conductance could be an important step forward for building a theory for discrete diffusion processes analogous to the mixing time of Markov chains. We also believe our results have close connections to randomized, distributed optimization algorithms for flow and decomposition problems [BMV12, Bec+13, Meh13, SV16a, SV16b, SV16c, Chu+18], where the dynamics of self-adjusting sandpiles (a Physarum slime mold in their model) are governed by electrical flows and resistances.

4.1.2 Techniques

Our approach is motivated by the method of Choure and Vishwanathan [CV12] for bounding the transience class of the Abelian sandpile model on graphs using electrical potential theory and the

analysis of random walks. Viewing the graph as an electrical network with a voltage source at some vertex and a grounded sink, we give more accurate voltage estimates by carefully considering the geometry of the grid. We use several lines of symmetry to compare escape probabilities of random walks with different initial positions, resulting in a new technique for comparing vertex potentials. These geometric arguments can likely be generalized to other lattice-based graphs. As a result, we get empirically tight inequalities for the sum of all vertex potentials in the grid and the voltage drop between opposite corners of the network.

For many of our voltage bounds, we interpret a vertex potential as an escape probability and decouple the corresponding two-dimensional random walks on the grid into independent one-dimensional random walks on a path graph. Decoupling is the standout technique in this chapter because it allows us to apply classical results about simple symmetric random walks on $\mathbb{Z}$ (for example, the reflection principle), which we extend as needed using conditional probability arguments. By reducing from two-dimensional random walks to one-dimensional walks, we utilize standard probabilistic tools including Stirling's approximation, Chernoff bounds, and the negative binomial distribution. Since we consider many different kinds of events in our analysis, Section 4.5 is an extensive collection of probability inequalities for symmetric t-step random walks on $\mathbb{Z}$ with various boundary conditions. We noticed that some of these inequalities are directly related to problems in enumerative combinatorics without closed-form solutions [ES77].

Lastly, we leverage well-known results about effective resistances of the $n \times n$ grid when viewed as an electrical network. We follow Choure and Vishwanathan in using the potential reciprocity theorem to swap the voltage source with any other non-sink vertex, but we use this theorem repeatedly with the fact that the effective resistance between any non-sink vertex and the sink is bounded between a constant and $O(\log n)$. This approach enables us to analyze tractable one-dimensional random walk problems at the expense of polylogarithmic factors.

4.2 Preliminaries

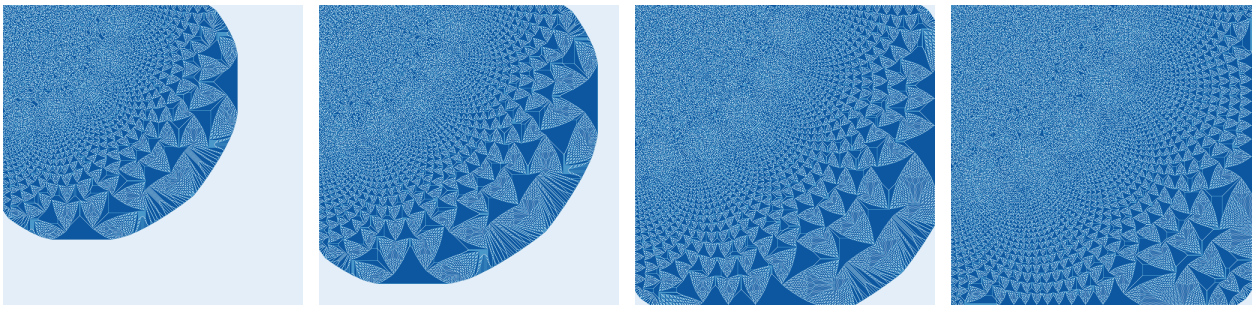
4.2.1 Abelian Sandpile Model

Let $G = (V, E)$ be an undirected multigraph. Throughout this chapter, all of the graphs we consider have a sink vertex denoted by v_{sink}. The *Abelian sandpile model* is a dynamical system on a graph G used to study the phenomenon of self-organized criticality. A *configuration* σ on G in the Abelian sandpile model is a vector of nonnegative integers indexed by the non-sink vertices such that $\sigma(v)$ denotes the number of grains of sand on vertex v. We say that a configuration is *stable* if $\sigma(v) < \deg(v)$ for all non-sink vertices and *unstable* otherwise. An unstable configuration σ moves towards stabilization by selecting a vertex v such that $\sigma(v) \geq \deg(v)$ and sending one grain of sand from v to each of its neighboring vertices. This event is called a *toppling* of v, and it creates a new configuration σ' such that $\sigma'(v) = \sigma(v) - \deg(v)$, $\sigma'(u) = \sigma(u) + 1$ for all vertices u adjacent to v, and $\sigma'(u) = \sigma(u)$ for all remaining vertices. This procedure eventually reaches a stable state because G has a sink. Moreover, the order in which vertices topple does not affect the final stable state. The initial configuration of the Abelian sandpile model is typically the zero vector, and in each iteration a grain of sand is placed on a vertex (chosen either deterministically or uniformly at random). The system evolves by stabilizing the configuration and then receiving another grain of sand.

A stable configuration σ is *recurrent* if the process can eventually return to σ. Any state that is not recurrent is *transient*. Note that once the system enters a recurrent state, it can never visit a transient state. Babai and Gorodezky [BG07] introduced the following notion to upper bound on the number of steps for the Abelian sandpile model to reach self-organized criticality.

Definition 4.2.1. The *transience class* of the Abelian sandpile model of G is the maximum number of grains that can be added to the empty configuration before the configuration necessarily becomes recurrent. We denote this quantity by $\text{tcl}(G)$.

In Section 4.3 we illustrate the transient configurations in the *transient period* of the Abelian



(a) (b) (c) (d)

Figure 4.1: Configurations of the Abelian sandpile model on the 500×500 grid during its transience period after placing (a) 10^{10}, (b) $2 \cdot 10^{10}$, (c) $4 \cdot 10^{10}$, and (d) $8 \cdot 10^{10}$ grains of sand on the vertex $(1, 1)$.

sandpile model as it advances towards its critical state. We specifically show in this chapter that by repeatedly placing grains of sand in the top-left corner of the grid, we maximize the length of the transience period up to a polylogarithmic factor.

In earlier related works, Björner, Lovász, and Shor [BLS91] studied a variant of this process without a sink and characterized the conditions needed for stabilization to terminate. They also related the spectrum of the underlying graph to the rate at which the system converges. In the model we study, an observation by Dhar [Dha90] and Kirchoff's theorem show that the stable recurrent states of the system are in bijection with the spanning trees of G. Choure and Vishwanathan [CV12] show that if every vertex in a configuration has toppled then the configuration is necessarily recurrent, which we use to bound the transience class. The Abelian sandpile model also has broad applications to algorithms and statistical physics, including a direct relation to the q-state Potts model and Markov chain Monte Carlo algorithms for sampling random spanning trees [Dha90, Wil10, JLP19, BCFR17, RS17]. For a comprehensive survey on the Abelian sandpile model, see [Hol+08].

4.2.2 Random Walks on Graphs

A walk w on G is a sequence of vertices $w^{(0)}, w^{(1)}, \ldots, w^{(t_{\max})}$ such that every $w^{(t+1)}$ is a neighbor of $w^{(t)}$. We let $t_{\max} = |w|$ denote the number of steps in the walk. A random walk is a process

that begins at vertex $w^{(0)}$, and at each time step t transitions from $w^{(t)}$ to $w^{(t+1)}$ such that $w^{(t+1)}$ is chosen uniformly at random from the neighbors of $w^{(t)}$. Note that this definition naturally extends to walking on a multigraph. We consider walks that continue until reaching a set of sink vertices, so it will be convenient for our analysis to formally define these families of walks.

Definition 4.2.2. For any set of starting vertices S and terminating vertices T in the graph G, let

$$\mathcal{W}(S \longrightarrow T) \stackrel{\text{def}}{=} \left\{ w \,:\, w^{(0)} \in S,\, w^{(i)} \notin T \cup \{v_{\text{sink}}\} \text{ for } 0 \le i < |w|, \text{ and } w^{(|w|)} \in T \right\}$$

be the set of finite walks from S to T.

Observe that this definition allows for walks w of length 0 if we have $w^{(0)} \in S \cap T$. Throughout the chapter, it will be convenient to consider random walks from one vertex u to another vertex v or the pair $\{v, v_{\text{sink}}\}$. We denote these cases by the notation $\mathcal{W}(u \longrightarrow v) = \mathcal{W}(\{u\} \longrightarrow \{v\})$. If walks on multiple graphs are being considered, we use $\mathcal{W}^G(u \longrightarrow v)$ to denote the underlying graph. Lastly, we consider the set of nonterminating walks in our analysis, so it will be useful to define

$$\mathcal{W}(S) \stackrel{\text{def}}{=} \left\{ w \in \prod_{i=0}^{\infty} V \,:\, w^{(0)} \in S \text{ and } w^{(i)} \neq v_{\text{sink}} \text{ for any } i \ge 0 \right\},$$

which is the set of infinite walks from S. An analogous definition follows when $S = \{u\}$.

The focus of our study is the $n \times n$ grid graph, denoted by $\textsc{Grid}_n$. Similar to previous works, we do not follow the usual graph-theoretic convention of using n to denote the vertex count. Start by denoting $[n] = \{1, 2, \ldots, n\}$. We formally define the one-dimensional projection of $\textsc{Grid}_n$ to be $\textsc{Path}_n$, which has the vertex set $[n] \cup \{v_{\text{sink}}\}$ and edges between i and $i+1$ for every $i \in [n-1]$, as well as two edges connecting v_{sink} to 1 and n. Thus, v_{sink} can be viewed as the set $\{0, n+1\}$. If we remove the sink (which can be thought of as letting $v_{\text{sink}} = \pm\infty$) then the resulting graph is the one-dimensional line with vertices $i \in \mathbb{Z}$ and edges between every pair $(i, i+1)$. We denote this graph by $\textsc{Line}$ and use the indices i, j, and k to represent its vertices. Analyzing random walks on

LINE is critical to our analysis. It will be useful to record the minimum and maximum position of truncated t-step walks, so we define the following functions.

Definition 4.2.3. For an initial position $s \in \mathbb{Z}$ and walk $w \in \mathcal{W}(s)$ on LINE, let the t-step minimum and maximum positions be

$$\min_{\leq t}(w) \stackrel{\text{def}}{=} \min\left(w^{(0)}, w^{(1)}, \dots, w^{(t)}\right)$$

and

$$\max_{\leq t}(w) \stackrel{\text{def}}{=} \max\left(w^{(0)}, w^{(1)}, \dots, w^{(t)}\right).$$

We construct GRID_n similarly. Its vertices are $[n]^2 \cup \{v_{\text{sink}}\}$, and its edges connect any pair of vertices that differ in one coordinate. Vertices on the boundary have edges connected to v_{sink} so that every non-sink vertex has degree 4. With this definition of GRID_n, each corner vertex has two edges to v_{sink} and non-corner vertices on the boundary share one edge with v_{sink}. Since all vertices correspond to pairs of coordinates, we use the vector notation $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2)$ to denote coordinates on the grid, as it easily extends to higher dimensions. Throughout the chapter, boldfaced variables denote vectors. A t-step random walk on GRID_n naturally induces a $(t_{\max} + 1) \times 2$ matrix. We can decouple such a walk w into its horizontal and vertical components, using the notation w_1 for the change in position of the first coordinate and w_2 for the change in position of the second coordinate. In general we use the notation w_i to index into one of the dimensions $i \in [d]$ of a d-dimensional walk. We do not record duplicate positions when the walk takes a step in a dimension different than i, so we have $|w| = |w_1| + |w_2| - 1$ when $d = 2$ since the initial vertex is present in both w_1 and w_2.

4.2.3 Electrical Networks

Vertex potentials are central to our analysis. They have close connections with electrical voltages and belong to the class of harmonic functions [DS84]. We analyze their relation to the transience

class of general graphs. For any non-sink vertex u, we can define a unique potential function π_u such that $\pi_u(u) = 1$, $\pi_u(v_{\text{sink}}) = 0$, and for all other vertices $v \in V \setminus \{u, v_{\text{sink}}\}$ we have

$$\pi_u(v) = \frac{1}{\deg(v)} \sum_{x \in N(v)} \pi_u(x),$$

where the sum is over the neighbors of v. Thus, $\pi_u(v)$ denotes the potential at v when the boundary conditions are set to 1 at u and 0 at the sink. Since we analyze potential vectors in both $\textsc{Path}_n$ and $\textsc{Grid}_n$, we use superscripts to denote the graph when context is unclear.

Choure and Vishwanathan showed that we can give upper and lower bounds on the transience class using potentials, which we rephrase in the following theorem.

Theorem 4.2.4 ([CV12])**.** *If G is a graph such that the degree of every non-sink vertex is bounded by a constant, then*

$$\text{tcl}(G) = O\left(\max_{u,v \in V \setminus \{v_{\text{sink}}\}} \left(\sum_{x \in V} \pi_u(x)\right) \pi_u(v)^{-1}\right)$$

and

$$\text{tcl}(G) = \Omega\left(\max_{u,v \in V \setminus \{v_{\text{sink}}\}} \pi_u(v)^{-1}\right).$$

All non-sink vertices have degree 4, so we can apply Theorem 4.2.4 to $\textsc{Grid}_n$.

The following combinatorial interpretations of potentials as random walks is fundamental to our investigation of the transience class of $\textsc{Grid}_n$. Note that we use boldfaced vector variables for non-sink vertices in $\textsc{Grid}_n$ as they can be identified by their coordinates.

Fact 4.2.5 ([DS84, Chapter 1.3.2])**.** *For any graph G and non-sink vertex u, the potential $\pi_u(v)$ is the probability of a random walk starting at v and reaching u before v_{sink}.*

It follows from the escape probability interpretation of $\pi_u(v)$ and the law of total probability that

we can decompose the voltage at a vertex as a sum of disjoint weighted walks.

Lemma 4.2.6. *Let* $\mathbf{u}$ *be a non-sink vertex of* $\textsc{Grid}_n$*. For any vertex* $\mathbf{v}$*, we have*

$$\pi_{\mathbf{u}}(\mathbf{v}) = \sum_{w \in \mathcal{W}(\mathbf{v} \to \mathbf{u})} 4^{-|w|}.$$

Proof. By definition, we have

$$\pi_{\mathbf{u}}(\mathbf{v}) = \frac{\sum_{w \in \mathcal{W}(\mathbf{v} \to \mathbf{u})} 4^{-|w|}}{\sum_{w \in \mathcal{W}(\mathbf{v} \to \{\mathbf{u}, v_{\text{sink}}\})} 4^{-|w|}}.$$

For any $\mathbf{v} \in V(\textsc{Grid}_n)$, let

$$f(\mathbf{v}) = \sum_{w \in \mathcal{W}(\mathbf{v} \to \{\mathbf{u}, v_{\text{sink}}\})} 4^{-|w|}$$

be the normalizing constant for $\pi_{\mathbf{u}}(\mathbf{v})$. It follows that $f(\mathbf{u}) = 1$ and $f(v_{\text{sink}}) = 1$, because the only such walk for each has length 0. For all other $\mathbf{v} \in V(\textsc{Grid}_n) \setminus \{\mathbf{u}, v_{\text{sink}}\}$, we have

$$f(\mathbf{v}) = \frac{1}{4} \sum_{\mathbf{x} \in N(\mathbf{v})} f(\mathbf{x}).$$

Therefore, $f(\mathbf{v})$ is a harmonic function with constant boundary values. It follows that $f(\mathbf{v}) = 1$ for all vertices $\mathbf{v} \in V(\textsc{Grid}_n)$. □

A systematic treatment of the connection between random walks and electrical networks can be found in the monograph by Doyle and Snell [DS84] or the survey by Lovász [Lov93].

The effective resistance between a pair of vertices u and v, denoted as $R_{\text{eff}}(u, v)$, is a widely-studied quantity in the design and analysis of algorithms, and it can be formalized in several ways. In the electrical interpretation [DS84], effective resistance can be viewed as the voltage needed to send one unit of current from u to v if every edge in G is a unit resistor. For a linear algebraic definition of effective resistance, we direct the reader to [Ell+11].

We first present a classic result that states the effective resistance between opposite corners

in an $n \times n$ resistor network is $\Theta(\log n)$. Then we use Thompson's principle of the electrical flow [DS84] and a triangle inequality for effective resistances to show that the effective resistance between v_{sink} and any internal vertex is bounded between a constant and $O(\log n)$. We note that our proof of Lemma 4.2.8 easily generalizes to any pair of vertices in GRID_n.

Proposition 4.2.7 ([LPW17, Proposition 9.16])**.** *Let G be an $n \times n$ network of unit resistors. If u and v are vertices at opposite corner vertices, then* $\log(n-1)/2 \le R_{\text{eff}}(u, v) \le 2 \log n$.

Lemma 4.2.8. *For any non-sink vertex* $\mathbf{u}$ *in* GRID_n*, we have* $1/4 \le R_{\text{eff}}(v_{\text{sink}}, \mathbf{u}) \le 2 \log n + 1$.

Proof. We start by proving the lower bound $1/4 \le R_{\text{eff}}(v_{\text{sink}}, \mathbf{u})$. The effective resistance between vertices v_{sink} and $\mathbf{u}$ is the reciprocal of the total current flowing into the circuit when $\pi_{\mathbf{u}}(\mathbf{u}) = 1$ and $\pi_{\mathbf{u}}(v_{\text{sink}}) = 0$. Since $\pi_{\mathbf{u}}$ is a harmonic function, we have $\pi_{\mathbf{u}}(\mathbf{v}) \ge 0$ for all $\mathbf{v} \in V(\text{GRID}_n)$. Moreover, we know that $\deg(\mathbf{u}) = 4$, so

$$R_{\text{eff}}(v_{\text{sink}}, \mathbf{u}) = \left(\sum_{\mathbf{v} \sim \mathbf{u}} \pi_{\mathbf{u}}(\mathbf{u}) - \pi_{\mathbf{u}}(\mathbf{v}) \right)^{-1} \ge \frac{1}{4}.$$

For the upper bound, we use Rayleigh's monotonicity law, Proposition 4.2.7, and the triangle inequality for effective resistances to show that $R_{\text{eff}}(v_{\text{sink}}, \mathbf{u}) \le 2 \log n + 1$, for n sufficiently large. Rayleigh's monotonicity law [DS84] states that if the resistances of a circuit are increased, the effective resistance between any two points can only increase. The following triangle inequality for effective resistances is given in [Tet91]:

$$R_{\text{eff}}(u, v) \le R_{\text{eff}}(u, x) + R_{\text{eff}}(x, v).$$

Define H to be the subgraph of GRID_n obtained by deleting v_{sink} and all edges incident to v_{sink}. Let m be the largest positive integer such that $\mathbf{u}_1 + i \le n$ and $\mathbf{u}_2 + j \le n$ for all $0 \le i, j < m$, and let $H(\mathbf{u})$ be the subgraph of H induced by the vertex set $\{(\mathbf{u}_1 + i, \mathbf{u}_2 + j) : 0 \le i, j < m\}$. We

can view $H(\mathbf{u})$ as the largest square resistor network in H such that $\mathbf{u}$ is the top-left vertex. Let $\mathbf{v} = [\mathbf{u}_1 + m - 1, \mathbf{u}_2 + m - 1]$ be the bottom-right vertex in $H(\mathbf{u})$. Using infinite resistors to remove every edge in $E(\textsc{Grid}_n) \setminus E(H(\mathbf{u}))$, we have

$$R_{\text{eff}}^{\textsc{Grid}_n}(\mathbf{v}, \mathbf{u}) \leq R_{\text{eff}}^{H(\mathbf{u})}(\mathbf{v}, \mathbf{u})$$

by Rayleigh's monotonicity law. Proposition 4.2.7 implies that $R_{\text{eff}}^{H(\mathbf{u})}(\mathbf{v}, \mathbf{u}) \leq 2 \log n$ since $m \leq n$. The vertex $\mathbf{v}$ is incident to v_{sink} in $\textsc{Grid}_n$, so Rayleigh's monotonicity law gives $R_{\text{eff}}^{\textsc{Grid}_n}(v_{\text{sink}}, \mathbf{v}) \leq 1$. By the triangle inequality for effective resistances, we have

$$R_{\text{eff}}(v_{\text{sink}}, \mathbf{u}) \leq R_{\text{eff}}(v_{\text{sink}}, \mathbf{v}) + R_{\text{eff}}(\mathbf{v}, \mathbf{u}) \leq 2 \log n + 1,$$

which completes the proof. □

Now we present the potential reciprocity lemma in Choure and Vishwanathan [CV12] to analyze the transience class of the Abelian sandpile model. This result is particularly powerful for $\textsc{Grid}_n$ graphs because we can combine it with Lemma 4.2.8 to swap the source vertex when computing vertex potentials at the expense of a $O(\log n)$ factor. We note that reciprocity has been studied in more general contexts and has been shown to be equivalent in some sense to reversibility in ergodic Markov chains [Tet94].

Lemma 4.2.9 ([CV12, Potential Reciprocity])**.** *Let G be a graph (not necessarily degree-bounded) with sink v_{sink}. For any pair of vertices u and v, we have*

$$R_{\text{eff}}(v_{\text{sink}}, u)\, \pi_u(v) = R_{\text{eff}}(v_{\text{sink}}, v)\, \pi_v(u).$$

Lemma 4.2.10. *For any non-sink vertices $\mathbf{u}$ and $\mathbf{v}$ in $\textsc{Grid}_n$, we have $\pi_{\mathbf{u}}(\mathbf{v}) \leq (8 \log n + 4)\, \pi_{\mathbf{v}}(\mathbf{u})$.*

Proof. This is a direct consequence of Lemma 4.2.8 and Lemma 4.2.9. □

Voltages and flows on electrical networks are central to many recent developments in algorithmic graph theory (e.g. modern maximum flow algorithms and interior point methods [Chr+11, Mad13]). The convergence of many of these algorithms depend on the extremal voltage values of the electrical flow that they construct. As a result, we believe some of our techniques are relevant to the grid-based instantiations of these algorithms.

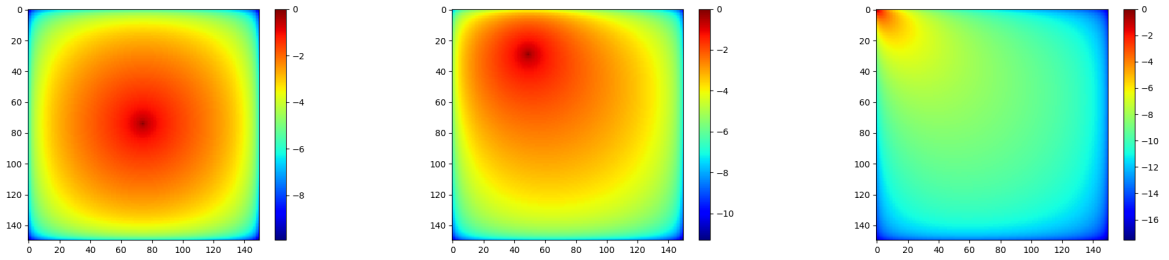
4.3 Upper Bounding the Transience Class

In this section we prove the upper bound in Theorem 4.1.2 for the transience class of the Abelian sandpile model on the square grid. Our proof follows the framework of Choure and Vishwanathan in that we use Theorem 4.2.4 to reduce the proof to bounding the following two quantities for any non-sink vertex $\mathbf{u} \in V(\textsc{Grid}_n)$:

- We upper bound the sum over all potentials $\sum_{\mathbf{v} \in V} \pi_{\mathbf{u}}(\mathbf{v})$.
- We lower bound the potential $\pi_{\mathbf{u}}(\mathbf{v})$ for all non-sink vertices $\mathbf{v}$.

By symmetry we assume without loss of generality that $\mathbf{u}$ is in the top-left quadrant of $\textsc{Grid}_n$ (i.e., we have $1 \le \mathbf{u}_1, \mathbf{u}_2 \le \lceil n/2 \rceil$). The principal idea is to use reciprocity from Lemma 4.2.9 and effective resistance bounds from Lemma 4.2.8 to swap source vertices and bound $\pi_{\mathbf{v}}(\mathbf{u})$ instead, at the expense of a $O(\log n)$ factor. The second key idea is to interpret potentials as random walks using Fact 4.2.5 and then decouple two-dimensional walks on $\textsc{Grid}_n$ into separate horizontal and vertical one-dimensional walks on $\textsc{Path}_n$. Using well-studied properties of one-dimensional random walks, we achieve nearly tight bounds on $\mathrm{tcl}(\textsc{Grid}_n)$.

We note that there is a natural trade-off in the choice of the source vertex $\mathbf{u}$. Choosing $\mathbf{u}$ near the boundary decreases vertex potentials because a random walk has a higher probability of escaping to v_{sink} instead of $\mathbf{u}$. This improves the upper bound of the sum of vertex potentials, but it weakens the lower bound of the minimum vertex potential. For vertices $\mathbf{u}$ that are not near the



(a) (b) (c)

Figure 4.2: Log-scale heat maps of the vertex potentials $\log \pi_{\mathbf{u}}(\mathbf{v})$ on the 150×150 grid with the boundary conditions $\pi_{\mathbf{u}}(\mathbf{u}) = 1$ and $\pi_{\mathbf{u}}(v_{\text{sink}}) = 0$, where (a) $\mathbf{u} = (75, 75)$, (b) $\mathbf{u} = (30, 50)$, and (c) $\mathbf{u} = (1, 1)$.

boundary, the opposite is true. Therefore, we need to accurately account for the choice of $\mathbf{u}$ in our bounds. We illustrate this trade-off phenomenon below in Figure 4.2.

4.3.1 Upper Bounding the Sum of Vertex Potentials

Lemma 4.3.1. *For any non-sink vertex* $\mathbf{u}$ *in* $\textsc{Grid}_n$*, we have*

$$\sum_{\mathbf{v} \in V} \pi_{\mathbf{u}}(\mathbf{v}) = O\left(\mathbf{u}_1 \mathbf{u}_2 \log^3 n\right).$$

Proof. We use Fact 4.2.5 and Lemma 4.2.6 to interpret vertex potentials as random walks. We can omit v_{sink} because any random walk starting there immediately terminates. By Lemma 4.2.10,

$$\pi_{\mathbf{u}}(\mathbf{v}) = O\left(\pi_{\mathbf{v}}(\mathbf{u}) \log n\right),$$

so we can apply the random walk interpretation to potentials starting at $\mathbf{u}$ instead of $\mathbf{v}$. Consider one such walk $w \in \mathcal{W}(\mathbf{u} \to \mathbf{v})$ and its one-dimensional decompositions w_1 and w_2. The probability of a walk from $\mathbf{u}$ reaching $\mathbf{v}$ is equal to the probability that two interleaved walks in $\textsc{Path}_n$ starting at $\mathbf{u}_1$ and $\mathbf{u}_2$ are present on $\mathbf{v}_1$ and $\mathbf{v}_2$, respectively, at the same time before either hits their one-dimensional sink $v_{\text{sink}} = \{0, n + 1\}$.

If we remove the restriction that these walks are present on $\mathbf{v}_1$ and $\mathbf{v}_2$ at the same time and only require that they visit $\mathbf{v}_1$ and $\mathbf{v}_2$ before hitting v_{sink}, then each of these less restricted walks w_d belongs to the class $\mathcal{W}^{\textsc{Path}_n}(\mathbf{u}_d \longrightarrow \mathbf{v}_d)$. Viewing a walk w on $\textsc{Grid}_n$ as infinite walk on the lattice $\mathbb{Z}^2$ induces independence between w_1 and w_2. Thus, we obtain the upper bound

$$\begin{aligned}
\pi_{\mathbf{v}}(\mathbf{u}) &= \Pr_{w\sim\mathcal{W}^{\mathbb{Z}^2}(\mathbf{u})}(w \text{ hits } \mathbf{v} \text{ before leaving } \textsc{Grid}_n) \\
&\leq \Pr_{w\sim\mathcal{W}^{\mathbb{Z}^2}(\mathbf{u})}[w_1 \text{ hits } \mathbf{v}_1 \text{ before } v_{\text{sink}} \text{ and } w_2 \text{ hits } \mathbf{v}_2 \text{ before } v_{\text{sink}}] \\
&= \Pr_{w\sim\mathcal{W}^{\mathbb{Z}^2}(\mathbf{u})}(w_1 \text{ hits } \mathbf{v}_1 \text{ before } v_{\text{sink}}) \cdot \Pr_{w\sim\mathcal{W}^{\mathbb{Z}^2}(\mathbf{u})}(w_2 \text{ hits } \mathbf{v}_2 \text{ before } v_{\text{sink}}) \\
&= \pi_{\mathbf{v}_1}^{\textsc{Path}_n}(\mathbf{u}_1) \cdot \pi_{\mathbf{v}_2}^{\textsc{Path}_n}(\mathbf{u}_2).
\end{aligned}$$

Summing over all choices of $\mathbf{v} = (\mathbf{v}_1, \mathbf{v}_2)$, it follows that

$$\sum_{\mathbf{v}\in V} \pi_{\mathbf{v}}(\mathbf{u}) \leq \left(\sum_{\mathbf{v}_1=1}^{n} \pi_{\mathbf{v}_1}^{\textsc{Path}_n}(\mathbf{u}_1)\right)\left(\sum_{\mathbf{v}_2=1}^{n} \pi_{\mathbf{v}_2}^{\textsc{Path}_n}(\mathbf{u}_2)\right).$$

The potentials of vertices in $\textsc{Path}_n$ have the following closed-form solution, as shown in [DS84]:

$$\pi_{\mathbf{v}_1}^{\textsc{Path}_n}(\mathbf{u}_1) = \begin{cases} \frac{n+1-\mathbf{u}_1}{n+1-\mathbf{v}_1} & \text{if } \mathbf{v}_1 \leq \mathbf{u}_1, \\ \frac{\mathbf{u}_1}{\mathbf{v}_1} & \text{if } \mathbf{v}_1 > \mathbf{u}_1. \end{cases}$$

Splitting the sum at $\mathbf{u}_1$ and using the fact that potentials are escape probabilities, we have

$$\sum_{\mathbf{v}_1=1}^{n} \pi_{\mathbf{v}_1}^{\textsc{Path}_n}(\mathbf{u}_1) \leq \mathbf{u}_1 + \sum_{\mathbf{v}_1=\mathbf{u}_1+1}^{n} \frac{\mathbf{u}_1}{\mathbf{v}_1} = O(\mathbf{u}_1 \log n).$$

We obtain the upper bound of $O(\mathbf{u}_2 \log n)$ for the other dimension similarly. These bounds along with the initial $O(\log n)$ overhead from swapping $\mathbf{u}$ and $\mathbf{v}$ gives the desired upper bound. □

4.3.2 Lower Bounding the Minimum Vertex Potential

The more involved part of this chapter proves a lower bound for the minimum vertex potential $\min_{\mathbf{v} \in V \setminus \{v_{\text{sink}}\}} \pi_{\mathbf{u}}(\mathbf{v})$ as a function of a fixed vertex $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2)$. Recall that we assumed without loss of generality that $\mathbf{u}$ is in the top-left quadrant of $\textsc{Grid}_n$. We first prove that the minimum potential occurs at vertex (n, n), the corner farthest from $\mathbf{u}$. Using Lemma 4.2.10 to swap $\mathbf{u}$ and (n, n) at the expense of a $\Omega(1/\log n)$ factor, we reduce the problem to giving a lower bound for $\pi_{(n,n)}(\mathbf{u})$. Next, we decompose walks $w \in \mathcal{W}(\mathbf{u} \longrightarrow \{(n, n), v_{\text{sink}}\})$ into their one-dimensional walks $w_1 \in \mathcal{W}^{\textsc{Path}_n}(\mathbf{u}_1)$ and $w_2 \in \mathcal{W}^{\textsc{Path}_n}(\mathbf{u}_2)$, and we interpret $\pi_{(n,n)}(\mathbf{u})$ as the probability that the individual processes w_1 and w_2 are present on n at the same time before either walk leaves the set $[n]$. Walks on $\textsc{Line}$ that meet at n before escaping $[n]$ are equivalent to walks on $\textsc{Path}_n$ that meet at n before terminating at v_{sink}. Lastly, we use conditional probabilities to analyze walks on $\textsc{Line}$ instead of walks on $\textsc{Path}_n$ so that we can leverage well-known facts about simple symmetric random walks.

To lower bound the desired probability $\pi_{(n,n)}(\mathbf{u})$, we show that a subset of $\mathcal{W}(\mathbf{u} \longrightarrow (n, n))$ of interleaved one-dimensional walks starting from $\mathbf{u}_1$ and $\mathbf{u}_2$ that first reach n in approximately the same number of steps has a sufficient amount of probability mass. We prove this by observing that the distributions of the number of steps for the walks to first reach n without leaving the set $[n]$ are concentrated around $(n - \mathbf{u}_1)^2$ and $(n - \mathbf{u}_2)^2$, respectively. Consequently, we show that this distribution is approximately uniform in an $\Theta(n^2)$ length interval, with each t-step walk having probability $\Omega(\mathbf{u}_1 n^{-3})$ and $\Omega(\mathbf{u}_2 n^{-3})$. We then use Chernoff bounds to show that both walks take approximately the same number of steps with constant probability. Combining these facts allows us to achieve the desired lower bound $\Omega(\mathbf{u}_1 \mathbf{u}_2 n^{-4})$.

We first show that the corner vertex (n, n) has the minimum potential up to a constant factor. Viewing potentials as escape probabilities, we utilize the geometry of the grid to construct maps between sets of random walks that show the potential of an interior vertex is greater than its

axis-aligned projection to the boundary of the grid.

Lemma 4.3.2. *If* $\mathbf{u}$ *is a vertex in the top-left quadrant of* $\textsc{Grid}_n$*, then for any non-sink vertex* $\mathbf{v}$*,*

$$\pi_{\mathbf{u}}(\mathbf{v}) \geq \frac{1}{16}\pi_{\mathbf{u}}((n,n)).$$

Proof. We use Lemma 4.2.6 to decompose $\pi_{\mathbf{u}}(\mathbf{v})$ as a sum of probabilities of walks, and then construct maps for all $\mathbf{v}_1, \mathbf{v}_2 \in [n]$ to show

$$\pi_{\mathbf{u}}((\mathbf{v}_1,\mathbf{v}_2)) \geq \max\left\{\frac{1}{4}\pi_{\mathbf{u}}((n,\mathbf{v}_2)),\ \frac{1}{4}\pi_{\mathbf{u}}((\mathbf{v}_1,n))\right\}.$$

We begin by considering the first dimension:

$$\pi_{\mathbf{u}}((\mathbf{v}_1,\mathbf{v}_2)) \geq \frac{\pi_{\mathbf{u}}((n,\mathbf{v}_2))}{4}.$$

Let ℓ_{hor} be the horizontal line of reflection passing through $(\lceil(\mathbf{v}_1+n)/2\rceil, 1)$ and $(\lceil(\mathbf{v}_1+n)/2\rceil, n)$ in $\mathbb{Z}^2$, and let $\mathbf{u}^*$ be the reflection of $\mathbf{u}$ over ℓ_{hor}. Note that $\mathbf{u}^*$ may be outside of the $n \times n$ grid. Next, define the map

$$f : \mathcal{W}((n,\mathbf{v}_2) \to \mathbf{u}) \to \mathcal{W}((\mathbf{v}_1,\mathbf{v}_2) \to \mathbf{u})$$

as follows. For any walk $w \in \mathcal{W}((n,\mathbf{v}_2) \to \mathbf{u})$:

1. Start the walk $f(w)$ at $(\mathbf{v}_1,\mathbf{v}_2)$, and if $n - \mathbf{v}_1$ is odd move to $(\mathbf{v}_1+1,\mathbf{v}_2)$.
2. Perform w but make opposite vertical moves before the walk hits ℓ_{hor}, so that the partial walk is a reflection over ℓ_{hor}.
3. After hitting ℓ_{hor} for the first time, continue performing w, but now use the original vertical moves.
4. Terminate this walk when it first reaches $\mathbf{u}$.

Denote the preimage of a walk $w' \in \mathcal{W}((\mathbf{v}_1, \mathbf{v}_2) \longrightarrow \mathbf{u})$ under f to be

$$f^{-1}(w') = \{w \in \mathcal{W}((n, \mathbf{v}_2) \longrightarrow \mathbf{u}) : f(w) = w'\}.$$

We claim that for any $w' \in \mathcal{W}^{\text{GRID}_n}((\mathbf{v}_1, \mathbf{v}_2) \longrightarrow \mathbf{u})$,

$$\frac{1}{4} \sum_{w \in f^{-1}(w')} 4^{-|w|} \leq 4^{-|w'|}.$$

If $f^{-1}(w') = \emptyset$ the claim is true, so assume $f^{-1}(w') \neq \emptyset$. We analyze two cases. First, if w' hits ℓ_{hor}, then $f^{-1}(w')$ contains exactly one walk w of length $|w'|$ or $|w'| - 1$. If w' does not hit ℓ_{hor}, then

$$f^{-1}(w') = \{w \in \mathcal{W}((n, \mathbf{v}_2) \longrightarrow \mathbf{u}) : w \text{ is a reflection of } w' \text{ over } \ell_{\text{hor}} \text{ before } w \text{ hits } \mathbf{u}^*\}.$$

It follows that any walk $w \in f^{-1}(w')$ can be split into $w = w_1 w_2$, where w_1 is the unique walk from $(n, \mathbf{v}_2)$ to $\mathbf{u}^*$ that is a reflection of w', and w_2 is a walk from $\mathbf{u}^*$ to $\mathbf{u}$ that avoids v_{sink} and hits $\mathbf{u}$ exactly once upon termination. Clearly w_1 has length $|w'|$ or $|w'| - 1$, and the set of admissible w_2 is $\mathcal{W}(\mathbf{u}^* \longrightarrow \mathbf{u})$. Therefore, we have

$$\begin{aligned}
\frac{1}{4} \sum_{w \in f^{-1}(w')} 4^{-|w|} &= 4^{-|w_1|-1} \sum_{w_2 \in \mathcal{W}(\mathbf{u}^* \to \mathbf{u})} 4^{-|w_2|} \\
&= 4^{-|w_1|-1} \pi_{\mathbf{u}}(\mathbf{u}^*) \\
&\leq 4^{-|w'|},
\end{aligned}$$

since $\pi_{\mathbf{u}}(\mathbf{u}^*)$ is an escape probability. Summing over all $w' \in \mathcal{W}((\mathbf{v}_1, \mathbf{v}_2) \longrightarrow \mathbf{u})$, it follows from

Lemma 4.2.6 and the previous inequality that

$$\begin{aligned}\pi_{\mathbf{u}}\left((\mathbf{v}_1, \mathbf{v}_2)\right) &= \sum_{w' \in \mathcal{W}((\mathbf{v}_1,\mathbf{v}_2)\to\mathbf{u})} 4^{-|w'|} \\ &\geq \sum_{w' \in \mathcal{W}((\mathbf{v}_1,\mathbf{v}_2)\to\mathbf{u})} \frac{1}{4} \sum_{w \in f^{-1}(w')} 4^{-|w|} \\ &\geq \frac{1}{4}\pi_{\mathbf{u}}\left((n, \mathbf{v}_2)\right),\end{aligned}$$

because every $w \in \mathcal{W}((n, \mathbf{v}_2) \to \mathbf{u})$ is the preimage of a $w' \in \mathcal{W}((\mathbf{v}_1, \mathbf{v}_2) \to \mathbf{u})$.

Similarly, we have $\pi_{\mathbf{u}}\left((\mathbf{v}_1, \mathbf{v}_2)\right) \geq \pi_{\mathbf{u}}\left((\mathbf{v}_1, n)\right)/4$ for all $\mathbf{v}_1 \in [n]$ by reflecting walks over the vertical line from $(1, \lceil (n + \mathbf{v}_2)/2 \rceil)$ to $(n, \lceil (n + \mathbf{v}_2)/2 \rceil)$. Combining inequalities proves the claim. □

By decomposing two-dimensional walks on $\textsc{Grid}_n$ that start at $\mathbf{u}$ into one-dimensional walks on $\textsc{Line}$, our lower bound relies on showing that there is an $\Theta(n^2)$ length interval such that each one-dimensional walk of a fixed length in this interval has probability $\Omega(\mathbf{u}_1 n^{-3})$ or $\Omega(\mathbf{u}_2 n^{-3})$, respectively, of remaining above 0 and reaching n for the first time upon termination. Lower bounds for this probability suffice for our purpose, and they are a consequence of the following key property for one-dimensional walks that we prove in Section 4.5.

Lemma 4.3.3. *Let $n \in \mathbb{Z}_{\geq 1}$ and $i \in [\lceil n/2 \rceil]$ be any starting position. For any constant $c > 4$ and $t \in \mathbb{Z}$ such that $n^2/c \leq t \leq n^2/4$ with $t \equiv n - i \pmod 2$, a simple symmetric random walk w on $\mathbb{Z}$ satisfies*

$$\Pr_{w \sim \mathcal{W}^{\textsc{Line}}(i)}\left(w^{(t)} = n, \max_{\leq t}(w) = n, \text{and } \min_{\leq t}(w) \geq 1\right) \geq e^{-2c-2}\frac{i}{n^3}.$$

Next, we give a constant lower bound for the probability of an n-step simple symmetric walk being sufficiently close to its starting position. We prove this by using the recursive definition of binomial coefficients and a Chernoff bound for symmetric random variables.

Lemma 4.3.4. *For all $n \geq 10$, we have*

$$\frac{1}{2^n} \min \left\{ \sum_{\substack{k=\lceil 1/4 n \rceil \\ k \text{ odd}}}^{\lfloor 3/4 n \rfloor} \binom{n}{k}, \sum_{\substack{k=\lceil 1/4 n \rceil \\ k \text{ even}}}^{\lfloor 3/4 n \rfloor} \binom{n}{k} \right\} \geq \frac{2}{5}.$$

Proof. First observe that for $n \geq 10$, we have

$$\frac{1}{2^n} \sum_{\substack{k=\lceil 1/4 n \rceil \\ k \text{ odd}}}^{\lfloor 3/4 n \rfloor} \binom{n}{k} \geq \frac{1}{2^n} \sum_{k \in \left(\frac{n-1}{4}, \frac{3(n-1)}{4}\right)} \binom{n-1}{k}$$

and

$$\frac{1}{2^n} \sum_{\substack{k=\lceil 1/4 n \rceil \\ k \text{ even}}}^{\lfloor 3/4 n \rfloor} \binom{n}{k} \geq \frac{1}{2^n} \sum_{k \in \left(\frac{n-1}{4}, \frac{3(n-1)}{4}\right)} \binom{n-1}{k}.$$

To see this, we exploit the parity restriction and expand the summands as $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$. Let $X_1, X_2, \ldots, X_{n-1}$ be independent Bernoulli random variables such that $\Pr(X_i = 0) = \Pr(X_i = 1) = 1/2$. Letting $S_{n-1} = X_1 + X_2 + \cdots + X_{n-1}$ and $\mu = E[S_{n-1}] = (n-1)/2$, it follows from a Chernoff bound that for $n \geq 60$, we have

$$\frac{1}{2^n} \sum_{k \in \left(\frac{n-1}{4}, \frac{3(n-1)}{4}\right)} \binom{n-1}{k} = \frac{1}{2} \left[1 - \Pr\left(|S_{n-1} - \mu| \geq \frac{1}{2} \cdot \mu \right) \right] \geq \frac{1}{2} - e^{-(n-1)/24} \geq \frac{2}{5}.$$

Checking the remaining cases numerically when $10 \leq n < 60$ proves the claim. □

Using Lemma 4.3.3 with Lemma 4.3.4, we give a lower bound for $\pi_{(n,n)}(\mathbf{u})$, the probability that a walk starting from $\mathbf{u}$ reaches (n, n) before v_{sink}.

Lemma 4.3.5. *For all $n \geq 10$ and any vertex $\mathbf{u}$ in the top-left quadrant of* $\textsc{Grid}_n$*, we have*

$$\pi_{(n,n)}(\mathbf{u}) \geq e^{-100} \frac{\mathbf{u}_1 \mathbf{u}_2}{n^4}.$$

Proof. We decouple each walk $w \in \mathcal{W}(\mathbf{u} \to (n, n))$ into its horizontal walk $w_1 \in \mathcal{W}^{\textsc{Line}}(\mathbf{u}_1)$ and

vertical walk $w_2 \in \mathcal{W}^{\textsc{Line}}(\mathbf{u}_2)$. The potential $\pi_{(n,n)}(\mathbf{u})$ can be interpreted as the probability that w_1 and w_2 visit n at the same time before either leaves the interval $[1, n]$. We can further decompose the t-step walks on $\textsc{Grid}_n$ into those that take t_1 steps in the horizontal direction and t_2 in the vertical direction. Considering restricted instances where w_1 and w_2 visit n exactly once, we obtain the following bound by Lemma 4.2.6:

$$\pi_{(n,n)}(\mathbf{u}) \geq \sum_{\substack{w \sim \mathcal{W}(\mathbf{u} \to (n,n)) \\ w_1 \text{ hits } n \text{ exactly once} \\ w_2 \text{ hits } n \text{ exactly once}}} 4^{-|w|}. \tag{4.1}$$

Accounting for all the ways that the two one-dimensional walks can be interleaved, the right hand side of Equation (4.1) is equal to

$$\sum_{t_1, t_2 \geq 0} \frac{\binom{t_1+t_2}{t_1}}{4^{t_1+t_2}} (\#\text{ of } t_1\text{-step walks from } \mathbf{u}_1 \text{ that stay in } [n-1] \text{ and terminate at } n) \\ \cdot (\#\text{ of } t_2\text{-step walks from } \mathbf{u}_2 \text{ that stay in } [n-1] \text{ and terminate at } n).$$

Observing that

$$\Pr_{w_1 \sim \mathcal{W}(\mathbf{u}_1)}\left(w_1^{(t_1)} = n, \max_{\leq t_1 - 1}(w) = n-1, \min_{\leq t_1 - 1}(w) \geq 1\right) \\ = \frac{(\#\text{ of } t_1\text{-step walks from } \mathbf{u}_1 \text{ that stay in } [n-1] \text{ and terminate at } n)}{2^{t_1}},$$

it follows from Equation (4.1) that

$$\pi_{(n,n)}(\mathbf{u}) \geq \sum_{t_1, t_2 \geq 0} \frac{\binom{t_1+t_2}{t_1}}{2^{t_1+t_2}} \Pr_{w_1 \sim \mathcal{W}(\mathbf{u}_1)}\left(w_1^{(t_1)} = n, \max_{\leq t_1 - 1}(w) = n-1, \min_{\leq t_1 - 1}(w) \geq 1\right) \\ \cdot \Pr_{w_2 \sim \mathcal{W}(\mathbf{u}_2)}\left(w_2^{(t_2)} = n, \max_{\leq t_2 - 1}(w) = n-1, \min_{\leq t_2 - 1}(w) \geq 1\right).$$

By our choice of n and $\mathbf{u}$, the right hand side of inequality above equals

$$\sum_{t_1,t_2\geq 5} \frac{\binom{t_1+t_2}{t_1}}{2^{t_1+t_2}} \left[\frac{1}{2}\Pr_{w_1\sim\mathcal{W}(\mathbf{u}_1)}\left(w_1^{(t_1-1)} = n-1, \max_{\leq t_1-1}(w) = n-1, \min_{\leq t_1-1}(w)\geq 1\right)\right] \cdot \left[\frac{1}{2}\Pr_{w_2\sim\mathcal{W}(\mathbf{u}_2)}\left(w_2^{(t_2-1)} = n-1, \max_{\leq t_2-1}(w) = n-1, \min_{\leq t_2-1}(w)\geq 1\right)\right]. \tag{4.2}$$

Letting $t = t_1 + t_2$, we can further refine the set of two-dimensional walks so that $t \in [1/5 n^2, 1/4 n^2]$ and $t_1, t_2 \in [1/4 t, 3/4 t]$ while capturing a sufficient amount of probability mass for a useful lower bound. Note that the parities of t_1 and t_2 satisfy $t_1 \equiv n - \mathbf{u}_1 \pmod 2$ and $t_2 \equiv n - \mathbf{u}_2 \pmod 2$ for valid walks. Let I be an indexing of all such pairs (t_1, t_2). Working from Equation (4.2), we have

$$\begin{aligned}
\pi_{(n,n)}(\mathbf{u}) &\geq \sum_{(t_1,t_2)\in I} \frac{\binom{t_1+t_2}{t_1}}{2^{t_1+t_2}} \left(\frac{1}{2}e^{-2(20)-2}\frac{\mathbf{u}_1}{n^3}\right)\left(\frac{1}{2}e^{-2(20)-2}\frac{\mathbf{u}_2}{n^3}\right) \\
&\geq e^{-84}\frac{\mathbf{u}_1\mathbf{u}_2}{4n^6} \sum_{\substack{t\in\left[1/5 n^2, 1/4 n^2\right] \\ t\equiv \mathbf{u}_1+\mathbf{u}_2 \pmod 2}} \frac{2}{5} \\
&\geq e^{-84}\frac{\mathbf{u}_1\mathbf{u}_2}{4n^6}\cdot\frac{n^2}{50}\cdot\frac{2}{5} \\
&\geq e^{-100}\frac{\mathbf{u}_1\mathbf{u}_2}{n^4}.
\end{aligned}$$

For the first inequality, we can apply Lemma 4.3.3 because $t_1, t_2 \in [1/20 n^2, 3/16 n^2]$. For the second inequality, we group pairs (t_1, t_2) by their sum $t = t_1 + t_2$ and apply Lemma 4.3.4. The number of step sizes $t \in [1/5 n^2, 1/4 n^2]$ with either of the parity restriction is at least $\lfloor 1/40 n^2 \rfloor \geq 1/50 n^2$. □

We now combine the upper bound for the sum of potentials given by Lemma 4.3.1 and the lower bounds in Section 4.3.2 to obtain the overall upper bound for the transience class of the grid.

Theorem 4.1.2. *The transience class of the Abelian sandpile model on the $n \times n$ grid is $O(n^4 \log^4 n)$.*

Proof. For any $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2)$ in the top-left quadrant of $\textsc{Grid}_n$, we have

$$\begin{aligned}
\max_{\mathbf{u},\mathbf{v}\in V\setminus\{v_{\text{sink}}\}} \left(\sum_{\mathbf{x}\in V} \pi_{\mathbf{u}}(\mathbf{x})\right) \pi_{\mathbf{u}}(\mathbf{v})^{-1} &\le \max_{\mathbf{u}\in V\setminus\{v_{\text{sink}}\}} \left(\sum_{\mathbf{x}\in V} \pi_{\mathbf{u}}(\mathbf{x})\right) \frac{16}{\pi_{\mathbf{u}}((n,n))} \\
&= \max_{\mathbf{u}\in V\setminus\{v_{\text{sink}}\}} \left(\sum_{\mathbf{x}\in V} \pi_{\mathbf{u}}(\mathbf{x})\right) \frac{O(\log n)}{\pi_{(n,n)}(\mathbf{u})} \\
&= \max_{\mathbf{u}\in V\setminus\{v_{\text{sink}}\}} O\left(\mathbf{u}_1\mathbf{u}_2 \log^3 n\right) O\left(\frac{n^4 \log n}{\mathbf{u}_1\mathbf{u}_2}\right) \\
&= O\left(n^4 \log^4 n\right).
\end{aligned}$$

The first inequality follows from Lemma 4.3.2, the second from Lemma 4.2.10, and the third from Lemma 4.3.5 and Lemma 4.3.1. Therefore, the result follows from Theorem 4.2.4, which bounds the transience class in terms of vertex potentials. □

4.4 Lower Bounding the Transience Class

In this section we lower bound $\text{tcl}(\textsc{Grid}_n)$ using techniques similar to those in Section 4.3. Since the lower bound in Theorem 4.2.4 considers the maximum inverse vertex potential over all pairs of non-sink vertices u and v, it suffices to upper bound $\pi_{(n,n)}((1,1))$. We lower bound vertex potentials by decomposing two-dimensional walks on $\textsc{Grid}_n$ into one-dimensional walks on $\textsc{Line}$ and then upper bound the probability that a t-step walk on $\textsc{Line}$ starting at 1 and ending at n does not leave the set $[n]$. More specifically, our upper bound for $\pi_{(n,n)}((1,1))$ follows from Lemma 4.4.1 (which we prove in Section 4.5) and Fact 4.4.2.

Lemma 4.4.1. *For all $n \ge 20$ and $t \ge n - 1$, we have*

$$\Pr_{w\sim\mathcal{W}^{\textsc{Line}}(1)}\left(w^{(t)} = n, \max_{\le t}(w) = n, \text{and } \min_{\le t}(w) \ge 1\right) \le \min\left\{\frac{e^{25}}{n^3}, 64\left(\frac{n}{t}\right)^3\right\}.$$

Fact 4.4.2. *For any nonnegative integer $t_1 \in \mathbb{Z}_{\geq 0}$, we have*

$$\sum_{t_2 \geq 0} \binom{t_1 + t_2}{t_2} \frac{1}{2^{t_1+t_2}} = 2.$$

Proof. This follows from the negative binomial distribution. In particular, observe that

$$\sum_{t_2 \geq 0} \binom{t_1 + t_2}{t_2} \frac{1}{2^{t_1+t_2}} = 2 \sum_{t_2 \geq 0} \binom{(t_1 + 1) - 1 + t_2}{t_2} \frac{1}{2^{t_1+1}} \cdot \frac{1}{2^{t_2}} = 2,$$

as desired. □

By decoupling the two-dimensional walks in a way similar to the proof of Lemma 4.3.5, we apply Lemma 4.4.1 to the resulting one-dimensional walks to achieve the desired upper bound.

Lemma 4.4.3. *For all $n \geq 20$, we have*

$$\begin{aligned}\pi_{(n,n)}((1,1)) \leq 2 \max\Big\{ \Pr_{w \sim \mathcal{W}(1)} \left(w^{(t)} = n, \max_{\leq t}(w) = n, \min_{\leq t}(w) \geq 1 \right) \,:\, t \in \mathbb{Z}_{\geq 0} \Big\} \\ \cdot \sum_{t \geq 0} \Pr_{w \sim \mathcal{W}(1)} \left(w^{(t)} = n, \max_{\leq t}(w) = n, \min_{\leq t}(w) \geq 1 \right).\end{aligned}$$

Proof. Analogous to our lower bound for $\pi_{(n,n)}((1,1))$, decouple each walk $w \in \mathcal{W}((1,1) \longrightarrow (n,n))$ into its horizontal walk $w_1 \in \mathcal{W}^{\textsc{Line}}(1)$ and its vertical walk $w_2 \in \mathcal{W}^{\textsc{Line}}(1)$. We view $\pi_{(n,n)}((1,1))$ as the probability that the walks w_1 and w_2 are present on n at the same time before either leaves the set $[n]$. Letting t_1 be the length of w_1 and t_2 be the length of w_2, we relax the conditions on the one-dimensional walks and only require that w_1 and w_2 both be at vertex n on the final step $t = t_1 + t_2$. Note that both of the walks could have previously been present on n at the same time before terminating. This gives the following upper bound:

$$\begin{aligned}\pi_{(n,n)}((1,1)) \leq \sum_{t_1,t_2 \geq 0} \frac{\binom{t_1+t_2}{t_1}}{2^{t_1+t_2}} \Pr_{w_1 \sim \mathcal{W}(1)} \left(w_1^{(t_1)} = n, \max_{\leq t_1}(w) = n, \min_{\leq t_1}(w) \geq 1 \right) \\ \cdot \Pr_{w_2 \sim \mathcal{W}(1)} \left(w_2^{(t_2)} = n, \max_{\leq t_2}(w) = n, \min_{\leq t_2}(w) \geq 1 \right).\end{aligned}$$

Nesting the summation above gives us

$$\pi_{(n,n)}((1,1)) \le \sum_{t_1 \ge 0} \Pr_{w_1 \sim \mathcal{W}(1)}\left(w_1^{(t_1)} = n, \max_{\le t_1}(w) = n, \min_{\le t_1}(w) \ge 1\right) \cdot \sum_{t_2 \ge 0} \frac{\binom{t_1+t_2}{t_1}}{2^{t_1+t_2}} \Pr_{w_2 \sim \mathcal{W}(1)}\left(w_2^{(t_2)} = n, \max_{\le t_2}(w) = n, \min_{\le t_2}(w) \ge 1\right).$$

Using Fact 4.4.2, we can upper bound the inner sum by

$$\sum_{t_2 \ge 0} \frac{\binom{t_1+t_2}{t_1}}{2^{t_1+t_2}} \Pr_{w_2 \sim \mathcal{W}(1)}\left(w_2^{(t_2)} = n, \max_{\le t_2}(w) = n, \min_{\le t_2}(w) \ge 1\right) \le 2 \max\left\{\Pr_{w_2 \sim \mathcal{W}(1)}\left(w_2^{(t_2)} = n, \max_{\le t_2}(w) = n, \min_{\le t_2}(w) \ge 1\right) : t_2 \in \mathbb{Z}_{\ge 0}\right\}.$$

Factoring out the inner upper bound and using the symmetric of the walks completes the proof. □

The upper bound on the maximum term in the right hand side of Lemma 4.4.3 follows immediately from Lemma 4.4.1. Now we upper bound the summation in the right hand side of Lemma 4.4.3 using a simple application of Lemma 4.4.1.

Lemma 4.4.4. *For $n \ge 20$ and $w \sim \mathcal{W}^{\text{LINE}}(1)$, we have*

$$\sum_{t \ge 0} \Pr\left(w^{(t)} = n, \max_{\le t}(w) = n, \min_{\le t}(w) \ge 1\right) \le \frac{e^{25}}{n}.$$

Proof. We start by partitioning the summation as

$$\sum_{t \ge 0} \Pr\left(w^{(t)} = n, \max_{\le t}(w) = n, \min_{\le t}(w) \ge 1\right) = \sum_{0 \le t \le n^2} \Pr\left(w^{(t)} = n, \max_{\le t}(w) = n, \min_{\le t}(w) \ge 1\right) + \sum_{t > n^2} \Pr\left(w^{(t)} = n, \max_{\le t}(w) = n, \min_{\le t}(w) \ge 1\right).$$

Our goal is to bound both terms by $O(1/n)$. The upper bound for the first term follows immediately

from Lemma 4.4.1 and the fact that we are summing $n^2 + 1$ terms starting with $t = 0$:

$$\sum_{0 \le t \le n^2} \Pr\left(w^{(t)} = n, \max_{\le t}(w) = n, \min_{\le t}(w) \ge 1\right) \le \frac{e^{25}}{n}.$$

To upper bound the second sum, we use the other component of Lemma 4.4.1. If $t > n^2$, then

$$\Pr\left(w^{(t)} = n, \max_{\le t}(w) = n, \min_{\le t}(w) \ge 1\right) \le 64\left(\frac{n}{t}\right)^3.$$

Since $64(n/t)^3$ is a decreasing function in t, we have

$$64\left(\frac{n}{t}\right)^3 \le \int_{t-1}^{t} 64\left(\frac{n}{t}\right)^3 \mathrm{d}t.$$

Therefore, we can bound the infinite sum by the integral

$$\sum_{t > n^2} \Pr\left(w^{(t)} = n, \max_{\le t}(w) = n, \min_{\le t}(w) \ge 1\right) \le \int_{n^2}^{\infty} 64\left(\frac{n}{t}\right)^3 \mathrm{d}t = \frac{32}{n}.$$

This completes the proof since both parts of the sum are bounded by $O(1/n)$. □

We can now easily combine the lemmas in this section with the bounds that relate vertex potentials to the lower bound for the transience class of $\textsc{Grid}_n$.

Theorem 4.1.3. *The transience class of the Abelian sandpile model on the $n \times n$ grid is $\Omega(n^4)$.*

Proof. Applying Lemma 4.4.3 and then Lemma 4.4.1 and Lemma 4.4.4, it follows that

$$\begin{aligned}
\pi_{(n,n)}((1,1)) &\le 2 \max\left\{\Pr_{w \sim \mathcal{W}(1)}\left(w^{(t)} = n, \max_{\le t}(w) = n, \min_{\le t}(w) \ge 1\right) \,:\, t \in \mathbb{Z}_{\ge 0}\right\} \\
&\qquad \cdot \sum_{t \ge 0} \Pr_{w \sim \mathcal{W}(1)}\left(w^{(t)} = n, \max_{\le t}(w) = n, \min_{\le t}(w) \ge 1\right) \\
&\le \frac{2e^{25}}{n^3} \cdot \frac{e^{25}}{n} \\
&\le \frac{e^{100}}{n^4}.
\end{aligned}$$

Therefore, we have $\pi_{(n,n)}((1,1))^{-1} = \Omega(n^4)$. Since the voltage at (n, n) gives a bound on the global minimum, it follows from Lemma 4.3.2 and Theorem 4.2.4 that we have $\mathrm{tcl}(\textsc{Grid}_n) = \Omega(n^4)$. □

4.5 Simple Symmetric Random Walks

Our proofs for upper and lower bounding the transience class on the grid heavily relies on decoupling two-dimensional walks into independent one-dimensional walks, as they are much easier to analyze. This idea is immediately apparent when working with vertex potentials for one-dimensional walks on the path, which we used in the proof of Lemma 4.3.1.

We assumed, however, two essential lemmas about one-dimensional walks to prove the lower and upper bound of the minimum vertex potential. Therefore, in this section we examine the probability

$$\Pr_{w\sim\mathcal{W}^{\textsc{Line}}(i)}\left(w^{(t)} = n, \max_{\le t}(w) = n, \min_{\le t}(w) \ge 1\right), \tag{4.3}$$

and we prove the required lower and upper bounds in Lemma 4.3.3 and Lemma 4.4.1 by extending previously known properties of simple symmetric random walks on $\mathbb{Z}$. The key ideas behind these proofs are that: the position of a walk in one dimension follows the binomial distribution, the number of walks reaching a maximum position in a fixed number of steps has an explicit formula, and there are tight bounds for binomial coefficients via Stirling's approximation.

The properties we need do not immediately follow from previously known facts because we assume strict conditions on both the minimum and maximum positions. In Section 4.5.2 we give proofs of known explicit expressions for the maximum and minimum position of a walk, along with several useful facts that follow from this proof. In Section 4.5.3 we apply Stirling's bound to give accurate lower bounds on a range of binomial coefficients. In Section 4.5.4 and Section 4.5.5 we prove several necessary preliminary lower bound lemmas. In Section 4.5.5 we prove Lemma 4.3.3, and in Section 4.5.6 we develop upper bound lemmas that are necessary to prove Lemma 4.4.1.

4.5.1 Lower and Upper Bounding Equation 4.3

Lower Bound. In order to lower bound Equation (4.3), we split the desired probability into the product of two probabilities using the definition of conditional probability, and then we lower bound each term:

- In Lemma 4.5.6 we show for lengths $t \in \Theta(n^2)$ that the probability of a random walk on $\mathbb{Z}$ starting at $i \in [\lceil n/2 \rceil]$ satisfies

$$\Pr_{w \sim \mathcal{W}(i)} \left(\min_{\leq t}(w) \geq 1 \right) = \Omega\left(\frac{i}{n} \right).$$

- In Lemma 4.5.8 and Lemma 4.5.7 we bound the probability that a walk starting at $i \in [\lceil n/2 \rceil]$ of length $t \in \Theta(n^2)$ reaches n at step t without ever having gone above n, conditioned on never dropping below 1:

$$\Pr\left(w^{(t)} = n, \max_{\leq t}(w) = n \,\middle|\, \min_{\leq t}(w) \geq 1 \right) = \Omega\left(\frac{1}{n^2} \right).$$

Lemma 4.3.3 immediately follows by multiplying these two bounds together. This approach allows us to separate proving a minimum and maximum, and in turn simplifies applying known bounds on binomial distributions. Specifically, Lemma 4.5.6 is an immediate consequence of explicit expressions for the minimum point of a walk and bounds on binomial coefficients, both of which we rigorously explore in Section 4.5.2.

Our results yield a known explicit expression for the probability of the walk reaching n at step t, while always staying to the left of n. All that remains from here is to condition the walk to not go to the left of 1. Note that 1 is in the opposite direction of n with respect to the starting position i. Formally, we show that the probability of reaching n without going above n only improves if the walk cannot move too far in the opposite direction—but only for $t \leq (n - i + 1)^2$, thus explaining the reason why we need to upper bound t by $n^2/4$.

Upper Bound. The desired lemma only concerns walks starting at $i = 1$, which are critical for our proofs. The key idea is to split the walk in half and consider the probability that the necessary conditions are satisfied in the first $t/2$ steps and in the second $t/2$ steps. The midpoint of the walk at $t/2$ steps can be any point in the set $[n]$, so we must sum over all these possible midpoints. Removing the upper and lower bound conditions, respectively, then gives the upper bound in Lemma 4.5.9:

$$\Pr_{w\sim\mathcal{W}(1)}\left(w^{(t)} = n, \max_{\leq t}(w) = n, \min_{\leq t}(w) \geq 1\right) \leq \sum_{i=1}^{n} \Pr_{w\sim\mathcal{W}(1)}\left(w^{(\lfloor t/2\rfloor)} = i, \min_{\leq\lfloor t/2\rfloor}(w) \geq 1\right) \cdot \Pr_{w\sim\mathcal{W}(i)}\left(w^{(\lceil t/2\rceil)} = n, \max_{\leq\lceil t/2\rceil}(w) = n\right).$$

Given that the first $(t/2)$-step walk starts at 1 and the second $(t/2)$-step walk ends at n, the conditions $\min_{\leq t}(w) \geq 1$ for the first walk and $\max_{\leq\lceil\frac{t}{2}\rceil}(w) = n$ for the second walk are the difficult properties for each walk to satisfy. Next we apply facts proved in Section 4.5.2 to obtain expressions for each term in the summation. The remainder of the upper bound analysis focuses on bounding these expressions.

4.5.2 Maximum Position of a Random Walk

As previously mentioned, our proofs leverage well-known facts about the maximum and minimum position of a random walk, along with tight bounds for these probabilities. This section first gives the result about the maximum and minimum positions of random walks, as well as a connection to Stirling's approximation.

If we are only concerned with a single end point, we can fix the starting location to be 0 by shifting accordingly. In these cases, the following bounds are well-known from probability theory.

Fact 4.5.1 ([RB79, Chapter 3.21]). *For any $t, n \in \mathbb{Z}_{\geq 0}$, we have*

$$\Pr_{w\sim\mathcal{W}(0)}\left(\max_{\leq t}(w) = n\right) = \begin{cases} \Pr\left(w^{(t)} = n\right) = \binom{t}{\frac{t+n}{2}}\frac{1}{2^t} & \text{if } t+n \equiv 0 \pmod 2, \\ \Pr\left(w^{(t)} = n+1\right) = \binom{t}{\frac{t+n+1}{2}}\frac{1}{2^t} & \text{if } t+n \equiv 1 \pmod 2. \end{cases}$$

Proof. For any $k \leq n$, consider a walk $w \in \mathcal{W}(0)$ that satisfies $w^{(t)} = k$ and $\max_{\leq t}(w) \geq n$. Let t^* be the first time that $w^{(t)} = n$, and construct the walk m ending at $2n - k$ such that

$$m^{(i)} = \begin{cases} w^{(i)} & \text{if } 0 \leq i \leq t^*, \\ 2n - w^{(i)} & \text{if } t^* < i \leq t. \end{cases}$$

This reflection map is a measure-preserving bijection, so for $k \leq n$ we have

$$\Pr_{w\sim\mathcal{W}(0)}\left(w^{(t)} = k, \max_{\leq t}(w) \geq n\right) = \Pr_{w\sim\mathcal{W}(0)}\left(w^{(t)} = 2n - k\right). \tag{4.4}$$

Subtracting the probability of the maximum position being at least $n+1$ gives us

$$\Pr_{w\sim\mathcal{W}^{\text{LINE}}(0)}\left(w^{(t)} = k, \max_{\leq t}(w) = n\right) = \Pr_{w\sim\mathcal{W}^{\text{LINE}}(0)}\left(w^{(t)} = 2n - k\right) - \Pr_{w\sim\mathcal{W}^{\text{LINE}}(0)}\left(w^{(t)} = 2(n+1) - k\right).$$

Summing over all $k \leq n$, it follows that

$$\Pr_{w\sim\mathcal{W}^{\text{LINE}}(0)}\left(\max_{\leq t}(w) = n\right) = \Pr_{w\sim\mathcal{W}^{\text{LINE}}(0)}\left(w^{(t)} = n\right) + \Pr_{w\sim\mathcal{W}^{\text{LINE}}(0)}\left(w^{(t)} = n+1\right).$$

Considering the parity of t and n completes the proof. □

This proof of Fact 4.5.1 contains two intermediate expressions for probabilities similar to the ones we want to bound.

Fact 4.5.2. *For any integers $n \geq 0$ and $k \leq n$, we have*

$$\Pr_{w\sim\mathcal{W}(0)}\left(w^{(t)} = k, \max_{\leq t}(w) \geq n\right) = \Pr_{w\sim\mathcal{W}(0)}\left(w^{(t)} = 2n - k\right).$$

Proof. This is the intermediate Equation (4.4) in the proof of Fact 4.5.1. □

Fact 4.5.3. *Let $t, n \in \mathbb{Z}_{\geq 0}$. For any integer $k \leq n$, we have*

$$\Pr_{w\sim\mathcal{W}(0)}\left(w^{(t)} = k, \max_{\leq t}(w) = n\right) = \begin{cases} \binom{t}{\frac{t+2n-k}{2}}\frac{1}{2^t} \cdot \frac{4n-2k+2}{t+2n-k+2} & \text{if } t + k \equiv 0 \pmod 2, \\ 0 & \text{if } t + k \equiv 1 \pmod 2. \end{cases}$$

Proof. Using Fact 4.5.1 and analyzing the parity of the walks gives us

$$\begin{aligned} \binom{t}{\frac{t+2n-k}{2}}\frac{1}{2^t} - \binom{t}{\frac{t+2n-k+2}{2}}\frac{1}{2^t} &= \binom{t}{\frac{t+2n-k}{2}}\frac{1}{2^t} - \frac{t-2n+k}{t+2n-k+2}\binom{t}{\frac{t+2n-k}{2}}\frac{1}{2^t} \\ &= \binom{t}{\frac{t+2n-k}{2}}\frac{1}{2^t} \cdot \frac{4n-2k+2}{t+2n-k+2}, \end{aligned}$$

as desired. □

4.5.3 Lower Bounding Binomial Coefficients

Ultimately, our goal is to give tight lower bounds on closely related probabilities to the ones in Section 4.5.2. To do so, we use various bounds for binomial coefficients that are consequences of Stirling's approximation.

Fact 4.5.4 (Stirling's Approximation)**.** *For any positive integer n, we have*

$$\sqrt{2\pi}n^{n+1/2}e^{-n} \leq n! \leq en^{n+1/2}e^{-n}.$$

An immediate consequence of this is a concentration bound for the binomial coefficients.

Fact 4.5.5. *Let $c, n \in \mathbb{R}_{>0}$ satisfy $c < \sqrt{n}$. For any integer $k \in [\frac{1}{2}(n - c\sqrt{n}), \frac{1}{2}(n + c\sqrt{n})]$, we have*

$$\binom{n}{k} \geq e^{-1-c^2} \frac{2^n}{\sqrt{n}}.$$

Proof. Substituting Stirling's approximation to the definition of binomial coefficients gives

$$\begin{aligned}
\binom{n}{\frac{n-c\sqrt{n}}{2}} &= \frac{n!}{\left(\frac{n-c\sqrt{n}}{2}\right)!\left(\frac{n+c\sqrt{n}}{2}\right)!} \\
&\geq \frac{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n}{e\sqrt{\frac{n-c\sqrt{n}}{2}}\left(\frac{n-c\sqrt{n}}{2e}\right)^{\frac{n-c\sqrt{n}}{2}} e\sqrt{\frac{n+c\sqrt{n}}{2}}\left(\frac{n+c\sqrt{n}}{2e}\right)^{\frac{n+c\sqrt{n}}{2}}} \\
&\geq \frac{2\sqrt{2\pi}}{e^2\sqrt{n}} \cdot \frac{2^n}{\left(1-\frac{c^2}{n}\right)^{\frac{n}{2}}\left(1-\frac{c}{\sqrt{n}}\right)^{-\frac{c\sqrt{n}}{2}}\left(1+\frac{c}{\sqrt{n}}\right)^{\frac{c\sqrt{n}}{2}}} \\
&\geq \frac{2\sqrt{2\pi}}{e^{2+c^2}} \cdot \frac{2^n}{\sqrt{n}} \\
&\geq e^{-1-c^2} \cdot \frac{2^n}{\sqrt{n}},
\end{aligned}$$

as desired. □

4.5.4 Lower Bounding the Minimum Position

Now we bound the probability of the minimum position of a walk in $\mathcal{W}(i)$ being at least 1 after taking exactly t steps.

Lemma 4.5.6. *For any positive integer n, initial position $i \in [\lceil n/2 \rceil]$, and constant $c > 4$, if the step size $t \in [n^2/c, n^2/4]$ then*

$$\Pr_{w \sim \mathcal{W}(i)}(\min_{\leq t}(w) \geq 1) \geq e^{-1-c} \frac{i}{n}.$$

Proof. First observe that

$$\Pr_{w \sim \mathcal{W}(i)}(\min_{\leq t}(w) \geq 1) = \sum_{k=1}^{i} \Pr_{w \sim \mathcal{W}(i)}(\min_{\leq t}(w) = k).$$

By the symmetry of reflection, this sum is equal to

$$\sum_{k=0}^{i-1} \Pr_{w\sim\mathcal{W}(0)}\left(\max_{\le t}(w) = k\right).$$

For each integer $0 \le k \le i-1$, Fact 4.5.1 implies that

$$\Pr_{w\sim\mathcal{W}(0)}\left(\max_{\le t}(w) = k\right) \ge \min\left\{\binom{t}{\frac{t+k}{2}}\frac{1}{2^t}, \binom{t}{\frac{t+k+1}{2}}\frac{1}{2^t}\right\}.$$

Using the assumption that $k \le k+1 \le i \le n \le \sqrt{ct}$, we can apply Fact 4.5.5 to give us

$$\min\left\{\binom{t}{\frac{t+k}{2}}\frac{1}{2^t}, \binom{t}{\frac{t+k+1}{2}}\frac{1}{2^t}\right\} \ge \binom{t}{\frac{t+\sqrt{ct}}{2}}\frac{1}{2^t} \ge e^{-1-c}\frac{1}{\sqrt{t}} \ge e^{-1-c}\frac{1}{n}.$$

The final inequality follows from $t \le n^2/4$. Summing over $0 \le k \le i-1$ gives the desired bound. □

4.5.5 Lower Bounding the Final and Maximum Position

We can similarly use binomial coefficient approximations to bound the probability of a t-step walk terminating at n while never advancing to a position greater than n.

Lemma 4.5.7. *For any initial position $i \in [\lceil n/2 \rceil]$ and step size $\max\{n, n^2/c\} \le t \le n^2/4$ satisfying $t \equiv n - i \pmod 2$, we have*

$$\Pr_{w\sim\mathcal{W}(i)}\left(\max_{\le t}(w) = n, w^{(t)} = n\right) \ge e^{-1-c}\frac{1}{n^2}.$$

Proof. By symmetry, we can rewrite the probability as

$$\Pr_{w\sim\mathcal{W}(0)}\left(\max_{\le t}(w) = n - i, w^{(t)} = n - i\right).$$

Fact 4.5.3 gives that this probability equals to

$$\Pr_{w\sim\mathcal{W}(0)}\left(\max_{\le t}(w) = n-i, w^{(t)} = n-i\right) = \frac{1}{2^t}\binom{t}{\frac{t+n-i}{2}}\frac{2(n-i+1)}{t+n-i+2}.$$

We bound the last two terms separately using our assumptions on t and i. Setting $i = 0$ minimizes the binomial $\binom{t}{(t+n-i)/2}$ for all $i \ge 0$. Setting $i = \lceil n/2 \rceil$ in the numerator and $i = 0$ and $t = n^2/4$ in the denominator minimizes the term $2(n-i+1)/(t+n-i+2)$. It follows that

$$\frac{1}{2^t}\binom{t}{\frac{t+n}{2}}\frac{2(\lfloor n/2\rfloor+1)}{n^2/4+n+2} \ge \frac{1}{2^t}\binom{t}{\frac{t+n}{2}}\frac{n}{n^2}.$$

We reapply Fact 4.5.5 with the observation that $n \le \sqrt{ct}$, which gives us

$$\frac{1}{2^t}\binom{t}{\frac{t+\sqrt{ct}}{2}}\frac{1}{n} \ge \frac{2\sqrt{2\pi}}{e^{2+c}}\cdot\frac{\sqrt{c}}{n^2} \ge e^{-1-c}\frac{1}{n^2},$$

as desired. □

It remains to condition on the minimum position of a walk. This hinges on the following statement, which shows that moving in the opposite direction can only decrease the probability of a walk starting at some $i \in [\lceil n/2 \rceil]$ and ending at n without ever exceeding n.

Lemma 4.5.8. *Let $i \in [\lceil n/2 \rceil]$ be the starting position. At each step $t \le n^2/4$ with $t \equiv n-i \pmod 2$, we have*

$$\Pr_{w\sim\mathcal{W}(i)}\left(w^{(t)} = n, \max_{\le t}(w) = n\right) \ge \Pr_{w\sim\mathcal{W}(i)}\left(w^{(t)} = n, \max_{\le t}(w) = n \mid \min_{\le t}(w) \le 0\right).$$

Proof. Condition on the even $\min_{\le t}(w) \le 0$ and consider the first time t^* the walk hits 0. This means that $i \equiv t^* \pmod 2$ and in turn $n \equiv t - t^* \pmod 2$. The probability of the event $\max_{\le t}(w) = w^{(t)} = n$

during the steps $t^* + 1, \dots, t$ is then at most

$$\Pr_{w \sim \mathcal{W}(0)} \left(w^{(t-t^*)} = \max_{\leq t-t^*}(w) = n \right).$$

Note that we have the inequality because it is possible that we already have $\max_{\leq t^*}(w) > n$. Therefore, it suffices to show for any n and any $t^* \in [t]$ that

$$\Pr_{w \sim \mathcal{W}(0)} \left(w^{(t-t^*)} = \max_{\leq t-t^*}(w) = n \right) \leq \Pr_{w \sim \mathcal{W}(i)} \left(w^{(t)} = \max_{\leq t}(w) = n \right).$$

There are two variables that are shifted from one side of the inequality to the other: the starting position of the walk and the number of steps. In order to prove the inequality, we show that taking more steps and starting further to the right of i can only improve the probability of ending at n and not going above n.

We start by showing that taking more steps only improves this probability. Concretely, we show that

$$\begin{aligned} &\Pr_{w \sim \mathcal{W}(0)} \left(w^{(t-t^*)} = \max_{\leq t-t^*}(w) = n \right) \\ &\qquad \leq \max \left\{ \Pr_{w \sim \mathcal{W}(0)} \left(w^{(t-1)} = \max_{\leq t-1}(w) = n \right), \Pr_{w \sim \mathcal{W}(0)} \left(w^{(t)} = \max_{\leq t}(w) = n \right) \right\}. \end{aligned}$$

There is no guarantee that $t \equiv n \pmod 2$, so we need to consider t or $t-1$ steps depending on the parity. We are guaranteed, however, that $t - 1 \geq t - t^*$ since $t^* \geq 1$, so without loss of generality, we assume $t \equiv t^* \pmod 2$ and show

$$\Pr_{w \sim \mathcal{W}(0)} \left(w^{(t-t^*)} = \max_{\leq t-t^*}(w) = n \right) \leq \Pr_{w \sim \mathcal{W}(0)} \left(w^{(t)} = \max_{\leq t}(w) = n \right).$$

Note that the proof is equivalent when $t - 1 \equiv t^* \pmod 2$.

Using Fact 4.5.3, we have an explicit expression for this probability:

$$\Pr_{w\sim\mathcal{W}(0)}\left(\max_{\le t}(w) = w^{(t)} = n\right) = \frac{1}{2^t}\binom{t}{\frac{t+n}{2}}\frac{2n+2}{t+n+2}.$$

Substituting t by $t-2$ into the equation above and comparing both sides implies that

$$\Pr_{w\sim\mathcal{W}(0)}\left(w^{(t-2)} = \max_{\le t-2}(w) = n\right) \le \Pr_{w\sim\mathcal{W}(0)}\left(w^{(t)} = \max_{\le t}(w) = n\right),$$

because we know from the assumption that

$$\begin{aligned}
\frac{1}{2^{t-2}}\binom{t-2}{\frac{t+n}{2}-1}\frac{2n+2}{t+n} \le \frac{1}{2^t}\binom{t}{\frac{t+n}{2}}\frac{2n+2}{t+n+2} &\iff \frac{(t-2)!}{\left(\frac{t+n-2}{2}\right)!\left(\frac{t-n-2}{2}\right)!(t+n)} \le \frac{t!}{4\left(\frac{t+n}{2}\right)!\left(\frac{t-n}{2}\right)!(t+n+2)} \\
&\iff \frac{1}{t+n} \le \frac{t(t-1)}{(t+n)(t-n)(t+n+2)} \\
&\iff 3t \le n^2 + 2n.
\end{aligned}$$

Inductively applying this argument inductively for $t-2$ proves the inequality.

To complete the proof, it now suffices to consider the starting position i and show that

$$\begin{aligned}
\max\left\{\Pr_{w\sim\mathcal{W}(0)}\left(w^{(t-1)} = \max_{\le t-1}(w) = n\right), \Pr_{w\sim\mathcal{W}(0)}\left(w^{(t)} = \max_{\le t}(w) = n\right)\right\} \\
\le \Pr_{w\sim\mathcal{W}(i)}\left(w^{(t)} = \max_{\le t}(w) = n\right).
\end{aligned}$$

We approach this inequality similarly. First, rewrite the right hand side using the fact that

$$\Pr_{w\sim\mathcal{W}(i)}\left(w^{(t)} = \max_{\le t}(w) = n\right) = \Pr_{w\sim\mathcal{W}(0)}\left(w^{(t)} = \max_{\le t}(w) = n-i\right),$$

and assume that $t \equiv n \pmod 2$, which implies $n \equiv n-i \pmod 2$. Again, using the explicit formula

from Fact 4.5.3 and substituting n by $n-2$ gives us

$$\Pr_{w\sim\mathcal{W}(0)}\left(w^{(t)}=\max_{\leq t}(w)=n\right)\leq\Pr_{w\sim\mathcal{W}(0)}\left(w^{(t)}=\max_{\leq t}(w)=n-2\right),$$

for $t+2\leq n^2$. This is true because of our initial assumption, and it can be inductively applied until $n-i+2$ because $n-i+2\geq\lceil n/2\rceil+1$. Unfortunately, we cannot apply the same proof when $t-1\equiv n \pmod 2$ because this would imply that $n\not\equiv n-i \pmod 2$. Applying the same proof as for $t\equiv n \pmod 2$ allows us to obtain the inequality

$$\Pr_{w\sim\mathcal{W}(0)}\left(w^{(t-1)}=\max_{\leq t-1}(w)=n\right)\leq\Pr_{w\sim\mathcal{W}(0)}\left(w^{(t-1)}=\max_{\leq t-1}(w)=n-i+1\right),$$

because $(t-1)+2\leq(n-i+3)^2$. Therefore, we can conclude the proof by showing that

$$\Pr_{w\sim\mathcal{W}(0)}\left(w^{(t-1)}=\max_{\leq t-1}(w)=n-i+1\right)\leq\Pr_{w\sim\mathcal{W}^{\textsc{Line}}(0)}\left(w^{(t)}=\max_{\leq t}(w)=n-i\right).$$

This is then true if and only if

$$n-i\leq\frac{t}{t-(n-i)}\cdot(n-i+1),$$

which holds for all integers $n-i\geq 0$. □

An immediate corollary of Lemma 4.5.8 is that if we condition on the walk not going to the left of 1, it only becomes more probable to reach n without going above n. Now we prove the main result of this section.

Lemma 4.3.3. *Let $n\in\mathbb{Z}_{\geq 1}$ and $i\in[\lceil n/2\rceil]$ be any starting position. For any constant $c>4$ and $t\in\mathbb{Z}$ such that $n^2/c\leq t\leq n^2/4$ with $t\equiv n-i \pmod 2$, a simple symmetric random walk w on $\mathbb{Z}$ satisfies*

$$\Pr_{w\sim\mathcal{W}^{\textsc{Line}}(i)}\left(w^{(t)}=n,\max_{\leq t}(w)=n,\text{and }\min_{\leq t}(w)\geq 1\right)\geq e^{-2c-2}\frac{i}{n^3}.$$

Proof. Consider any starting position $i \in [\lceil n/2 \rceil]1$ and time $n^2/c \leq t \leq n^2/4$ with $t \equiv n - i \pmod 2$. From the definition of conditional probability, we have

$$\Pr_{w\sim\mathcal{W}(i)}\left(w^{(t)} = n, \max_{\leq t}(w) = n, \min_{\leq t}(w) \geq 1\right) = \Pr_{w\sim\mathcal{W}(i)}\left(w^{(t)} = n, \max_{\leq t}(w) = n \,\middle|\, \min_{\leq t}(w) \geq 1\right) \cdot \Pr_{w\sim\mathcal{W}(i)}\left(\min_{\leq t}(w) \geq 1\right).$$

Lemma 4.5.6 shows that the second term is at least $\exp(-1-c)i/n$. Considering the probability of the event $\min_{\leq t}(w) \geq 1$ (i.e., the complement of $\min_{\leq t}(w) < 1$) and using Lemma 4.5.8 allows us to upper bound the first term by Lemma 4.5.7. It follows from these results that

$$\begin{aligned}\Pr_{w\sim\mathcal{W}(i)}\left(w^{(t)} = n, \max_{\leq t}(w) = n \,\middle|\, \min_{\leq t}(w) \geq 1\right) &\geq \Pr_{w\sim\mathcal{W}(i)}\left(w^{(t)} = n, \max_{\leq t}(w) = n\right) \\ &\geq e^{-1-c}\frac{1}{n^2}.\end{aligned}$$

Putting both inequalities together then gives us

$$\begin{aligned}\Pr_{w\sim\mathcal{W}(i)}\left(w^{(t)} = n, \max_{\leq t}(w) = n, \min_{\leq t}(w) \geq 1\right) &\geq e^{-1-c}\frac{i}{n} \cdot e^{-1-c}\frac{1}{n^2} \\ &= e^{-2-2c}\frac{i}{n^3},\end{aligned}$$

which completes the proof. □

4.5.6 Upper Bounding the Final, Maximum, and Minimum Position

We begin by splitting every t-step walk in half, and instead consider the probability of each walk satisfying the given conditions. In order to give upper bounds for these probabilities, we relax the requirements, which allows us to more easily relate the probabilities to previously known facts about one-dimensional walks that we proved in Section 4.5.2. Additionally, we have to consider all the possible midpoints in $[n]$ because we partitioned the walks.

Lemma 4.5.9. *For any integer $n \in [t]$, we have*

$$\Pr_{w\sim\mathcal{W}(1)}\left(w^{(t)} = n, \max_{\le t}(w) = n, \min_{\le t}(w) \ge 1\right) \le \sum_{i=1}^{n} \Pr_{w\sim\mathcal{W}(1)}\left(w^{(\lfloor t/2\rfloor)} = i, \min_{\le \lfloor t/2\rfloor}(w) \ge 1\right) \cdot \Pr_{w\sim\mathcal{W}(i)}\left(w^{(\lceil t/2\rceil)} = n, \max_{\le \lceil t/2\rceil}(w) = n\right).$$

Proof. By subdividing the length of the walk in half, we consider all possible positions of the walk after half of its steps such that the walk satisfies the maximum and minimum conditions. The second half of the walk must end at n, which implies the maximum position of the walk must be at least n. Thus, the first half of the walk only needs to not go above n. Accordingly, we can write

$$\begin{aligned}\Pr_{w\sim\mathcal{W}(1)}&\left(w^{(t)} = n, \max_{\le t}(w) = n, \min_{\le t}(w) \ge 1\right) \\ &= \sum_{i=1}^{n} \Pr_{w\sim\mathcal{W}(1)}\left(w^{(\lfloor t/2\rfloor)} = i, \max_{\le \lfloor t/2\rfloor}(w) \le n, \min_{\le \lfloor t/2\rfloor}(w) \ge 1\right) \\ &\quad \cdot \Pr_{w\sim\mathcal{W}(i)}\left(w^{(\lceil t/2\rceil)} = n, \max_{\le \lceil t/2\rceil}(w) = n, \min_{\le \lceil t/2\rceil}(w) \ge 1\right).\end{aligned}$$

Relaxing the requirements that the random walks must satisfy cannot decrease the probability, so our upper bound follows. □

Using Fact 4.5.3 we can obtain explicit expressions for each inner term of the summation. We then simplify these expression to achieve a tight bound for the summation in the following lemma.

Lemma 4.5.10. *For all integers $n \in [t]$, we have*

$$\begin{aligned}\Pr_{w\sim\mathcal{W}(1)}&\left(w^{(t)} = n, \max_{\le t}(w) = n, \min_{\le t}(w) \ge 1\right) \\ &\le \sum_{i=1}^{n} \frac{16i(n-i+1)}{t^2} \binom{\lfloor t/2\rfloor}{\frac{\lfloor t/2\rfloor + i - 1}{2}} \frac{1}{2^{\lfloor t/2\rfloor}} \binom{\lceil t/2\rceil}{\frac{\lceil t/2\rceil + (n-i+1) - 1}{2}} \frac{1}{2^{\lceil t/2\rceil}}.\end{aligned}$$

Proof. Apply the upper bound in Lemma 4.5.9 and examine each inner term in the summation. By the symmetry of the walks, there must be an equivalent number of $\lfloor t/2\rfloor$ step walks with endpoints 1 and i that never walk below the position 1 versus those that never walk above i.

Therefore, we have

$$\Pr_{w\sim\mathcal{W}(1)}\left(\min_{\le\lfloor t/2\rfloor}(w)\ge 1, w^{(\lfloor t/2\rfloor)}=i\right)=\Pr_{w\sim\mathcal{W}(1)}\left(\max_{\le\lfloor t/2\rfloor}(w)\le i, w^{(\lfloor t/2\rfloor)}=i\right).$$

Shifting the start of the walk to 0 allows us to apply Fact 4.5.3 because $\max_{\le\lfloor t/2\rfloor}(w)\le i$ is equivalent to $\max_{\le\lfloor t/2\rfloor}(w)=i$ if the walk is required to end at i. It follows that

$$\Pr_{w\sim\mathcal{W}(1)}\left(\min_{\le t}(w)\ge 1, w^{(\lfloor t/2\rfloor)}=i\right)=\frac{2i}{\lfloor t/2\rfloor+i+1}\binom{\lfloor t/2\rfloor}{\frac{\lfloor t/2\rfloor+i-1}{2}}\frac{1}{2^{\lfloor t/2\rfloor}},$$

if the parity is correct and 0 otherwise. This works well for us as an upper bound. Similarly, by shifting the starting position to 0 and applying Fact 4.5.3, we also have

$$\Pr_{w\sim\mathcal{W}(i)}\left(w^{(\lceil t/2\rceil)}=n, \max_{\le\lceil t/2\rceil}(w)=n\right)=\frac{2(n-i+1)}{\lceil t/2\rceil+(n-i+1)+1}\binom{\lceil t/2\rceil}{\frac{\lceil t/2\rceil+(n-i+1)-1}{2}}\frac{1}{2^{\lceil t/2\rceil}}.$$

Applying Lemma 4.5.9, we now have expressions for the term inside the summation, so

$$\begin{aligned}
&\Pr_{w\sim\mathcal{W}(1)}\left(w^{(\lfloor t/2\rfloor)}=i, \min_{\le\lfloor t/2\rfloor}(w)\ge 1\right)\cdot\Pr_{w\sim\mathcal{W}(i)}\left(w^{(\lceil t/2\rceil)}=n, \max_{\le\lceil t/2\rceil}(w)=n\right)\\
&\quad=\frac{2i}{\lfloor t/2\rfloor+i+1}\binom{\lfloor t/2\rfloor}{\frac{\lfloor t/2\rfloor+i-1}{2}}\frac{1}{2^{\lfloor t/2\rfloor}}\cdot\frac{2(n-i+1)}{\lceil t/2\rceil+(n-i+1)+1}\binom{\lceil t/2\rceil}{\frac{\lceil t/2\rceil+(n-i+1)-1}{2}}\frac{1}{2^{\lceil t/2\rceil}}\\
&\quad\le\frac{16i(n-i+1)}{t^2}\binom{\lfloor t/2\rfloor}{\frac{\lfloor t/2\rfloor+i-1}{2}}\frac{1}{2^{\lfloor t/2\rfloor}}\binom{\lceil t/2\rceil}{\frac{\lceil t/2\rceil+(n-i+1)-1}{2}}\frac{1}{2^{\lceil t/2\rceil}}.
\end{aligned}$$

The first equality follows from the fact that

$$\left(\left\lfloor\frac{t}{2}\right\rfloor+i+1\right)\left(\left\lceil\frac{t}{2}\right\rceil+(n-i+1)+1\right)\ge\frac{t^2}{4}.$$

This allows us to derive the desired upper bound and complete the proof. □

The following lemma gives an upper bound for the inner expression in Lemma 4.5.10 by bounding the binomial coefficients via the central binomial coefficients and Stirling's approximation.

Lemma 4.5.11. *For any integer $i \in [n]$, we have*

$$\frac{16i(n-i+1)}{t^2}\binom{\lfloor t/2 \rfloor}{\frac{\lfloor t/2 \rfloor + i - 1}{2}}\frac{1}{2^{\lfloor t/2 \rfloor}}\binom{\lceil t/2 \rceil}{\frac{\lfloor t/2 \rfloor + (n-i+1)-1}{2}}\frac{1}{2^{\lceil t/2 \rceil}} \leq 64\frac{n^2}{t^3}.$$

Proof. Given that $i \in [n]$, we can crudely upper bound $i(n-i+1)$ by n^2. Additionally, we can will use Stirling's approximation for the central binomial coefficient to upper bound our binomial coefficients by

$$\binom{\lceil t/2 \rceil}{\frac{\lceil t/2 \rceil + (n-i+1)-1}{2}} \leq 2^{\lceil t/2 \rceil} \cdot \frac{1}{\sqrt{\lceil t/2 \rceil}},$$

and

$$\binom{\lfloor t/2 \rfloor}{\frac{\lfloor t/2 \rfloor + i - 1}{2}} \leq 2^{\lfloor t/2 \rfloor} \cdot \frac{1}{\sqrt{\lfloor t/2 \rfloor}}.$$

The exponential terms cancel each other out, so we are left with

$$\frac{1}{\sqrt{\lceil t/2 \rceil}} \cdot \frac{1}{\sqrt{\lfloor t/2 \rfloor}} \leq \frac{4}{t},$$

which leads us to the desired upper bound. □

The upper bound in Lemma 4.5.11 is not sufficient for t that are asymptotically less than n^2, so for these values of t we need to give a more detailed analysis. Therefore, we more carefully examine the binomial coefficients that are significantly smaller than the central coefficient for small t. Consequently, the exponential terms will not cancel out in the same way as before in this case. More specifically, we show that the function of t on the right hand side of Lemma 4.5.10 is increasing in t until approximately n^2.

In the following lemma we consider even length walks for simplicity, but the proof for odd length walks follows analogously.

Lemma 4.5.12. *Let $n \geq 20$ and $i \in [n]$ be integers. For all $t \leq n^2/40$, we have*

$$\frac{16i(n-i+1)}{(2t)^2} \cdot \frac{1}{2^{2t}} \binom{t}{\frac{t+i-1}{2}} \binom{t}{\frac{t+(n-i+1)-1}{2}} \leq \frac{16i(n-i+1)}{(2t+4)^2} \cdot \frac{1}{2^{2t+4}} \binom{t+2}{\frac{t+2+i-1}{2}} \binom{t+2}{\frac{t+2+(n-i+1)-1}{2}},$$

where we consider lengths $2t$ and $2t+4$ to ensure that $(2t)/2$ and $(2t+4)/2$ have the same parity.

Proof. Canceling like terms implies that the desired inequality is equivalent to

$$\frac{1}{t^2} \binom{t}{\frac{t+i-1}{2}} \binom{t}{\frac{t+(n-i+1)-1}{2}} \leq \frac{1}{(t+2)^2} \cdot \frac{1}{16} \binom{t+2}{\frac{t+2+i-1}{2}} \binom{t+2}{\frac{t+2+(n-i+1)-1}{2}}.$$

Carefully examining the binomial coefficients shows us that

$$\binom{t}{\frac{t+i-1}{2}} \frac{(t+2)(t+1)}{\frac{t+1+i}{2} \cdot \frac{t+3-i}{2}} = \binom{t+2}{\frac{t+2+i-1}{2}}$$

and

$$\binom{t}{\frac{t+(n-i+1)-1}{2}} \frac{(t+2)(t+1)}{\frac{t+2+(n-i)}{2} \cdot \frac{t+2-(n-i)}{2}} = \binom{t+2}{\frac{t+2+(n-i+1)-1}{2}}.$$

Using these identities, the desired inequality is equal to

$$\frac{1}{t^2} \leq \frac{16^{-1}}{(t+2)^2} \cdot \frac{(t+2)(t+1)}{\frac{t+1+i}{2} \cdot \frac{t+3-i}{2}} \cdot \frac{(t+2)(t+1)}{\frac{t+2+(n-i)}{2} \cdot \frac{t+2-(n-i)}{2}}.$$

Further cancelling like terms and moving the denominators on each side to the numerators of the other side implies that our desired inequality is equivalent to

$$(t+1+i)(t+3-i)(t+2+(n-i))(t+2-(n-i)) \leq t^2(t+1)^2.$$

It is straightforward to see that $(t+1+i)(t+3-i)$ is maximized by $i=1$, and

$$(t+2+(n-i))(t+2-(n-i))$$

is maximized by $n-i=0$. Furthermore, it must be true that either $i \geq n/2$ or $n-i \geq n/2$, so we can upper bound the left hand side of our inequality by substituting $n/2$ for i or $n-i$, and then setting the other terms to the value that maximizes the product. Therefore, we have

$$(t+1+i)(t+3-i)(t+2+(n-i))(t+2-(n-i)) \leq (t+2)^2\,(t+3+n/2)\,(t+3-n/2)\,.$$

To prove our desired inequality it now suffices to show $(t+2)^2\,(t+3+n/2)\,(t+3-n/2) \leq t^2(t+1)^2$, which is equivalent to

$$\left(t+3+\frac{n}{2}\right)\left(t+3-\frac{n}{2}\right) \leq t^2\left(1-\frac{1}{t+2}\right)^2.$$

Expanding both sides of the inequality and rearranging terms yields

$$6t+9+\frac{2t^2}{t+2}-\left(\frac{t}{t+2}\right)^2 \leq \frac{n^2}{4}.$$

Since $2t^2/(t+2) \leq 2t$, it suffices to show that $8t+9 \leq n^2/4$. This is true if $t \leq n^2/40$ and $n \geq 20$. □

We can now prove the main upper bound in this section using the previously derived bounds for the right hand side of the expression in Lemma 4.5.10.

Lemma 4.4.1. *For all $n \geq 20$ and $t \geq n-1$, we have*

$$\Pr_{w\sim\mathcal{W}^{\textsc{Line}}(1)}\left(w^{(t)}=n, \max_{\leq t}(w)=n, \text{and } \min_{\leq t}(w)\geq 1\right) \leq \min\left\{\frac{e^{25}}{n^3}, 64\left(\frac{n}{t}\right)^3\right\}.$$

Proof. Applying Lemmas 4.5.10 and 4.5.11 gives us

$$\Pr_{w\sim\mathcal{W}(1)}\left(w^{(t)} = n, \max_{\le t}(w) = n, \min_{\le t}(w) \ge 1\right) \le \sum_{i=1}^{n} 64\left(\frac{n^2}{t^3}\right),$$

which immediately gives the upper bound $64(n/t)^3$. Similarly, Lemmas 4.5.10 and 4.5.12 imply that for step sizes $t \le n^2/40$ and the value $T = n^2/40$, we have

$$\Pr_{w\sim\mathcal{W}(1)}\left(w^{(t)} = n, \max_{\le t}(w) = n, \min_{\le t}(w) \ge 1\right) \le \sum_{i=1}^{n} \frac{16i(n-i+1)}{T^2}\binom{\lfloor T/2\rfloor}{\frac{\lfloor T/2\rfloor+i-1}{2}}\frac{1}{2^{\lfloor T/2\rfloor}}\binom{\lceil T/2\rceil}{\frac{\lceil T/2\rceil+(n-i+1)-1}{2}}\frac{1}{2^{\lceil T/2\rceil}}.$$

Next, we use Lemma 4.5.11 and sum from 1 to n to obtain the inequality

$$\Pr_{w\sim\mathcal{W}(1)}\left(w^{(t)} = n, \max_{\le t}(w) = n, \min_{\le t}(w) \ge 1\right) \le \sum_{i=1}^{n} \frac{64n^2}{T^3} \le \frac{64(40)^3}{n^3} \le \frac{e^{25}}{n^3},$$

for all $t \le n^2/40$. Using the fact that $64(n/t)^3$ is a decreasing function in t, we have

$$64\left(\frac{n}{t}\right)^3 \le \frac{e^{25}}{n^3},$$

for all $t \ge n^2/40$, which completes the proof. □

4.6 Extension to Higher Dimensions

In this section we show how to extend our analysis of the upper and lower bounds on the transience class to d-dimensional cubic hypergrids. We combine both of our results in the following theorem.

Theorem 4.1.4. *For any integer $d \ge 2$, the transience class of the Abelian sandpile model on the d-dimensional cubic hypergrid with n^d vertices is $O(n^{3d-2}\log^{d+2} n)$ and $\Omega(n^{3d-2})$.*

We denote by $\textsc{Hypergrid}_{d,n}$ the d-dimensional cubic hypergrid with n^d vertices and construct it analogously to $\textsc{Grid}_n$. Its vertex set is $[n]^d \cup \{v_{\text{sink}}\}$ and its edges connect any pair of vertices

that differ in one coordinate. Vertices on the boundary have additional edges connecting to v_{sink} so that every non-sink vertex has degree $2d$. We use the vector notation $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_d)$ to identify non-sink vertices. We can decouple a walk w on $\textsc{Hypergrid}_{d,n}$ into one-dimensional walks $w_1, w_2, \ldots, w_d$ so that each step of a random walk on $\textsc{Hypergrid}_{d,n}$ can be understood as choosing a random direction with probability $1/d$ and then a step in the corresponding one-dimensional walk with probability 1/2.

Our bounds for the two-dimensional grid heavily relied on decoupling walks into interleaved one-dimensional walks, and applying bounds from Section 4.5 for simple symmetric walks. Generalizing these bounds to d-dimensional hypercubes follows similarly and only requires simple extensions of our lemmas for two-dimensional grids. Therefore, we will list the necessary lemmas from previous sections and comment on the minor modifications needed to give analogous lemmas for hypergrids. The upper bound proof requires several key lemmas and is more involved, whereas extending the lower bound only requires one simple addition to our proof in Section 4.4.

4.6.1 Upper Bounding the Transience Class

Since Theorem 4.2.4 from [CV12] relies on non-sink vertices having constant degree, we assume that d is constant and that all non-sink vertices have degree $2d$. In addition to utilizing properties of one-dimensional walks, specifically Lemma 4.3.3 proven in Section 4.5, the proof of our higher-dimensional upper bound relies on four key lemmas:

- Lemma 4.2.10 — The source vertex can be swapped with a any non-sink vertex at the expense of a $O(\log n)$ approximation factor in the potential.
- Lemma 4.3.1 — An upper bound for the sum of all vertex potentials that we achieve by factoring the expression into one-dimensional vertex potentials.
- Lemma 4.3.2 — For any interior vertex, the opposite corner vertex minimizes the potential up to a constant factor.

- Lemma 4.3.5 — A lower bound for the voltage $\pi_{(n,n)}(\mathbf{u})$, for any vertex $\mathbf{u}$ in the upper-right quadrant of $\textsc{Grid}_n$.

Now we describe how to extend each of these lemmas to a constant number of dimensions. These results essentially follow from decoupling walks into one-dimensional walks.

Lemma 4.6.1. *For any pair of non-sink vertices* $\mathbf{u}$, $\mathbf{v}$ *in* $\textsc{Hypergrid}_{d,n}$, $\pi_{\mathbf{u}}(\mathbf{v}) \leq (8 \log n + 4)\, \pi_{\mathbf{v}}(\mathbf{u})$.

Proof. This is a consequence of Rayleigh's monotonicity theorem. Fix an underlying $n \times n$ subgraph of the hypergrid with corners at the source and sink, and set the rest of the resistors to infinity. The upper bound for this $n \times n$ grid is an upper bound for the hypergrid. □

Our d-dimensional generalization of Lemma 4.3.1 follows from Lemma 4.6.1 and projecting coordinate walks into one-dimension.

Lemma 4.6.2. *For any non-sink vertex* $\mathbf{u}$ *in* $\textsc{Hypergrid}_{d,n}$, *we have*

$$\sum_{\mathbf{v} \in V} \pi_{\mathbf{u}}(\mathbf{v}) = O\left(\log^{d+1}(n) \prod_{i=1}^{d} \mathbf{u}_i\right).$$

Proof. We follow the proof structure of Lemma 4.3.1. First, take the reciprocal by Lemma 4.6.1:

$$\pi_{\mathbf{u}}^{\textsc{Hypergrid}_{d,n}}(\mathbf{v}) = O\left(\pi_{\mathbf{v}}^{\textsc{Hypergrid}_{d,n}}(\mathbf{u}) \log n\right).$$

Performing the decoupling step across all d dimensions gives us

$$\begin{aligned}
\pi_{\mathbf{v}}^{\textsc{Hypergrid}_{d,n}}(\mathbf{u}) &= \Pr_{w \sim \mathcal{W}^{\mathbb{Z}^d}(\mathbf{u})}(w \text{ hits } \mathbf{v} \text{ before leaving } \textsc{Hypergrid}_{d,n}) \\
&\leq \Pr_{w \sim \mathcal{W}^{\mathbb{Z}^d}(\mathbf{u})}\left(\bigcap_{i=1}^{d} w_i \text{ hits } \mathbf{v}_i \text{ before } v_{\text{sink}}\right) \\
&= \prod_{i=1}^{d} \Pr_{w \sim \mathcal{W}^{\mathbb{Z}^d}(\mathbf{u})}(w_i \text{ hits } \mathbf{v}_i \text{ before } v_{\text{sink}}) \\
&= \prod_{i=1}^{d} \pi_{\mathbf{v}_i}^{\textsc{Path}_n}(\mathbf{u}_i).
\end{aligned}$$

Similar to before, factoring the sum over all choices of $\mathbf{v}$ gives us

$$\sum_{\mathbf{v} \in V(\textsc{Hypergrid}_{d,n})} \pi_{\mathbf{v}}^{\textsc{Hypergrid}_{d,n}}(\mathbf{u}) \leq \prod_{i=1}^{d} \left(\sum_{\mathbf{v}_i=1}^{n} \pi_{\mathbf{v}_i}^{\textsc{Path}_n}(\mathbf{u}_i) \right).$$

The remainder of the proof proceeds in the same way as the proof of Lemma 4.3.1, where we use closed-form expressions for the one-dimensional voltages. □

Next, we generalize our proof of Lemma 4.3.2 to higher dimensions. This is a direct consequence of the orthogonality between dimensions in our analysis.

Lemma 4.6.3. *If* $\mathbf{u}$ *is a non-sink vertex of* $\textsc{Hypergrid}_{d,n}$ *such that* $\mathbf{u}_i \in [\lceil n/2 \rceil]$*, then for all* $i \in [d]$ *we have*

$$\pi_{\mathbf{u}}(\mathbf{v}) \geq \left(\frac{1}{2d} \right)^{d} \pi_{\mathbf{u}}((n, n, \ldots, n)).$$

Proof. Extend the proof of Lemma 4.3.2 by reflecting walks across the $(d-1)$-dimensional hyperplane perpendicular to the chosen axis instead of a line. □

Last, we generalize Lemma 4.3.5. The key idea of this result is to considers walks of length $\Theta(n^2)$ and show that there is a constant fraction such that both dimensions have taken $\Theta(n^2)$ steps, which then allows us to apply Lemma 4.3.3 to each possible walk. To do this, we essentially union bound Lemma 4.3.4 over all d dimensions, which shows that $\Theta(n^2)$ walk lengths take $\Theta(n^2)$ steps in each direction with probability at least 2^{-d}.

Lemma 4.6.4. *For* $n \geq 10$ *and* $\mathbf{u} \in V(\textsc{Hypergrid}_{d,n})$ *such that* $1 \leq \mathbf{u}_i \leq \lceil n/2 \rceil$ *for* $1 \leq i \leq d$*, we have*

$$\pi_{(n,n,\ldots,n)}(\mathbf{u}) = \Omega\left(\frac{\prod_{i=1}^{d} \mathbf{u}_i}{n^{3d-2}} \right).$$

Proof. Decouple walks $w \in \mathcal{W}(\mathbf{u} \to (n, n, \ldots, n))$ into one-dimensional walks $w_i \in \mathcal{W}^{\textsc{Line}}(\mathbf{u}_i)$, and view $\pi_{(n,n,\ldots,n)}(\mathbf{u})$ as the probability that each walk w_i is present on n at the same time before any leaves the set $[n]$. If each walk takes $t_1, t_2, \ldots, t_d$ steps, respectively, then the total number of

possible interleavings of these walks is the multinomial coefficient

$$\binom{t_1 + t_2 + \cdots + t_d}{t_1, t_2, \ldots, t_d} = \frac{(t_1 + t_2 + \cdots + t_d)!}{t_1! t_2! \cdots t_d!}.$$

Just as before, we can obtain the lower bound

$$\pi_{(n,n,\ldots,n)}(\mathbf{u}) \geq \sum_{t_1,t_2,\ldots,t_d \geq 0} \frac{\binom{t_1+t_2+\cdots+t_d}{t_1,t_2,\ldots,t_d}}{d^{t_1+t_2+\cdots+t_d}} \prod_{i=1}^{d} \frac{1}{2} \Pr\left(w_i^{(t_i-1)} = n-1, \max_{\leq t_i-1}(w) = n-1, \min_{\leq t_i-1}(w) \geq 1\right).$$

To apply Lemma 4.3.3 to each walk, we need each t_i to be in the interval $[1/c\, n^2, 1/4\, n^2]$ for $c = 16d$. Then we consider all walks of length $t \in [1/8\, n^2, 1/4\, n^2]$, where $t = t_1 + t_2 + \cdots + t_d$, and show that a constant fraction of these walks satisfy $t_i \geq n^2/c$ with t_i having the correct parity. Note that we can ignore the parity conditions by simply lower bounding the probability of all having correct parity by 4^{-d}. Now it remains to show that all walks satisfy the inequality $t_i \geq n^2/c$ with constant probability.

Consider the probability that $t_1 \geq n^2/c$. The other dimensions follow identically. Letting each dimension take at least n^2/c steps introduces dependence, so we instead consider the probability that $t_1 \geq n^2/c$ and condtion on $t_2, t_3, \ldots, t_d \geq n^2/c$ (which can only decrease the probability of the event $t_1 \geq n^2/c$). This is equivalent to fixing n^2/c steps in each of those directions and randomly choosing all remaining steps with probability $1/d$ for each direction. The remaining number of steps is then at least dn^2/c by our assumption that $t \geq n^2/8$. Therefore, the expected number of steps in the first dimension is at least n^2/c, which implies $t_1 \geq n^2/c$ with probability at least $1/2$. Multiplying this probability over all dimensions gives $t_i \geq n^2/c$ with probability at least 2^{-d}.

Thus, there are $O(n^2)$ values of t that we can decompose into one-dimensional walks, each occurring with constant probability. Applying Lemma 4.3.3 to each decomposition and summing

$$\Omega\left(\prod_{i=1}^{d} \frac{\mathbf{u}_i}{n^3}\right)$$

over $O(n^2)$ possible walk lengths proves the claim. □

Proof of Theorem 4.1.4 (Upper Bound). We prove that $\mathrm{tcl}(\textsc{Hypergrid}_{d,n}) = O(n^{3d-2}\log^{d+2} n)$ using Theorem 4.2.4. For any $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_d)$ in the top-left orthant of $\textsc{Hypergrid}_{d,n}$, it follows that

$$\begin{aligned}
\max_{\mathbf{u},\mathbf{v}\in V\setminus\{v_{\text{sink}}\}} \left(\sum_{\mathbf{x}\in V} \pi_{\mathbf{u}}(\mathbf{x})\right) \pi_{\mathbf{u}}(\mathbf{v})^{-1} &\le \max_{\mathbf{u}\in V\setminus\{v_{\text{sink}}\}} \left(\sum_{\mathbf{x}\in V} \pi_{\mathbf{u}}(\mathbf{x})\right) \frac{(2d)^d}{\pi_{\mathbf{u}}((n,n,\dots,n))} \\
&= \max_{\mathbf{u}\in V\setminus\{v_{\text{sink}}\}} \left(\sum_{\mathbf{x}\in V} \pi_{\mathbf{u}}(\mathbf{x})\right) \frac{O(\log n)}{\pi_{(n,n,\dots,n)}(\mathbf{u})} \\
&= \max_{\mathbf{u}\in V\setminus\{v_{\text{sink}}\}} O\left(\log^{d+1}(n) \prod_{i=1}^{d} \mathbf{u}_i\right) O\left(\frac{n^{3d-2}\log n}{\prod_{i=1}^{d}\mathbf{u}_i}\right) \\
&= O\left(n^{3d-2}\log^{d+2} n\right).
\end{aligned}$$

This completes this part of the proof. □

4.6.2 Lower Bounding the Transience Class

Extending our lower bound to d-dimensional hypergrids is a simple consequence of decoupling d-dimensional walks into one-dimensional walks since we only need to generalize the upper bound in Lemma 4.4.3 to

$$\begin{aligned}
\pi_{(n,n,\dots,n)}((1,1,\dots,1)) \le\ & d \max\left\{\Pr_{w\sim\mathcal{W}(1)}\left(w^{(t)} = n, \max_{\le t}(w) = n, \min_{\le t}(w) \ge 1\right) \,:\, t \in \mathbb{Z}_{\ge 0}\right\} \\
& \cdot \sum_{t\ge 0} \Pr_{w\sim\mathcal{W}(1)}\left(w^{(t)} = n, \max_{\le t}(w) = n, \min_{\le t}(w) \ge 1\right).
\end{aligned}$$

We achieve this by replacing the negative binomial distribution with the negative multinomial distribution.

Fact 4.6.5. *For any nonnegative integer t_1, we have*

$$\sum_{t_2,\dots,t_d\ge 0} \binom{t_1 + t_2 + \cdots + t_d}{t_1, t_2, \dots, t_d} \frac{1}{d^{t_1+t_2+\cdots+t_d}} = d.$$

Proof. Consider the proof of Fact 4.4.2 using the negative multinomial distribution. □

We are now ready to complete the proof of Theorem 4.1.4.

Proof of Theorem 4.1.4 (Lower Bound). Applying Lemma 4.4.1 and Lemma 4.4.4, it follows that

$$\pi_{(n,n,\ldots,n)}((1,1,\ldots,1)) = O\left(\left(\frac{1}{n^3}\right)^{d-1}\frac{1}{n}\right) = O\left(n^{-3d+2}\right).$$

By Theorem 4.2.4, it follows that $\mathrm{tcl}(\textsc{Hypergrid}_{d,n}) = \Omega(n^{3d-2})$, as desired. □

4.7 Summary and Discussion

In this chapter we prove bounds that are tight up to polylogarithmic factors for the transience class of the Abelian sandpile model on the $n \times n$ grid, solving an open problem originally posed by Babai and Gorodezky in [BG07]. Building on previous work of Choure and Vishwanathan [CV12] that shows how to bound the transience class in terms of vertex potentials when the underlying graph is viewed as an electrical network, we interpret voltages of internal vertices as escape probabilities of random walks on the grid and apply a decoupling technique to achieve tight inequalities for the vertex potentials. Our analysis heavily relies on potential reciprocity (Lemma 4.2.9) and the low effective resistance of the grid (Lemma 4.2.8). Combining these ideas allows us to swap voltage sources with minimal overhead, which then reduces our approach to analyzing one-dimensional random walks conditioned on various boundary events. The techniques we present in this chapter suggest that low effective resistance captures a different but possibly similar behavior to high conductance and edge expansion for random walks on graphs. Investigating the distinction between the roles of effective resistance and conductance could be an important step forward for building a theory of discrete diffusion processes analogous to the mixing time of Markov chains.

CHAPTER 5

APPROXIMATELY SAMPLING ELEMENTS WITH FIXED RANK IN GRADED POSETS

In this chapter we return our focus to the design and analysis of Markov chain Monte Carlo algorithms. We show that for a particular class of graded posets, biased Markov chains that walk along the edges of *Hasse diagrams* allow us to approximately generate uniform samples with a given rank in polynomial time. We present a new technique called Boltzmann sampling with *balanced bias*, which allows us to bound rejection rates using only mixing times of the underlying Markov chains. We then investigate how this method can be applied to sampling integer partitions.

5.1 Introduction

Graded posets are partially ordered sets equipped with a unique rank function that both respects the partial order and such that neighboring elements in the Hasse diagram of the poset have ranks that differ by ±1. Graded posets arise throughout combinatorics, including permutations ordered by numbers of inversions, geometric lattices ordered by volume, and independent sets and matchings ordered by cardinality. Sometimes we find rich underlying structures that allow us to directly count, and therefore sample, fixed rank elements of a graded poset. In other cases, efficient methods are unlikely to exist, so Markov chains offer the best approach to sampling and approximate counting.

Jerrum and Sinclair [JS96] showed that we can sample matchings of a fixed size by introducing a bias parameter λ that assigns weight proportional to $\lambda^{|m|}$ to each matching m. In particular, they use the fact that for any graph G, the sequence $(a_i)_{i=0}^n$ of the number of matchings in G of size i is log-concave [HL72]. Letting $b_i = a_i\lambda^i$, this implies the sequence $(b_i)_{i=0}^n$ is also log-concave and therefore unimodal for all λ. Setting $\lambda = a_k/a_{k+1}$ forces the mode of the distribution $(b_i)_{i=0}^n$ to be k,

so samples drawn from this distribution are of the desired size with probability at least $1/(n+1)$. Jerrum and Sinclair showed that the matching Markov chain is rapidly mixing for all λ, so it can find matchings of fixed size k efficiently whenever $1/\text{poly}(n) < \lambda < \text{poly}(n)$. This condition is not always satisfied, but the more sophisticated algorithm of Jerrum, Sinclair, and Vigoda circumvents this issue [JSV06]. Log-concavity is critical to this argument in order to conclude that there is a value of λ for which samples of the desired size occur with high enough probability.

This follows a common approach used in physics for which we would like to sample from a *microcanonical ensemble*, i.e., the states with a fixed energy, from a much larger *canonical* (or *grand canonical*) *ensemble*, where the energies are allowed to vary due to interactions with the external environment. In particular, given an input parameter λ that is often related to temperature, a configuration $x \in \Omega$ has *Boltzmann* weight $\pi(x) = \lambda^{r(\sigma)}/Z$, where $r(x)$ is the rank of x and Z is the normalizing constant. Elements x sampled from this distribution are uniformly distributed, conditioned on their rank. The choice of λ controls the expected rank of the distribution, so simulations of the Markov chain at various λ can be useful for understanding properties of configurations with a fixed energy. Typically, however, there is no a priori guarantee that this approach will enable us to sample configurations of a given size efficiently.

Our main example throughout this chapter will be sampling and counting *integer partitions*, possibly restricted to boundary conditions. An integer partition of nonnegative integer n is a decomposition of n into a nonincreasing sequence of positive integers that sum to n. For example, the seven partitions of 5 are: (5), $(4,1)$, $(3,2)$, $(3,1,1)$, $(2,2,1)$, $(2,1,1,1)$, and $(1,1,1,1,1)$. Integer partitions are commonly represented by staircase walks in $\mathbb{Z}^2$ known as *Young diagrams*, where the heights of the columns represent distinct pieces of the partition. Partitions of n have exactly n squares, i.e., the *area* of the diagram, and their column heights are nonincreasing. Partitions arise in many contexts, include exclusion processes [CMO07], random matrices [Oko02], representation theory [Jam06], juggling patterns [ABCN14], and growth processes [GPR09] (see, e.g., [And98]).

Several general approaches have been developed to sample elements of fixed rank from a graded

poset, with varying success. The three main approaches for sampling are dynamic programming algorithms using self-reducibility, Boltzmann samplers using geometric random variables, and Markov chain Monte Carlo. The first two approaches require methods to estimate the number of configurations of each size. Therefore, Markov chains offer the most promising approach for sampling when these estimates are unavailable.

All three of these approaches have been studied extensively in the context of sampling integer partitions. The first approach uses dynamic programming and generating functions to iteratively count the number of partitions of a given type. Nijinhuis and Wilf [NW78] gave a recursive algorithm that uses dynamic programming to compute a table of exact values. Their algorithm takes $O(n^{5/2})$ time and space for preprocessing and $O(n^{3/2})$ time per sample. Squire [Squ93] improved on their algorithm and showed that $O(n^2)$ time for preprocessing and $O(n^{3/2}\log(n))$ time per sample can be achieved using Euler's pentagonal recurrence and a more efficient search method. We note that the time and space complexity of these algorithms account for the fact that each value of $p(n)$ and the intermediate summands require $O(n^{1/2})$ bits by the Hardy-Ramanujan formula. Therefore, even if time is available, dynamic programming approaches for exact sampling break down in practice on a single machine when $n \geq 10^6$ due to space constraints.

Boltzmann samplers offer a more direct method for sampling that avoids the computationally expensive task of counting partitions. A Boltzmann sampler generates samples from a larger combinatorial class with probability proportional to the Boltzmann weight $\lambda^{|x|}$, where $|x|$ is the size of the partition. Samples of the same size are drawn uniformly at random, and the algorithm rejects those that fall outside of the target size [DFLS04, FFP07]. The value λ is chosen to maximize the yield of samples of our target size n. Fristedt [Fri93] proposed an approach that quickly generates a random partition using appropriate independent geometric random variables. His approach exploits the factorization of the generating function for $p(n)$ and can be interpreted as sampling Young diagrams x in the $n \times \infty$ grid with probability proportional to the Boltzmann weight $\lambda^{|x|}$. Recently, Arratia and DeSalvo [AD16] gave a probabilistic approach that is

substantially more efficient than previous algorithms, thus allowing for fast generation of random partitions for significantly larger numbers (e.g., $n \geq 10^6$). Building on the work of Fristedt, they introduce the *probabilistic divide-and-conquer* (PDC) method to generate random partitions of n in optimal $\tilde{O}(n^{1/2})$ expected time and space, where $\tilde{O}$ suppresses $\log n$ factors. Their PDC algorithm also uses independent geometric random variables to generate a partition, but does so recursively in phases. It is worth noting that PDC achieves superior performance relative to conventional Boltzmann sampling by rejecting impossible configurations in early phases.

Lastly, stochastic approaches that utilize Markov chains have produced a similarly rich corpus of work, but until now have not provided rigorous polynomial bounds. One popular direction uses Markov chains based on coagulation and fragmentation processes that allow pieces of the partition to be merged and split [Ald99, BP07]. Ayyer et al. [ABCN14] recently proposed several natural Markov chains on integer partitions in order to study juggling patterns. In all of these works, most of the effort has been to show that the Markov chains converge to the uniform distribution over partitions and often use stopping rules in order to generate samples. Experimental evidence suggests these chains converge quickly to the correct equilibrium, but they lack explicit bounds.

5.1.1 Main Results

For any graded poset, let Ω_k be the elements of rank k and let $\Omega = \bigcup_{i=0}^{n} \Omega_i$ be the entire poset. We show that *provably efficient* Boltzmann samplers on Ω_k can be easily constructed from certain rapidly mixing Markov chains on the Hasse diagram of the entire poset Ω under mild conditions. We apply this technique to design the first provably efficient Markov chain based algorithms for sampling integer partitions of an integer n, permutations with a fixed number of inversions, and lozenge tilings with fixed average height. Unlike all other methods for sampling that depend on efficient counting techniques, our results extend to interesting subintervals of these posets, such as partitions with at least k pieces with size greater than ℓ, or partitions into pieces with distinct sizes, or many other such restricted classes. For these restrictions, our results provide the first

sampling algorithms that do not require the space-expensive task of counting.

We focus on the example of integer partitions of n and prove that there is a Markov chain Monte Carlo algorithm for uniformly sampling partitions of n from a large family of *region-restricted partitions* (i.e., Young diagrams restricted to any simply-connected bounding region). The Markov chain on the Hasse diagram for partitions is the natural "mountain-valley" chain studied for staircase walks, tilings, and permutations [LRS01]. The transition probabilities are designed such that we generate a diagram x with weight proportional to $\lambda^{|x|}$. Previous work on biased card shuffling [BBHM05] and growth processes [BBHM05, LPW17, GPR09] shows that this chain is rapidly mixing for any constant λ on well-behaved regions.

In the general setting of sampling from a graded poset, our algorithm is similar to Boltzmann samplers that heuristically sample elements of a given size, but often without rigorous analysis. We establish, however, conditions under which these algorithms can be shown to be efficient, including restricted settings for which no other methods provide guarantees on both efficiency and accuracy. For example, we show that our approach can generate uniformly random partitions of n in $O(n^{9/4})$ expected time with only $O(n^{1/2}\log(n))$ space. Using *coupling from the past* [PW96], we can generate samples of the target size exactly from the uniform distribution.

Although our algorithm is slower than recent results for sampling unrestricted partitions using independent geometric random variables [AD16, Fri93] (in the settings where those methods apply), our method is significantly more versatile. The Markov chain algorithm readily adapts to various restricted state spaces, such as sampling partitions with bounded size and numbers of parts, partitions whose Durfee square is bounded (i.e., the largest square contained in the Young diagram), and partitions with prescribed gaps between successive pieces including partitions into pieces with distinct sizes. For bounded regions, our algorithm uses $O(n^{1/2}\log(n))$ space, and thus is usually more suitable than other approaches with substantially larger space requirements.

Finally, we achieve similar results for sampling from fixed a rank in other graded posets. These include permutations with a fixed number of inversions and lozenge tilings with a given average

height (i.e., the height function representation of the tilings as explained in [LRS01]). Kenyon and Okounkov [KO07] explored the limit shapes of tilings with fixed volume and showed that such constraints simplified some arguments, but there has not been rigorous work addressing the sampling problem.

5.1.2 Techniques

We start by presenting a new argument that shows how to construct Boltzmann samplers with provable performance guarantees, even in cases where the underlying distributions are not known to be unimodal, provided that the Markov chain is rapidly mixing on the entire Hasse diagram. We prove that there always exists a *balanced bias* parameter λ^* that we can find efficiently and that allows us to generate configurations of the target size with probability at least $1/\text{poly}(n)$. The target set is no longer guaranteed to be the mode of the distribution, as generally required, but we show that rejection probability will not be too high. We carefully define a polynomial-sized set from which the bias parameter λ will be chosen. Then we show that at least one bias parameter in this set will define a distribution satisfying $\sum_{i=1}^{k} \Pr(\Omega_i) \geq 1/c(n)$ and $\sum_{i=k+1}^{n} \Pr(\Omega_i) \geq 1/c(n)$, for some polynomial $c(n)$. Since the Markov chain $\mathcal{M}$ changes the rank by at most 1 in each step, we use a conductance-based argument to show that we necessarily generate samples of size exactly k with probability at least $1/\tau(\varepsilon)$, where $\tau(\varepsilon)$ is the mixing time of $\mathcal{M}$. Therefore, when the chain is rapidly mixing for all such values of λ, samples of size k must occur with non-negligible probability. This method based on balanced bias parameters is quite general and circumvents the need to make strong assumptions about the underlying distributions.

Our main tool is to use biased Markov chains and Boltzmann sampling to generate samples of the targets size k. We assign probability $\lambda^{r(x)}/Z$ to every element $x \in \Omega$, where $r(x)$ is the rank of x in the poset and Z is the normalizing constant. When the underlying sequence $f(i) = |\Omega_i|\lambda^i$ is known to be log-concave in i (e.g., unrestricted integer partitions or permutations with a fixed number of inversions), we can provide stronger guarantees than what is achievable by our general

balanced bias algorithm. For integer partitions specifically, several observations allow us to improve the running time of our algorithm. First, instead of sampling Young diagrams in an $n \times n$ lattice region, we restrict to diagrams lying in the first quadrant of $\mathbb{Z}^2$ below the curve $y = 2n/x$, since this region contains all the Young diagrams of interest and has area $\Theta(n \log n)$, allowing the Markov chain to converge faster. Second, we improve the bounds on the mixing time for our particular choice of λ given by path coupling with exponential metrics in [GPR09] by instead carefully analyzing a recent result in [LPW17]. Lastly, we show how to *salvage* a substantial fraction of the samples rejected by Boltzmann sampling, which ultimately allows us to increase the success probability to at least $\Omega(1/n^{1/4})$. Putting these observations together, we are able to design a uniform sampling algorithm for integer partitions of size n such that the underlying Markov chain converges in $O(n^2)$ time and the number of trials needed to generate a partition of size n is $O(n^{1/4})$ in expectation. Finally, we show how to optimize the space required to implement the Markov chain. Since Young diagrams have at most $O(n^{1/2})$ corners, each diagram can be stored in $O(n^{1/2} \log n)$ space.

5.2 Preliminaries

We start by formally defining partial orders, graded posets, and rank generating functions. Then we briefly comment on designing Markov chains that walk on Hasse diagrams. A *partial order* is a binary relation $\leq$ over a set Ω that satisfies the following three axioms for all elements $x, y, z \in \Omega$:

1. $x \leq x$ (reflexivity).
2. If $x \leq y$ and $y \leq x$, then $x = y$ (antisymmetry).
3. If $x \leq y$ and $y \leq z$, then $x \leq z$ (transitivity).

A set equipped with a partial order is called a *partially ordered set* (also known as a *poset*). We say that two elements x and y in a poset Ω are *comparable* if $x \leq y$ or $y \leq x$; otherwise x and y

are *incomparable*. An element x is *strictly less than* y if $x \leq y$ and $x \neq y$. Further, we say that x is *covered* by y if x is strictly less than y and there does not exist and element $z \in \Omega \setminus \{x, y\}$ such that $x \leq z \leq y$. We denote the property x covered by y by $x \lessdot y$. The cover structure of a poset induces a directed graph on the elements of Ω where there is an edge from x to y if $x \lessdot y$. The *Hasse diagram* of a poset is an undirected version of the graph given by covering relations.

A *graded poset* is a partially ordered set equipped with a rank function $r : \Omega \to \mathbb{Z}_{\geq 0}$ that satisfies the following properties for all elements $x, y \in \Omega$:

1. $r(x) = 0$ for all minimal elements of Ω.
2. If $x \leq y$, then $r(x) \leq r(y)$ (compatible with the partial order).
3. If $x \lessdot y$, then $r(y) = r(x) + 1$ (consistent with the covering relation).

Some examples of graded posets and their rank functions are the Boolean lattice of finite subsets ordered by inclusion where the rank function is the cardinality of the subset and the positive integers ordered by divisibility where the rank function is the number of prime factors counted with multiplicity.

We define the rank of a graded poset Ω to be $R = \max\{r(x) : x \in \Omega\}$, and we denote its *rank generating function* by

$$F(\lambda) = \sum_{x \in \Omega} \lambda^{r(x)}.$$

Let $\Omega_k = \{x \in \Omega : r(x) = k\}$ be the set of elements with rank k, and let $a_k = |\Omega_k|$ denote the number of such elements. For any $\lambda \geq 0$, the Gibbs measure of each $x \in \Omega$ is $\pi(x) = \lambda^{r(x)}/Z$, where

$$Z = F(\lambda) = \sum_{i=0}^{R} a_i \lambda^i$$

is the normalizing constant for the Boltzmann distribution.

The Markov chains that we consider in this chapter walk along the edges of the Hasse diagram of the corresponding graded poset. It follows that the rank of the current configuration can change

by at most 1 in each step. Therefore, we use a Metropolis filter [Met+53] in each step so that the stationary distribution efficiently converges to the Boltzmann distribution parameterized by λ. For a thorough review of Markov chains and mixing times, we refer the reader to Section 2.1.

5.3 Balanced Bias and Rejection Rates

Our ultimate goal is to uniformly sample an element $x \in \Omega_k$, for a given $k \in \{0, 1, \ldots, R\}$. To achieve this, we repeatedly sample from favorable Boltzmann distributions over all of Ω until we encounter an element of rank k. We show that if mild conditions on the coefficients of the rank generating function are satisfied and that the Markov chain $\mathcal{M}$ on Ω is rapidly mixing for all of the values of λ we consider, then the uniform sampling procedure from the set Ω_k will be efficient. Specifically, we only require that $R = O(\text{poly}(n))$ and $1 \leq a_i \leq c(n)^i$ for some polynomial $c(n)$.

To formalize our claim, we first assume that the polynomial $c(n) \geq 2$. Then for $t \geq 0$, we define

$$\beta_t \stackrel{\text{def}}{=} \log\left(c(n)^{-1}\right) + \frac{t \log(c(n))}{R} \tag{5.1}$$

and

$$\lambda_t \stackrel{\text{def}}{=} e^{\beta_t} = c(n)^{t/R-1}, \tag{5.2}$$

where $\log x$ denotes the natural logarithm. Next, for every $x \in \Omega$, let

$$\Pr_t(x) \stackrel{\text{def}}{=} \frac{\lambda_t^{r(x)}}{Z_t} = \frac{\lambda_t^{r(x)}}{F(\lambda_t)}. \tag{5.3}$$

The sequence $(\lambda_t)_{t=0}^{\infty}$ is constructed in such a way that at most R^2 values need to be considered.

Lemma 5.3.1. *For all $x \in \Omega$ and $t \geq 0$, we have*

$$\frac{\Pr_{t+1}(x)}{\Pr_t(x)} \geq \frac{1}{c(n)}.$$

Proof. From the definition of β_{t+1} in Equation (5.1), we know that

$$1 \geq \frac{e^{\beta_t r(x)}}{e^{\beta_{t+1} r(x)}} = e^{-\log(c) r(x)/R} \geq \frac{1}{c(n)}.$$

Therefore, it follows that

$$\frac{\Pr_{t+1}(x)}{\Pr_t(x)} = \frac{e^{\beta_{t+1} r(x)}}{e^{\beta_t r(x)}} \cdot \frac{Z_t}{Z_{t+1}} \geq \frac{Z_t}{Z_{t+1}} = \frac{\sum_{x \in \Omega} e^{\beta_t r(x)}}{\sum_{x \in \Omega} e^{\beta_{t+1} r(x)}} \geq \frac{\sum_{x \in \Omega} e^{\beta_t r(x)}}{\sum_{x \in \Omega} c e^{\beta_t r(x)}} = \frac{1}{c(n)},$$

as desired. □

The following lemma is the key to our argument and states that there exists a *balanced bias* parameter λ_{t^*} relative to our target set Ω_k that assigns nontrivial probability mass to elements with rank at most k and elements with rank greater than k.

Lemma 5.3.2. *Let Ω be a graded poset with rank $R \geq 1$ such that $1 \leq a_i \leq c(n)^i$ for all $i \in \{0, 1, \ldots, R\}$ and for some polynomial $c(n) \geq 2$. If $k \in \{0, 1, \ldots, R-1\}$, then there exists a $t \in [R^2]$ for which*

$$\Pr_t(r(x) \leq k) \geq \frac{1}{c(n)+1}$$

and

$$\Pr_t(r(x) > k) \geq \frac{1}{c(n)+1}.$$

Proof. Suppose there exists a minimum $t^* \in \mathbb{Z}_{\geq 1}$ satisfying the property

$$\Pr_{t^*}(r(x) > k) > \frac{1}{c(n)+1}.$$

By minimality, it follows that

$$\Pr_{t^\star-1}(r(x) > k) \le \frac{1}{c(n)+1} \implies \Pr_{t^\star-1}(r(x) \le k) \ge \frac{c(n)}{c(n)+1}.$$

Therefore, since Lemma 5.3.1 states that $\Pr_{t^\star}(x) \ge \Pr_{t^\star-1}(x)/c(n)$, we can conclude that

$$\Pr_{t^\star}(r(x) \le k) \ge \frac{1}{c(n)+1}.$$

To prove the existence of such a t^*, recall that $\Pr_t(r(x) > k) > 1/(c(n)+1)$ if and only if

$$(c+1)\sum_{i=k+1}^{R} a_i \lambda_t^i > 1. \tag{5.4}$$

To prove Equation (5.4) it suffices to prove $(c(n)+1)\lambda_t^R > 1$. Setting $t = R^2$, we have $\lambda_t = c(n)^{R-1} \ge 1$ since $R \ge 1$. It follows that $(c(n)+1)\lambda_t^R > 1$, as desired. To finish the proof, let $t^\star$ be the minimum integer $t \in [R^2]$ satisfying the inequality

$$\Pr_t(r(x) > k) > \frac{1}{c(n)+1}.$$

This completes the proof. □

Now that we have proved Lemma 5.3.2, we can present our main result. Let $\tau_t(\varepsilon)$ denote the mixing time of the Markov chain $\mathcal{M}$ with bias parameter λ_t as defined in Equation (5.2). Our proof uses a characterization of the mixing time in terms of the *conductance* of the Markov chain, which we describe in detail in Section 2.1.4. For convenience, recall that for an ergodic, reversible Markov chain $\mathcal{M}$ with stationary distribution π, the conductance of a nonempty set $S \subseteq \Omega$ is defined as

$$\Phi(S) = \sum_{x\in S, y\in\overline{S}} \frac{\pi(x)P(x,y)}{\pi(S)}.$$

The global conductance of the Markov chain is $\Phi^* = \min_{S\subseteq\Omega:\, 0<\pi(S)\le 1/2} \Phi(S)$, and it captures the mixing time up to polynomial factors as shown in Theorem 2.1.5. Analogous to the notation $\tau_t(\varepsilon)$, let Φ_t^* denote the conductance of the Markov chain with bias parameter λ_t.

Theorem 5.3.3 (Balanced Bias)**.** *Let $\mathcal{M}_t$ be a Markov chain whose state space is the graded poset Ω with rank function $r : \Omega \longrightarrow \mathbb{Z}_{\ge 0}$ and rank R. If the stationary distribution satisfies $\pi_t(x) \propto \lambda_t^{r(x)}$ for all $x \in \Omega$ and if there exists a polynomial $c(n) \ge 2$ such that $1 \le a_i \le c(n)^i$ for all $i \in \{0, 1, \dots, R\}$, then for any $k \in \{0, 1, \dots, R\}$ there exists a $t^* \in [R^2]$, and hence balanced bias parameter λ_{t^*}, such that*

$$\pi_{t^*}(\Omega_k) \ge \frac{1}{2(c(n)+1)(\tau_{t^*}(\varepsilon)+1)}.$$

Furthermore, since $\mathcal{M}_t$ can be used to approximately sample from π_t in $O(\tau_t(\varepsilon))$ time, we can approximately sample from Ω_k in expected $O(c(n)\tau_{t^*}(\varepsilon)^2)$ time, for any $\varepsilon \le 1/e$.

Proof. Let Φ_t^* denote the conductance of $\mathcal{M}_t$. Assume that $k < R$ and consider the cut $S = \Omega_{\le k}$. We have $\min(\pi_{t^*}(S), \pi_{t^*}(\overline{S})) \ge 1/(c(n)+1)$ for the balanced bias parameter λ_{t^*} by Lemma 5.3.2. Therefore, we can relate the conductance of the set S to the probability mass of the set Ω_k by

$$\Phi_{t^*}(S) \le \frac{\sum_{x\in S, y\in\overline{S}} \pi_{t^*}(x)P_{t^*}(x,y)}{\min(\pi_{t^*}(S), \pi_{t^*}(\overline{S}))} \le (c(n)+1) \sum_{x\in\Omega_k, y\in\Omega_{k+1}} \pi_{t^*}(x)P_{t^*}(x,y) \le (c(n)+1)\pi_{t^*}(\Omega_k). \tag{5.5}$$

Using Theorem 2.1.5 and setting $\varepsilon = 1/e$, it follows that

$$\Phi_t^* \ge \frac{1}{2(\tau_t(1/e)+1)}. \tag{5.6}$$

The minimality of Φ_t^* together with Equation (5.6) and Equation (5.5) imply that

$$\frac{1}{2(\tau_{t^*}(1/e)+1)} \le \Phi_{t^*}^* \le \Phi_{t^*}(S) \le (c(n)+1)\pi_{t^*}(\Omega_k).$$

Therefore, we can lower bound the target probability mass by

$$\pi_{t^*}(\Omega_k) \geq \frac{1}{2(c(n)+1)(\tau(1/e)+1)}. \tag{5.7}$$

Now we show that $O(c(n)\tau_{t^*}(1/e))$ samples from π_{t^*} are needed in expectation in order to generate an approximately uniform sample from the target set Ω_k. Let X be a random variable denoting the number of samples needed to generate an object of size k. Since X is a geometric random variable parameterized by $p = \pi_{t^*}(\Omega_k)$, Equation (5.7) implies that

$$\mathbb{E}[X] = p + 2(1-p)p + 3(1-p)^2 p + \cdots = \frac{1}{p} \leq 2(c(n)+1)(\tau_{t^*}(1/e)+1) = O(c(n)\tau_{t^*}(1/e)).$$

Lastly, since each sample can be generated in $O(\tau_{t^*}(\varepsilon))$ time and because we assume that $\varepsilon \leq 1/e$, we can approximately sample from Ω_k in expected $O(c(n)\tau_{t^*}(\varepsilon)^2)$ time. This analysis for $k = R - 1$ directly extends to the case $k = R$ by the detailed balance equation, which completes the proof. □

5.4 Sampling Integer Partitions

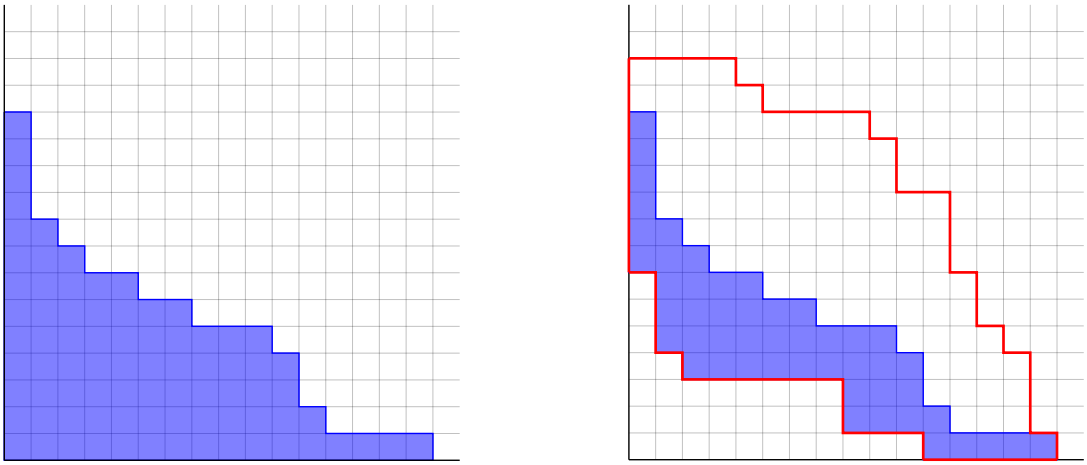
In this section we demonstrate how to use the balanced bias technique to efficiently sample from general classes of restricted integer partitions. Integer partitions have a natural representation as *Young diagrams*, which formally are finite subsets $S \subseteq \mathbb{Z}^2_{\geq 0}$ with the property that if $(a, b) \in S$, then

$$\{(i, j) \in \mathbb{Z}^2_{\geq 0} : 0 \leq i \leq a \text{ and } 0 \leq j \leq b\} \subseteq S.$$

Young diagrams can be visualized as a connected set of unit squares on the integer lattice with a corner at $(0, 0)$ and a nonincreasing upper boundary from left to right. Each square in the Young diagram must be supported below by the x-axis or another square and supported to the left by the y-axis or another square. We are interested in *region-restricted Young diagrams*, a variant of Young diagrams whose squares are restricted to lie in a connected region $R \subseteq \mathbb{Z}^2_{\geq 0}$ such that each

square is supported below and to the left by the boundary of R or another square. Note that we use R in this section to denote a region instead of the rank of a poset. We will see that the rank of the poset induced by the natural partial order on R-restricted Young diagrams is $|R|$.

We call Young diagrams $S \subseteq \mathbb{Z}^2_{\geq 0}$ such that $|S| = n$ *unrestricted integer partitions* of n and use this term interchangeably with integer partitions. Many well-studied classes of restricted integer partitions have natural interpretations as region-restricted Young diagrams. For example, the set of integer partitions of n with at most k parts and with each part at most size ℓ give rise to the Gaussian binomial coefficients and can be thought of as the set of Young diagrams of size n contained in a $k \times \ell$ box.



(a) (b)

Figure 5.1: Examples of (a) unrestricted and (b) region-restricted integer partitions.

5.4.1 Biased Markov Chain on Integer Partitions

Let the state space Ω be the set of all Young diagrams restricted to lie in a region R. Young diagrams have a natural graded partial order via inclusion, where $x \leq y$ if and only if $x \subseteq y$. It follows that the rank of a diagram x is $r(x) = |x|$. The following Markov chain on the Hasse diagram of this

graded poset makes local transitions in each step that add or remove a square on the boundary of the Young diagram according to the Metropolis–Hastings algorithm [Has70]. It follows that the stationary distribution is the Boltzmann distribution parameterized by the bias parameter λ.

Algorithm 5.1 Biased Markov chain for sampling integer partitions restricted to the region R.

procedure IntegerPartitionMarkovChain(R, λ, ε)
    Set $x_0 \leftarrow R$
    for $t = 1$ to $\tau_\lambda(\varepsilon)$ **do**
        Choose a neighbor y of x_t uniformly at random with probability $1/(2\Delta)$
        Set $x_{t+1} \leftarrow y$ with probability $\min(1, \lambda^{|y|-|x_t|})$
        Set $x_{t+1} \leftarrow x_t$ with all remaining probability
    return x_t

Note that in Algorithm 5.1 we let Δ be an upper bound for the number of neighboring configurations of x in the state space induced by R, for all partitions $x \in \Omega$.

Throughout the rest of the chapter, we use $\mathcal{M}$ as an alias for the Markov chain in Algorithm 5.1. The state space of the integer partitions Markov chain Ω is connected because any configuration can eventually reach the maximum configuration $x = R$ with positive probability. By construction, the Markov chain is lazy (i.e., $P(x, x) \geq 1/2$ for all $x \in \Omega$) and reversible, so it follows that the Markov chain is ergodic. Therefore, $\mathcal{M}$ has a unique stationary distribution π. Using the detailed balance equation in Section 2.1, we see that $\pi(x) \propto \lambda^{|x|}$ for all $x \in \Omega$, as desired.

Remark 5.4.1. Once we show that this Markov chain is rapidly mixing for the necessary values of λ, we can use Algorithm 5.1 to efficiently approximate the number of integer partitions of size n restricted to the region R to within an arbitrarily small relative error and failure probability. This is a consequence of *self-reducibility* [Jer03]. To see this, observe that we can run $\mathcal{M}$ restricted to R polynomially many times to compute an empirical mean for the height m of the first column in R. Then we can use $\mathcal{M}$ to recursively approximate the number of partitions of size $n - m$ restricted to the new region $R' = \{(i, j) \in R : 1 \leq i \text{ and } j \leq m\}$ and return the corresponding product of the conditional probabilities.

5.4.2 Boltzmann Sampling with Balanced Bias

We start by presenting a very general and liberal theorem that describes how the Markov chain in Algorithm 5.1 can be used to sample restricted integer partitions of a target size k. The bounds on the running time of the algorithm are far from tight in most cases, but it does readily provide a framework for obtaining polynomial-time sampling algorithms for a target set, provided one can bound the mixing time of the Markov chain at all values of λ_t on the entire state space.

Theorem 5.4.2. *Let $\tau^*(\varepsilon) = \max_{1 \le t \le |R|^2} \tau_t(\varepsilon)$ be the maximum mixing time of $\mathcal{M}$ for the region R. There exists an algorithm to approximately uniformly sample restricted integer partitions of size k restricted to the region R in expected $O(|R|^2 \Delta \tau^*(\varepsilon)^2)$ time.*

Proof. If $k = 0$ or $k = |R|$ then there is only one such partition, so we assume that $k \in [|R| - 1]$. By definition we have $|\Omega_{i+1}|/|\Omega_i| \le \Delta$, so it follows by induction that $1 \le |\Omega_i| \le \Delta^i$ for all $i \in \{0, 1, \dots, |R|\}$. Lemma 5.3.2 implies that there exists a balanced bias parameter λ_{t^*} for some value $t^* \in [|R|^2]$. Since $\tau^*(\varepsilon)$ is an upper bound on the mixing time for all $t \in [|R|^2]$, we can run Algorithm 5.1 independently with all the candidate inputs λ_t. The result follows from Theorem 5.3.3. □

If more is known about the number of elements at each rank or the geometry of R, then we can give explicit bounds on the running time of this algorithm. For example, if R is the region of a *skew Young diagram*, a region contained between two Young diagrams (see Figure 5.1), then we can adapt a recent mixing result of Levin and Peres [LP16] about biased exclusion processes for this setting. We note that Theorem 5.4.3 uses the path coupling proof given in [GPR09] for sampling biased lattice structures using exponential metrics.

Theorem 5.4.3 ([LP16])**.** *Consider the biased exclusion process with bias $\beta = \beta_n = 2p_n - 1 > 0$ on the segment of length $2n$ and with n particles. Set $\alpha = \sqrt{p_n/(1 - p_n)}$. For $\varepsilon > 0$, if n is large enough, then*

$$\tau(\varepsilon) \le \frac{2n}{\beta^2} \left[\log(1/\varepsilon) + \log \left[\alpha \left(\frac{\alpha^n - 1}{\alpha - 1} \right)^2 \right] \right].$$

It will also be useful in our analysis to use Wilson's tight bound on the mixing time of the unbiased exclusion process.

Theorem 5.4.4 ([Wil04])**.** *The mixing time of the unbiased exclusion process on a segment of length $2n$ with n particles satisfies*

$$\tau(\varepsilon) \leq \frac{2}{\pi^2}[1 + o(1)]n^3 \log(n/\varepsilon).$$

Corollary 5.4.5. *If the region R is a skew Young diagram contained in an $n \times n$ square, then we can uniformly sample partitions of size k restricted to fitting in R in expected $O(n^{20} \log \varepsilon^{-1})$ time.*

Proof. The biased exclusion process on a path of length $2n$ with n particles is the same as the Markov chain $\mathcal{M}$ when the restricting region is an $n \times n$ box. The proof of Theorem 5.4.3 uses a path coupling argument (i.e., Theorem 2.1.4) that immediately extends to regions that are defined by a skew Young diagram since the expected change in distance between any two adjacent states in the restricted setting can be at most the change in distance in the original dynamics. Therefore, we can obtain an upper bound for the mixing time of R when it is a skew Young diagram. Note that this is really just a subinterval of the underlying poset.

Start by letting $\lambda_{n,t}$ denote λ_t in an instance of size n. We analyze the cases $\lambda_{n,t} < 1$, $\lambda_{n,t} = 1$, and $\lambda_{n,t} > 1$ separately, and then we bound the mixing time of $\mathcal{M}$ for all candidate values of $\lambda_{n,t}$. A balanced bias exists by Lemma 5.3.2 because $a_i \leq p(i) \leq 2^i$ for all $i \in \{0, 1, \ldots, n^2\}$, where $p(k)$ is the number of unrestricted integer partitions of k.

In the first case, we assume $\lambda_{t,n} < 1$. It follows that $t \in [n^2 - 1]$ since $|R| = n^2$ and $\lambda_{n,t} = 2^{t/n^2 - 1}$. Translating the Metropolis transition probabilities in $\mathcal{M}$ to the exclusion process notation gives us

$$p_{n,t} = \frac{1}{1 + \lambda_{n,t}}. \tag{5.8}$$

Therefore, it follows from the definitions in the hypothesis of Theorem 5.4.3 and Equation (5.8)

that

$$\beta_{n,t} = 2p_{n,t} - 1 = \frac{1 - \lambda_{n,t}}{1 + \lambda_{n,t}},$$

and

$$\alpha_n = \sqrt{\frac{p_{n,t}}{1 - p_{n,t}}} = \sqrt{\frac{1}{\lambda_{n,t}}}.$$

We start by proving that $1/\beta_{n,t}^2 \leq 10n^4$ in order to use Theorem 5.4.3, First observe that in this case, setting $t = n^2 - 1$ minimizes $\beta_{n,t}$, and hence maximizes the desired quantity. Since the extreme value of the bias parameter is

$$\lambda_{n,n^2-1} = 2^{\frac{n^2-1}{n^2}-1} = (1/2)^{1/n^2},$$

it follows that

$$\left(\frac{1}{\beta_{n,t}}\right)^2 \leq \left(\frac{1 + \lambda_{n,n^2-1}}{1 - \lambda_{n,n^2-1}}\right)^2 = \left(\frac{1 + (1/2)^{1/n^2}}{1 - (1/2)^{1/n^2}}\right)^2 \leq 10n^2, \tag{5.9}$$

where the final inequality in Equation (5.9) can be achieved by taking the logarithm of both sides. To upper bound the $\alpha_{n,t}$ term in Theorem 5.4.3, we expand the implicit geometric series and use the fact that $\alpha_{n,t} > 1$ to get

$$\alpha_{n,t}\left(\frac{\alpha_{n,t}^n - 1}{\alpha_{n,t} - 1}\right)^2 \leq \alpha_{n,t}\left(n\alpha_{n,t}^n\right)^2 \leq \frac{n^2}{\lambda_{n,t}^{n+1/2}} \leq n^2 2^{n+1/2}. \tag{5.10}$$

The final inequality in Equation (5.10) is a consequence of the assumption that $\lambda_{n,t} \geq 1/2$, for all values of $t \geq 1$. Taking the logarithm of Equation (5.10) shows that

$$\log\left[\alpha_{n,t}\left(\frac{\alpha_{n,t}^n - 1}{\alpha_{n,t} - 1}\right)^2\right] \leq 2\log(n) + (n + 1/2)\log(2).$$

Therefore, it follows that the mixing time in this case satisfies $\tau(\varepsilon) = O(n^3(\log \varepsilon^{-1} + n))$ for all candidate bias parameters $\lambda_{n,t}$ by Theorem 5.4.3.

In the unbiased case when $\lambda_{n,t} = 1$, we use Theorem 5.4.4 to show that the mixing time of $\mathcal{M}$ is $\Theta(n^3 \log(n\varepsilon^{-1}))$. Lastly, in the third case when $\lambda_{t,n} > 1$, we have $t \in \{n^2 + 1, n^2 + 2, \dots, n^4\}$, so then

$$p_{n,t} = \frac{\lambda_{n,t}}{1 + \lambda_{n,t}}.$$

Therefore, translating to the exclusion process notation gives us

$$\beta_{n,t} = \frac{\lambda_{n,t} - 1}{\lambda_{n,t} + 1}$$

and

$$\alpha_{n,t} = \sqrt{\lambda_{n,t}}.$$

By a similar analysis to before, $t = n^2 + 1$ now maximizes the expression $1/\beta_{n,t}^2$, so it follows that

$$\left(\frac{1}{\beta_{n,t}}\right)^2 \le \left(\frac{1 + \lambda_{n,n^2+1}}{1 - \lambda_{n,n^2+1}}\right)^2 = \left(\frac{1 + 2^{1/n^2}}{1 - 2^{1/n^2}}\right)^2 \le 10n^4. \tag{5.11}$$

Furthermore, since $\lambda_{n,t} \le 2^{n^2-1}$ and $\alpha_{n,t} > 1$, we have

$$\alpha_{n,t}\left(\frac{\alpha_{n,t}^n - 1}{\alpha_{n,t} - 1}\right)^2 \le \alpha_{n,t}\left(n\alpha_{n,t}^n\right)^2 \le n^2\lambda_{n,t}^{n+1/2} \le n^2\left(2^{n^2-1}\right)^{n+1/2}. \tag{5.12}$$

Taking the logarithm of Equation (5.12), we have

$$\log\left[\alpha_{n,t}\left(\frac{\alpha_{n,t}^n - 1}{\alpha_{n,t} - 1}\right)^2\right] \le 2\log(n) + (n^2 - 1)(n + 1/2)\log(2).$$

Using Theorem 5.4.3, it follows that the mixing time is $\tau(\varepsilon) = O(n^5(\log \varepsilon^{-1} + n^3))$. Therefore, for all values of λ_t we have $\tau_t(\varepsilon) = O(n^8 \log \varepsilon^{-1})$. Applying Theorem 5.4.2 using the fact that $|R| \le n^2$ and observing that $a_i \le 2^i$ instead of the generous upper bound of Δ completes the proof. □

5.4.3 Sampling from Log-Concave Distributions

If more is known about the stationary distribution π, in particular the sequence $(|\Omega_i|)_{i=0}^{\infty}$, we can often dramatically improve the bound for the running time of a rejection sampling-based algorithm. In this section we contrast the balanced bias framework to a tighter analysis and modified algorithm that exploits log-concavity for uniformly sampling integer partitions in expected $O(n^{9/4})$ time. Our main techniques involve a compressed representation of partitions and using log-concavity to prove concentration of measure around partitions of the target size. This approach extends to a variety of settings where log-concavity or concentration can be shown.

To efficiently sample integer partitions of size n, we set $\lambda_n = p(n-1)/p(n)$, where $p(n)$ is the number of partitions of n, which forces the stationary distribution to concentrate near n. Since the integer sequence $(p(k))_{k=26}^{\infty}$ is log-concave [Nic78, DP14], it follows that the sequence $(p(k)\lambda_n^k)_{k=26}^{\infty}$ is also log-concave, for any value of $\lambda \in \mathbb{R}_{\geq 0}$. Log-concave sequences of positive terms are unimodal, and it can easily be shown that the mode of the stationary distribution is $k = n$ for this value of λ_n. Furthermore, we show how log-concavity forces an exponential decay on both sides of the mode, which in turn gives us strong concentration.

First, we argue that we need only consider Young diagrams that lie under the curve $y = 2n/x$ when sampling partitions of n. This is a simple consequence of the fact that all Young diagrams with squares above this curve necessarily have more than $2n$ squares total. We will elaborate on our choice of using $2n$ instead of n later, but for now it suffices to say that this choice can be used to improve the rejection rate of our Markov chain Monte Carlo algorithm.

Proposition 5.4.6. *A Young diagram that lies under the curve $y = 2n/x$ can be stored in $O(n^{1/2} \log n)$ space.*

Proof. For any square in the Young diagram, both of its coordinates are not greater than $\sqrt{2n}$, otherwise it would lie above $y = 2n/x$. We can record the height of each column and the width of each row in the range $\{0, 1, \ldots, \lfloor\sqrt{2n}\rfloor\}$ to capture the position of every square in the diagram.

Therefore, we can represent the diagram using exactly these $2\lfloor\sqrt{2n}\rfloor$ heights and widths, each of which require $O(\log n)$ bits to store. □

Using the compressed representation in the previous proposition, it follows that there are not more than $O(n^{1/2})$ possible transitions at any state. To see this, note that our algorithms adds or removes at most one square on the upper boundary in each step. We can further adapt this technique in the case of any general region R that lies under the curve $y = 2n/x$.

Proposition 5.4.7. *There are at most $4\sqrt{2n}$ potential transitions for any Young diagram that lies under the curve $y = 2n/x$.*

Proof. Observe that since the squares in any row or column must be connected, there are at most two valid moves in any particular row or column. Therefore, by Proposition 5.4.6 there are at most $4\lfloor\sqrt{2n}\rfloor$ possible transitions from any such Young diagram. □

We now shift our attention to bounding the bias parameter λ_n and analyzing its effect on the mixing time of $\mathcal{M}$ and the concentration of the stationary distribution π. In 1918, Hardy and Ramanujan [HR18] proved the classic asymptotic formula for the number of integer partitions:

$$p(n) \sim \frac{1}{4\sqrt{3}n} e^{\pi\sqrt{2n/3}}. \tag{5.13}$$

We use closely related bounds proved in [DP14] to give the following tight inequality for our choice of bias λ_n. The proof of this result is deferred to Section 5.4.4. Then we bound the mixing time of the Markov chain $\mathcal{M}$ in Algorithm 5.1 using our new bounds for λ_n and Theorem 5.4.3.

Lemma 5.4.8. *Let $\lambda_n = p(n-1)/p(n)$. For all integers $n \geq 30$, we have*

$$1 - \frac{2}{\sqrt{n}} < \lambda_n < 1 - \frac{1}{\sqrt{n}}.$$

Theorem 5.4.9. *The Markov chain $\mathcal{M}$ in Algorithm 5.1 with bias $\lambda_n = p(n-1)/p(n)$ restricted to the region R bounded by the curve $y = 2n/x$ has mixing time $\tau(\varepsilon) = O(n^{3/2}(\log \varepsilon^{-1} + n^{1/2}))$.*

Proof. We modify Theorem 5.4.3 and its proof in [LP16]. In this biased exclusion process we have, we have $\lambda = \lambda_n$, $\beta = (1-\lambda)/(1+\lambda)$, and $\alpha = \sqrt{1/\lambda}$. By Proposition 5.4.7, there are at most $4\sqrt{2n}$ transitions from any state, so for n large enough we have

$$\tau(\varepsilon) \leq \frac{8\sqrt{2n}}{\beta^2} \left[\log(1/\varepsilon) + \log(\operatorname{diam}(\Omega))\right], \tag{5.14}$$

where $\operatorname{diam}(\Omega)$ is the maximum length path between any two states in the underlying path coupling metric, as defined in [LPW17]. Therefore, it follows that $\operatorname{diam}(\Omega) \leq 2|R|\alpha^{2n}$, since this is an upper bound on the distance between the maximum and minimum configurations. Observing that

$$|R| \leq 2nH_{2n} \leq 2n(\log(2n) + 1),$$

where H_n is the n-th harmonic number, it follows that we can upper bound the diameter by

$$\operatorname{diam}(\Omega) \leq 4n(\log(2n) + 1)\alpha^{2n} = 4n(\log(2n) + 1)\lambda^{-n}.$$

Using Lemma 5.4.8 along with the inequality $1 + x \leq e^x$, we have

$$\log(\lambda^{-n}) \leq \log\left[\left(1 - \frac{2}{\sqrt{n}}\right)^{-n}\right] = \log\left[\left(1 + \frac{2}{\sqrt{n} - 2}\right)^{n}\right] \leq 3\sqrt{n},$$

for sufficiently large values of n. It also follows from Lemma 5.4.8 that

$$\frac{1}{\beta} = \frac{1+\lambda}{1-\lambda} < \frac{1 + \left(1 - \frac{1}{\sqrt{n}}\right)}{1 - \left(1 - \frac{1}{\sqrt{n}}\right)} = 2\sqrt{n} - 1. \tag{5.15}$$

Applying Equation (5.15) and our upper bound for $\operatorname{diam}(\Omega)$ in terms of $\log(\lambda^{-n})$ to Equation (5.14),

it follows that $\tau(\varepsilon) = O(n^{3/2}(\log \varepsilon^{-1} + n^{1/2}))$, as desired. □

Another key observation we make for sampling integer partitions of size n is to try to salvage integer partitions of size $k > n$ instead of rejecting them. We need to be careful to ensure that we preserve the uniform distribution when recycling partitions, but it turns out this can be done relatively easily. For the rest of the section, we describe how this idea can be used to substantially boost the acceptance rate of our Markov chain Monte Carlo algorithm.

For any $k \geq 0$, consider the function $f_k : \Omega_n \longrightarrow \Omega_{n+k}$ that maps a partition $x = (x_1, x_2, \ldots, x_m)$ of size n to the partition $f(x) = (x_1 + k, x_2, \ldots, x_m)$ of size $n + k$. Recall that $x_1 \geq x_2 \geq \cdots \geq x_m$ since x is a Young diagram. Clearly f_k is injective, so we can consider the inverse map $f_k^{-1}(y)$, when it is well-defined, that subtracts k from y_1 if $y_1 - k \geq y_2$. We can perform this inversion procedure for all $k \geq 0$ in aggregate by defining the function $g_n : \Omega_{\geq n} \longrightarrow \Omega_n \cup \{\texttt{null}\}$ such that for any integer partition $x = (x_1, x_2, \ldots, x_\ell) \in \Omega_{\geq n}$, we have

$$g_n(x) = \begin{cases} (x_1 - k, x_2, \ldots, x_\ell) & \text{if } x_1 + x_2 + \cdots + x_\ell = n + k \text{ and } x_1 - k \geq x_2, \\ \texttt{null} & \text{otherwise.} \end{cases} \tag{5.16}$$

Note that for any $k \geq 0$ that we use in this recycling technique, as long as the image set $f_k(\Omega_n) \subseteq \Omega$, then the inversion $g_n(f_k(\Omega_n))$ preserves the uniform distribution on the target set Ω_n.

The following result demonstrates the benefit of using the inverse map $g_n(x)$ instead of simply rejecting samples of size $k \neq n$. The proof of Lemma 5.4.10 is a consequence of log-concavity, so we defer the details to Section 5.4.4. For contrast, recall that standard rejection sampling gives us an expected acceptance rate of $O(n)$ using only the fact that the distribution is unimodal.

Lemma 5.4.10. *Let x be a Young diagram sampled from the stationary distribution of $\mathcal{M}$, and let g be the function defined in Equation* (5.16)*. For sufficiently large values of n, we have*

$$\Pr(g_n(x) \textit{ is an integer partition of size } n) \geq \frac{1}{160n^{1/4}}.$$

Putting together the ideas for reducing the space complexity and improving the rejection rate, we now formally present our Markov chain Monte Carlo algorithm for approximately sampling integer partitions of n uniformly at random.

Algorithm 5.2 Markov chain Monte Carlo algorithm for uniformly sampling integer partitions.

```
1: procedure RandomIntegerPartition(n, ε)
2:     Let R to be the region bounded below the curve y = 2n/x
3:     Set λ ← 1 − 1.5n^{−1/2}
4:     repeat
5:         Sample x using IntegerPartitionMarkovChain(R, λ, ε)
6:     until n ≤ |x| ≤ 2n and g_n(x) ≠ null
7:     return g_n(x)
```

Note that we have the condition $|x| \leq 2n$ instead of $|x| \leq 2n \log n$ in Algorithm 5.2 so that g_n maps to the target set Ω_n uniformly. All partitions of size $2n$ are elements of Ω_{2n}, but the same is not true for larger partitions since the bounding region R is the curve $y = 2n/x$. We also note that coupling from the past can be used efficiently here to generate uniform samples exactly from the uniform distribution (as opposed to ε-close) because the underlying path coupling is monotone and there is a single minimum and maximum configuration [PW96, GPR09]. We conclude this section with our main theorem for uniformly sampling integer partitions.

Theorem 5.4.11. *The algorithm* RandomIntegerPartition *generates an integer partition of n approximately uniformly at random in expected $O(n^{7/4}(\log \varepsilon^{-1} + n^{1/2}))$ time using $O(n^{1/2} \log n)$ space.*

Proof. The proof is a direct consequence of Theorem 5.4.9, Lemma 5.4.10, and Proposition 5.4.6. □

5.4.4 Bounding the Improved Rejection Rate

In this section we prove Lemma 5.4.8 and Lemma 5.4.10. Our analysis relies heavily on inequalities for $p(n)$ given in [DP14] and the Hardy–Ramanujan formula in Equation (5.13). In order to give tight lower and upper bounds for λ_n, we start by restating using explicit bounds for $p(n)$ in [DP14].

For all integers $n \geq 2$, define the quantities

$$\mu_n = \frac{\pi\sqrt{24n-1}}{6}, \tag{5.17}$$

$$\nu_n = \frac{\sqrt{12}}{24n-1}, \tag{5.18}$$

$$T(n) = \nu_n \left[\left(1 - \frac{1}{\mu_n}\right) e^{\mu_n} + \frac{(-1)^n}{\sqrt{2}} e^{\mu_n/2} \right]. \tag{5.19}$$

The function $T(n)$ is the sum of the three largest terms in the Hardy-Ramanujan formula and the explicit error bounds we use from [DP14] were first proved by Lehmer [Leh38]. In what follows, we only prove the upper bound in Lemma 5.4.8 since the lower bound follows from an analogous argument.

Lemma 5.4.12. *For all integers $n \geq 2$, we have*

$$\left| p(n) - \nu_n \left(1 - \frac{1}{\mu_n}\right) e^{\mu_n} \right| < 1 + e^{\mu_n/2}.$$

Proof. If follows from [DP14, Lemma 2.3 and Proposition 2.4] that

$$p(n) < T(n) + 1 + \frac{16}{\mu_n^3} e^{\mu_n/2} < \nu_n \left(1 - \frac{1}{\mu_n}\right) e_n^{\mu} + 1 + e^{\mu_n/2}.$$

The lower bound follows similarly, and therefore completes the proof. □

Now that we have a reasonably tight bound for $p(n)$ in Lemma 5.4.12, we can prove Lemma 5.4.8 and obtain convenient lower and upper bounds for $\lambda_n = p(n-1)/p(n)$.

Proof of Lemma 5.4.8. It follows from Lemma 5.4.12 that for all $n \geq 14$ we have

$$\lambda_n < \frac{1 + e^{\mu_{n-1}/2} + \nu_{n-1}\left(1 - \frac{1}{\mu_{n-1}}\right) e^{\mu_{n-1}}}{-\left(1 + e^{\mu_n/2}\right) + \nu_n \left(1 - \frac{1}{\mu_n}\right) e^{\mu_n}} = \frac{e^{\mu_{n-1}}}{e^{\mu_n}} \left(\frac{e^{-\mu_{n-1}} + e^{-\mu_{n-1}/2} + \nu_{n-1}\left(1 - \frac{1}{\mu_{n-1}}\right)}{-\left(e^{-\mu_n} + e^{-\mu_n/2}\right) + \nu_n \left(1 - \frac{1}{\mu_n}\right)} \right).$$

We note that the lower bound for $p(n)$ is initially negative, hence the assumption that $n \geq 14$. Furthermore, for all $n \geq 65$ we have the inequality

$$e^{-\mu_n} + e^{-\mu_n/2} < \frac{\nu_n}{n}\left(1 - \frac{1}{\mu_n}\right).$$

Therefore, it follows that

$$\lambda_n < \frac{e^{\mu_{n-1}}}{e^{\mu_n}}\left(\frac{\left(1 + \frac{1}{n-1}\right)\nu_{n-1}\left(1 - \frac{1}{\mu_{n-1}}\right)}{\left(1 - \frac{1}{n}\right)\nu_n\left(1 - \frac{1}{\mu_n}\right)}\right) = \frac{e^{\mu_{n-1}}}{e^{\mu_n}}\left(\frac{n}{n-1}\right)^2\left(\frac{\mu_n^3(\mu_{n-1}-1)}{\mu_{n-1}^3(\mu_n - 1)}\right).$$

Observing that for all $n \geq 2$ we have $\mu_{n-1} - \mu_n < -\pi/\sqrt{6n}$ and

$$\frac{\mu_n^3(\mu_{n-1}-1)}{\mu_{n-1}^3(\mu_n - 1)} < \frac{n}{n-1},$$

we use the inequality $e^x \leq 1 + x + x^2/2$, for all $x \leq 0$, to show that

$$\lambda_n < \frac{e^{\mu_{n-1}}}{e^{\mu_n}}\left(\frac{n}{n-1}\right)^3 < e^{-\frac{\pi}{\sqrt{6n}}}\left(\frac{n}{n-1}\right)^3 \leq \left(1 - \frac{\pi}{\sqrt{6n}} + \frac{\pi^2}{12n}\right)\left(\frac{n}{n-1}\right)^3 < 1 - \frac{1}{\sqrt{n}}.$$

The final inequality is true for all values of $n \geq 160$. Verifying the claim numerically for $30 \leq n < 160$ completes the proof. □

Now we analyze the rejection rate of Algorithm 5.2, which attempts to repurpose partitions of size $k \geq n$ by truncating their largest part according to the function $g_n(x)$ defined in Equation (5.16). By Hardy-Ramanujan formula in Equation (5.13) and the definition of asymptotic equivalence, for any constant $c > 0$ and n sufficiently large, it follows that

$$\frac{1-c}{4\sqrt{3}n}e^{\pi\sqrt{2n/3}} \leq p(n) \leq \frac{1+c}{4\sqrt{3}n}e^{\pi\sqrt{2n/3}}.$$

In particular, the Hardy-Ramanujan formula implies that for all $n \geq 20$, we have

$$e^{-\pi k/\sqrt{6n}} \leq \lambda_n^k \leq e^{k/n - \pi k/\sqrt{6n}}. \tag{5.20}$$

Lemma 5.4.13. *Let Z_n be the partition function of the Boltzmann distribution parameterized by λ_n from which Algorithm 5.2 samples. For sufficiently large values of n, we have $Z_n \leq 40n^{3/4}\lambda_n^n p(n)$.*

Proof. We can clearly use the generating function for integer partitions to upper bound Z_n by

$$Z_n \leq \sum_{k=0}^{\infty} p(k)\lambda_n^k.$$

Letting $f(k) = p(k)\lambda_n^k$, the sequence $(f(k))_{k=0}^{\infty}$ is log-concave for $n \geq 26$ because $(p(k))_{k=0}^{\infty}$ is log-concave. Therefore, $(f(k))_{k=0}^{\infty}$ is unimodal and reaches its maximum at $k = n$ by our choice of λ_n. Log-concavity further implies for all $n \geq 26$ and $k \geq 1$, we have the inequalities

$$\frac{f(n+k)}{f(n)} \geq \frac{f(n+2k)}{f(n+k)} \geq \frac{f(n+3k)}{f(n+2k)} \geq \dots \tag{5.21}$$

and

$$\frac{f(n-k)}{f(n)} \geq \frac{f(n-2k)}{f(n-k)} \geq \frac{f(n-3k)}{f(n-2k)} \geq \dots \tag{5.22}$$

These two sets of inequalities give us a strategy for upper bounding Z_n. Observe that by centering the sum at n and considering blocks of k consecutive terms, we have

$$kf(n) \geq f(n) + f(n+1) + \dots f(n+k-1)$$

$$kf(n+k) \geq f(n+k) + f(n+k+1) + \dots + f(n+2k-1)$$

$$kf(n+2k) \geq f(n+2k) + f(n+2k+1) + \dots + f(n+3k-1),$$

and so forth. A similar set of inequalities can be derived to the left side of the mode. Normalizing these block terms by $f(n)$, we can exploit telescoping and Equation (5.21) to show that the contribution of each block decreases geometrically. Concretely, log-concavity implies that

$$\frac{f(n+2k)}{f(n)} = \frac{f(n+k)}{f(n)} \cdot \frac{f(n+2k)}{f(n+k)} \leq \left(\frac{f(n+k)}{f(n)}\right)^2.$$

Therefore, for any block size $k \geq 1$, it follows that we can bound Z_n geometrically by

$$Z_n \leq kf(n)\left(\frac{1}{1-\frac{f(n+k)}{f(n)}} + \frac{1}{1-\frac{f(n-k)}{f(n)}}\right). \tag{5.23}$$

In particular, if both $f(n+k)/f(n)$ and $f(n-k)/f(n)$ are at most some fixed constant less than 1, then we have $Z_n = O(kf(n))$. Using the Hardy–Ramanujan formula and Equation (5.20), we have

$$f(k) = p(k)\lambda_n^k \leq \frac{1+c}{4\sqrt{3}k} e^{k/n - \pi(k-2\sqrt{kn})/\sqrt{6n}} = \frac{1+c}{4\sqrt{3}k} e^{k/n - \pi(\sqrt{k}-\sqrt{n})^2/\sqrt{6n} + \pi\sqrt{n/6}}.$$

Setting $n + k = n + 2n^{3/4} + \sqrt{n} = (\sqrt{n} + n^{1/4})^2$, it follows for large enough values of n that

$$f\left(\left[\sqrt{n} + n^{1/4}\right]^2\right) \leq \frac{1+c}{4\sqrt{3}n} e^{1.1 - \pi/\sqrt{6} + \pi\sqrt{n/6}}.$$

Therefore, we can then bound the density value at $(\sqrt{n} + n^{1/4})^2$ relative to the maximum by

$$\frac{f\left(\left[\sqrt{n} + n^{1/4}\right]^2\right)}{f(n)} \leq \frac{\frac{1+c}{4\sqrt{3}n} e^{1.1 - \pi/\sqrt{6} + \pi\sqrt{n/6}}}{\frac{1-c}{4\sqrt{3}n} e^{\pi\sqrt{2n/3} - \pi\sqrt{n/6}}} = \frac{1+c}{1-c} e^{1.1 - \pi/\sqrt{6}}. \tag{5.24}$$

Setting $c = 0.01$, it immediately follows from Equation (5.24) that for large enough values of n,

$$\frac{f\left(\left[\sqrt{n} + n^{1/4}\right]^2\right)}{f(n)} \leq \frac{1.01}{0.99} e^{1.1 - \frac{\pi}{\sqrt{6}}} < 0.85.$$

Similarly, on the left side of the mode we also can derive the inequality

$$\frac{f\left(\left[\sqrt{n} - n^{1/4}\right]^2\right)}{f(n)} < 0.85.$$

Therefore, we have $Z_n \leq 40n^{3/4}\lambda_n^n p(n)$ since $n + k \leq n + 3n^{3/4}$, as desired. □

We are now equipped to bound the rejection rate of Algorithm 5.2. The improved probability that we obtain is a consequence of the concentration of log-concave distributions, our choice of λ_n, and the truncation scheme used to validate some partitions of size $k \geq n$.

Proof of Lemma 5.4.10. We use Lemma 5.4.13 and Lemma 5.4.8 to bound the probability that $g_n(x)$ generates a partition of n successfully. In particular, observe that

$$\begin{aligned}
\Pr(g_n(x) \text{ is an integer partition of size } n) &= \sum_{k=0}^{n} \frac{\lambda_n^{n+k} p(n)}{Z_n} \\
&\geq \frac{1}{40n^{3/4}} \sum_{k=0}^{n} \lambda_n^k \\
&\geq \frac{1}{40n^{3/4}} \sum_{k=0}^{n} \left(1 - \frac{2}{\sqrt{n}}\right)^k \\
&= \frac{1}{40n^{3/4}} \cdot \frac{\sqrt{n}}{2} \left(1 - \left(1 - \frac{2}{\sqrt{n}}\right)^{n+1}\right) \\
&\geq \frac{1}{160n^{1/4}}.
\end{aligned}$$

The second equality uses the closed-form for a finite geometric series, and the last inequality uses the fact that the rightmost term is at least 1/2. □

5.5 Applications to Other Graded Posets

In this section we experimentally demonstrate the versatility of using a Markov chain that walks along the edges of the Hasse diagram of a graded poset in order to uniformly sample elements of a fixed rank. When this chain is rapidly mixing for all λ_t with $t \in [R^2]$, we can use Theorem 5.3.3 to

generate approximately uniform samples of a target size in polynomial time. There are analogous notions of self-reducibility for restricted families of permutations and lozenge tilings, similar to region-restricted integer partitions. It follows that there exist fully polynomial-time approximation schemes for these enumerations problems when we can efficiently sample elements of a given rank from their respective posets [Jer03].

5.5.1 Permutations with Fixed Inversion Number

Our first example considers permutations of n elements with a fixed number of inversions. The Hasse diagram in this setting connects permutations that differ by one adjacent transposition, and the partial order is the weak Bruhat order on the symmetric group. It follows that the Markov chain walks along edges of the permutohedron. In the unbiased case when $\lambda = 1$, the nearest neighbor Markov chain mixes in $\Theta(n^3 \log n)$ time [Wil04]. Furthermore, if the bias of this chain is a constant then the dynamics converge in $\Theta(n^2)$ time [BBHM05, GPR09]. The number of permutations of n with k inversions is known to be log-concave in k, so standard Boltzmann sampling techniques can be used. However, in restricted settings (i.e., subintervals of the poset), we can potentially use our balanced bias method and avoid the need to bound the number permutations with k inversions in the restricted setting.

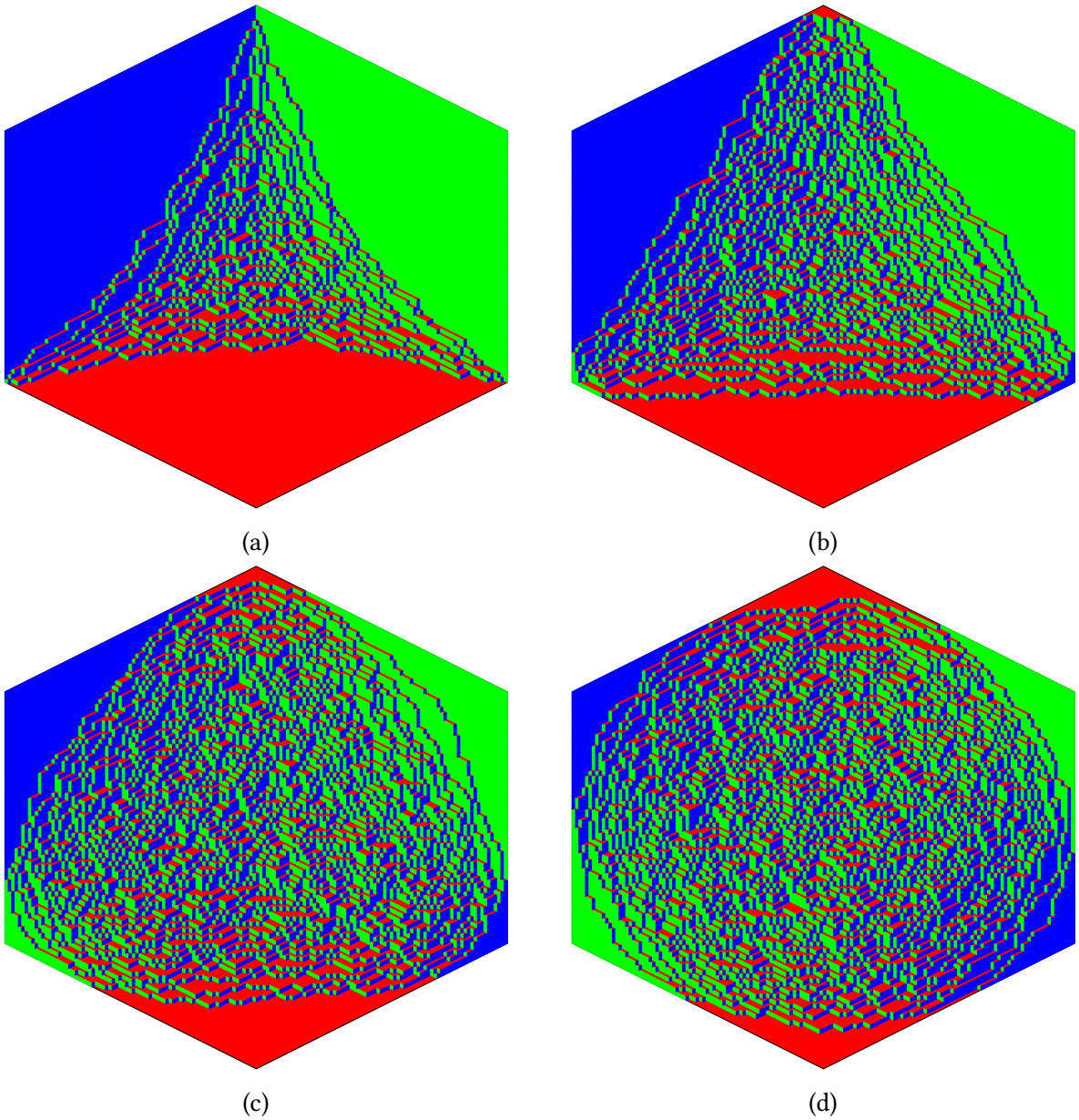
5.5.2 Lozenge Tilings with Fixed Average Height

The next example we consider are Lozenge tilings of a triangular lattice region with fixed average height. Lozenge tilings are pairs of adjacent equilateral regions in the triangular lattice, and the height function maps lozenge tilings bijectively to plane partitions lying in an $n \times n \times n$ box (e.g., see [LRS01]). It follows that lozenge tilings with a fixed average height of k are precisely the plane partitions with volume k. The Markov chain that adds or remove a single cube on the surface of the plane partition (which corresponds to rotating three nested lozenges by 180 degrees) is known to mix rapidly in the unbiased case. Caputo et al. [CMT11] studied the biased version of this

Markov chain with a preference towards removing cubes and showed that this chain converges in $O(n^3)$ time. Applying the balanced bias method, we can use Boltzmann sampling to generate random lozenge tilings with any target average height in polynomial time, as shown in Figure 5.2.

5.6 Summary and Discussion

Graded posets arise frequently throughout computer science, discrete mathematics, and statistical physics. While it is natural to try to count the number of elements with a given rank, these problems are often #**P**-complete. In this chapter, we work towards addressing these obstacles by designing Markov chain Monte Carlo algorithms for approximate counting and uniform sampling. We show that for certain classes of graded posets, biased Markov chains that walk along the edges of Hasse diagrams allow us to approximately generate uniform samples from a given rank in expected polynomial time. Our main contribution in this work is the method of Boltzmann sampling with a balanced bias, which shows that in certain situations one can bound the rejection rate of a Markov chain Monte Carlo algorithm using only its mixing time. We apply this technique to the problem of sampling integer partitions of n subject to a variety of constraints for which generating function-based methods break down. We also give an optimized Markov chain Monte Carlo algorithm for uniformly sampling unrestricted integer partitions of n and explore how this technique can be used for sampling permutations and lozenge tilings. An important future direction of this work is to characterize the mixing time of biased Markov chains on Hasse diagrams using only structural properties of the underlying graded poset. The design of provably efficient Markov chains on graded posets that perform global moves would also be of independent interest.



(a) (b) (c) (d)

Figure 5.2: Random lozenge tilings with average height (a) 5, (b) 15, (c) 35, (d) 50 percent of 75^3.

CHAPTER 6

ANALYZING BOLTZMANN SAMPLERS FOR BOSE–EINSTEIN CONDENSATES

In this chapter we take a noticeably different approach to Boltzmann sampling and use probabilistic interpretations of ordinary generating functions instead of Markov chains to generate samples. Specifically, we use techniques from analytic combinatorics to give provably efficient sampling algorithms for a broad class of weighted partitions and selections. This work primarily focuses on a family of weighted partitions related to *Bose–Einstein condensation* from statistical physics.

6.1 Introduction

Bose–Einstein condensation is a quantum phenomena that occurs when subatomic particles called bosons are cooled to nearly absolute zero. These particles behave as waves due to their quantum nature, and their wavelength increases as their temperature decreases. At a low enough temperature, the size of the waves exceeds the average distance between two particles and a constant fraction of bosons enter their ground state. These particles then coalesce into a single collective quantum wave called a *Bose–Einstein condensate*, which can, incredibly, be observed at the macroscopic scale.

In statistical physics, such thermodynamic systems are often modeled in one of three settings: the *microcanonical ensemble*, where both the number of particles and the total energy in the system are kept constant; the *canonical ensemble*, where the number of particles is constant but the energy is allowed to vary; and the *grand canonical ensemble*, where both the number of particles and energy can vary. A substantial amount of research has focused on understanding the asymptotic behavior of Bose–Einstein condensates in the microcanonical and canonical ensembles [BP83, CD14]. In this chapter, we give a provably efficient algorithm for uniformly sampling Bose–

Einstein condensate configurations from the microcanonical ensemble in the *low-temperature regime* when the number of particles is greater than or equal to the total energy in the system. This approach allows us to explore the thermodynamic properties of systems with millions of noninteracting bosons instead of relying solely on its limiting behavior.

Random sampling is widely used across scientific disciplines when exact solutions are unavailable. In many settings, *Boltzmann samplers* have proven to be particularly useful for sampling combinatorial objects of a fixed size. The state space $\mathcal{C}$ includes configurations of all sizes, and the Boltzmann distribution assigns a configuration $\gamma \in \mathcal{C}$ probability $\Pr_\lambda(\gamma) \propto \lambda^k$, where k is the size of γ and $\lambda \in \mathbb{R}_{>0}$ is a parameter of the system that biases the distribution toward configurations of the desired size. Boltzmann sampling is most effective if the sampling procedure on $(\mathcal{C}, \Pr_\lambda)$ is efficient on and *rejection sampling* (i.e., outputting objects of the desired size and rejecting all others) succeeds with high enough probability to produce samples of the desired size in expected polynomial time.

Generating random integer partitions is an illustrative example for demonstrating the effectiveness of Boltzmann sampling. Integer partitions arise throughout many areas of mathematics and physics (e.g., random matrices [Oko02], representation theory [Jam06], and self-assembly processes [GPR09]), and they are closely related to Bose–Einstein condensates. An integer partition of n is a nonincreasing sequence of positive integers that sums to n, and the simplest method for uniformly sampling partitions is based on the dynamic programming algorithm of Nijenhuis and Wilf [NW78]. This approach relies on exact counting and requires $O(n^{2.5})$ time and space. Alternatively, one can use Boltzmann sampling to generate partitions biased to have size close to n, and then use rejection sampling to only output objects of the desired size. Somewhat surprisingly, this approach can lead to substantially more efficient algorithms. Arratia and Tavaré [AT94] showed that integer partitions and many other objects with multiplicative generating functions can be sampled from Boltzmann distributions using independent random processes. Duchon et al. [DFLS04] turned this idea into a systematic Boltzmann sampling framework using techniques from analytic

combinatorics. Their approach offers a Boltzmann sampler for integer partitions with time and space complexity that is linear in the size of partition produced. Arratia and DeSalvo [AD16] then leveraged additional symmetries to develop an algorithm for sampling integer partitions of n in expected $O(\sqrt{n})$ time. Taking a rather different approach, Bhakta et al. [BCFR17] also recently used Boltzmann distributions and rejection sampling to design the first rigorous Markov chain Monte Carlo algorithm for sampling partitions.

While Boltzmann sampling is known to be quite effective for a vast collection of problems in statistical physics and combinatorics, many applications lack rigorous arguments that show the rejection sampling is efficient. In this chapter, we give a provably efficient Boltzmann sampler for Bose–Einstein condensates and we rigorously bound the rejection rate through singularity analysis of Dirichlet generating functions. Our techniques naturally extend to a broad class of multiplicative objects known as *weighted partitions*, which generalize integer partitions, plane partitions, and Bose–Einstein condensates.

The primary focus of this work is the design and analysis of an algorithm for uniformly sampling Bose–Einstein condensates in an idealized microcanonical setting with limited interactions between particles. We can view Bose–Einstein condensates combinatorially as weighted partitions with $b_k = \binom{k+3-1}{3-1}$ types of summands of size k. Each of the b_k summands corresponds to a degenerate energy state of a boson with energy k. We think of each of these energy states as multisets on three different colors with cardinality k. Since Bose-Einstein configurations are unordered collections of bosons, these configurations can be understood as multisets of bosons, or equivalently weighted partitions. It follows that we can visualize Bose–Einstein condensates by coloring Young diagrams. Every column corresponds to the energy state of a particle, and the columns are sorted lexicographically to produce a partition. In a microcanonical ensemble with m particles and energy n, the number of particles in their ground state is m minus the width of the Young diagram. Bose–Einstein condensation occurs when the width of the expected Young diagram is at most a constant fraction of m, the number of particles.

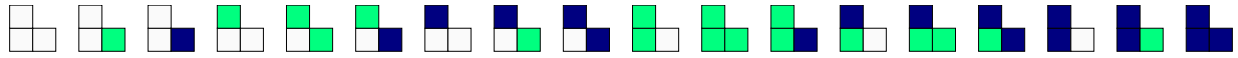


Figure 6.1: Young diagrams corresponding to Bose–Einstein condensates with shape $(2, 1)$, where the colors (gray, green, blue) correspond to the numbers $(1, 2, 3)$.

In the language of analytic combinatorics, Bose–Einstein condensates are given by the combinatorial class $\textsc{MSet}(\textsc{MSet}_{\geq 1}(3\mathcal{Z}))$. For example, when $n = 2$ there are 12 possible configurations: if there is one particle with energy 2 we have the states $\{\{1, 1\}\}$, $\{\{1, 2\}\}$, $\{\{1, 3\}\}$, $\{\{2, 2\}\}$, $\{\{2, 3\}\}$, $\{\{3, 3\}\}$, and if there are two particles each with energy 1 we have the states $\{\{1\}, \{1\}\}$, $\{\{1\}, \{2\}\}$, $\{\{1\}, \{3\}\}$, $\{\{2\}, \{2\}\}$ $\{\{2\}, \{3\}\}$, $\{\{3\}, \{3\}\}$. When $n = 3$ there are 38 such configurations: 10 if there is one particle with energy 3, 18 corresponding to the Young diagram with shape $(2, 1)$ where one particle has energy 2 and one has energy 1 (see, e.g., Figure 6.1), and 10 when there are three particles with energy 1. We remark that Bendkowski, Bodini, and Dovgal [BBD18] also recently used techniques from analytic combinatorics to investigate algorithms for sampling Bose–Einstein condensates and weighted partitions. Their work, however, focuses on a fine-grained tuning algorithm based on convex optimization for the nonuniform problem of targeted multiparametric Boltzmann sampling, and it does not fully analyze the corresponding rejection rates.

6.1.1 Main Results

Our complexity analysis follows the conventions in [BFP10, BLR14, Fla07, FFP07] and assumes the real-arithmetic model of computation, an oracle that evaluates a generating function within its radius of convergence in constant time, and a root-finding oracle. The main contribution of our work is a new approach for rigorously analyzing algorithms that sample from the Bose–Einstein distribution in low-temperature microcanonical ensembles when the number of particles exceeds the total energy. In particular, we give a provably efficient algorithm for uniformly sampling Bose–Einstein condensates by constructing a linear-time Boltzmann sampler using the framework established in [FFP07], and then we bound its rejection rate through singularity analysis of an associated Dirichlet generating function.

Theorem 6.1.1. *There exists a uniform sampling algorithm for Bose–Einstein condensates of size n that runs in expected $O(n^{1.625})$ time and uses $O(n)$ space.*

Our algorithm efficiently generates samples of size n exactly from the uniform distribution in expected polynomial time, which allows us to rigorously study the expected width of Young diagrams arising from random configurations (or, equivalently, the fraction of particles in a Bose–Einstein condensate occupying their ground state) without relying on the limiting properties given in [Yak12].

The singularity analysis used in the proofs generalizes to a broader family of weighted partitions, including integer partitions and plane partitions [BFP10]. To see this, call a *positive integer sequence of degree r* be a sequence of positive integers $(b_k)_{k=1}^{\infty}$ such that $b_k = p(k)$ for some polynomial $p(x) = a_0 + a_1 x + \cdots + a_r x^r \in \mathbb{R}[x]$, with $\deg(p) = r$. We show how the rightmost pole of the Dirichlet generating function for the sequence $(b_k)_{k=1}^{\infty}$ and its residue are related to a_r, the leading coefficient of $p(x)$, which we then use to establish rigorous rejection rates for Boltzmann sampling.

Our second main result gives guarantees for algorithms that generate uniform samples from a broad class of weighted partitions and selections. The running time of these algorithms is given as a function of the degree of the parameterizing polynomial.

Theorem 6.1.2. *There exists a uniform sampling algorithm for any class of weighted partitions or selections parameterized by a positive integer sequence of degree r for objects of size n that runs in expected time $O(n^{r+1+(r+3)/(2r+4)})$ and uses $O(n)$ space.*

In particular, the number of samples needed in expectation is $O(n^{(r+3)/(2r+4)})$, which is asymptotically tight by Theorem 6.2.6. If $r = 0$ (e.g., integer partitions) we need $O(n^{3/4})$ samples, and as $r \longrightarrow \infty$ the number of required samples converges to $O(\sqrt{n})$. We note that DeSalvo and Menz [DM16] independently developed a recent probabilistic model that gives a central limit theorem for the same family of weighted partitions and selections that circumvents the singularity analysis of

Dirichlet generating functions.

6.1.2 Techniques

The Boltzmann sampling framework in [DFLS04, Fla07, FFP07] provides an approach for sampling from Boltzmann distributions in polynomial time by interpreting the ordinary generating functions of these objects as probabilistic processes. However, this does not guarantee the rejection sampling component will be efficient. In order to lower bound the probability of generating an object of the target size, we rely on the *Khintchine–Meinardus probabilistic method* [GSE08, GS12, Gra18] and show that it holds for a broad class of weighted partitions and selections. For sampling Bose–Einstein condensates in particular, we give an improved algorithm based on the symbolic method of analytic combinatorics that exploits the underlying combinatorial structure instead of simply using independent geometric random variables for each energy state (i.e., Theorem 6.1.1 versus Theorem 6.1.2 with $r = 2$).

The goal of the Khintchine–Meinardus probabilistic is to asymptotically enumerate combinatorial objects through singularity analysis of Dirichlet generating functions. To analyze our algorithms, we use an intermediate local limit theorem from this theory. Only recently were these enumeration techniques extended to handle Dirichlet series with multiple poles on the real axis [GS12]. This allows for the analysis of Boses–Einstein condensates and other weighted partitions parameterized by nonconstant integer sequences. In this chapter, we show that the Dirichlet series for classes of weighted partitions parameterized by positive integer sequences are linear combinations of shifted and scaled Riemann zeta functions, and thus amenable to singularity analysis. Using bounds for the zeta function from analytic number theory, we show that the local limit theorem in [GS12] holds for Bose–Einstein condensates. Then we bound the residue of the rightmost pole of the Dirichlet generating function by viewing the parameterizing polynomial $p(x)$ as a Newtown-interpolating polynomial, which yields a convenient lemma for rejection rates.

In the process of analyzing a subroutine of our enhanced Boltzmann sampler for Bose–

Einstein condensates, we also develop a new tail inequality for the negative binomial distribution (Lemma 6.5.4) that is tighter than the standard Chernoff-type inequality for this use case. We believe the singularity analysis of Dirichlet series in this chapter will be valuable for a wide variety of sampling problems in combinatorics and statistical physics, where the objects of interest can be decomposed into noninteracting components and when transfer theorems for their ordinary generating functions are not applicable.

6.2 Preliminaries

We start by presenting several tools and key concepts for our sampling algorithms. First, we introduce the fundamental ideas of Boltzmann sampling for unlabeled objects and the symbolic method from analytic combinatorics. Next, we use the symbolic method to define two families of multiplicative objects called weighted partitions (a generalization of Bose–Einstein condensates) and selections. Then we finish by presenting a local limit theorem for these objects developed in the Khintchine–Meinardus probabilistic method, which we ultimately use to bound the rejection rate of our Boltzmann samplers.

6.2.1 Boltzmann Sampling

A combinatorial class $\mathcal{C}$ is a finite or countably infinite set equipped with a size function $|\cdot| : \mathcal{C} \longrightarrow \mathbb{Z}_{\geq 0}$ such that the number of elements of any given size is finite. For a class $\mathcal{C}$, let c_k denote the number of elements of size k. The counting sequence of $\mathcal{C}$ is the integer sequence $(c_k)_{k=0}^{\infty}$, and the ordinary generating function of $\mathcal{C}$ is

$$C(z) = \sum_{k=0}^{\infty} c_k z^k = \sum_{\gamma \in \mathcal{C}} z^{|\gamma|}.$$

Definition 6.2.1. The *Boltzmann distribution* of a class $\mathcal{C}$ parameterized by $\lambda \in (0, \rho_C)$ is the

probability distribution, for all $\gamma \in \mathcal{C}$, defined as

$$\Pr_\lambda(\gamma) = \frac{\lambda^{|\gamma|}}{C(\lambda)},$$

where ρ_C is the radius of convergence of $C(z)$. A *Boltzmann sampler* $\Gamma C(\lambda)$ for a class $\mathcal{C}$ is an algorithm that generates objects from $\mathcal{C}$ according to the Boltzmann distribution with parameter λ.

The size of an object generated by $\Gamma C(\lambda)$ is a random variable denoted by U with the probability distribution

$$\Pr_\lambda(U = n) = \frac{c_n \lambda^n}{C(\lambda)}.$$

All objects of size n occur with equal probability, so if $\Gamma C(\lambda)$ returns an object γ of size n, then γ is a uniform random sample among all size n objects in $\mathcal{C}$. This means we can use *rejection sampling* to generate objects of size n uniformly at random. However, for this technique to be effective, we need an efficient sampling algorithm $\Gamma C(\lambda)$ as well as a provably low rate of rejection for the target size n.

We assume that $\mathcal{C}$ contains infinitely many objects, n is feasible (i.e., $c_n \geq 1$), and that n is not the smallest feasible size in $\mathcal{C}$. In order to maximize the probability of generating an object of size n, we tune the Boltzmann sampler so that $\mathbb{E}_\lambda[U] = n$. To see why this strategy works, observe that the quotient rule gives us

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\Pr_\lambda(U = n) = \frac{c_n \lambda^{n-1}}{C(\lambda)}(n - \mathbb{E}_\lambda[U]), \tag{6.1}$$

where $\mathbb{E}_\lambda[U] = \lambda C'(\lambda)/C(\lambda)$. Since $\mathcal{C}$ contains objects of different sizes, it follows that

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\mathbb{E}_\lambda[U] = \frac{\mathrm{Var}_\lambda(U)}{\lambda} > 0. \tag{6.2}$$

The equality in Equation (6.2) is a property of Boltzmann distributions that can easily be seen

by differentiating the series $\mathbb{E}_\lambda[U] = \sum_{k=0}^{\infty} k c_k \lambda^k / C(\lambda)$, and the inequality holds because U is a nonconstant random variable and $\lambda \in (0, \rho_C)$. Therefore, $\mathbb{E}_\lambda[U]$ is strictly increasing, which implies that λ_n is unique. It follows from Equation (6.1) that λ_n maximizes $\Pr_\lambda(U = n)$.

6.2.2 Symbolic Method

The symbolic method from analytic combinatorics is a convenient language for describing Bose–Einstein condensates and constructing classes of weighted partitions and selections. There are two primitive combinatorial classes in the symbolic method: the neutral class $\mathcal{E}$ and the atomic class $\mathcal{Z}$. The class $\mathcal{E}$ contains a single element of size 0 called the neutral object, and the class $\mathcal{Z}$ contains a single element of size 1 called an atom. Neutral objects are used to mark objects as different and atoms are combined to form larger combinatorial objects. For example, atoms are often vertices in a tree or characters in a word. We can express a rich family of discrete structures using these primitive classes along with the following operators.

Cartesian Product. The Cartesian product of $\mathcal{A}$ and $\mathcal{B}$ is the class

$$\mathcal{C} = \mathcal{A} \times \mathcal{B} = \{(\alpha, \beta) : \alpha \in \mathcal{A}, \beta \in \mathcal{B}\}.$$

The size of the tuple $\gamma = (\alpha, \beta) \in \mathcal{C}$ is defined to be $|\gamma|_C = |\alpha|_\mathcal{A} + |\beta|_\mathcal{B}$, and the generating function for $\mathcal{C}$ is the discrete convolution $C(z) = A(z)B(z)$.

Combinatorial Sum. The combinatorial sum (disjoint union) of $\mathcal{A}$ and $\mathcal{B}$ is

$$\mathcal{C} = \mathcal{A} + \mathcal{B} = (\mathcal{E}_1 \times \mathcal{A}) \cup (\mathcal{E}_2 \times \mathcal{B}),$$

where $\mathcal{E}_1$ and $\mathcal{E}_2$ are two different neutral classes. The size of an element in $\mathcal{C}$ is the same as in its class of origin, so the ordinary generating function for $\mathcal{C}$ is $C(z) = A(z) + B(z)$.

Sequence Construction. The sequence of a class $\mathcal{B}$ with $b_0 = 0$ is the infinite sum

$$\mathcal{C} = \textsc{Seq}(\mathcal{B}) = \mathcal{E} + \mathcal{B} + (\mathcal{B} \times \mathcal{B}) + (\mathcal{B} \times \mathcal{B} \times \mathcal{B}) + \cdots,$$

and the corresponding generating function for $\mathcal{C}$ is

$$C(z) = 1 + B(z) + B(z)^2 + B(z)^3 + \cdots = \frac{1}{1 - B(z)}.$$

Power Set Construction. The power set of $\mathcal{B}$ is the class of all subsets of $\mathcal{B}$. Formally, the power set is given by the product

$$\mathcal{C} = \textsc{PSet}(\mathcal{B}) = \prod_{\beta \in \mathcal{B}} (\mathcal{E} + \{\beta\}).$$

Each factor can be interpreted as an independent decision about whether or not to include the object $\beta \in \mathcal{B}$ in the subsets. The generating function for $\mathcal{C}$ is

$$C(z) = \prod_{\beta \in \mathcal{B}} \left(1 + z^{|\beta|}\right) = \prod_{k=1}^{\infty} \left(1 + z^k\right)^{b_k}.$$

Multiset Construction. The multiset of a class $\mathcal{B}$ with $b_0 = 0$ is defined to be

$$\mathcal{C} = \textsc{MSet}(\mathcal{B}) = \prod_{\beta \in \mathcal{B}} \textsc{Seq}(\{\beta\}).$$

The elements of $\mathcal{C}$ are multisets of the objects in $\mathcal{B}$, and we can think of them as finite tuples of objects $\beta \in \mathcal{B}$ (possibly repeated) sorted in a canonical order. It follows from the previous operators that the generating function for $\mathcal{C}$ can be written as

$$C(z) = \prod_{\beta \in \mathcal{B}} \left(1 - z^{|\beta|}\right)^{-1} = \prod_{k=1}^{\infty} \left(1 - z^k\right)^{-b_k}.$$

It is also useful to consider the exp-log transform of $C(z) = \exp(\log C(z))$, which gives us

$$\begin{aligned} C(z) &= \exp\left(\sum_{k=1}^{\infty} b_k \log\left(1 - z^k\right)^{-1}\right) \\ &= \exp\left(\sum_{k=1}^{\infty} b_k \sum_{j=1}^{\infty} \frac{z^{kj}}{j}\right) \\ &= \prod_{k=1}^{\infty} \exp\left(\frac{1}{k} B\left(z^k\right)\right). \end{aligned} \tag{6.3}$$

The second equality follows from expanding the logarithm

$$\log(1 + z) = z - \frac{z^2}{2} + \frac{z^3}{3} - \dots,$$

and the third equality results from exchanging the order the summations [FS09]. Both interpretations of the multiset construction are essential to our algorithm and analysis because they demonstrate exactly how Bose–Einstein condensates decompose into combinatorial atoms.

6.2.3 Bose–Einstein Condensates and Weighted Partitions

The central object of study in the Khintchine–Meinardus probabilistic method are weighted partitions, which are unlabeled multiplicative objects that generalize integer partitions and Bose–Einstein condensates.

Definition 6.2.2. The class $\mathcal{C}^{(1)}$ of *weighted partitions* with b_k different types of summands of size $k \geq 1$ is implicitly defined by the generating function

$$C^{(1)}(z) \stackrel{\text{def}}{=} \sum_{k=0}^{\infty} c_k^{(1)} z^k = \prod_{k=1}^{\infty} \left(1 - z^k\right)^{-b_k}.$$

Equivalently, $\mathcal{C}^{(1)} = \textsc{MSet}(\mathcal{B})$ is parameterized by the class $\mathcal{B}$ of permissible summands.

Setting $b_k = 1$ for all $k \geq 1$ recovers the generating function for integer partitions since integer partitions consider only one type of summand of size k. Now that we have defined weighted

partitions, we can formally define Bose–Einstein condensates.

Definition 6.2.3. *Bose–Einstein condensates* are weighted partitions with the parameters

$$b_k = \binom{k+2}{2}.$$

In the language of analytic combinatorics, they are the class $\textsc{MSet}(\textsc{MSet}_{\geq 1}(3\mathcal{Z}))$.

The parameterizing class $\textsc{MSet}_{\geq 1}(3\mathcal{Z})$ is the set of all nonempty multisets of 3 different colored atoms. There are $\binom{k+3-1}{3-1}$ such multisets of size k, each corresponding to a different type of summand. From a physics point of view, multisets of size k in $\textsc{MSet}_{\geq 1}(3\mathcal{Z})$ are isomorphic to three-dimensional degenerage energy states of a boson with energy k. Since Bose–Einstien condensates are unordered collections of bosons, an object of size n in $\textsc{MSet}(\textsc{MSet}_{\geq 1}(3\mathcal{Z}))$ uniquely corresponds to a Bose–Einstein condensate with total energy n.

For the problem of uniform sampling, it is beneficial to work with the truncated class of weighted partitions $\mathcal{C}_n^{(1)}$ whose generating function is

$$C_n^{(1)}(z) \stackrel{\text{def}}{=} \prod_{k=1}^{n} \left(1 - z^k\right)^{-b_k},$$

since it completely contains the target set of objects of size n.

Next, we define an analog of weighted partitions called selections, which are fundamental to Fermi–Dirac statistics from physics and correspond to the power set construction.

Definition 6.2.4. The class $\mathcal{C}^{(2)}$ of *selections* with b_k different types of summands of size $k \geq 1$ is implicitly defined by the generating function

$$C^{(2)}(z) \stackrel{\text{def}}{=} \sum_{k=0}^{\infty} c_k^{(2)} z^k = \sum_{k=1}^{\infty} \left(1 + z^k\right)^{b_k}.$$

Equivalently, the class $\mathcal{C}^{(2)} = \textsc{PSet}(\mathcal{B})$ is parameterized by the class $\mathcal{B}$ of permissible summands.

The truncated class $\mathcal{C}_n^{(2)}$ and its generating function $C_n^{(2)}$ are defined similarly.

Finally, we define a random variable for the size of a weighted partition or selection when generated from the Boltzmann distribution of a truncated class.

Definition 6.2.5. Let $U_n^{(i)}$ be the random variable for the size of an object generated by $\Gamma C_n^{(i)}(\lambda)$.

6.2.4 Khintchine–Meinardus Probabilistic Method

We use the Khintchine–Meinardus probabilistic method to lower bound $\Pr(U_n^{(i)} = n)$, that is, the probability that our algorithms sample an object of the target size. Meinardus first established an asymptotic equivalence between the number of weighted partitions $c_n^{(1)}$ and the analytic behavior of the Dirichlet series

$$D(s) \stackrel{\text{def}}{=} \sum_{k=1}^{\infty} b_k k^{-s},$$

where $s = \sigma + it$ is a complex variable, in 1954 by using the saddle-point method under somewhat restrictive conditions [Mei54]. Several decades later, Granovsky, Stark, and Erlihson [GSE08] extended Meinardus' theorem by using Khintchine's probabilistic method to include new multiplicative combinatorial objects such as selections and assemblies [Khi11]. Granovsky and Stark [GS12] generalized their results to include families of weighted partitions such that the corresponding series $D(s)$ has multiple singularities on the positive real axis. This includes the class of Bose–Einstein condensates. In short, to use the Khintchine–Meinardus probabilistic method for weighted partitions, one must show that the Dirichlet series $D(s)$ satisfies the following conditions [GS12]:

(I) The series $D(s)$ has $r \geq 1$ simple poles at real positions $0 < \rho_1 < \rho_2 < \cdots < \rho_r$ with positive residues $A_1, A_2, \ldots, A_r$, respectively, and is analytic in the half-plane $\sigma > \rho_r > 0$. Moreover, there is a constant $0 < C_0 \leq 1$ such that the function $D(s)$ has a meromorphic continuation to the half-plane

$$\mathcal{H} = \{s : \sigma \geq -C_0\},$$

on which it is analytic except for above the r simple poles.

(II) There is a constant $C_1 > 0$ such that

$$D(s) = O\left(|t|^{C_1}\right),$$

uniformly for $s = \sigma + it \in \mathcal{H}$, as $t \longrightarrow \infty$.

(III) For $\delta > 0$ small enough and some $\varepsilon > 0$, we have

$$2\sum_{k=1}^{\infty} b_k e^{-k\delta} \sin^2(\pi k\alpha) \geq \mathcal{M}^{(i)}\left(1 + \frac{\rho_r}{2} + \varepsilon\right)|\log \delta|,$$

for all $\sqrt{\delta} \leq |\alpha| \leq 1/2$, where the constant $\mathcal{M}^{(i)}$ is defined by

$$\mathcal{M}^{(i)} = \begin{cases} \frac{4}{\log 5} & \text{if } i = 1, \\ 4 & \text{if } i = 2 \end{cases}$$

Now we present a local limit theorem for weighted partitions and selections. We use this result to prove asymptotically tight rejection rates for our Boltzmann sampler as a function of the rightmost pole of $D(s)$. In this statement, $\Gamma(z)$ is the gamma function and $\zeta(s)$ is the Riemann zeta function.

Theorem 6.2.6 ([GS12, Local Limit Theorem])**.** *If conditions (I)–(III) above hold, then*

$$\Pr_{\lambda_n}\left(U_n^{(i)} = n\right) \sim \frac{1}{\sqrt{2\pi \mathrm{Var}\left(U_n^{(i)}\right)}} \sim \frac{1}{\sqrt{2\pi K_2^{(i)}}} \left(\frac{K_2^{(i)}}{\rho_r + 1}\right)^{\frac{2+\rho_r}{2(\rho_r+1)}} n^{-\frac{2+\rho_r}{2(\rho_r+1)}},$$

as $n \longrightarrow \infty$*, where the constants* $K_2^{(i)}$ *are defined by*

$$K_2^{(i)} = \begin{cases} A_r \Gamma(\rho_r + 2)\zeta(\rho_r + 1) & \text{if } i = 1, \\ A_r (1 - 2^{-\rho_r})\Gamma(\rho_r + 2)\zeta(\rho_r + 1) & \text{if } i = 2. \end{cases}$$

6.3 Sampling Algorithms

In this section we present our sampling algorithms for Bose–Einstein condensates and a broad class of weighted partitions and selections. The core idea behind our algorithms is that these random objects can be generated by systematically combining the results of independent random processes that draw from simpler probability distributions. We begin by defining several of the primitive distributions that are fundamental to the subroutines of our algorithms.

We let Geometric(λ) denote the *geometric distribution* with success probability λ and probability density function $\Pr_\lambda(k) = (1 - \lambda)^k \lambda$, for all $k \in \mathbb{Z}_{\geq 0}$. Similarly, let Bernoulli(λ) denote the *Bernoulli distribution* with success probability λ. We heavily rely on the next two distributions in order to fully exploit the structure of Bose–Einstein condensates. These allow us to design a substantially more efficient and better-tailored algorithm.

Poisson Distribution. Let Poisson(λ) denote the *Poisson distribution* with rate parameter λ and probability density function

$$\Pr_\lambda(k) = \frac{\lambda^k}{e^\lambda k!},$$

for all $k \in \mathbb{Z}_{\geq 0}$. The zero-truncated Poisson distribution $\text{Poisson}_{\geq 1}(\lambda)$ (i.e., conditioned on the event $k \geq 1$) with rate parameter λ has the probability density function

$$\Pr_\lambda(k) = \frac{\lambda^k}{(e^\lambda - 1)k!}.$$

Negative Binomial Distribution. Let NegativeBinomial(r, λ) denote the *negative binomial distribution* with r failures, success probability λ, and probability density function

$$\Pr_\lambda(k) = \binom{k + r - 1}{r - 1} \lambda^k (1 - \lambda)^r,$$

for all $k \in \mathbb{Z}_{\geq 0}$. The zero-truncated negative binomial distribution is NegativeBinomial$_{\geq 1}(r, \lambda)$.

6.3.1 Sampling Bose–Einstein Condensates

Our sampling algorithm RandomBoseEinsteinCondensate for Bose–Einstein condensates of size n is given in Algorithm 6.1. Recall that the truncated class of Bose–Einstein condensates that we consider is $\textsc{MSet}(\textsc{MSet}_{1..n}(3\mathcal{Z}))$. The first of the two main subroutines in Algorithm 6.1 is the templated Boltzmann sampler for the class $\textsc{MSet}(\mathcal{A})$, where $\mathcal{A}$ is any combinatorial class with $a_0 = 0$. The second subroutine $\Gamma\textsc{MSet}_{1..n}[d\mathcal{Z}](\lambda)$ is a Boltzmann sampler for the nonempty multisets of d different colored atoms of cardinality at most n. We show in Section 6.5 that we can efficiently implement $\Gamma\textsc{MSet}_{1..n}[d\mathcal{Z}](\lambda)$ for values of λ_n^k by using $\Gamma\textsc{MSet}_{\geq 1}[d\mathcal{Z}](\lambda)$ and rejection sampling.

Algorithm 6.1 Algorithm for uniformly sampling Bose–Einstein condensates of size n.

procedure RandomBoseEinsteinCondensate(n)
  $\lambda_n \leftarrow$ Solution to $\sum_{k=1}^{n} k\binom{k+2}{2}\lambda^k/(1-\lambda^k) = n$
  repeat
    $\gamma \leftarrow \Gamma\text{MSet}[\text{MSet}_{1..n}(3\mathcal{Z})](\lambda_n)$
  until $|\gamma| = n$
  return γ
procedure $\Gamma\text{MSet}[\mathcal{A}](\lambda)$
  $\gamma \leftarrow$ Empty associative array
  $k_0 \leftarrow \text{MaxIndex}(A, \lambda)$
  for $k = 1$ **to** k_0 **do**
    if $k < k_0$ **then**
      $m \leftarrow \text{Poisson}(A(\lambda^k)/k)$
    else
      $m \leftarrow \text{Poisson}_{\geq 1}(A(\lambda^k)/k)$
    for $j = 1$ **to** m **do**
      $\alpha \leftarrow \Gamma A(\lambda^k)$
      $\gamma[\alpha] \leftarrow \gamma[\alpha] + k$
  return γ
procedure $\Gamma\text{MSet}_{1..n}[d\mathcal{Z}](\lambda)$
  repeat
    $m \leftarrow \text{NegativeBinomial}_{\geq 1}(d, \lambda)$
  until $m \leq n$
  return RandomMultiset(m, d)

The templated subroutine $\Gamma\text{MSet}[\mathcal{A}](\lambda)$ repeatedly makes calls to the sampler of the input class $\Gamma A(\lambda^k)$ for different values of k and is part of the Boltzmann sampling framework for combinatorial classes that can be constructed using the symbolic method [FFP07]. The algorithm $\Gamma\text{MSet}[\mathcal{A}](\lambda)$ is a manifestation of the exp-log transform of the generating function for the multiset construction in Equation (6.3) that leverages the observation that a geometric random variable can be decomposed into an infinite sum of independent, scaled Poisson random variables. Concretely, if $(Y_k)_{k=1}^{\infty}$ is a sequence of independent random variables such that $Y_k \sim \text{Poisson}(\lambda^k/k)$ with $\lambda < 1$ and we let $X = \sum_{k=1}^{\infty} kY_k$, then the random variable $X \sim \text{Geometric}(1-\lambda)$. We direct the reader to the proof of [FFP07, Proposition 2.1] for more details. This implementation of $\text{MSet}[\mathcal{A}](\lambda)$ samples from the distribution $\text{MaxIndex}(A, \lambda)$, which is defined to have the cumulative density

function

$$\Pr_\lambda(U \le k) = \frac{\prod_{j=1}^{k} \exp\left(\frac{1}{j}A\left(\lambda^j\right)\right)}{\prod_{j=1}^{\infty} \exp\left(\frac{1}{j}A(\lambda^j)\right)} = \prod_{j=k+1}^{\infty} \exp\left(-\frac{1}{j}A\left(\lambda^j\right)\right), \tag{6.4}$$

with support $k \in \mathbb{Z}_{\ge 1}$. In our model of computation, sampling from the distribution $\textsc{MaxIndex}(A, \lambda)$ is efficient. The following result gives guarantees for the time and space complexity of the $\Gamma\textsc{MSet}[\mathcal{A}](\lambda)$ subroutine.

Proposition 6.3.1 ([FFP07])**.** *The algorithm* $\Gamma\textsc{MSet}[\mathcal{A}](\gamma)$ *is a valid Boltzmann sampler for* $\textsc{MSet}(\mathcal{A})$*. Moreover, if the time and space complexities of* $\Gamma A(\lambda)$ *are, in the worst case, linear in the size of the object produced, the time and space complexities of* $\Gamma\textsc{MSet}[\mathcal{A}](\gamma)$ *are also linear in the size of the object produced.*

For the second subroutine $\Gamma\textsc{MSet}_{1..n}[d\mathcal{Z}](\lambda)$, our crucial observation is that a negative binomial experiment with d failures can be interpreted as a multiset of d different colored atoms using the classic combinatorial idea of stars and bars. In such an experiment, a successful trial adds a new atom of the current color and a failing trial inserts a new bar that separates atoms of different colors. This process allows us to efficiently use the zero-truncated negative binomial distribution to implement $\Gamma\textsc{MSet}_{\ge 1}[d\mathcal{Z}](\lambda)$. In particular, once m is determined, the function $\textsc{RandomMultiset}(m, d)$ returns one of the $\binom{m+d-1}{d-1}$ multisets of size m uniformly at random.

6.3.2 Sampling Weighted Partitions and Selections

Next, we consider the more general problem of uniformly sampling weighted partitions or selections of size n. The primary subroutine for the algorithms RandomWeightedPartition and RandomSelection in Algorithm 6.2 is the Boltzmann sampler $\Gamma C_n^{(i)}(\lambda)$. This procedure is the natural algorithmic interpretation of the truncated generating functions for weighted partitions

and selections that iterates over all types of summands of size at most n and determines how many parts of each type appear in the final object. The complete factorization of these generating functions corresponds to the decomposable structure that allows us to leverage independent random processes.

Algorithm 6.2 Sampling Algorithms for weighted partitions and selections of size n.

procedure RandomWeightedPartition(n)
    $\lambda_n \leftarrow$ Solution to $\sum_{k=1}^{n} k b_k \lambda^k/(1-\lambda^k) = n$
    repeat
        $\gamma \leftarrow \Gamma C_n^{(1)}(\lambda_n)$
    until $|\gamma| = n$
    return γ
procedure RandomSelection(n)
    $\lambda_n \leftarrow$ Solution to $\sum_{k=1}^{n} k b_k \lambda^k/(1+\lambda^k) = n$
    repeat
        $\gamma \leftarrow \Gamma C_n^{(2)}(\lambda_n)$
    until $|\gamma| = n$
    return γ
procedure $\Gamma C_n^{(i)}(\lambda)$
    $\gamma \leftarrow$ Empty associative array
    for $k = 1$ **to** n **do**
        for $j = 1$ **to** b_k **do**
            if $i = 1$ **then** ▷ Weighted partitions
                $m \leftarrow$ Geometric$(1-\lambda^k)$
            else ▷ Selections
                $m \leftarrow$ Bernoulli$(\lambda^k/(1+\lambda^k))$
            if $m \geq 1$ **then**
                $\gamma[(k,j)] \leftarrow m$
    return γ

6.4 Analysis

Now we analyze the running time and space complexity of the procedures in Algorithm 6.1 and Algorithm 6.2. In particular, we bound the rejection rates of these Boltzmann samplers through the singularity analysis of an associated Dirichlet generating function and using the Khintchine–

Meinardus probabilistic method. Once we establish guarantees for these rejection rates, we prove our two main theorems about algorithms for uniformly sampling Bose–Einstein condensates, weighted partitions, and selections.

6.4.1 Tuning the Boltzmann Samplers

The product form of the generating functions for weighted partitions and selections has a useful probabilistic interpretation in the context of Boltzmann sampling. For the two truncated classes, we have

$$U_n^{(i)} = \sum_{k=1}^{n} \sum_{j=1}^{b_k} k Y_{k,j}^{(i)}, \tag{6.5}$$

where

$$Y_{k,j}^{(i)} \sim \begin{cases} \text{Geometric}\left(1 - \lambda^k\right) & \text{if } i = 1, \\ \text{Bernoulli}\left(\frac{\lambda^k}{1+\lambda^k}\right) & \text{if } i = 2. \end{cases}$$

This decomposition follows from iterating over all of the types of summands of size at most n and viewing the number of times each part appears in the final object as a random variable. The key observation here is that these processes are independent of one another.

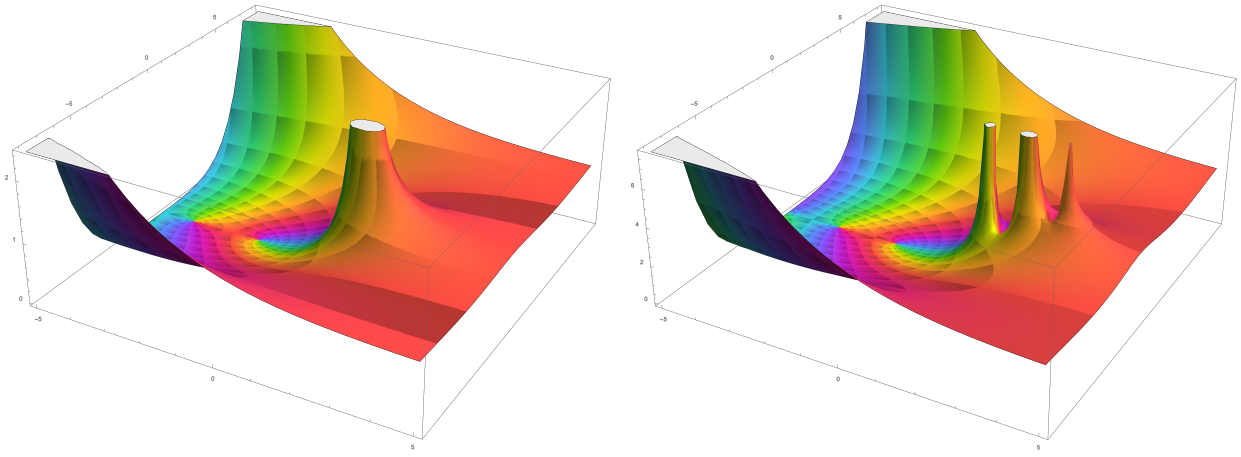
Lemma 6.4.1. *For any $0 \le \lambda < 1$, we have*

$$\mathbb{E}_\lambda\left[U_n^{(i)}\right] = \sum_{k=1}^{n} k b_k \left(\frac{\lambda^k}{1 + (-1)^i \lambda^k}\right).$$

Proof. The result follows from the linearity of expectation and the means of the variables $Y_{k,j}$. □

6.4.2 Bounding the Rejection Rate for Weighted Partitions and Selections

In this subsection we show that the local limit theorem in Theorem 6.2.6 holds for a broad family of weighted partitions and selections, which in turn yields a rigorous rejection rate for all of our sampling algorithms. In particular, for any class of weighted partitions or selections parameterized



(a) (b)

Figure 6.2: Complex three-dimensional plots of (a) the Riemann zeta function and (b) the Dirichlet generating function for Bose–Einstein condensates in the neighborhood of their singularities.

by a positive integer sequence of degree r, we bound the residue of the rightmost pole of the Dirichlet series $D(s)$ and therefore obtain a tight lower bound for $\Pr_{\lambda_n}(U_n^{(i)} = n)$ as a function of the degree r.

Definition 6.4.2. Let a *positive integer sequence of degree r* be a sequence of positive integers $(b_k)_{k=1}^{\infty}$ such that $b_k = p(k)$ for the polynomial

$$p(x) = a_0 + a_1 x + \cdots + a_r x^r \in \mathbb{R}[x],$$

where $\deg(p) = r$.

In order to interface with the Khintchine–Meinardus probabilistic method, we need to use results from analytic number theory about the Riemann zeta function. The Riemann zeta function $\zeta(s)$, with $s = \sigma + it$, is defined as the analytic continuation of the convergent series

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s},$$

for all $\sigma > 1$, into the entire complex plane. The only singularity of $\zeta(s)$ is a simple pole at $s = 1$.

The following lemma shows that the Dirichlet generating function $D(s)$ for the sequence $(b_k)_{k=1}^{\infty}$ is a linear combination of shifted Riemann zeta functions scaled by the coefficients of their parameterizing polynomial $p(x)$. We illustrate this With this, we can easily compute the residues of the poles of $D(s)$ and satisfy conditions (I)–(III) in Section 6.2.4.

Lemma 6.4.3. *If $(b_k)_{k=1}^{\infty}$ is a positive integer sequence of degree r, its Dirichlet generating series is*

$$D(s) = \sum_{k=0}^{r} a_k \zeta(s-k),$$

and it satisfies conditions (I)–(III). Moreover, $D(s)$ has at most $r+1$ simple poles on the positive real axis at positions $\rho_k = k+1$ with residue $A_k = a_k$ if and only if $a_k \neq 0$, for $k \in \{0, 1, \ldots, r\}$.

Proof. The Riemann zeta function converges uniformly and is analytic on $\mathbb{C} \setminus \{1\}$, so

$$\begin{aligned} D(s) &= \sum_{k=1}^{\infty} \frac{b_k}{k^s} \\ &= \sum_{k=1}^{\infty} \frac{a_0 + a_1 k + \cdots + a_r k^r}{k^s} \\ &= \sum_{k=0}^{r} a_k \zeta(s-k). \end{aligned}$$

Since $\zeta(s)$ has a simple pole at $s = 1$ with residue 1, the result about the poles and residues of $D(s)$ immediately follows. We can satisfy condition (I) by setting $C_0 = 1$. For condition (II), we set $C_1 = 2 + r$ since $D(s)$ is a linear combination of shifted zeta functions and use the follow fact in [BD04, Section 8.2]:

$$\zeta(s) = \begin{cases} O\left(t^{1/2-\sigma}\right) & \text{if } \sigma < 0, \\ O(t) & \text{if } 0 \leq \sigma \leq 1, \\ O(1) & \text{if } 1 < \sigma. \end{cases}$$

To show condition (III), we follow an approach similar to the proof of [GSE08, Lemma 1], which uses the following inequality of Karatsuba and Voronin [KV92, Section 4.2, Lemma 1] for

trigonometric sums related to the Riemann zeta function. This result states that for all positive integers m, we have

$$2\sum_{k=1}^{m}\sin^2(\pi k\alpha) \geq m\left(1-\min\left(1,\frac{1}{2m|\alpha|}\right)\right). \tag{6.6}$$

By assumption, $(b_k)_{k=1}^{\infty}$ is a sequence of positive integers and $0 < \sqrt{\delta} \leq |\alpha| \leq 1/2$, so

$$\begin{aligned}2\sum_{k=1}^{\infty} b_k e^{-k\delta}\sin^2(\pi k\alpha) &\geq 2\sum_{k=1}^{m} e^{-k\delta}\sin^2(\pi k\alpha)\\ &\geq e^{-m\delta}m\left(1-\min\left(1,\frac{1}{2m|\alpha|}\right)\right).\end{aligned}$$

Using Equation (6.6), we set $m = \lceil 1/(2|\alpha|) + 1/\delta\rceil \geq 1$ so that

$$\begin{aligned}2\sum_{k=1}^{\infty} b_k e^{-k\delta}\sin^2(\pi k\alpha) &\geq e^{-m\delta}\left(m-\frac{1}{2|\alpha|}\right)\\ &\geq e^{-\left(\frac{1}{2|\alpha|}+\frac{1}{\delta}+1\right)\delta}\delta^{-1}\\ &\geq e^{-\left(\frac{\sqrt{\delta}}{2}+1+\delta\right)}\delta^{-1}.\end{aligned}$$

Recall that the position of the pole $\rho_r = r + 1$ is fixed and $\mathcal{M}^{(i)}$ is constant. Therefore, it follows that

$$e^{-\left(\frac{\sqrt{\delta}}{2}+1+\delta\right)}\frac{1}{\delta|\log\delta|} \geq \mathcal{M}^{(i)}\left(1+\frac{\rho_r}{2}+\varepsilon\right),$$

for a sufficiently small value of δ and taking $\varepsilon = 1$, which completes the proof. □

Now we use Lemma 6.4.3 and the method of finite differences to lower bound the rightmost residue A_r, which allows us to more conveniently analyze the rejection rates of our algorithms using the local limit theorem. Specifically, we bound the coefficients of $p(x)$ in the binomial basis centered at $x = 0$, as illustrated in [BFT16].

Definition 6.4.4. Let the *forward difference* operator Δ denote

$$\Delta p(x) = p(x+1) - p(x).$$

We recursively define higher order differences by

$$\Delta^n p(x) = \Delta^{n-1} p(x+1) - \Delta^{n-1} p(x).$$

Viewing $p(x)$ as a Newton interpolating polynomial, we can recover the change of basis

$$p(x) = \sum_{j=0}^{r} \Delta^j p(0) \binom{x}{j}.$$

The following result is a simple consequence of this binomial basis representation and states that integer-valued polynomials necessarily have integer-valued forward differences.

Lemma 6.4.5 ([Sta11])**.** *Let $p(x) \in \mathbb{R}[x]$ be a polynomial of degree r. We have the property $p(n) \in \mathbb{Z}$, for all $n \in \mathbb{Z}$, if and only if $\Delta^j p(0) \in \mathbb{Z}$, for all $0 \le j \le r$.*

Now we can present our lower bound on the rightmost residue A_r, which we use with the local limit theorem (Theorem 6.2.6) to obtain a more convenient rejection rate.

Lemma 6.4.6. *If $(b_k)_{k=1}^{\infty}$ is a positive integer sequence of degree r, then for n sufficiently large, we have*

$$\Pr_{\lambda_n}\left(U_n^{(i)} = n\right) \ge \frac{1}{2\sqrt{2\pi}}((r+2)n)^{-\frac{r+3}{2(r+2)}}.$$

Proof. Theorem 6.2.6 holds for $U_n^{(i)}$ by Lemma 6.4.3. Since $(b_k)_{k=1}^{\infty}$ is a positive integer sequence of degree r, Lemma 6.4.5 implies that $p(x)$ has integral coefficients $\Delta^j p(0)$ in the binomial basis. Therefore, the leading coefficient in the binomial basis $\Delta^r p(0)$ is a positive integer, so the leading coefficient in the monomial basis a_r satisfies

$$a_r = \Delta^r p(0) \cdot [x^r] \binom{x}{r} = \Delta^r p(0) \cdot [x^r] \frac{x(x-1)\ldots(x-r+1)}{r!} \ge \frac{1}{r!}.$$

In the equation above, the $[x^k]$ operator extracts the coefficient of x^k. The residue $A_r = a_r$ by

Lemma 6.4.3, so we can lower bound $K_2^{(i)}$ by

$$K_2^{(i)} \geq \frac{A_r}{2}\Gamma(\rho_r + 2)\zeta(\rho_r + 1) \geq \frac{1}{2r!}(r+2)! \geq 1,$$

since $\rho_r = r + 1$ and $\zeta(n) \geq 1$, for all $n \geq 2$. It follows that for $\varepsilon = 1/2$ and n sufficiently large, Theorem 6.2.6 gives us

$$\begin{aligned}
\Pr_{\lambda_n}\left(U_n^{(i)} = n\right) &\geq (1-\varepsilon)\frac{1}{\sqrt{2\pi K_2^{(i)}}}\left(\frac{K_2^{(i)}}{(\rho_r+1)n}\right)^{\frac{2+\rho_r}{2(\rho_r+1)}} \\
&\geq \frac{1}{2\sqrt{2\pi}}\left(K_2^{(i)}\right)^{\frac{1}{2(r+2)}}((r+2)n)^{-\frac{r+3}{2(r+2)}} \\
&\geq \frac{1}{2\sqrt{2\pi}}((r+2)n)^{-\frac{r+3}{2(r+2)}},
\end{aligned}$$

which completes the proof. □

6.4.3 Proof of the Main Theorems

Recall that we follow the convention of using the real-arithmetic model of computation and an oracle that can evaluate a generating function within its radius of convergence in constant time [BFP10, FFP07]. A consequence of this assumption is that we can sample from the distributions Geometric(λ), Poisson(λ), MaxIndex(A, λ), etc., iteratively in $O(m)$ time, where m is the value of the generated sample.

We now proceed by proving our two main theorems, which we restate for convenience.

Theorem 6.1.1. *There exists a uniform sampling algorithm for Bose–Einstein condensates of size n that runs in expected $O(n^{1.625})$ time and uses $O(n)$ space.*

Proof. We analyze the time and space complexity of RandomBoseEinsteinCondensate and its subroutines in Algorithm 6.1. We can compute the exact value of λ_n in the tuning step in $O(1)$ time using the root-finding oracle and Lemma 6.4.1. The Boltzmann sampler $\Gamma\text{MSet}[\text{MSet}_{1..n}(3\mathcal{Z})](\lambda_n)$

is called at most $O(n^{5/8})$ times in expectation before generating an object of size n by Lemma 6.4.6 since Bose–Einstein condensates are parameterized by a positive integer sequence of degree $r = 2$. We implement this Boltzmann sampler using the subroutines $\Gamma\textsc{MSet}[\mathcal{A}](\lambda)$ and $\Gamma\textsc{MSet}_{1..n}[d\mathcal{Z}](\lambda)$. Lemma 6.5.1 ensures that the time and space complexities of $\Gamma\textsc{MSet}_{1..n}[d\mathcal{Z}](\lambda)$ are linear in the size of the object produced. Proposition 6.3.1 and our choice of λ_n guarantees that $\Gamma\textsc{MSet}[\mathcal{A}](\lambda)$ runs in expected $O(n)$ time and space. Therefore, RandomBoseEinsteinCondensate runs in $O(n^{1.625})$ time and uses $O(n)$ space in expectation.

In order to use deterministic $O(n)$ space, we can modify the Boltzmann sampler to reject partially constructed objects if their size is at least $2n$. By Markov's inequality, this refined Boltzmann sampler outputs objects of size less than $2n$ from the new conditional Boltzmann distribution with probability at least

$$\begin{aligned} 1 - \Pr_{\lambda_n}\left(U_n^{(1)} \geq 2n\right) &\geq 1 - \frac{\mathbb{E}_{\lambda_n}\left[U_n^{(1)}\right]}{2n} \\ &= \frac{1}{2}. \end{aligned}$$

Therefore, at most a constant number of trials are needed in expectation to sample from this tail-truncated Boltzmann distribution. □

The singularity analysis in Lemma 6.4.3 and Lemma 6.4.6 holds for all classes of weighted partitions and selections parameterized by a positive integer sequence of degree r. However, we need a more general Boltzmann sampler to generate these multiplicative objects. The truncated class $\mathcal{C}_n^{(1)}$ of weighted partitions has the generating function

$$C_n^{(1)}(z) = \prod_{k=1}^{n}(1 - z^k)^{-b_k},$$

so we can use independent geometric random variables to randomly sample the number of parts each type of summand contributes to the final configuration. For selections, the truncated

generating function $C_n^{(2)}$ analogously implies that we can use independent Bernoulli random variables to sample from the corresponding Boltzmann distribution. See Boltzmann samplers for the Cartesian product, sequence construction, and power set operator in [FFP07] for more details.

Theorem 6.1.2. *There exists a uniform sampling algorithm for any class of weighted partitions or selections parameterized by a positive integer sequence of degree r for objects of size n that runs in expected time $O(n^{r+1+(r+3)/(2r+4)})$ and uses $O(n)$ space.*

Proof. We start by considering weighted partitions and the procedure RandomWeightedPartition in Algorithm 6.2. The tuning step is the same as in the proof of Theorem 6.1.1 and takes constant time in our model of computation. The Boltzmann distribution for the truncated class $\mathcal{C}_n^{(1)}$ is the product of

$$\sum_{k=1}^{n} b_k = O\left(n^{r+1}\right)$$

geometric distributions, and sampling from each of these geometric distributions takes time and space proportional to the number they output. The total number of positive geometric trials performed across all of the $O(n^{r+1})$ distributions is $O(n)$ by our choice of λ_n. Therefore, the time and space complexities of $\Gamma C_n^{(1)}(\lambda_n)$ are $O(n^{r+1})$ and $O(n)$ in expectation, respectively. Using Markov's inequality as in the proof of Theorem 6.1.1, we can also guarantee that the algorithm uses deterministic $O(n)$ space. Since the number of summands $(b_k)_{k=1}^{\infty}$ is given by a positive integer sequence of degree r, we can use Lemma 6.4.6 to bound the rejection rate. We expect to generate a weighted partition of size n in $O(n^{(r+3)/(2r+4)})$ calls to $\Gamma C_n^{(1)}(\lambda_n)$, so the result for weighted partitions follows. The analysis for selections is similar, but we sample from Bernoulli distributions instead of from geometric distributions. □

6.5 Sampling from Truncated Negative Binomial Distributions

In this section we investigate properties of negative binomial distributions in the context of rejection sampling. Specifically, our main result (Lemma 6.5.1) shows that the rejection sampling

in the subroutine $\Gamma\text{MSET}_{1..n}[d\mathcal{Z}](\lambda)$ of Algorithm 6.1 requires a constant number of trials in expectation each time it is called by the multiset sampler $\Gamma\text{MSET}[\mathcal{A}](\lambda)$. We generalize our analysis from the case $d = 3$ (i.e., Bose–Einstein condensates) to $d \geq 1$ so that our arguments can be used in other contexts. One of the highlights of our analysis is a simple but effective tail inequality for negative binomial random variables parameterized by high success probabilities (Lemma 6.5.4) that outperforms the standard Chernoff-type inequality in this setting.

For the remainder of the chapter, let $[n] = \{1, 2, \ldots, n\}$ and let V be the random variable for the size of an object drawn from $\Gamma\text{MSET}_{\geq 1}(d\mathcal{Z})(\lambda)$. The symbolic method implies that the generating function for $\mathcal{B} = \text{MSET}(d\mathcal{Z})$ (the multisets of d distinct atoms) is

$$B(z) = \sum_{k=0}^{\infty} \binom{k+d-1}{d-1} z^k = \frac{1}{(1-z)^d},$$

so it follows that $V \sim \text{NegativeBinomial}_{\geq 1}(d, \lambda)$. We also let $W \sim \text{NegativeBinomial}(d, \lambda)$.

Lemma 6.5.1. *For n sufficiently large and all $k \in [n]$, we have*

$$\Pr_{\lambda_n^k}(V \leq n) \geq \frac{1}{2}.$$

We prove Lemma 6.5.1 using three prerequisite lemmas. First, recall that the probability mass function for the negative binomial distribution is

$$\Pr_\lambda(W = k) = \binom{k+d-1}{d-1} \lambda^k (1-\lambda)^d, \tag{6.7}$$

The cumulative distribution function for W can be written as

$$\Pr_\lambda(W \leq k) = 1 - I_\lambda(k+1, d), \tag{6.8}$$

where $I_\lambda(a, b)$ is the regularized incomplete beta function defined as

$$I_\lambda(a, b) \stackrel{\text{def}}{=} \frac{B_\lambda(a, b)}{B_1(a, b)},$$

with

$$B_\lambda(a, b) \stackrel{\text{def}}{=} \int_0^\lambda t^{a-1}(1-t)^{b-1}\mathrm{d}t.$$

The first of the three lemmas (Lemma 6.5.2) shows that it suffices to lower bound the success probability in the subroutine $\Gamma\text{MSET}_{1..n}[d\mathcal{Z}](\lambda^k)$ for $k = 1$ instead of all $k \in [n]$.

Lemma 6.5.2. *For all $k \in [n]$, we have*

$$\Pr_{\lambda_n^k}(V \le n) \ge \Pr_{\lambda_n}(V \le n).$$

Proof. Using the probability mass function Equation (6.7) and cumulative density function Equation (6.8), observe that the conditional probability for the zero-truncated negative binomial distribution is

$$\begin{aligned}
\Pr_{\lambda_n^k}(V \le n) &= \frac{\Pr_{\lambda_n^k}(W \le n) - \Pr_{\lambda_n^k}(W = 0)}{1 - \Pr_{\lambda_n^k}(W = 0)} \\
&= \frac{1 - I_{\lambda_n^k}(n+1, d) - \left(1-\lambda_n^k\right)^d}{1 - (1-\lambda_n^k)^d} \\
&= 1 - \frac{I_{\lambda_n^k}(n+1, d)}{1 - (1-\lambda_n^k)^d}.
\end{aligned}$$

The integrand of the beta function $t^{a-1}(1-t)^{b-1}$ is positive on $(0, 1)$ and $\lambda_n \in (0, 1)$, so it follows that

$$1 - \frac{I_{\lambda_n^k}(n+1, d)}{1 - (1-\lambda_n^k)^d} \ge 1 - \frac{I_{\lambda_n}(n+1, d)}{1 - (1-\lambda_n)^d},$$

for all $k \in [n]$, which concludes the proof. □

Therefore, it is sufficient to consider the case $k = 1$ when analyzing the rejection rate for sampling from the zero-truncated distribution $\text{NegativeBinomial}_{\geq 1}(d, \lambda_n^k)$.

The second prerequisite lemma (Lemma 6.5.3) gives an upper bound and lower bound for λ_n using an asymptotic formula from the Khintchine–Meinardus probabilsitic method.

Lemma 6.5.3. *In the algorithm* $\textsc{MSet}(\textsc{MSet}_{1..n}(d\mathcal{Z}))$*, for n sufficiently large, the value of* λ_n *satisfies*

$$\exp\left(-2n^{-\frac{1}{d+1}}\right) \leq \lambda_n \leq \exp\left(-\frac{1}{2}n^{-\frac{1}{d+1}}\right).$$

Proof. Let $\lambda_n = e^{-\delta_n}$. Equation (32) in [GS12] asserts that as $n \longrightarrow \infty$, δ_n satisfies

$$\delta_n \sim (A_r \Gamma(\rho_r)\zeta(\rho_r + 1)\rho_r)^{\frac{1}{\rho_r+1}} n^{-\frac{1}{\rho_r+1}}.$$

In this setting, the degree of the polynomial parameterizing the sequence $b_k = \binom{k+d-1}{d-1}$ is $r = d - 1$. Therefore, Lemma 6.4.3 implies that the rightmost pole is $\rho_r = d$ and its residue is $A_r = 1/(d-1)!$. Since $\Gamma(\rho_r) = (d-1)!$, we have

$$\delta_n \sim (\zeta(d+1)d)^{\frac{1}{d+1}} n^{-\frac{1}{d+1}},$$

as $n \longrightarrow \infty$. Therefore, for any $\varepsilon > 0$ and n sufficiently large, we have the inequalities

$$\delta_n \geq (1-\varepsilon)(\zeta(d+1)d)^{\frac{1}{d+1}} n^{-\frac{1}{d+1}},$$

and

$$\delta_n \leq (1+\varepsilon)(\zeta(d+1)d)^{\frac{1}{d+1}} n^{-\frac{1}{d+1}}.$$

Since $d \geq 1$, we know that the Riemann zeta function satisfies $1 \leq \zeta(d+1) \leq \zeta(2) = \pi^2/6$. Using these bounds and setting $\varepsilon = 1/3$ completes the proof. □

The third prerequisite lemma (Lemma 6.5.4) is a new tail inequality for negative binomial random variables. Although our derivation is easily understood using standard techniques from enumerative combinatorics, this inequality captures an ample amount of probability mass for all integers $n \geq 0$. It is empirically tighter than the Chernoff-type inequality as $n \longrightarrow \infty$ and $\lambda_n \longrightarrow 1$.

Lemma 6.5.4 (Negative Binomial Tail Inequality)**.** *If* $W \sim \text{NegativeBinomial}(d, \lambda)$*, then for all integers* $n \geq 0$*, we have*

$$\Pr(W > n) \leq 1 - \left(1 - \lambda^{n/d}\right)^d.$$

Proof. Let $m = \lfloor n/d \rfloor$ and observe that

$$\sum_{k=0}^{n} \binom{k+d-1}{d-1} \lambda^k \geq \left(\sum_{k=0}^{m} \lambda^k\right)^d.$$

This inequality has a direct combinatorial interpretation in terms of weak d-compositions. The left-hand side is the truncated ordinary generating function for weak compositions of k into d parts, while the right-hand side is the ordinary generating function for weak compositions of k into d parts of size at most m. Using a property of geometric sequences, it follows that

$$\left(\sum_{k=0}^{m} \lambda^k\right)^d = \left(\frac{1-\lambda^{m+1}}{1-\lambda}\right)^d \geq \left(1 - \lambda^{n/d}\right)^d (1-\lambda)^{-d},$$

since $0 < \lambda < 1$. Therefore, the probability mass function in Equation (6.7) and the two inequalities above imply that

$$\begin{aligned}
\Pr(W > n) &= 1 - \Pr(W \leq n) \\
&= 1 - \sum_{k=0}^{n} \binom{k+d-1}{d-1} \lambda^k (1-\lambda)^d \\
&\leq 1 - \left(1 - \lambda^{n/d}\right)^d,
\end{aligned}$$

as desired. □

We can now prove Lamma 6.5.1 by combining the three prerequisite lemmas and using the definition of the probability mass function for the negative binomial distribution.

Proof of Lemma 6.5.1. We only need to consider the case when $k = 1$ by Lemma 6.5.2. Using Lemma 6.5.4 and Lemma 6.7, we have

$$\begin{aligned}\Pr_{\lambda_n}(V \leq n) &= \frac{\Pr_{\lambda_n}(W \in [n])}{\Pr_{\lambda_n}(W \geq 1)} \\ &\geq \Pr_{\lambda_n}(W \leq n) - \Pr_{\lambda_n}(W = 0) \\ &\geq \left(1 - \lambda_n^{n/d}\right)^d - (1 - \lambda_n)^d.\end{aligned}$$

The upper and lower bounds for λ_n obtained in Lemma 6.5.3 imply that

$$\left(1 - \lambda_n^{n/d}\right)^d - (1 - \lambda_n)^d \geq \left(1 - \exp\left(-\frac{1}{2d} n^{\frac{d}{d+1}}\right)\right)^d - \left(1 - \exp\left(-2n^{-\frac{1}{d+1}}\right)\right)^d \geq \frac{1}{2},$$

for sufficiently large values of n because we have the limits

$$\lim_{n \to \infty} \exp\left(-n^{\frac{d}{d+1}}\right) = 0,$$

and

$$\lim_{n \to \infty} \exp\left(-n^{-\frac{1}{d+1}}\right) = 1.$$

This completes the proof. □

6.6 Summary and Discussion

In this chapter we have shown how to analyze the time complexity of Boltzmann samplers for a family of weighted partitions (including Bose–Einstein condensates) and selections through the singularity analysis of Dirichlet generating functions. We do so by relating the degree of

the polynomial that parameterizes the sequence $(b_k)_{k=1}^{\infty}$ to the rejection rate of our sampling algorithms through the local limit theorem in [GS12]. The main observation in our analysis is that the Dirichlet generating function for a positive integer sequence of degree r is a linear combination of shifted and scaled Riemann zeta functions. This allows us to apply results from analytic number theory to analyze the poles of these functions. In our analysis, we also use Newton interpolating polynomials to bound residues and we develop a new negative binomial tail inequality to analyze an intermediate rejection rate.

Future directions of this work include analyzing these kinds of generating-function based sampling algorithms in the interval or floating-point arithmetic models of computation, as opposed to the idealized real-arithmetic model. The primary question to address here is how accurate the approximation of λ_n needs to be in order to achieve a similar rejection rate, since λ_n approaches an essential singularity of the generating function for weighted partitions $C_n^{(1)}(z)$.

CHAPTER 7

CONCLUSION

The research in this dissertation introduces several new probabilistic techniques for designing Markov chain Monte Carlo algorithms and analyzing the rate at which various lattice models in statistical physics converge to equilibrium. We start by showing that the mixing time of Glauber dynamics for the six-vertex model can be exponentially slow in its ordered phases, resolving a widely-believed conjecture about the dynamics of weighted, intersecting lattice path models. In particular, we show there exist boundary conditions for which local Markov chains are slowly mixing in all of the ferroelectric phase. We also analyze the Glauber dynamics in the antiferroelectric phase subject to free boundary conditions and significantly extend the subregion for which Glauber dynamics was previously known to be slow mixing [Liu18]. Next, we leverage connections between random walks and electrical networks to prove nearly tight bounds for the transience class of the Abelian sandpile model, closing an open problem of Babai and Gorodevsky [BG07]. In the following chapter, we present a new technique called Boltzmann sampling with balanced bias, which allows us to design Markov chain Monte Carlo algorithms for uniformly sampling elements with a fixed rank from a particular family of graded posets in polynomial time. We then explore how this approach applies to uniformly sampling integer partitions subject to a variety of geometric constraints. Lastly, we take a noticeably different approach to Boltzmann sampling and use ideas from analytic combinatorics to design uniform sampling algorithms for combinatorial objects with multiplicative structure. This chapter primarily focuses on sampling Bose–Einstein condensates and bounding rejection rates through the analysis of Dirichlet generating functions.

REFERENCES

[AHS85] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski. "A learning algorithm for Boltzmann machines". In: *Cognitive Science* 9.1 (1985), pp. 147–169.

[Ald99] D. Aldous. "Deterministic and stochastic models for coalescence (aggregation and coagulation): A review of the mean-field theory for probabilists". In: *Bernoulli* 5.1 (1999), pp. 3–48.

[Ald83] D. Aldous. "Random walks on finite groups and rapidly mixing Markov chains". In: *Séminaire de Probabilités XVII 1981/82*. Springer, 1983, pp. 243–297.

[AR05] D. Allison and N. Reshetikhin. "Numerical study of the 6-vertex model with domain wall boundary conditions". In: *Annales de l'institut Fourier* 55.6 (2005), pp. 1847–1869.

[And98] G. E. Andrews. *The Theory of Partitions*. Cambridge University Press, 1998.

[AD16] R. Arratia and S. DeSalvo. "Probabilistic Divide-and-Conquer: A new exact simulation method, with integer partitions as an example". In: *Combinatorics, Probability and Computing* 25.3 (2016), pp. 324–351.

[AT94] R. Arratia and S. Tavaré. "Independent process approximations for random combinatorial structures". In: *Advances in Mathematics* 104.1 (1994), pp. 90–154.

[Asc11] M. Aschwanden. *Self-Organized Criticality in Astrophysics: The Statistics of Nonlinear Processes in the Universe*. Springer Science & Business Media, 2011.

[ABCN14] A. Ayyer, J. Bouttier, S. Corteel, and F. Nunzi. "Multivariate juggling probabilities". In: *Electronic Journal of Probability* 20.5 (2014), pp. 1–29.

[BG07] L. Babai and I. Gorodezky. "Sandpile transience on the grid is polynomially bounded". In: *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA). Society for Industrial and Applied Mathematics. 2007, pp. 627–636.

[Bak96] P. Bak. *How Nature Works: The Science of Self-Organized Criticality*. Copernicus, 1996.

[BTW87] P. Bak, C. Tang, and K. Wiesenfeld. "Self-organized criticality: An explanation of the $1/f$ noise". In: *Physical Review Letters* 59.4 (1987), p. 381.

[BD04] P. T. Bateman and H. G. Diamond. *Analytic Number Theory: An Introductory Course*. World Scientific, 2004.

[Bec+13] L. Becchetti, V. Bonifaci, M. Dirnberger, A. Karrenbauer, and K. Mehlhorn. "Physarum can compute shortest paths: Convergence proofs and complexity bounds". In: *Proceedings of the 40th International Colloquium on Automata, Languages, and Programming* (ICALP). Springer. 2013, pp. 472–483.

[BBD18] M. Bendkowski, O. Bodini, and S. Dovgal. "Polynomial tuning of multiparametric combinatorial samplers". In: *Proceedings of the Fifteenth Workshop on Analytic Algorithmics and Combinatorics* (ANALCO). Society for Industrial and Applied Mathematics. 2018, pp. 92–106.

[BBHM05] I. Benjamini, N. Berger, C. Hoffman, and E. Mossel. "Mixing times of the biased card shuffling and the asymmetric exclusion process". In: *Transactions of the American Mathematical Society* 357.8 (2005), pp. 3013–3029.

[BP07] N. Berestycki and J. Pitman. "Gibbs distributions for random partitions generated by a fragmentation process". In: *Journal of Statistical Physics* 127 (2007), pp. 381–418.

[BFR18] M. Bernstein, M. Fahrbach, and D. Randall. "Analyzing Boltzmann samplers for Bose–Einstein condensates with Dirichlet generating functions". In: *Proceedings of the Fifteenth Workshop on Analytic Algorithmics and Combinatorics* (ANALCO). Society for Industrial and Applied Mathematics. 2018, pp. 107–117.

[BCFR17] P. Bhakta, B. Cousins, M. Fahrbach, and D. Randall. "Approximately sampling elements with fixed rank in graded posets". In: *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA). Society for Industrial and Applied Mathematics. 2017, pp. 1828–1838.

[BFT16] S. Billey, M. Fahrbach, and A. Talmage. "Coefficients and roots of peak polynomials". In: *Experimental Mathematics* 25.2 (2016), pp. 165–175.

[BPR15] A. E. Biondo, A. Pluchino, and A. Rapisarda. "Modeling financial markets by self-organized criticality". In: *Physical Review E* 92.4 (2015), p. 042814.

[BLS91] A. Björner, L. Lovász, and P. W. Shor. "Chip-firing games on graphs". In: *European Journal of Combinatorics* 12.4 (1991), pp. 283–291.

[BF06] P. Bleher and V. Fokin. “Exact solution of the six-vertex model with domain wall boundary conditions. Disordered phase”. In: *Communications in Mathematical Physics* 268.1 (2006), pp. 223–284.

[BL09] P. Bleher and K. Liechty. “Exact solution of the six-vertex model with domain wall boundary conditions. Ferroelectric phase”. In: *Communications in Mathematical Physics* 286.2 (2009), pp. 777–801.

[BL10] P. Bleher and K. Liechty. “Exact solution of the six-vertex model with domain wall boundary conditions: Antiferroelectric phase”. In: *Communications on Pure and Applied Mathematics* 63.6 (2010), pp. 779–829.

[BFP10] O. Bodini, É. Fusy, and C. Pivoteau. “Random sampling of plane partitions”. In: *Combinatorics, Probability and Computing* 19.2 (2010), pp. 201–226.

[BLR14] O. Bodini, J. Lumbroso, and N. Rolin. “Analytic samplers and the combinatorial rejection method”. In: *Proceedings of the Twelfth Workshop on Analytic Algorithmics and Combinatorics* (ANALCO). Society for Industrial and Applied Mathematics. 2014, pp. 40–50.

[BPZ02] N. M. Bogoliubov, A. G. Pronko, and M. B. Zvonarev. “Boundary correlation functions of the six-vertex model”. In: *Journal of Physics A: Mathematical and General* 35.27 (2002), p. 5525.

[BMV12] V. Bonifaci, K. Mehlhorn, and G. Varma. “Physarum can compute shortest paths”. In: *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA). Society for Industrial and Applied Mathematics, 2012, pp. 233–240.

[BKW73] H. J. Brascamp, H. Kunz, and F. Y. Wu. “Some rigorous results for the vertex model in statistical mechanics”. In: *Journal of Mathematical Physics* 14.12 (1973), pp. 1927–1932.

[Bro+16] L. Brochini, A. de Andrade Costa, M. Abadi, A. C. Roque, J. Stolfi, and O. Kinouchi. “Phase transitions and self-organized criticality in networks of stochastic spiking neurons”. In: *Scientific Reports* 6 (2016), p. 35831.

[BD97] R. Bubley and M. Dyer. “Path coupling: A technique for proving rapid mixing in Markov chains”. In: *Proceedings of the 38th Annual IEEE Symposium on Foundations of Computer Science* (FOCS). Institute of Electrical and Electronics Engineers. 1997, pp. 223–231.

[BP83] E. Buffet and J. V. Pulé. “Fluctuation properties of the imperfect Bose gas”. In: *Journal of Mathematical Physics* 24.6 (1983), pp. 1608–1616.

[CFX18] J.-Y. Cai, Z. Fu, and M. Xia. “Complexity classification of the six-vertex model”. In: *Information and Computation* 259 (2018), pp. 130–141.

[CLL19] J.-Y. Cai, T. Liu, and P. Lu. “Approximability of the six-vertex model”. In: *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA). Society for Industrial and Applied Mathematics. 2019, pp. 2248–2261.

[CR16] S. Cannon and D. Randall. “Sampling on lattices with free boundary conditions using randomized extensions”. In: *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA). Society for Industrial and Applied Mathematics. 2016, pp. 1952–1971.

[CMT11] P. Caputo, F. Martinelli, and F. L. Toninelli. “Convergence to equilibrium of biased plane partitions”. In: *Random Structures & Algorithms* 39.1 (2011), pp. 83–114.

[CD14] S. Chatterjee and P. Diaconis. “Fluctuations of the Bose–Einstein condensate”. In: *Journal of Physics A: Mathematical and Theoretical* 47.8 (2014), p. 085201.

[CV12] A. Choure and S. Vishwanathan. “Random walks, electric networks and the transience class problem of sandpiles”. In: *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA). Society for Industrial and Applied Mathematics, 2012, pp. 1593–1611.

[Chr+11] P. Christiano, J. A. Kelner, A. Madry, D. A. Spielman, and S.-H. Teng. “Electrical flows, Laplacian systems, and faster approximation of maximum flow in undirected graphs”. In: *Proceedings of the Forty-Third ACM Symposium on Theory of Computing* (STOC). Association for Computing Machinery, 2011, pp. 273–282.

[Chu+18] T. Chu, Y. Gao, R. Peng, S. Sachdeva, S. Sawlani, and J. Wang. “Graph sparsification, spectral sketches, and faster resistance computation, via short cycle decompositions”. In: *Proceedings of the 59th Annual IEEE Symposium on Foundations of Computer Science* (FOCS). Institute of Electrical and Electronics Engineers. 2018, pp. 361–372.

[CEP96] H. Cohn, N. Elkies, and J. Propp. “Local statistics for random domino tilings of the Aztec diamond”. In: *Duke Mathematics Journal* 85.1 (Oct. 1996), pp. 117–166.

[CMO07] A. Comtet, S. N. Majumdar, and S. Ouvry. “Integer partitions and exclusion statistics”. In: *Journal of Physics A: Mathematical and Theoretical* 40.37 (2007), pp. 11255–11269.

[CFF19] R. Cummings, M. Fahrbach, and A. Fatehpuria. *A fast minimum degree algorithm and matching lower bound*. Preprint, arXiv:1907.12119v1. 2019.

[DP14] S. DeSalvo and I. Pak. “Log-concavity of the partition function”. In: *The Ramanujan Journal* 38.1 (2014), pp. 61–73.

[DM16] S. DeSalvo and G. Menz. *A robust quantitative local central limit theorem with applications to enumerative combinatorics and random combinatorial structures*. Preprint, arXiv:1610.07664v1. 2016.

[Dha90] D. Dhar. “Self-organized critical state of sandpile automaton models”. In: *Physical Review Letters* 64.14 (1990), p. 1613.

[Dha06] D. Dhar. “Theoretical studies of self-organized criticality”. In: *Physica A: Statistical Mechanics and its Applications* 369.1 (2006), pp. 29–70.

[DRSV95] D. Dhar, P. Ruelle, S. Sen, and D.-N. Verma. “Algebraic aspects of Abelian sandpile models”. In: *Journal of Physics A: Mathematical and General* 28.4 (1995), p. 805.

[DS84] P. G. Doyle and J. L. Snell. *Random Walks and Electric Networks*. Mathematical Association of America, 1984.

[DFLS04] P. Duchon, P. Flajolet, G. Louchard, and G. Schaeffer. “Boltzmann samplers for the random generation of combinatorial structures”. In: *Combinatorics, Probability and Computing* 13.4-5 (2004), pp. 577–625.

[DFGX18] D. Durfee, M. Fahrbach, Y. Gao, and T. Xiao. “Nearly tight bounds for sandpile transience on the grid”. In: *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA). Society for Industrial and Applied Mathematics. 2018, pp. 605–624.

[DFK91] M. Dyer, A. Frieze, and R. Kannan. “A random polynomial-time algorithm for approximating the volume of convex bodies”. In: *Journal of the ACM* 38.1 (1991), pp. 1–17.

[DG98] M. Dyer and C. Greenhill. “A more rapidly mixing Markov chain for graph colorings”. In: *Random Structures & Algorithms* 13.3-4 (1998), pp. 285–317.

[Ell+11] W. Ellens, F. Spieksma, P. Van Mieghem, A. Jamakovic, and R. Kooij. "Effective graph resistance". In: *Linear Algebra and its Applications* 435.10 (2011), pp. 2491–2506.

[ES77] C. Everett and P. Stein. "The combinatorics of random walk with absorbing barriers". In: *Discrete Mathematics* 17.1 (1977), pp. 27–45.

[FMZ19a] M. Fahrbach, V. Mirrokni, and M. Zadimoghaddam. "Non-monotone submodular maximization with nearly optimal adaptivity and query complexity". In: *International Conference on Machine Learning* (ICML). 2019, pp. 1833–1842.

[FMZ19b] M. Fahrbach, V. Mirrokni, and M. Zadimoghaddam. "Submodular maximization with nearly optimal approximation, adaptivity and query complexity". In: *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA). Society for Industrial and Applied Mathematics. 2019, pp. 255–273.

[FR19] M. Fahrbach and D. Randall. "Slow mixing of Glauber dynamics for the six-vertex model in the ordered phases". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques* (RANDOM). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. 2019.

[Fah+18] M. Fahrbach, G. L. Miller, R. Peng, S. Sawlani, J. Wang, and S. C. Xu. "Graph sketching against adaptive adversaries applied to the minimum degree algorithm". In: *Proceedings of the 59th Annual IEEE Symposium on Foundations of Computer Science* (FOCS). Institute of Electrical and Electronics Engineers. 2018, pp. 101–112.

[FS06] P. L. Ferrari and H. Spohn. "Domino tilings and the six-vertex model at its free-fermion point". In: *Journal of Physics A: Mathematical and General* 39.33 (2006), p. 10297.

[Fla07] P. Flajolet. "Analytic combinatorics—A calculus of discrete structures". In: *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA). Society for Industrial and Applied Mathematics. 2007, pp. 137–148.

[FFP07] P. Flajolet, É. Fusy, and C. Pivoteau. "Boltzmann sampling of unlabelled structures". In: *Proceedings of the Fourth Workshop on Analytic Algorithmics and Combinatorics* (ANALCO). Society for Industrial and Applied Mathematics, 2007, pp. 201–211.

[FS09] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.

[Fri93] B. Fristedt. "The Structure of random partitions of large integers". In: *Transactions of the American Mathematical Society* 337.2 (1993), pp. 703–735.

[GJS74] M. R. Garey, D. S. Johnson, and L. Stockmeyer. “Some simplified NP-complete problems”. In: *Proceedings of the Sixth Annual ACM Symposium on Theory of Computing* (STOC). Association for Computing Machinery. 1974, pp. 47–63.

[Gil55] J. Gillis. “Correlated random walk”. In: *Mathematical Proceedings of the Cambridge Philosophical Society* 51.4 (1955), pp. 639–651.

[GMP04] L. A. Goldberg, R. Martin, and M. Paterson. “Random sampling of 3-colorings in $\mathbb{Z}^2$”. In: *Random Structures & Algorithms* 24.3 (2004), pp. 279–302.

[Gra18] B. L. Granovsky. “Asymptotic enumeration by Khintchine–Meinardus probabilistic method: Necessary and sufficient conditions for sub-exponential growth”. In: *The Ramanujan Journal* 45.3 (2018), pp. 739–765.

[GS12] B. L. Granovsky and D. Stark. “A Meinardus theorem with multiple singularities”. In: *Communications in Mathematical Physics* 314.2 (2012), pp. 329–350.

[GSE08] B. L. Granovsky, D. Stark, and M. Erlihson. “Meinardus’ theorem on weighted partitions: Extensions and a probabilistic proof”. In: *Advances in Applied Mathematics* 41.3 (2008), pp. 307–328.

[GPR09] S. Greenberg, A. Pascoe, and D. Randall. “Sampling biased lattice configurations using exponential metrics”. In: *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA). Society for Industrial and Applied Mathematics. 2009, pp. 76–85.

[GC01] A. J. Guttmann and A. R. Conway. “Square lattice self-avoiding walks and polygons”. In: *Annals of Combinatorics* 5.3-4 (2001), pp. 319–345.

[HF98] J. W. Hanneken and D. R. Franceschetti. “Exact distribution function for discrete time correlated random walks in one dimension”. In: *The Journal of Chemical Physics* 109.16 (1998), pp. 6533–6539.

[HR18] G. H. Hardy and S. Ramanujan. “Asymptotic formulae in combinatory analysis”. In: *Proceedings of the London Mathematical Society* s2-17 (1918), pp. 75–115.

[Has70] W. K. Hastings. “Monte Carlo sampling methods using Markov chains and their applications”. In: *Biometrika* 57 (1970), pp. 97–109.

[HL72] O. J. Heilmann and E. H. Lieb. “Theory of monomer-dimer systems”. In: *Statistical Mechanics*. Springer, 1972, pp. 45–87.

[Hol+08] A. E. Holroyd, L. Levine, K. Mészáros, Y. Peres, J. Propp, and D. B. Wilson. "Chip-firing and rotor-routing on directed graphs". In: *In and Out of Equilibrium 2* (2008), pp. 331–364.

[HLZ16] L. Huang, P. Lu, and C. Zhang. "Canonical paths for MCMC: From art to science". In: *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA). Society for Industrial and Applied Mathematics. 2016, pp. 514–527.

[ICK92] A. G. Izergin, D. A. Coker, and V. E. Korepin. "Determinant formula for the six-vertex model". In: *Journal of Physics A: Mathematical and General* 25.16 (1992), p. 4315.

[Jam06] G. D. James. *The Representation Theory of the Symmetric Groups*. Springer, 2006.

[JLP19] D. Jerison, L. Levine, and J. Pike. "Mixing time and eigenvalues of the Abelian sandpile Markov chain". In: *Transactions of the American Mathematical Society* (2019).

[Jer03] M. Jerrum. *Counting, Sampling and Integrating: Algorithms and Complexity*. Birkhäuser Basel, 2003.

[JS89] M. Jerrum and A. Sinclair. "Approximate counting, uniform generation and rapidly mixing Markov chains". In: *Information and Computation* 82 (1989), pp. 93–133.

[JSV06] M. R. Jerrum, A. J. Sinclair, and E. Vigoda. "A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries". In: *Journal of the ACM* 41 (2006), pp. 671–697.

[JVV86] M. R. Jerrum, L. Valiant, and V. Vazirani. "Random generation of combinatorial structures from a uniform distribution". In: *Theoretical Computer Science* 43 (1986), pp. 169–188.

[JS96] M. Jerrum and A. Sinclair. "The Markov chain Monte Carlo method: An approach to approximate counting and integration". In: *Approximation Algorithms for NP-Hard Problems* (1996), pp. 482–520.

[KV92] A. A. Karatsuba and S. M. Voronin. *The Riemann Zeta-Function*. Walter de Gruyter, 1992.

[KS18] D. Keating and A. Sridhar. "Random tilings with the GPU". In: *Journal of Mathematical Physics* 59.9 (2018), p. 091420.

[KO07] R. Kenyon and A. Okounkov. “Limit shapes and the complex Burgers equation”. In: *Acta Mathematica* 199 (2007), pp. 263–302.

[Khi11] A. Y. Khinchin. *Mathematical Foundations of Quantum Statistics*. Dover Publications, 2011.

[KGV83] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. “Optimization by simulated annealing”. In: *Science* 220.4598 (1983), pp. 671–680.

[KZJ00] V. Korepin and P. Zinn-Justin. “Thermodynamic limit of the six-vertex model with domain wall boundary conditions”. In: *Journal of Physics A: Mathematical and General* 33.40 (2000), p. 7053.

[KG09] T. Kron and T. Grund. “Society as a self-organized critical system”. In: *Cybernetics & Human Knowing* 16.1 (2009), pp. 65–82.

[Kup96] G. Kuperberg. “Another proof of the alternative-sign matrix conjecture”. In: *International Mathematics Research Notices* 1996.3 (1996), pp. 139–150.

[Leh38] D. H. Lehmer. “On the series for the partition function”. In: *Transactions of the American Mathematical Society* 43.2 (1938), pp. 271–295.

[LP16] D. A. Levin and Y. Peres. “Mixing of the exclusion process with small bias”. In: *Journal of Statistical Physics* 165.6 (2016), pp. 1036–1050.

[LPW17] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov Chains and Mixing Times*. Second Edition. American Mathematical Society, 2017.

[LHG07] A. Levina, J. M. Herrmann, and T. Geisel. “Dynamical synapses causing self-organized criticality in neural networks”. In: *Nature Physics* 3.12 (2007), pp. 857–860.

[Lie67a] E. H. Lieb. “Exact solution of the problem of the entropy of two-dimensional ice”. In: *Physical Review Letters* 18.17 (1967), p. 692.

[Lie67b] E. H. Lieb. “Exact solution of the two-dimensional Slater KDP model of a ferroelectric”. In: *Physical Review Letters* 19.3 (1967), p. 108.

[Lie67c] E. H. Lieb. “Residual entropy of square ice”. In: *Physical Review* 162.1 (1967), p. 162.

[Liu18] T. Liu. “Torpid mixing of Markov chains for the six-vertex Model on $\mathbb{Z}^2$”. In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques* (RANDOM). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. 2018.

[Lov93] L. Lovász. “Random walks on graphs: A survey”. In: *Combinatorics, Paul Erdős is Eighty* 2.1 (1993), pp. 1–46.

[LRS01] M. Luby, D. Randall, and A. Sinclair. “Markov chain algorithms for planar lattice structures”. In: *SIAM Journal on Computing* 31.1 (2001), pp. 167–192.

[LKRV18] I. Lyberg, V. Korepin, G. A. P. Ribeiro, and J. Viti. “Phase separation in the six-vertex model with a variety of boundary conditions”. In: *Journal of Mathematical Physics* 59.5 (2018), p. 053301.

[LKV17] I. Lyberg, V. Korepin, and J. Viti. “The density profile of the six vertex model with domain wall boundary conditions”. In: *Journal of Statistical Mechanics: Theory and Experiment* 2017.5 (2017), p. 053103.

[Mad13] A. Madry. “Navigating central path with electrical flows: From flows to matchings, and back”. In: *Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science* (FOCS). Institute of Electrical and Electronics Engineers, 2013, pp. 253–262.

[Man91] S. Manna. “Two-state model of self-organized criticality”. In: *Journal of Physics A: Mathematical and General* 24.7 (1991), p. L363.

[MO94a] F. Martinelli and E. Olivieri. “Approach to equilibrium of Glauber dynamics in the one phase region”. In: *Communications in Mathematical Physics* 161.3 (1994), pp. 447–486.

[MO94b] F. Martinelli and E. Olivieri. “Approach to equilibrium of Glauber dynamics in the one phase region. II. The general case”. In: *Communications in Mathematical Physics* 161.3 (1994), pp. 487–514.

[McQ13] C. McQuillan. *Approximating holant problems by winding*. Preprint, `arXiv:1301.2880v1`. 2013.

[Meh13] K. Mehlhorn. “Physarum computations”. In: *Proceedings of the 30th International Symposium on Theoretical Aspects of Computer Science* (STACS). Schloss Dagstuhl, 2013, pp. 5–6.

[Mei54] G. Meinardus. “Asymptotische aussagen über partitionen”. In: *Mathematische Zeitschrift* 59 (1954), pp. 388–398.

[Met+53] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. "Equation of state calculations by fast computing machines". In: *The Journal of Chemical Physics* 21.6 (1953), pp. 1087–1092.

[MTN94] S. Mineshige, M. Takeuchi, and H. Nishimori. "Is a black hole accretion disk in a self-organized critical state?" In: *The Astrophysical Journal* 435 (1994), pp. L125–L128.

[Nic78] J. L. Nicolas. "Sur les entiers N pour lesquels il y a beaucoup de groupes abéliens d'ordre N". In: *Annales de l'institut Fourier* 28.4 (1978), pp. 1–16.

[NW78] A. Nijenhuis and H. S. Wilf. *Combinatorial Algorithms*. Academic Press, 1978.

[Oko02] A. Okounkov. "Symmetric functions and random partitions". In: *Symmetric Functions 2001: Surveys of Developments and Perspectives*. Springer Netherlands, 2002, pp. 223–252.

[Ons44] L. Onsager. "Crystal statistics. I. A two-dimensional model with an order-disorder transition". In: *Physical Review* 65.3-4 (1944), p. 117.

[Pau35] L. Pauling. "The structure and entropy of ice and of other crystals with some randomness of atomic arrangement". In: *Journal of the American Chemical Society* 57.12 (1935), pp. 2680–2684.

[PS15] Y. Peres and P. Sousi. "Mixing times are hitting times of large sets". In: *Journal of Theoretical Probability* 28.2 (2015), pp. 488–519.

[Phi14] J. Phillips. "Fractals and self-organized criticality in proteins". In: *Physica A: Statistical Mechanics and Its Applications* 415 (2014), pp. 440–448.

[PW96] J. G. Propp and D. B. Wilson. "Exact sampling with coupled Markov chains and applications to statistical mechanics". In: *Random Structures & Algorithms* 9.1–2 (1996), pp. 223–252.

[RS17] A. Ramachandran and A. Schild. "Sandpile prediction on a tree in near linear time". In: *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA). Society for Industrial and Applied Mathematics. 2017, pp. 1115–1131.

[RAM09] O. Ramos, E. Altshuler, and K. Maløy. "Avalanche prediction in a self-organized pile of beads". In: *Physical Review Letters* 102.7 (2009), p. 078701.

[Ran06a] D. Randall. “Rapidly mixing Markov chains with applications in computer science and physics”. In: *Computing in Science & Engineering* 8.2 (2006), pp. 30–41.

[Ran06b] D. Randall. “Slow mixing of Glauber dynamics via topological obstructions”. In: *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA). Society for Industrial and Applied Mathematics. 2006, pp. 870–879.

[RT00] D. Randall and P. Tetali. “Analyzing Glauber dynamics by comparison of Markov chains”. In: *Journal of Mathematical Physics* 41.3 (2000), pp. 1598–1615.

[RH81] E. Renshaw and R. Henderson. “The correlated random walk”. In: *Journal of Applied Probability* 18.2 (1981), pp. 403–414.

[RB79] G.-C. Rota and K. A. Baclawski. *Introduction to Probability and Random Processes*. 1979.

[SMM14] H. Saba, J. Miranda, and M. Moret. “Self-organized critical phenomenon as a q-exponential decay–Avalanche epidemiology of dengue”. In: *Physica A: Statistical Mechanics and its Applications* 413 (2014), pp. 205–211.

[SW94] J. A. Scheinkman and M. Woodford. “Self-organized criticality and economic fluctuations”. In: *The American Economic Review* 84.2 (1994), pp. 417–421.

[SS89] A. Sornette and D. Sornette. “Self-organized criticality and earthquakes”. In: *Europhysics Letters* 9.3 (1989), p. 197.

[Squ93] M. Squire. *Efficient generation of integer partitions*. Unpublished manuscript. 1993.

[Sta11] R. P. Stanley. *Enumerative Combinatorics*. Second Edition. Vol. 1. Cambridge University Press, 2011.

[SV16a] D. Straszak and N. K. Vishnoi. *IRLS and slime mold: Equivalence and convergence*. Preprint, `arXiv:1601.02712v1`. 2016.

[SV16b] D. Straszak and N. K. Vishnoi. “Natural Algorithms for Flow Problems”. In: *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA). Society for Industrial and Applied Mathematics, 2016, pp. 1868–1883.

[SV16c] D. Straszak and N. K. Vishnoi. “On a Natural Dynamics for Linear Programming”. In: *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science* (ITCS). Association for Computing Machinery, 2016, p. 291.

[Sut67] B. Sutherland. “Exact solution of a two-dimensional model for hydrogen-bonded crystals”. In: *Physical Review Letters* 19.3 (1967), p. 103.

[Tet91] P. Tetali. “Random walks and the effective resistance of networks”. In: *Journal of Theoretical Probability* 4.1 (1991), pp. 101–109.

[Tet94] P. Tetali. “An extension of Foster’s network theorem”. In: *Combinatorics, Probability and Computing* 3.3 (1994), pp. 421–427.

[Tho89] L. E. Thomas. “Bound on the mass gap for finite volume stochastic Ising models at low temperature”. In: *Communications in Mathematical Physics* 126.1 (1989), pp. 1–11.

[Wat+16] N. W. Watkins, G. Pruessner, S. C. Chapman, N. B. Crosby, and H. J. Jensen. “25 years of self-organized criticality: Concepts and controversies”. In: *Space Science Reviews* 198.1-4 (2016), pp. 3–44.

[WWAM06] R. J. Wijngaarden, M. S. Welling, C. M. Aegerter, and M. Menghini. “Avalanches and self-organized criticality in superconductors”. In: *The European Physical Journal B–Condensed Matter and Complex Systems* 50.1 (2006), pp. 117–122.

[Wil10] D. B. Wilson. “Dimension of the loop-erased random walk in three dimensions”. In: *Physical Review E* 82.6 (2010), p. 062102.

[Wil04] D. B. Wilson. “Mixing times of lozenge tiling and card shuffling Markov chains”. In: *The Annals of Applied Probability* 14.1 (2004), pp. 274–325.

[Yak12] Y. Yakubovich. “Ergodicity of multiplicative statistics”. In: *Journal of Combinatorial Theory, Series A* 119.6 (2012), pp. 1250–1279.

[Yan67] C. P. Yang. “Exact solution of a model of two-dimensional ferroelectrics in an arbitrary external electric field”. In: *Physical Review Letters* 19.10 (1967), p. 586.

[Zei96] D. Zeilberger. “Proof of the alternating sign matrix conjecture”. In: *Electronic Journal of Combinatorics* 3.2 (1996), R13.