

MARKOV CHAINS FOR WEIGHTED LATTICE STRUCTURES

A Thesis
Presented to
The Academic Faculty

by

Prateek Bhakta

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
Algorithms, Combinatorics, and Optimization

Georgia Institute of Technology
June 2016

MARKOV CHAINS FOR WEIGHTED LATTICE STRUCTURES

Approved by:

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

Milena Mihail
School of Computer Science
Georgia Institute of Technology

Eric Vigoda
School of Computer Science
Georgia Institute of Technology

Prasad Tetali
School of Mathematics and School of
Computer Science
Georgia Institute of Technology

David Goldberg
School of Industrial Systems and
Engineering
Georgia Institute of Technology

Date Approved: Not Yet Approved

To my parents Tarulata and Jayesh who have all my life always been my strongest supporters. To my sister Smita, who was also there, I guess.

ACKNOWLEDGEMENTS

I first and foremost thank my advisor Dana Randall for supporting and encouraging me in my time at Georgia Tech. Without her patient mentorship, I would be greatly diminished as an academic. Her guidance was essential to my development as a researcher, communicator, and teacher.

I thank my co-authors, Ben Cousins, Matthew Fahrback, Sarah Miracle, and Amanda Streib, who were an integral part of the development of this thesis. Their partnership in my research endeavors has been invaluable to me, and I am thankful to have had the opportunity to work with them. I especially thank Sarah Miracle for being a great mentor and teaching partner.

Finally, I thank the other members of my dissertation committee, David Goldberg, Milena Mihail, Prasad Tetali, and Eric Vigoda, for taking the time to help me write and defend my thesis. I especially thank Milena Mihail for serving as my thesis reader, and Prasad Tetali and Eric Vigoda for their valuable advice during my proposal and their letters of support on my behalf.

TABLE OF CONTENTS

DEDICATION	iii
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	vii
LIST OF FIGURES	viii
SUMMARY	ix
I INTRODUCTION	1
1.1 Markov Chain Basics and Techniques	4
1.1.1 Phase Transitions in the Ising Model	5
1.1.2 Sampling Perfect Matchings of Lattices with Path Structures	6
1.1.3 Biased Shuffling and Plane Partitions	7
1.2 Markov chains on Weighted Lattice Structures	9
1.2.1 Sampling Integer Partitions	9
1.2.2 Perfect Matchings on the Square-Octagon Lattice	10
1.2.3 The Schelling Segregation Model	10
II MARKOV CHAIN BACKGROUND AND TECHNIQUES	12
2.1 Markov Chain Fundamentals	12
2.2 Coupling	15
2.2.1 Path Coupling	15
2.3 Conductance	17
III UNIFORMLY SAMPLING INTEGER PARTITIONS WITH BIASED MARKOV CHAINS	18
3.1 Integer Partitions	18
3.1.1 Results	22
3.1.2 Techniques	24
3.2 A Markov chain on Young Diagrams	25
3.3 Bounding the mixing time	28

3.4	Efficient Rejection Sampling	33
3.4.1	Main results	34
3.4.2	Adaptive sampling for restricted partitions	37
IV	MIXING TIMES OF THE WEIGHTED FORTRESS MODEL . . .	39
4.1	Fortresses	39
4.2	The Fortress Model	44
4.2.1	Weighted Turning Graphs	46
4.2.2	A Markov chain on weighted turning graphs	47
4.3	Mixing of the Markov Chain \mathcal{M} on $\pi_{\lambda,1}$	48
4.3.1	Slow mixing of \mathcal{M} on $\pi_{\lambda,1}$ for $\lambda < 1/2\sqrt{e}$	49
4.3.2	Slow mixing of \mathcal{M} on $\pi_{\lambda,1}$ for $\lambda > 2\sqrt{e}$	54
4.3.3	Polynomial mixing of \mathcal{M} on $\pi_{\lambda,1}$ when $\lambda = 1$	56
4.4	Mixing of \mathcal{M} on $\pi_{\lambda,\mu}$ for general $\mu > 1$	61
V	CLUSTERING AND MIXING TIMES FOR SEGREGATION MODELS ON \mathbb{Z}^2.	64
5.1	Introduction	65
5.1.1	Relation to spin systems.	66
5.1.2	Generalized segregation models.	67
5.2	Preliminaries	69
5.2.1	The General Influence Model.	69
5.2.2	Mixing and clustering.	71
5.3	Bounding the Mixing Time	72
5.3.1	Slow mixing at high λ	74
5.3.2	Rapid mixing at low λ	81
	REFERENCES	85

LIST OF TABLES

LIST OF FIGURES

1	Young and Ferrers diagrams for the partition $(3, 2, 2, 2, 1, 1)$	19
2	Cases for the path coupling.	32
3	The mapping between (a) perfect matchings of G and (b) turning graphs of G^*	42
4	Two possible orientations (a) and (b) for each free vertex in G^* (c).	43
5	Configurations in (a) Ω_L , (b) Ω_R , and (c) Ω_C respectively.	49
6	The mapping $\phi : \Omega_C \rightarrow \Omega$	52
7	Weight-preserving bijection between $\sigma \subset G$ at parameter λ and $\sigma' \subset G'$ at $\lambda' = 1/\lambda$	55
8	Coloring representations of boundary conditions and turning graphs.	58
9	(a) A configuration with a contour, (b) the corresponding fat contour, and (c) an R -cross.	73
10	(a) A configuration σ with a fault line, (b) the 1-extended fault, and (c) $\phi(\sigma)$	75

SUMMARY

Markov chains are an essential tool for sampling from large sets, and are ubiquitous across many scientific fields, including statistical physics, industrial engineering, and computer science. Our chief concern is bounding the mixing time, the number of steps needed for a Markov chain to converge to a suitably random sample. We study problems that arise from the design and analysis of Markov chains that sample from configurations of lattice structures. Specifically, we will be interested in settings where each state is sampled with a non-uniform weight that depends on the structure of the configuration. Our focus will be on exploiting these weightings both to develop new algorithms for sampling and to prove new mixing time bounds for existing Markov chains.

The first problem we study is that of sampling integer partitions of n . A partition of a whole number n is a way to represent n as a sum of other whole numbers. For example, $n = 11$ can be partitioned as $n = 2 + 3 + 6$. In this setting, we present the first provably efficient Markov chain based algorithm for generating random integer partitions of n . Our chain uses weights to generate a partition of n with high likelihood, and rejects samples of size other than n until one is generated. Our algorithm runs in $O(n^{9/4})$ expected time and uses optimal $\tilde{O}(n^{1/2})$ space. Furthermore, our Markov chain can be adapted to a broad set of restricted classes of integer partitions, and in these settings is guaranteed to converge in $\tilde{O}(n^{5/2})$ time and still use only $\tilde{O}(n^{1/2})$ space. This allows us to generate partitions near n efficiently in practice, and we can uniformly generate partitions of size exactly n when the restricted partition numbers are well behaved.

For our second problem, we study sampling perfect matchings of finite, simply connected regions of the square-octagon lattice. We are interested in extending the analysis of a well studied “rotation” Markov chain, whose analogues have been used in practice

to sample random perfect matchings of regions of other lattices. However, this Markov chain on the square-octagon lattice empirically appears to converge slowly. To understand why, we interpret perfect matchings as a weighted model of so-called *turning paths*, a related routing structure on \mathbb{Z}^2 , and prove that this chain can converge in exponential time at certain settings of the parameters, and in polynomial time at other settings. This provides the first rigorous slow mixing result for this model, and makes progress towards explaining the slow convergence behavior of the original rotation chain on the square-octagon lattice.

In our final problem, we introduce and analyze a generalized version of the Schelling segregation model, first proposed in 1971 by economist Thomas Schelling to explain possible causes of racial segregation in cities. In this model, Schelling considered residents of two types living on a housing grid, where everyone prefers that the majority of his or her neighbors are of the same type. He showed through simulations that even mild preferences of this type can lead to segregation if residents move whenever they are not happy with their local environments.

We present a generalized model that includes both a natural variant of the Schelling model as well as the Ising model of statistical physics, which we call the General Influence Model. Our generalization includes a broad class of bias functions that models the neighborhood preferences of individuals, and the effect that this has on their desire to move. We show that for any influence function in this class, the dynamics will be rapidly mixing and cities will be integrated (i.e., there will not be clustering) if the bias is sufficiently low. Next, we show that for two broad classes of influence functions, when the bias is sufficiently high, the dynamics take exponential time to mix and we will have segregation and large monochromatic neighborhoods will form.

CHAPTER I

INTRODUCTION

Markov chains are an essential tool for randomly sampling from large sets, and in many applications offer the only practical method for generating nearly random samples. As a result, they are critical for studying complex distributions, and have become integral to the study of many scientific fields, including statistical physics, industrial engineering, and computer science.

A Markov chain performs a random walk over the set of states of interest, which we call the state space. Often, the state space is not explicitly given, but is the set of configurations for an underlying model. Markov chains have proven to be a valuable tool for random sampling in these settings, as there are frequently far too many configurations to sample one at random directly. Instead, beginning at some initial configuration, a Markov chain will repeatedly make random, local changes until we arrive at an approximately random configuration. The set of possible changes must *connect* the state space; in other words it must be possible to eventually reach any state starting from any other state via transitions of this Markov chain. The probabilities with which the random walk makes these changes are designed to cause the Markov chain to eventually converge to a target distribution of interest over the state space.

To be a useful tool for sampling, Markov chains should converge relatively quickly to their target distribution. The number of steps needed for a Markov chain to converge, commonly called the *mixing time*, is well understood to be controlled by the spectral gap of its transition matrix, the difference between its second largest eigenvalue and 1. However, computing the spectral gap directly is impractical for many large, interesting state spaces. Over the last 25 years, there have been many advancements that have

led to the creation and application of many useful tools that allow us to infer bounds on the mixing time without the transition matrix [25, 28, 17, 56, 68]. These tools have allowed computer scientists to design efficient Markov chains to solve difficult algorithmic problems, notably those of estimating the permanent of a matrix [48] and estimating the volume of a convex body [27, 53, 23].

Although Markov chains with a large mixing time are not immediately useful for sampling, they are of great mathematical interest for a variety of reasons. A slow convergence rate tells us meaningful information about the underlying state space of the Markov chain, most importantly that it must have a “bottleneck”, a partition of the state space into two subsets with a relatively small probability of transitioning between the two. The Gibbs (or Boltzmann) distribution is natural in statistical physics, and has provided insight into conditions for which certain classes of Markov chains are or are not efficient. A deep understanding of the bottlenecks of a slow Markov chain can aid in the design of other, efficient Markov chains that avoid the same bottleneck [77].

One of the most interesting phenomenon that appears in the study of some Markov chains is the presence of a *phase transition*, where a small, continuous change in one underlying parameter of the model leads to a large, discontinuous change in the probability distribution of the state space. The phase transition phenomenon appears in many contexts throughout the sciences, and reflects real world phase transitions that are found in physics and chemistry. Phase transitions in statistical models reveal a great deal about the distribution of the model, and can offer insight into their related, real world counterparts. The most famous of these is the Ising model of ferromagnetism, a statistical physics model which describes the spin of electrons in a magnet. As with real world magnets, a small, continuous change in the “temperature” of the model leads to a dramatic change in its “magnetization”, the macroscopic alignment of spins. In many cases in statistical physics, as the distribution of the underlying model undergoes a phase transition, so too does the mixing time of natural Markov chains for sampling

from this model. For example, just as the underlying Ising model switches from unmagnetized to magnetized with a small change in the temperature, so does the mixing time of the natural, local Markov chain known as the *Glauber Dynamics* for the Ising model switches from polynomial in the size of the model to exponential.

In this thesis, we study problems that arise from the design and analysis of Markov chains where each state is sampled with a non-uniform weight that depends on the structure of the configuration. Specifically, we will be interested in Markov chains that generate samples from configurations of lattice structures. These weighted lattice models arise naturally in many contexts, including the Ising model, and are typically more difficult to analyze than their unweighted counterparts.

Our main interest is analyzing the effect of these weightings on the convergence of the Markov chain. For example, it is possible to design weighted Markov chains that converge faster than their corresponding unweighted Markov chains, as is shown in the case of a simple random walk on a set of states arranged in a line [9, 82]. However, introducing weights can also cause a Markov chain to become slower than the unweighted case, as in the Ising model. Our focus will be on exploiting these weightings to develop new, faster algorithms for sampling, as well as proving new bounds on weighted extensions of natural Markov chains. In the latter case, we are especially interested in exposing phase transitions in our weighted model.

The remainder of this introduction gives background for the study of Markov chains as it relates to this thesis. In Section 1.1, we give some technical background on some important concepts involved in the study of Markov chains, and some of the analysis techniques that we will later use in the main body of this thesis. We will present these techniques in the context of some notable and relevant previous work. Finally in Section 1.2, we briefly describe the novel results that constitute the main contributions of this thesis.

1.1 Markov Chain Basics and Techniques

A Markov chain is a random process that undergoes transitions from one state to another on a state space Ω . The defining feature of a Markov chain is its *memoryless property*, the fact that the transition probabilities at any given step depend only on the current state and not on the sequence of states that preceded it. Many real world processes in the sciences, engineering, economics, and finance can be thought of as memoryless, and thus Markov chains have proven to be valuable models in these settings.

Formally, a Markov chain is a sequence of random variables $\{X_i\}$ for $i \geq 0$, where each $X_t \in \Omega$, and the conditional distribution of X_t depends only on X_{t-1} . In other words for all $i > 0$,

$$X_t | \{X_s\}_{s < t} = X_t | X_{t-1}.$$

Thus it suffices to describe a Markov chain by its *transition probabilities* $P(u, v) = P(X_t = v | X_{t-1} = u)$. It is also often convenient to consider the weighted adjacency matrix of this graph, which we call the *transition matrix*. If the state of the Markov chain at time t has probability distribution v over Ω , the state of the Markov chain at time $t + 1$ will have distribution $P \cdot v$. Therefore we say that $P^t(u, v)$ is the *t-step transition probability*, or the probability that a Markov chain in state u will be at state v exactly t steps later.

The distance between the t-step distribution $P^t(\cdot)$ and the stationary distribution π is measured in terms of their *total variation distance*, which is defined to be

$$\|P^t, \pi\|_{tv} = \max_{\sigma \in \Omega} \frac{1}{2} \sum_{\rho \in \Omega} |P^t(\sigma, \rho) - \pi(\rho)|.$$

We measure the convergence rate of a Markov chain by the number of steps needed until the total variation distance between P^t and π is small. More precisely, for all $\varepsilon > 0$ the *mixing time* $\tau(\varepsilon)$ of \mathcal{M} is

$$\tau(\varepsilon) = \min\{t : \|P^{t'}, \pi\|_{tv} \leq \varepsilon, \forall t' \geq t\}.$$

For a Markov chain to be a useful and efficient tool for sampling, we want the mixing time to be small relative to the size n of the configuration. We say that a Markov chain is *rapidly mixing* if the mixing time is bounded above by a polynomial in n and $\log(\varepsilon^{-1})$. Similarly, we call it *slowly mixing* if it is bounded below by an exponential function in n .

For the remainder of this section, we will introduce examples of more advanced techniques used to bound the mixing time of a Markov chain. We will present them in the context of some notable previous work relevant to the main body of this thesis.

1.1.1 Phase Transitions in the Ising Model

One of the most studied and best understood statistical physics models is the Ising model, which describes the spins of electrons in a magnet. In the Ising model of ferromagnetism, the vertices of a graph, say a finite region $G = (V, E)$ of \mathbb{Z}^2 , are assigned + or - spins. To model the behavior of real world magnets, we weight each configuration so that neighboring vertices prefer to have the same spin; the more neighboring electrons that match, the more likely the configuration.

Since we are interested in the effect of temperature on magnets, we introduce a parameter λ that is related to the inverse “temperature” of the model. The effect of magnetic fields on the spins of real world electrons is less pronounced as the temperature increases. To model this, the weight of each electron interaction increases with λ . More precisely, the stationary probability of a configuration is defined to be $\sigma \in \{\pm 1\}^V$ is

$$\pi(\sigma) = \lambda^{|\{x,y: (x,y) \in E, \sigma(x)=\sigma(y)\}|} / Z,$$

where

$$Z = \sum_{\sigma \in \{\pm 1\}^V} \lambda^{|\{x,y: (x,y) \in E, \sigma(x)=\sigma(y)\}|}$$

is the normalizing constant known as the *partition function*. The *Glauber dynamics* of the Ising model is a natural, local Markov chain that changes one spin at a time using appropriate probabilities to force the chain to converge to π . This Markov chain

undergoes a *phase transition*, where a small change in the value of λ leads to a large change in the mixing time. In fact, this phase transition is known to be *sharp*, in that there is a known value λ_c such that when $\lambda < \lambda_c$, the Glauber dynamics for the Ising model mixes in time polynomial in $|V|$ and when $\lambda > \lambda_c$, it mixes in exponential time [46, 69, 54, 77]. Moreover, the phase transition in the mixing time is accompanied by a corresponding transition in the stationary distribution of the Markov chain; at low λ , an average sample from the steady state is “unmagnetized” with regards to the proportions of spins, while at high λ , an average sample is “magnetized”, and has large regions of predominantly one spin type.

The one of the key techniques developed for the analysis of the Ising model is the so-called “Peierls argument,” a method for proving that a candidate subset of the state space is exponentially small in weight [61]. A Peierls argument defines a map from the candidate set to the entire state space with exponential gain in weight, allowing us to conclude that the set of interest must have exponentially small weight. When this unlikely set separates the state space into two pieces, this implies that the chain must take exponential time to move from one side of this cut to the other, and is therefore slowly mixing. The Peierls argument has proven to be an effective tool for identifying bottlenecks in a state space and proving slow mixing of Markov chains.

1.1.2 Sampling Perfect Matchings of Lattices with Path Structures

We now turn our attention to intersecting path models on lattices, which were developed as a tool to analyze natural Markov chains on the set of perfect matchings on a graph, one of the most fundamental structures in computer science. A perfect matching in a graph $G = (V, E)$ is a set of edges such that every vertex is an endpoint of exactly one of the edges. They arise in many natural computational and combinatorial contexts, and are also studied by statistical physicists in the *dimer model*, a model representing diatomic molecules.

The seminal work of Edmonds established that deciding if a given graph has a perfect matching, and finding it if so, could be done in polynomial time [29]. Subsequently, Valiant showed that counting perfect matchings is #P-complete, so it is believed that there is no such polynomial time general solution [79]. As a consequence, there has been a great deal of interest in finding both efficient *approximate* counting algorithms, as well as efficient exact counting algorithms in restricted settings. Jerrum, Sinclair and Vigoda showed how to approximately count and sample perfect matchings in any bipartite graph efficiently, although the complexity remains open on general graphs [48].

There is a natural Markov chain for this problem on graphs that are sub regions of lattices. In these settings, perfect matchings can be represented as a tiling of the dual lattice. This Markov chain takes a random face of the lattice, and attempts to “rotate” the edges on the boundary of the face if this is possible. Propp and Wilson developed the “coupling-from-the-past” algorithm for this problem, and showed that it could be run on dimer covers of the Cartesian lattice Z^2 to generate perfectly uniform samples of perfect matchings, although there were no guarantees that the algorithm would terminate in expected polynomial time [65]. Subsequently, a proof of rapid mixing was provided by Luby et al. [55], Randall and Tetali [68] and further improved by Wilson [82]. These techniques made use of a useful bijection between the natural tiling model and a path routing model that allowed for a more natural analysis of the Markov chain.

1.1.3 Biased Shuffling and Plane Partitions

Sampling a random permutation from S_n is a fundamental problem from probability theory. A natural Markov chain, the nearest neighbor transposition chain, will take neighboring elements of the permutation and swap them with probability $1/2$. This chain generates permutations uniformly, and is known to converge in time $\Theta(n^3 \log n)$ [82]. A natural weighting that arises in the context of queueing theory leads to a *biased* model, where each permutation is not equally likely but instead prefers adjacent elements to be

in sorted order. Here the Markov chain puts adjacent elements in order with probability $p \geq 1/2$ and out of order with probability $1 - p$. This chain was recently shown to converge in time $\Theta(n^2)$ [9] for constant p . However, less is known about the more general setting where the probability to exchange elements i and j can vary depending on the elements themselves. The problem of bounding the mixing rate of this chain was posed by Fill [31, 32] and was motivated by the Move-Ahead-One self-organizing list update algorithm. It was conjectured that the chain would always be rapidly mixing if $1/2 \leq p_{i,j} \leq 1$ for all $i < j$, but this was only known in the two cases described, or when $p_{i,j}$ is equal to $1/2$ or 1 , a case that corresponds to sampling linear extensions of a partial order.

In joint work with Sarah Miracle, Amanda Strieb, and Dana Randall, we proved that this chain is rapidly mixing for two much larger classes of probabilities $p_{i,j}$: “Choose Your Weapon,” where we are given r_1, \dots, r_{n-1} with $r_i \geq 1/2$ and $p_{i,j} = r_i$ for all $i < j$ (so the dominant player chooses the game, thus fixing his or her probability of winning), and “League Hierarchies,” where there are two leagues and players from the A-league have a fixed probability of beating players from the B-league. In the second setting, players within each league are similarly divided into sub-leagues with a possibly different fixed probability, and so forth recursively. Both of these classes include permutations with constant bias as a special case.

Moreover, we also prove that the most general conjecture is false by constructing a counterexample where although $1/2 \leq p_{x,y} \leq 1$ for all $x < y$, the nearest neighbor transposition chain requires exponential time to converge. This counter-example assigns probabilities to the $p_{x,y}$ that allow us to reinterpret this Markov chain on permutations as a simpler biased Markov chain on *bounded plane partitions*, another fundamental and well studied combinatorial model. A plane partition is a finite sequence of non-increasing integers, and is often visualized as a set of supported “boxes,” in our case restricted to the $n \times n$ box. In the context of our counter example, Bhatnagar et. al., showed that

the Markov chain that adds or removes one box at a time with appropriate probabilities mixes rapidly if there is a uniform bias in every box [11]. For our counterexample, we create an exponentially small bottleneck in the state space by adding larger bias to the boxes in one corner.

1.2 *Markov chains on Weighted Lattice Structures*

We now present the problems of study that form the main body of this thesis. Each topic is related to the broader theme of bounding the mixing time of weighted structures of lattice configurations. Our analysis techniques are drawn from the interface of theoretical computer science with discrete mathematics, statistical physics, and economics.

1.2.1 **Sampling Integer Partitions**

For our first problem, we look at the problem of sampling integer partitions of n . A partition of a whole number n is a way to represent n as a sum of other whole numbers. For example, $n = 11$ can be partitioned as $n = 2 + 3 + 6$. In this setting, we present the first provably efficient Markov chain based algorithm for generating random integer partitions of n . Unlike other attempts at developing efficient Markov chains on the space of partitions of n , our Markov chain walks on a larger state space consisting of partitions of various size, not only those of size n . Our chain is *biased* to generate a partition of n with high likelihood, and rejects samples of size other than n until one is generated. Our algorithm runs in $O(n^{9/4})$ expected time and uses optimal $\tilde{O}(n^{1/2})$ space. Other efficient methods for generating random partitions based on Boltzmann sampling also adopt this high level strategy. However, they require advance knowledge of the partition number $p(n)$, the number of partitions of size n , and therefore cannot be easily adapted to restricted families of partitions, such as those with bounded numbers of pieces, bounded size, and/or bounded rank. Our Markov chain can be adapted to a broad set of restricted classes of integer partitions, and in these settings is guaranteed to converge in $\tilde{O}(n^{5/2})$ time and still use only $\tilde{O}(n^{1/2})$ space. This allows us to generate

partitions near n efficiently in practice, and we can uniformly generate partitions of size exactly n when the restricted partition numbers are well behaved.

1.2.2 Perfect Matchings on the Square-Octagon Lattice

For our second problem, we study the problem of sampling perfect matchings of finite, simply connected regions of the square-octagon lattice. We are interested in extending the analysis of the “rotation” Markov chain introduced in Section 1.1.2, which is known to be ergodic and has been used in practice to sample random perfect matchings to discover their properties. However unlike domino and lozenge tilings, this Markov chain on the square-octagon lattice appears to converge slowly. To understand why, we introduce and analyze a weighted Markov chain on *turning paths*, a related path structure on \mathbb{Z}^2 , and prove that this chain can converge in exponential time at certain settings of the parameters, and in polynomial time at other settings. This provides the first rigorous slow mixing result for this model. Our slow mixing results rely on Peierls arguments, which we use to identify a cut in the state space of a Markov chain that has small conductance [61]. Our rapid mixing result relies on novel bijection between turning paths and 3-colorings of regions of the grid.

1.2.3 The Schelling Segregation Model

In the final problem, we will introduce and analyze a generalized version of the Schelling segregation model. The Schelling segregation model attempts to explain possible causes of racial segregation in cities. In his model, Schelling considered residents of two types living on a housing grid, where everyone prefers that the majority of his or her neighbors are of the same type. He showed through simulations that even mild preferences of this type can lead to segregation if residents move whenever they are not happy with their local environments.

We present a general model that includes both a natural variant of the Schelling model as well as the Ising model, which we call the General Influence Model. Our

generalization includes a broad class of bias functions that models the neighborhood preferences of individuals, and the effect that this has on their desire to move. We show that for any influence function in this class, the dynamics will be rapidly mixing and cities will be integrated (i.e., there will not be clustering) if the bias is sufficiently low. Next we analyze two broad classes of influence functions. The first type, Increasing Bias Functions (IBF), has an individual's likelihood of moving increases each time someone of the same color leaves (this does not include Schelling's threshold models). The second type, Threshold Bias Functions (TBF) is reminiscent of the model Schelling originally proposed, and only causes an increase in the likelihood that someone leaves if the number of same-type neighbors dips below a certain threshold greater than $1/2$ of all neighbors. For both these classes (IBF and TBF), we show that when the bias is sufficiently high, the dynamics take exponential time to mix and we will have segregation and a large "ghetto" will form. As in our study of the fortress model, our proofs of slow mixing rely on Peierls arguments to show slow mixing by identifying an exponentially small cut in the state space.

CHAPTER II

MARKOV CHAIN BACKGROUND AND TECHNIQUES

We begin by formally describing some of the critical concepts and key techniques that have been developed in the analysis of Markov chains that we will use in this thesis.

2.1 Markov Chain Fundamentals

Markov chains on a state space Ω are defined by their *transition probabilities* $P(u, v)$. At any time t in the evolution of the Markov process, $P(u, v)$ is the probability that the Markov chain will transition from state u at time t to state v at time $t + 1$. We will often represent a Markov chain as a *transition graph* on the vertex set Ω , with edges (u, v) weighted by $P(u, v)$ if $P(u, v) > 0$ and absent otherwise. It is also often convenient to consider the weighted adjacency matrix of this graph, which we call the *transition matrix*. If the state of the Markov chain at time t has probability distribution v over Ω , the state of the Markov chain at time $t + 1$ will have distribution $P \cdot v$. Therefore we say that $P^t(u, v)$ is the *t-step transition probability*, or the probability that a Markov chain in state u will be at state v exactly t steps later.

For a Markov chain to be a viable tool for sampling from the state space Ω , it must be able to reach every possible state in Ω . We say that a Markov chain is *irreducible* if for any states u and v , there exists a time t such that

$$P^t(u, v) > 0.$$

In other words, a Markov chain is irreducible if the underlying transition graph is strongly connected. A Markov chain is *aperiodic* if for every state u ,

$$\gcd(\{t | P^t(u, u) > 0\}) = 1.$$

Together, a Markov chain that is both irreducible and aperiodic is called *ergodic*.

Lemma 2.1.1: For any ergodic Markov chain M on a finite state space with transition matrix P , there exists a *unique* distribution π over Ω such that

$$\lim_{t \rightarrow \infty} P^t \cdot v \rightarrow \pi$$

for any initial distribution v .

We call π the *stationary distribution* of M . By the lemma, any distribution π that satisfies $P\pi = \pi$ for an ergodic Markov chain's transition matrix P must be its stationary distribution. A common special case of this is described below.

Lemma 2.1.2: For any ergodic Markov chain M with transition probabilities P , if π is a distribution over Ω that satisfies

$$\forall u, v \quad \pi(u)P(u, v) = \pi(v)P(v, u),$$

then π must be the stationary distribution of M .

We say that when the above condition holds, that the Markov chain \mathcal{M} satisfies *detailed balance* with respect to π , and any Markov chain that satisfies detailed balance with respect to some distribution is called *reversible*.

Many natural Markov chains are reversible, and algorithmists often design Markov chains to be reversible to exploit the well known properties of reversible Markov chains. The most notable example is the famous Metropolis-Hastings Algorithm [57, 44]. The Metropolis-Hastings algorithm defines an assignment of probabilities to the transitions of a Markov chain designed to cause the Markov chain to converge to some “target” distribution π over Ω . More precisely,

The Metropolis-Hastings Markov chain \mathcal{M} with target distribution π .

Let Δ be the maximum degree of the underlying graph of \mathcal{M} . Starting at any initial state σ_0 , repeat for all $t \geq 1$:

- Choose neighbor τ of σ_t uniformly at random, each with probability $1/2\Delta$,

- Set $\sigma_{t+1} = \tau$ with probability $\min(1, \pi(\tau)/\pi(\sigma))$.
- With remaining probability, set $\sigma_{t+1} = \sigma_t$.

It is easy to verify that Metropolis-Hasting algorithm satisfies detailed balance with respect to π , and therefore converges to π as desired.

The Metropolis-Hastings algorithm has the crucial additional benefit that the target distribution need not be a true normalized distribution in order to execute the Markov chain. That is, for any positive weighting $w(\sigma)$ over Ω , we can define the transition probabilities in terms of w rather than π . This will cause the Markov chain to converge to the target distribution $\pi(\sigma) = w(\sigma)/Z$, where Z is the normalizing constant $Z = \sum_{\sigma \in \Omega} w(\sigma)$. In many contexts, this normalizing constant reveals a great deal of information about the Markov chain, and it is referred to as the *partition function* of the weighting.

Although these techniques ensure that the Markov chain converges to the target π , they say nothing about how long we need to run the Markov chain to obtain samples close to π . The closeness of the t -step distribution $P^t(\cdot)$ and the target π is measured in terms of their *total variation distance*, which is defined to be

$$\|P^t, \pi\|_{tv} = \max_{\sigma \in \Omega} \frac{1}{2} \sum_{\rho \in \Omega} |P^t(\sigma, \rho) - \pi(\rho)|.$$

For all $\varepsilon > 0$, the *mixing time* $\tau(\varepsilon)$ of \mathcal{M} is the number of steps until the distribution of the Markov chain is within ε from π , or

$$\tau(\varepsilon) = \min\{t : \|P^t, \pi\|_{tv} \leq \varepsilon, \forall t' \geq t\}.$$

For a Markov chain to be a useful and efficient tool for sampling, we want the mixing time to be small. We say that a Markov chain is *rapidly mixing* if the mixing time is bounded above by a polynomial in n and $\log(\varepsilon^{-1})$. Similarly, we call it *slowly mixing* if it is bounded below by an exponential function.

Much information about a Markov chain can be found from the eigenvalues of its transition matrix P [73]. In particular, the *spectral gap* $1 - |\lambda_2|$ provides a good and useful bound on the mixing time of \mathcal{M} . However, we will deal with Markov chains with exponentially large state spaces, and directly calculating λ_2 is infeasible in these settings. The remainder of this chapter discusses some key tools that we will use to study the mixing times of Markov chains with exponentially large state spaces.

2.2 Coupling

One of the most successful techniques for upper bounding the mixing time of a Markov chain is known as coupling. A coupling of a Markov chain \mathcal{M} is itself a Markov chain $(X_t, Y_t)_{t=0}^\infty$ on $\Omega \times \Omega$ such that each of the processes X_t and Y_t when viewed in isolation is a faithful representation of \mathcal{M} . By this we mean that

$$\Pr(X_{t+1} = \sigma | X_t = \tau) = \Pr(Y_{t+1} = \sigma | Y_t = \tau) = P(\sigma, \tau).$$

We add the requirement that if the two processes should ever “meet”, or both occupy the same state, then they will always continue to share the same state. In other words if $X_t = Y_t$, then $X_{t+1} = Y_{t+1}$. Given such a coupling, the *coupling time* T is a measure of the time needed until the two processes converge to the same state. Namely, define

$$T = \max_{x,y} E[\min\{t : X_t = Y_t | X_0 = x, Y_0 = y\}].$$

For any coupling, the following theorem (see, e.g. [1]) relates the coupling time and the mixing time.

Theorem 2.2.1: $\tau(\epsilon) \leq T e \lceil \ln \epsilon^{-1} \rceil$.

2.2.1 Path Coupling

Much of the difficulty in using the coupling theorem lies in carefully choosing a coupling between all pairs of states in such a way as to bound the expected coupling time. A popular technique is to define a metric ϕ on Ω such that the expected distance between

any two states does not increase in a single step. We can then bound the number of steps expected until this metric reaches 0. This task was made much more convenient with the Path coupling theorem of Bubley and Dyer [17]. They argue that the argument only needs to be considered for a relatively small subset U of all pairs of states; in practice often we need consider only the neighboring pairs in Ω .

We will use the version of the path coupling theorem due to Dyer and Greenhill [28].

Theorem 2.2.2 (Path Coupling): Let M be a Markov chain on Ω and let (X_t, Y_t) be a coupling of M . Let φ be an integer-valued metric defined on $\Omega \times \Omega$ which takes values in $\{0, \dots, B\}$. Let U be a subset of $\Omega \times \Omega$ such that for all $(X_t, Y_t) \in \Omega \times \Omega$ there exists a path $X_t = z_0, z_1, \dots, z_r = Y_t$ between X_t and Y_t such that $(z_i, z_{i+1}) \in U$ for $0 \leq i < r$ and $\sum_{i=0}^{r-1} \phi(z_i, z_{i+1}) = \phi(X_t, Y_t)$.

Suppose $E[\phi(X_{t+1}, Y_{t+1})] \leq \beta\phi(X_t, Y_t)$, for all $X_t, Y_t \in U$ and some $0 \leq \beta \leq 1$.

- If $\beta < 1$, then the mixing time of \mathcal{M} satisfies

$$\tau(\varepsilon) \leq \frac{\ln(B\varepsilon^{-1})}{1 - \beta}$$

- If $\beta = 1$, then let $\alpha > 0$ satisfy $\Pr[\phi(X_{t+1}, Y_{t+1}) \neq \phi(X_t, Y_t)] \geq \alpha$ for all t such that $X_t \neq Y_t$. The mixing time of M then satisfies

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

The above theorem is useful for proving rapid mixing of \mathcal{M} only if the upper bound on the metric B is polynomial in n . We will use the following convenient version of the path coupling theorem, proposed by Greenberg et al. [40], that handles exponential sized metrics to analyze the mixing time of \mathcal{M} but requires a stronger condition on the expected change in distance.

Theorem 2.2.3: Let M be a Markov chain on Ω and let (X_t, Y_t) be a coupling of M . Let $\varphi : \Omega \times \Omega \rightarrow \mathbb{R}_{\geq 0}$ be a metric that takes finitely many values in $\{0\} \cup [1, B]$.

Let U be a subset of $\Omega \times \Omega$ such that for all $(x_t, y_t) \in \Omega \times \Omega$ there exists a path $x_t = z_0, z_1, \dots, z_r = y_t$ such that $(z_i, z_{i+1}) \in U$ for $0 \leq i < r$ and $\sum_{i=0}^{r-1} \varphi(z_i, z_{i+1}) = \varphi(x_t, y_t)$.

Suppose there exists $\beta \leq 1$ such that $E[\varphi(x_{t+1}, y_{t+1})] \leq \beta \varphi(x_t, y_t)$ for all $(x_t, y_t) \in U$.

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

$$\tau(\varepsilon) \leq \frac{\ln(B\varepsilon^{-1})}{1 - \beta}.$$

2. If there exists $\kappa, \eta \in (0, 1)$ such that $\Pr[|\varphi(x_{t+1}, y_{t+1}) - \varphi(x_t, y_t)| \geq \eta \varphi(x_t, y_t)] \geq \kappa$ for all t such that $x_t \neq y_t$, then the mixing time satisfies

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

2.3 Conductance

Several of our results will center on the relationship between the *conductance* of a Markov chain and its mixing time [47, 73]. Introduced by Jerrum and Sinclair, the conductance of a Markov chain captures the intuitive notion of a bottleneck in the state space, and is a good measure of the mixing rate of a chain. For an ergodic Markov chain \mathcal{M} with stationary distribution π , the conductance of a subset $S \subseteq \Omega$ is defined as

$$\Phi(S) = \sum_{s_1 \in S, s_2 \in \bar{S}} \pi(s_1)P(s_1, s_2)/\pi(S).$$

The conductance of the chain \mathcal{M} is then the minimum conductance of all subsets,

$$\Phi = \min_{S \subseteq \Omega} \{\Phi(S) : \pi(S) \leq 1/2\}.$$

We will focus on the conductance as a *lower bound* of the mixing time, and show that Markov chains are slowly mixing by showing that they have an exponentially small conductance. To do this, we will make use of the following relationship between the conductance Φ of a Markov chain and its mixing time $\tau(\varepsilon)$ [47].

Theorem 2.3.1: The mixing time of a Markov chain with conductance Φ satisfies:

$$\left(\frac{1 - \Phi}{2\Phi} \right) \ln \varepsilon^{-1} \leq \tau(\varepsilon) \leq \frac{1}{\Phi^2} (\ln(\pi^{-1}_*) + \ln(\varepsilon^{-1}))$$

CHAPTER III

UNIFORMLY SAMPLING INTEGER PARTITIONS WITH BIASED MARKOV CHAINS

We begin with work on sampling integer partitions of n which was introduced in Section 1.2.1. Specifically, we present the first provably efficient Markov chain based algorithm for generating random integer partitions. Unlike other attempts at developing efficient Markov chains on the space of partitions of n , our Markov chain walks on a larger state space consisting of partitions of various size. Our chain is *biased* to generate a partition of n with high likelihood, and rejects samples of size other than n until one is generated. Our algorithm runs in $O(n^{9/4})$ expected time and uses optimal $\tilde{O}(n^{1/2})$ space. Other efficient methods for generating random partitions based on Boltzmann sampling also adopt this high level strategy. However, they require advance knowledge of the partition number $p(n)$, the number of partitions of size n , and therefore cannot be easily adapted to restricted families of partitions, such as those with bounded numbers of pieces, bounded size, and/or bounded rank. Our Markov chain can be adapted to a broad set of restricted classes of integer partitions, and in these settings is guaranteed to converge in $\tilde{O}(n^{5/2})$ time and still use only $\tilde{O}(n^{1/2})$ space. This allows us to generate partitions near n efficiently in practice, and we can uniformly generate partitions of size exactly n when the restricted partition numbers are well behaved.

3.1 Integer Partitions

The problem of sampling *integer partitions* of n , one of the most extensively researched combinatorial structures in discrete mathematics. A *partition* of a nonnegative integer n is a decomposition of n into a nonincreasing sequence of positive integers that sum

to n . For example, 5 has seven partitions: (5) , $(4, 1)$, $(3, 2)$, $(3, 1, 1)$, $(2, 2, 1)$, $(2, 1, 1, 1)$, and $(1, 1, 1, 1, 1)$. Partitions arise in mathematics and physics in many contexts, including exclusion processes [22], random matrices [60], representation theory [45], juggling patterns [7], and various growth processes [40]. For a comprehensive history, we refer the reader to any of many books on the topic (see, e.g., [4, 5]).

Common representations of partitions include *Young diagrams* and *Ferrers diagrams*, depending on one's proclivity for squares or circles (see Figure 1). Throughout this paper, we use French notation, which aligns diagrams in the lower-left corner. The heights of the columns represent distinct pieces of the partition; hence all partitions of n will have exactly n squares or circles in total, known as the *area of the diagram*, and each column is nonincreasing in height from left to right.

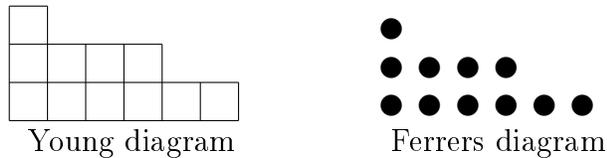


Figure 1: Young and Ferrers diagrams for the partition $(3, 2, 2, 2, 1, 1)$.

The *partition numbers*, denoted as $p(n)$, count the number of partitions of n , with the convention that $p(n) = 0$, for all $n < 0$. The history of the partition numbers begins with Euler, who discovered the simple generating function

$$\sum_{n=0}^{\infty} p(n)x^n = \prod_{n=1}^{\infty} (1 - x^n)^{-1}.$$

Together with his *pentagonal number theorem*, the generating function implies the recurrence relation $p(n) = p(n-1) + p(n-2) - p(n-5) - p(n-7) + \dots = \sum_{k \neq 0} (-1)^{k-1} p(n - g_k)$, where the sum is over $k = 1, -1, 2, -2, 3, \dots$ and $g_k = k(3k-1)/2$. More than a century later, Hardy and Ramanujan [43] quantified the rate of growth of the partition numbers and proved the asymptotic formula

$$p(n) \sim \frac{1}{4\sqrt{3n}} e^{\pi\sqrt{2n/3}},$$

pioneering the *circle method* in analytic number theory. Rademacher [66] subsequently improved the Hardy-Ramanujan formula to an exact formula using a convergent series in terms of modified Bessel functions of the first kind and Kloosterman sums. Bruinier and Ono [16] recently obtained an explicit finite formula for $p(n)$ as a sum of algebraic numbers, specifically singular moduli for a *weak Maass form*. Their discovery stems from the fractal structure of the partition numbers for every prime. Similar recursive and asymptotic formulae are known for families of restricted partitions, such as the number of partitions of n into at most k parts [3].

An equally active and important area of research concerns the design of efficient algorithms for generating partitions of n uniformly at random. There have been three main directions for sampling, those based on counting partitions, Boltzmann samplers, and those based on fixed-size Markov chains. The first class of approaches first counts the total number of partitions of a given type, and then uses self-reducability to generate a random partition. Nijinhuis and Wilf [59] give two recursive algorithms based on dynamic programming that require computing tables of exact values. The first is a straightforward application of the recurrence $p(n, m) = p(n - m, m) + p(n - 1, m - 1)$, where $p(n, m)$ is the number of partitions of n with largest part equal to m . This algorithm takes $O(n^{5/2})$ time and space for preprocessing and $O(n^{3/2})$ time per sample. Squire [74] improved the second algorithm of Nijinhuis and Wilf, which makes use of Euler's pentagonal recurrence, to $O(n^2)$ time and space for preprocessing and $O(n^{3/2} \log n)$ time per sample. The state time and space complexity bounds of these algorithms account for the fact that each value of $p(n)$, as well as the intermediate summands, requires $O(n^{1/2})$ space by the Hardy-Ramanujan formula. Therefore, dynamic programming approaches for exact sampling break down in practice when $n \geq 10^6$.

By avoiding the computationally expensive task of counting partitions, *Boltzmann samplers* offer a more direct method for sampling. A Boltzmann sampler generates samples from a larger combinatorial class with probability proportional to the Boltzmann

weight $\lambda^{|\sigma|}$, where $|\sigma|$ 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 [26, 33]. The value λ is chosen to maximize the yield of samples of our target size n . Fristedt [34] suggested an approach which 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 σ in the $n \times \infty$ grid with probability proportional to the Boltzmann weight $\lambda^{|\sigma|}$. A recent breakthrough by Arratia and DeSalvo [6] gives a probabilistic approach that is substantially more efficient than previous algorithms, thus allowing fast generation of random partitions for significantly larger numbers (e.g., $n \geq 10^6$). Building on the work of Fristedt [34], 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 factors). Their PDC algorithm also uses appropriate independent geometric random variables to generate a partition, but does so recursively in phases. This approach achieves a superior runtime over previous Boltzmann sampling approaches because they reject these phases early when possible.

Finally, stochastic approaches using Markov chains have produced a similarly rich corpus of work, but until now have not provided algorithms with rigorous polynomial bounds. One popular approach designs Markov chains based on *coagulation and fragmentation processes* that allow pieces of the partition to be merged and split [2, 10]. Recently Ayer et al. [7] 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 that these chains converge quickly to equilibrium but they lack explicit polynomial bounds.

3.1.1 Results

We give the first provably efficient Markov chain algorithm for uniformly sampling integer partitions. A key contribution of this algorithm is that it can be applied to sample efficiently from the set of partitions *restricted* to lie between *arbitrary* upper and lower envelopes. This includes commonly studied classes of restricted partitions, namely partitions with restricted numbers and/or size of parts, partitions with bounded rank, and more.

For the special case of sampling unrestricted partitions, we prove

Theorem 3.1.1: There is a Markov chain Monte Carlo algorithm for sampling partitions of n that runs in $O(n^{9/4})$ expected time and uses $O(n^{1/2} \log n)$ space.

We apply Boltzmann sampling techniques [6, 34] and insights from biased card shuffling [9, 40] to sample Young diagrams of varying sizes restricted to a simply connected region in the first quadrant of \mathbb{Z}^2 that includes all Young diagrams of area n . We use a natural biased “mountain-valley” Markov chain on staircase walks representing the upper envelope of the Young diagram to choose a diagram σ with weight proportional to $\lambda^{|\sigma|}$. Setting $\lambda_n = p(n-1)/p(n)$, we show that we will hit our target set consisting of partitions of n with probability at least $\Omega(1/n^{1/4})$. Applying this rigorous analysis along with the *coupling from the past* variant of the sampling algorithm, we can in fact generate partitions of n *exactly* from the uniform distribution in expected time $O(n^{9/4})$.

Although using a Markov chain for random generation of a partition is slower than the approaches using independent geometric random variables taken by Fristedt’s Boltzmann Sampler and Arratio and Delavso’s PDC algorithm, our method is much more versatile and can be adapted to efficiently sample general classes of restricted partitions. Techniques using geometric random variables could be easily extended to sample from sets of restricted partitions that have generating functions of the form $G(x) = x^a \prod_{b \in S} (1-x^b)^{-1}$, where a is a nonnegative integer and S is a countable multiset of positive integers. This

class includes partitions with a bounded number of parts, partitions with part sizes restricted to a countable set of positive integers, and partitions with fixed *rank*. There are, however, several natural restrictions on partitions that do not have generating functions of this form, such as those whose Young diagram fits inside an $M \times N$ rectangle and partitions with bounded rank. Although we are unable to adapt generation using geometric random variables in these settings, we can sample these restricted partitions in polynomial time with our Markov chain Monte Carlo (MCMC) algorithm.

Our Markov Chain algorithm can be adapted to sample partitions of a given size from families of “region-restricted partitions,” i.e., Young diagrams restricted to lie within a general simply-connected “bounding region.” For any choice of the bias parameter λ and any simply-connected bounding region R , we show the Markov chain will converge in $O(n^{5/2})$ time, and, for λ defined as a ratio of successive partition numbers, the chain will provably converge in time $O(n^2)$, even for arbitrarily region-restricted Young diagrams.

Importantly, our approach does not rely on estimates of the numbers of partitions of each size in our restricted region, or on assumptions about the log-concavity of those numbers. We then show that with probability at least $\Omega(1/n^2)$, we will obtain a sample of the desired size, no matter the shape of the bounding region. If there is log-concavity or concentration for our restricted class, as is the case for unrestricted partitions, our algorithm will produce samples of area n with a significantly higher frequency than our bounds imply, reducing the time to produce a random restricted partition of n .

Finally, although dynamic programming techniques can also be extended to sample from these general regions, they are space prohibitive for large values of $n > 10^6$. Our algorithm only uses $\tilde{O}(n^5)$ space, and therefore is more suitable for sampling restricted partitions for large n .

3.1.2 Techniques

We use a *biased Markov chain* on Young diagrams to sample a configuration σ with probability proportional to $\lambda^{|\sigma|}$, for some λ . This chain was studied in the context of biased growth processes and biased card shuffling and is known to converge in at most $O(|R|^2)$ time on any simply connected lattice region R , where $|R|$ is the area of R [40]. DeSalvo and Pak [24] recently showed that $p(n)$ is log-concave when $n > 25$, so the function $p(n)\lambda^n$ is also log-concave, and hence is unimodal. By setting $\lambda_n = p(n-1)/p(n)$ when sampling unrestricted partitions, the mode of the weighted distribution will be concentrated around partitions of n . This gives a polynomial time stochastic algorithm.¹

Several observations allow us to improve the running time. 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 the Young diagrams of all partitions of n and has area $\Theta(n \log n)$. Next, we show that we can derive better bounds on the convergence time of the Markov chain than given in [40] using a more careful analysis of the exponential metric for our particular choice of λ . Last, there is a simple mapping from partitions of size greater than n to partitions of size n that will succeed with probability at least $\Omega(1/n^{1/4})$. From this we conclude that the chain will converge in $O(n^2)$ time and $O(n^{1/4})$ trials are needed in expectation before generating a sample corresponding to a partition of n .² Moreover, all Young diagrams that are contained within the region R have at most $O(n^{1/2})$ “corners,” so each configuration can be stored in $O(n^{1/2} \log n)$ space.

¹ The function $p(n)$ fails to be log-concave for small n because of oscillations due to parity, but the function $q(n) = p(2n) + p(2n + 1)$ is log-concave for all n . This suffices for our applications, and we suspect that $q_R(n) = p_R(2n) + p_R(2n + 1)$ will be log-concave for most natural sets of restricted partitions as well.

²Note that if we restrict our state space to Young diagrams in R , we can no longer infer that the number of diagrams of each size is log-concave. However, diagrams with area n will continue to be the mode since we are reducing the number of diagrams of size greater than n while still including all diagrams with area exactly n .

The following analysis of the mixing time holds when the Young diagrams are restricted to lie in any simply connected subregion of R . For example, if R is intersected with the half-plane $x < k$, diagrams correspond to restricted partitions of n with at most k parts.

Even in more complicated cases when the number of Young diagrams in a class are not known a priori to be log-concave, we can still use this Markov chain, but we rely on well-believed conjectures to bound the rejection rates. Regardless of the class of restricted partitions we wish to sample, the chain always converges in polynomial time for any λ , and samples with area n are guaranteed to be chosen uniformly. Note that for applications where we need uniform partitions with size approximately n , our algorithm always produces useful samples in polynomial time.

3.2 A Markov chain on Young Diagrams

We begin by presenting background information about partitions and Markov chains before presenting our Markov chain. A *Young diagram* is a finite subset σ of $\mathbb{Z}_{\geq 0}^2$ with the property that if $(a, b) \in \sigma$, then $\{(x, y) \in \mathbb{Z}_{\geq 0}^2 : 0 \leq x \leq a \text{ and } 0 \leq y \leq b\} \subseteq \sigma$. They are 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 unit square in the diagram is labeled with the coordinate of its lower-left corner. Partitions of n are in bijection with Young diagrams σ with $|\sigma| = n$, so we give an algorithm to uniformly sample Young diagrams of area n .

Our Markov chain \mathcal{M} samples from the state space Ω of all Young diagrams in the first quadrant whose upper boundary lies under the curve $y = 2n/x$. Observe that all partitions of n lie in this region. We could sample Young diagrams from an $n \times n$ grid, but this square region has area $\Theta(n^2)$ instead of $\Theta(n \log n)$, which increases the runtime of our algorithm, as we will see later. Additionally, we use the curve $y = 2n/x$ as an upper limit instead of $y = n/x$ to reduce the rejection rate of our algorithm by

mapping Young diagrams with area greater than n to diagrams of area n while preserving uniformity.

The chain repeatedly chooses a square on the upper envelope of the Young diagram and attempts to either add or remove that square. Using the Metropolis-Hastings algorithm to define transition probabilities, the stationary probability of each diagram σ is $\pi(\sigma) = \lambda^{|\sigma|}/Z$, where $Z = \sum_{\sigma \in \Omega} \lambda^{|\sigma|}$ is the normalizing constant. Any configuration σ drawn from this distribution with area $|\sigma|$ will be uniformly chosen from the set $\{\rho \in \Omega : |\rho| = |\sigma|\}$ when the chain is in equilibrium. For sampling partitions of n , we set $\lambda_n = p(n-1)/p(n)$ to force the stationary distribution to concentrate at n . This works because $p(k)$ is log-concave for $k > 25$, which was recently shown in [24]. It follows that the sequence $p(k)\lambda^k$ is log-concave for any $\lambda > 0$ and $k > 25$, and so $p(k)\lambda_n^k$ is log-concave with the dual mode $n-1$ and n .

Although we are sampling Young diagrams lying under $y = 2n/x$, the analysis of our algorithm holds for $y = (n + \sqrt{n})/x$. The choice of $y = 2n/x$, however, is more convenient and does not affect the complexity. The following claim shows that we can store a Young diagram in $O(n^{1/2} \log n)$ space.

Claim 3.2.1: A Young diagram σ restricted to lie under the curve $y = 2n/x$ can be stored in $O(n^{1/2} \log n)$ space.

PROOF: For any square in σ , both of its coordinates cannot be greater than $\sqrt{2n}$, for then it would lie above $y = 2n/x$. We may record the height of each column and width of each row in the range $\{0, 1, \dots, \lfloor \sqrt{2n} \rfloor - 1\}$ and capture every square in σ at least once. Therefore, we can represent σ using exactly these $2\lfloor \sqrt{2n} \rfloor$ heights and widths. ■

Using the representation in the previous claim, we see that there will not be more than $O(n^{1/2})$ possible transitions at any possible state, since our algorithm adds or removes at most one square on the upper boundary in each step.

Claim 3.2.2: For any Young diagram σ restricted to lie under the curve $y = 2n/x$, there will be at most $4\sqrt{2n}$ potential transitions.

PROOF: We 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: that of removing the current furthest square, or that of adding the next one. Therefore, by Claim 3.2.1, there are at most $4\lfloor\sqrt{2n}\rfloor$ valid transitions possible from any σ . ■

We now formally present the Markov chain \mathcal{M} that samples Young diagrams σ that lie under the curve $y = 2n/x$ with probability proportional to $\lambda_n^{|\sigma|}$. When we say a diagram ρ is valid, we mean that it is a Young diagram that lies under the curve $y = 2n/x$.

The Young diagram Markov chain \mathcal{M} with bias λ_n .

Starting at any Young diagram σ_0 , repeat for all $t \geq 1$:

- Select laziness $\ell \in \{0, 1\}$, index $i \in \{0, 1, \dots, \lfloor\sqrt{2n}\rfloor - 1\}$, axis $a \in \{“x”, “y”\}$, and bit $b \in \{-1, 1\}$ uniformly at random.
- If laziness $\ell = 0$:
 - If axis $a = “x”$ and bit $b = 1$, let ρ be the diagram obtained by adding a square to column i .
 - If axis $a = “x”$ and bit $b = -1$, let ρ be the diagram obtained by removing a square from column i .
 - If axis $a = “y”$, bit $b = 1$, and row i has width at least $\lfloor\sqrt{2n}\rfloor$, let ρ be the diagram obtained by adding a square to row i .
 - If axis $a = “y”$, bit $b = -1$, and row i has width at least $\lfloor\sqrt{2n}\rfloor + 1$, let ρ be the diagram obtained by removing a square from row i .
 - If bit $b = 1$ and ρ is valid, set $\sigma_{t+1} = \rho$ with probability λ_n .
 - If bit $b = -1$ and ρ is valid, set $\sigma_{t+1} = \rho$ with probability 1.

- With all remaining probability, set $\sigma_{t+1} = \sigma_t$.

The Markov chain \mathcal{M} is constructed so that every valid transition occurs with the uniform probability $1/(8\lfloor\sqrt{2n}\rfloor)$, and each state update takes $O(1)$ time to execute using our representation of a Young diagram as a series of rows widths and column heights. The state space Ω is connected, since any configuration can eventually reach the “bottom” configuration $\sigma = \emptyset$ with positive probability. We defined \mathcal{M} to be lazy, i.e., it is always possible that $\sigma_t = \sigma_{t+1}$, so it follows that \mathcal{M} is an *ergodic* Markov chain. This Markov chain transitions according to the Metropolis-Hastings algorithm with respect to the probability distribution $\pi(\sigma) = \lambda^{|\sigma|}/Z$, and therefore π is its stationary distribution. As \mathcal{M} is ergodic, the process σ_t eventually converges to the stationary distribution π starting from any σ_0 [50].

3.3 Bounding the mixing time

We now prove our main result that the Markov chain \mathcal{M} mixes in $O(n^2 \log(\varepsilon^{-1}))$ steps using a path coupling argument.

We follow the approach of Greenberg et al. [40] to show that \mathcal{M} is rapidly mixing. For convenience, let $\lambda = \lambda_n$ from here onwards. Define the *distance* between any two $\sigma, \rho \in \Omega$ as

$$\varphi(\sigma, \rho) = \sum_{(x,y) \in \sigma \oplus \rho} \lambda^{-(x+y)/2},$$

where $\sigma \oplus \rho$ is the symmetric difference of σ and ρ . In other words, $\varphi(\sigma, \rho)$ is a weighted sum over all squares on which σ and ρ disagree, where each square (x, y) is weighted by $\lambda^{-(x+y)/2}$.

We define the path coupling $\xi_t = (\sigma_t, \rho_t)$ as follows. Let σ_t and ρ_t be two Young diagrams that differ by one square, i.e., $|\sigma_t \oplus \rho_t| = 1$, and let the coordinates of this square be (x, y) . Choose the same index, axis, and bit (i, a, b) uniformly at random for both σ_t and ρ_t . Let the coordinates of the square that would be added or removed in σ_t by move (i, a, b) be (x', y') . If $|x - x'| + |y - y'| \geq 2$, then we choose the same

laziness ℓ for both configurations uniformly at random, so that we accept or reject the move with the same probability in σ_t and ρ_t . In this case, $|\sigma_{t+1} \oplus \rho_{t+1}| = 1$ for the resulting configurations. For all transitions (i, a, b) such that $|x - x'| + |y - y'| \leq 1$, we allow at most one of σ_t or ρ_t to transition by coupling the probability that $\ell = 0$ in one configuration with the probability that $\ell = 1$ in the other. Specifically, choose ℓ_{σ_t} uniformly from $\{0, 1\}$ and set $\ell_{\rho_t} = 1 - \ell_{\sigma_t}$, and proceed with the move (i, a, b) .

To prove our main mixing result, we first bound λ_n using bounds for $p(n)$ given by DeSalvo and Pak [24]. Then we analyze the previously defined path coupling and apply Theorem 2.2.2 to bound the mixing time of \mathcal{M} .

Lemma 3.3.1: For n sufficiently large, we have

$$1 - \frac{2}{\sqrt{n}} < \lambda_n < 1 - \frac{1}{\sqrt{n}}.$$

PROOF: For convenience and clarity, let $\mu(n) = \frac{\pi}{6}\sqrt{24n-1}$.

Proposition 3.3.1 ([24, Proposition 2.4]): For $n \geq 2$,

$$\left| p(n) - \frac{\pi^2}{6\mu(n)^2\sqrt{3}} \left[\left(1 - \frac{1}{\mu(n)}\right) e^{\mu(n)} + \frac{(-1)^n}{\sqrt{2}} e^{\mu(n)/2} \right] \right| < 1 + \frac{16}{\mu(n)^3} e^{\mu(n)/2}.$$

Using this asymptotic bound, it follows for $n \geq 565$ that

$$\begin{aligned} \left| p(n) - \frac{\pi^2}{6\mu(n)^2\sqrt{3}} \left(1 - \frac{1}{\mu(n)}\right) e^{\mu(n)} \right| &< 1 + \frac{16}{\mu(n)^3} e^{\mu(n)/2} + \frac{\pi^2}{6\mu(n)^2\sqrt{6}} |(-1)^n e^{\mu(n)/2}| \\ &\leq 1 + e^{\mu(n)/2} \left(\frac{\pi^2}{6\mu(n)^2\sqrt{6}} + \frac{16}{\mu(n)^3} \right) \\ &\leq \frac{\pi^2 e^{\mu(n)/2}}{6\mu(n)^2\sqrt{3}} \left(1 - \frac{1}{\mu(n)}\right). \end{aligned}$$

The final step follows from the fact the second to last term is asymptotically equal to

$$\frac{\pi^2 e^{\mu(n)/2}}{6\mu(n)^2\sqrt{6}} (1 + o(1)),$$

while the final term is

$$\frac{\pi^2 e^{\mu(n)/2}}{6\mu(n)^2 \sqrt{6}} \left(\sqrt{2} - o(1) \right).$$

We now use the following basic fact about real numbers, which we state without proof.

Fact 3.3.1: Let $a, a', b, b', \varepsilon_a, \varepsilon_b$ be positive real numbers with $b' > \varepsilon_b$. If $|a - a'| \leq \varepsilon_a$ and $|b - b'| \leq \varepsilon_b$, then

$$\left| \frac{a}{b} - \frac{a'}{b'} \right| \leq \frac{a'}{b'} \left(\frac{\varepsilon_b/b' + \varepsilon_a/a'}{1 - \varepsilon_b/b'} \right).$$

For convenience, denote our current estimate for $\lambda_n = p(n-1)/p(n)$ by

$$E(n) = \left(\frac{\mu(n)^3}{\mu(n-1)^3} \right) \left(\frac{\mu(n-1) - 1}{\mu(n) - 1} \right) e^{\mu(n-1) - \mu(n)}$$

Applying Fact 3.3.1, we have

$$\left| \frac{p(n-1)}{p(n)} - E(n) \right| \leq E(n) \left[\frac{e^{-\mu(n)/2} + e^{-\mu(n-1)/2}}{1 - e^{-\mu(n)/2}} \right]$$

Expressed differently, $\lambda_n = E(n)(1 + O(e^{-\mu(n)/2}))$. We now focus on bounding each term of $E(n)$ to bound λ_n . Firstly,

$$\frac{\mu(n)}{\mu(n-1)} = \sqrt{\frac{24n-1}{24n-25}} = \sqrt{1 + \frac{24}{24n-25}} \leq 1 + \frac{12}{24n-25}$$

We then use the fact that for $n \geq 3$, we have

$$\frac{\pi}{\sqrt{6(n-1)}} < \mu(n) - \mu(n-1) < \frac{\pi}{\sqrt{6n}}.$$

Put differently, $\mu(n) - \mu(n-1) = \frac{\pi}{\sqrt{6n}}(1 - O(n^{-3/2}))$. Therefore,

$$\frac{\mu(n-1) - 1}{\mu(n) - 1} = 1 + \frac{\mu(n-1) - \mu(n)}{\mu(n) - 1} \geq 1 - \frac{\pi}{\sqrt{6n}\mu(n)}(1 - O(n^{-3/2})) = 1 - O(n^{-1}),$$

and

$$\begin{aligned} e^{\mu(n-1) - \mu(n)} &= e^{-\frac{\pi}{\sqrt{6n}}(1 - o(n^{-3/2}))} \\ &= e^{-\frac{\pi}{\sqrt{6n}}} e^{-o(n^{-2})} \\ &= e^{-\frac{\pi}{\sqrt{6n}}(1 - o(n^{-2}))}. \end{aligned} \tag{Taylor series}$$

We now have all the pieces needed to prove our bounds on λ_n .

Lemma 3.3.2: We have

$$\left| \lambda_n - e^{-\frac{\pi}{\sqrt{6n}}} \right| \leq O(n^{-1}).$$

PROOF: Using the previously calculated bounds, observe that

$$\begin{aligned} \lambda_n &= E(n)(1 + O(e^{-\mu(n)/2})) \\ &= \left(e^{-\frac{\pi}{\sqrt{6n}}} \right) (1 + O(n^{-1})) (1 - O(n^{-2})) (1 - O(n^{-1})) (1 + O(n^{-1}))^3 (1 + O(e^{-\mu(n)/2})) \\ &= \left(e^{-\frac{\pi}{\sqrt{6n}}} \right) (1 + O(n^{-1})). \end{aligned}$$

The desired result follows. ■

Lemma 3.3.3: We have

$$\left| \lambda_n - \left(1 - \frac{\pi}{\sqrt{6n}} \right) \right| \leq O(n^{-1}).$$

PROOF: This is an immediate consequence of Lemma 3.3.2 and the fact that

$$e^{-\frac{\pi}{\sqrt{6n}}} = \left(1 - \frac{\pi}{\sqrt{6n}} + O(n^{-1}) \right),$$

by considering the Taylor series expansion. ■

PROOF (LEMMA 3.3.1): The result follows immediately from Lemma 3.3.3, as $1 < \pi/\sqrt{6} < 2$. ■

Lemma 3.3.4: For Young diagrams σ_t and ρ_t satisfying $|\sigma_t \oplus \rho_t| = 1$, the coupling $\xi_t = (\sigma_t, \rho_t)$ has the property

$$E[\varphi(\sigma_{t+1}, \rho_{t+1})] < \left(1 - \frac{1}{32\sqrt{2}n^{3/2}} \right) \varphi(\sigma_t, \rho_t)$$

under the distance metric φ , for sufficiently large n .

PROOF: Let (σ_t, ρ_t) differ at square (x, y) , and assume $(x, y) \notin \sigma_t$ and $(x, y) \in \rho_t$, without loss of generality. Then $\varphi_t = \varphi(\sigma_t, \rho_t) = \lambda^{-(x+y)/2}$. If the square chosen by ξ is (x', y') and $|x - x'| + |y - y'| \geq 2$, then $\sigma_{t+1} \oplus \rho_{t+1} = \sigma_t \oplus \rho_t$ and $\varphi_{t+1} - \varphi_t = 0$. Otherwise, we have the following three cases by symmetry:

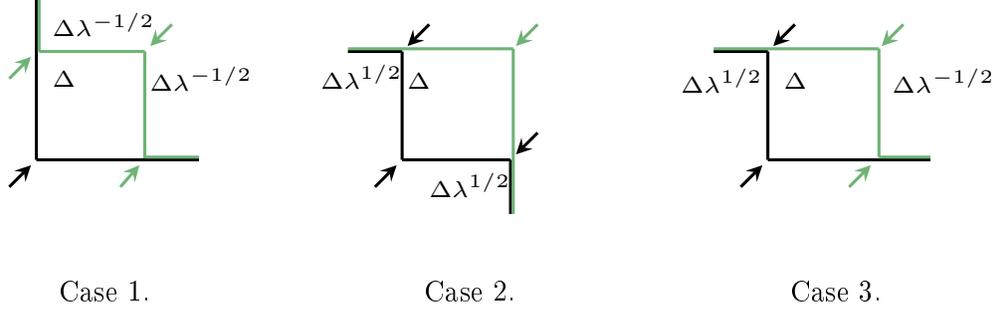


Figure 2: Cases for the path coupling.

There are two moves in all three cases that reduce the distance to 0: that of removing (x, y) from ρ_t , and that of adding (x, y) to σ_t . The probability that this happens is $(1 + \lambda)p$, where $p = 1/(8\lfloor\sqrt{2n}\rfloor)$ is the probability that any configuration transitions to a neighboring state. We consider moves that may increase the distance in each case:

- Case 1: There are two moves with probability λp that increase the distance by $\varphi_t \lambda^{-1/2}$.
- Case 2: There are two moves with probability p that increase the distance by $\varphi_t \lambda^{1/2}$.
- Case 3: There is one move with probability λp that increases the distance by $\varphi_t \lambda^{-1/2}$, and one other move with probability p that increases the distance by $\varphi_t \lambda^{1/2}$.

Upper bounding $E[\varphi_{t+1}]$ for Case 1, we have

$$E[\varphi_{t+1}] \leq 2\lambda p (\varphi_t + \varphi_t \lambda^{-1/2}) + (1 - (1 + \lambda)p - 2\lambda p) \varphi_t = \left(1 - p(1 - \lambda^{1/2})^2\right) \varphi_t,$$

and bounding the expectations for the other two cases, we see they are all equal. Next, we use Lemma 3.3.1 to bound $\lambda^{1/2}$ by

$$\sqrt{\lambda_n} < \sqrt{1 - \frac{1}{\sqrt{n}}} < \sqrt{1 - \frac{1}{\sqrt{n}} + \frac{1}{4n}} = 1 - \frac{1}{2\sqrt{n}},$$

for sufficiently large n . It follows that

$$1 - p(1 - \lambda^{1/2})^2 < 1 - \frac{1}{8\lfloor\sqrt{2n}\rfloor} \left(\frac{1}{2\sqrt{n}}\right)^2 \leq 1 - \frac{1}{32\sqrt{2n^{3/2}}},$$

since $1/(2\sqrt{n}) < 1 - \lambda^{1/2} < 1$, which completes the proof. \blacksquare

Theorem 3.3.5: The Markov chain \mathcal{M} mixes in $O(n^2 \log(\varepsilon^{-1}))$ time.

PROOF: An upper bound for the maximum possible distance between any two configurations is the maximum weight of any square λ^{-n} times the number of squares under the curve $y = 2n/x$, which is $\sum_{k=1}^{2n} \lfloor 2n/k \rfloor \leq 2nH_{2n} \leq 2n(\ln(2n) + 1)$. It follows from Theorem 2.2.2 and Theorem 3.3.4 that the mixing time of \mathcal{M} satisfies

$$\begin{aligned} \tau(\varepsilon) &\leq \frac{\ln(2n(\ln(2n) + 1)\lambda^{-n}\varepsilon^{-1})}{1 - \left(1 - \frac{1}{32\sqrt{2n^{3/2}}}\right)} \leq cn^{3/2} \ln(\varepsilon^{-1}) \ln(\lambda^{-n}) \\ &\leq cn^{3/2} \ln(\varepsilon^{-1}) \ln\left(\left(1 + \frac{2}{\sqrt{n} - 2}\right)^n\right) = O(n^2 \log(\varepsilon^{-1})), \end{aligned}$$

for a sufficiently large constant c . \blacksquare

3.4 Efficient Rejection Sampling

In this section, we give our Markov chain Monte Carlo algorithm for sampling partitions of n that runs in expected time $O(n^{9/4})$ and uses $O(n^{1/2} \log n)$ space. We prove Theorem 3.1.1, and then show how to adapt our algorithm to sample restricted partitions in polynomial time using techniques introduced in [67] for self-avoiding walks.

3.4.1 Main results

Our MCMC algorithm uses coupling from the past [65] to sample *perfectly* from the stationary distribution of \mathcal{M} . We can implement coupling from the past in expected time $O(n^2)$, our bound on the mixing time of \mathcal{M} , because our coupling is monotone with respect to the natural partial order for Young diagrams. This means we only need to consider the “bottom” state $\sigma = \emptyset$ and the “top” state $\rho = \{(x, y) \in \mathbb{Z}_{\geq 0}^2 : (x+1)(y+1) \leq n\}$ of the partial order on the state space Ω to determine when all states have coalesced, because monotonicity ensures that when σ and ρ coalesce, all other configurations will have too.

Next, we show how to use random samples with area greater than n to reduce the rejection rate of our algorithm. Suppose our algorithm produces a Young diagram σ such that $|\sigma| > n$. Instead of rejecting σ and resampling until we obtain a diagram with area n , we attempt to map σ to a partition of n while preserving uniformity over all partitions of n . Let Y_n denote the set of Young diagrams with area n , and consider the function $f_k : Y_n \rightarrow Y_{n+k}$ that maps a partition $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_m) \in Y_n$ to $f_k((\sigma_1, \sigma_2, \dots, \sigma_m)) = (\sigma_1 + k, \sigma_2, \dots, \sigma_m) \in Y_{n+k}$. Note that $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_m$ since σ is a Young diagram. Clearly f_k is injective, so we can consider the inverse map $f_k^{-1}((q_1, q_2, \dots, q_\ell))$ that subtracts k from q_1 if $q_1 - k \geq q_2$, and is undefined otherwise. Then, define $g : \bigcup_{k \geq 0} Y_{n+k} \rightarrow Y_n \cup \{0\}$ as

$$g((q_1, q_2, \dots, q_\ell)) = \begin{cases} (q_1 - k, q_2, \dots, q_\ell) & \text{if } q_1 + q_2 + \dots + q_\ell = n + k \text{ and } q_1 - k \geq q_2, \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 3.4.1: Let X be a random sample from the stationary distribution of \mathcal{M} , and let g be the function defined above. Then for n sufficiently large, we have

$$\Pr[g(X) \text{ yields partition of size } n] \geq \frac{1}{100n^{1/4}}.$$

PROOF: From the Hardy-Ramanujan asymptotic formula for $p(n)$ given in [43], we have that for some constant $c > 0$,

$$\frac{1-c}{4n\sqrt{3}}e^{\pi\sqrt{\frac{2n}{3}}} \leq p(n) \leq \frac{1+c}{4n\sqrt{3}}e^{\pi\sqrt{\frac{2n}{3}}}. \quad (3.4.1)$$

Letting $\lambda = \lambda_n$ for convenience, Lemma 3.3.2 implies that

$$(1-c)e^{-\pi k \frac{1}{\sqrt{6n}}} \leq \lambda^k \leq (1+c)e^{-\pi k \frac{1}{\sqrt{6n}}}. \quad (3.4.2)$$

Note that we can make c arbitrarily small in both of the above inequalities by selecting n large enough.

Claim 3.4.2: Let Z_n be the normalizing constant for the target distribution. Then we have

$$Z_n \leq 12n^{3/4}\lambda^n p(n).$$

PROOF: We have that

$$Z_n \leq \sum_{k=0}^{\infty} \lambda^k p(k).$$

We further know that $f(k) = \lambda^k p(k)$ has a maximum at $k = n$. Note that by the log-concavity of f , we have that for $k > 0$,

$$\frac{f(n+k)}{f(n)} \geq \frac{f(n+2k)}{f(n+k)} \quad \text{and} \quad \frac{f(n-k)}{f(n)} \geq \frac{f(n-2k)}{f(n-k)}.$$

Therefore, for any $k > 0$, we can bound Z_n as

$$Z_n \leq kf(n) \left(\frac{1}{1 - f(n+k)/f(n)} + \frac{1}{1 - f(n-k)/f(n)} \right). \quad (3.4.3)$$

Specifically, if both $f(n+k)/f(n)$ and $f(n-k)/f(n)$ are at most some fixed constant less than 1, then $Z_n = O(kf(n))$. Using (3.4.1) and (3.4.2), we see that

$$\begin{aligned} f(k) &= \lambda^k p(k) \\ &\leq (1+c)^2 \frac{1}{4k\sqrt{3}} e^{-\frac{\pi}{\sqrt{6n}}(k-2\sqrt{kn})} \\ &= (1+c)^2 \frac{1}{4k\sqrt{3}} e^{-\frac{\pi}{\sqrt{6n}}(\sqrt{k}-\sqrt{n})^2} e^{\pi\sqrt{\frac{n}{6}}}. \end{aligned}$$

Setting $k = (\sqrt{n} + n^{1/4})^2$ gives

$$f((\sqrt{n} + n^{1/4})^2) \leq (1+c)^2 \frac{1}{4n\sqrt{3}} e^{\pi\sqrt{\frac{n}{6}}} e^{-\frac{\pi}{\sqrt{6}}}.$$

We can then bound the density value at $(\sqrt{n} + n^{1/4})^2$ relative to the maximum:

$$\begin{aligned} \frac{f((\sqrt{n} + n^{1/4})^2)}{f(n)} &\leq \frac{(1+c)^2 \frac{1}{4n\sqrt{3}} e^{\pi\sqrt{\frac{n}{6}}} e^{-\frac{\pi}{\sqrt{6}}}}{(1-c)^2 e^{-\pi\sqrt{\frac{n}{6}}} \frac{1}{4n\sqrt{3}} e^{\pi\sqrt{\frac{2n}{3}}}} \\ &\leq \frac{(1+c)^2}{(1-c)^2} e^{-\frac{\pi}{\sqrt{6}}}. \end{aligned}$$

Assuming that $c \leq 0.1$, we have

$$\frac{f((\sqrt{n} + n^{1/4})^2)}{f(n)} \leq \frac{1.1^2 e^{-\frac{\pi}{\sqrt{6}}}}{0.9^2} < \frac{1}{2}.$$

Similarly,

$$f((\sqrt{n} - n^{1/4})^2) \leq (1+c)^2 \frac{1}{4n\sqrt{3}} e^{\pi\sqrt{\frac{n}{6}}} e^{-\frac{\pi}{\sqrt{6}}},$$

and thus

$$\frac{f((\sqrt{n} - n^{1/4})^2)}{f(n)} < \frac{1}{2}.$$

Therefore, by (3.4.3) and using the fact that $k \leq 3n^{3/4}$, we have that

$$Z_n \leq 12n^{3/4} \lambda^n p(n), \tag{3.4.4}$$

which proves the claim. ■

PROOF (LEMMA 3.4.1): We can bound the probability of generating a partition of size n as

$$\begin{aligned} \Pr[g(X) \text{ yields partition of size } n] &= \sum_{k=0}^n \frac{\lambda^{n+k} p(n+k)}{Z_n} \cdot \frac{p(n)}{p(n+k)} \\ &\geq \frac{1}{12n^{3/4}} \sum_{k=0}^n \lambda^k \quad (\text{by Claim 3.4.2}) \\ &\geq \frac{1}{12n^{3/4}} \sum_{k=0}^n \left(1 - \frac{2}{\sqrt{n}}\right)^k \\ &\geq \frac{1}{12n^{3/4}} \sqrt{n} \left(1 - \frac{2}{\sqrt{n}}\right)^{\sqrt{n}} \\ &\geq \frac{1}{100n^{1/4}}, \quad \blacksquare \end{aligned}$$

which proves the lemma. ■

Now we formally present our MCMC algorithm for generating a uniformly random partition of n in expected time $O(n^{9/4})$ and $O(n^{1/2} \log n)$ space, and give the proof of Theorem 3.1.1.

Integer Partition Sampling Algorithm.

Repeat until success:

- Sample $\sigma \in \Omega$ with \mathcal{M} using coupling from the past.
- If $n \leq |\sigma| \leq 2n$ and $g(\sigma) \neq 0$, return $g(\sigma)$.

PROOF (THEOREM 3.1.1): Using coupling from the past, the Markov chain \mathcal{M} mixes in expected time $O(n^2)$ by Theorem 3.3.5. The expected number of trials until a successful sample is $O(n^{1/4})$ by Lemma 3.4.1 and properties of the geometric distribution. Therefore, this algorithm runs in expected time $O(n^{9/4})$. By recycling space in each iteration, \mathcal{M} requires $O(n^{1/2} \log n)$ space by Claim 3.2.1. ■

3.4.2 Adaptive sampling for restricted partitions

Here we outline how to modify our Markov chain to sample from families of restricted partitions in polynomial time. We adapt the approach used by Randall and Sinclair [67] in the context of self-avoiding walks for our setting. For the remainder of this section, we define \mathcal{M} to be the Markov chain given in Section 3.3, but restricted to Young diagrams that are contained in an arbitrary simply connected region $R \subseteq \mathbb{Z}_{\geq 0}^2$ containing $(0, 0)$, instead of the region below $y = 2n/x$. We define $p_R(n)$ to be the number of Young diagrams of area n contained in R .

Two particular classes of restricted partitions are of particular interest, because they demonstrate the versatility of our algorithm. The first is the set of partitions of n into at most M parts each of size at most N . Let $p(M, N; n)$ denote the number of such partitions. To sample from this set, we let R be an $M \times N$ rectangle. Almkvist and Andrews

[3] extended the Hardy-Ramanujan-Rademacher formula to $p(M, N; n)$, so we should be able to obtain similar results about log-concavity and rejection rates for this class. The second family of restricted partitions is those with bounded rank, or equivalently a bounded *Durfee square*. We let $R = (\{0, 1, \dots, r-1\} \times \mathbb{Z}_{\geq 0}) \cup (\mathbb{Z}_{\geq 0} \times \{0, 1, \dots, r-1\})$ to sample partitions of n with rank at most r , and we rely on the well-believed conjecture that $p_R(n)$ is log-concave for all r to achieve an expected polynomial runtime.

Even if $p_R(n)$ is known to be log-concave, the primary obstacle to modifying our algorithm for arbitrary regions is computing $\lambda_n = p_R(n-1)/p_R(n)$, since we do not have estimates of $p_R(n)$ for general regions R . To overcome this, we use an adaptive, *self-testing* strategy introduced by Randall et al. [67], which runs a sequence of Markov chains $\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_n$ to discover and iteratively improve an estimate of λ_n . The self-testing aspect of the algorithm guarantees polynomial run time, and if the bootstrapping process is successful, which it will be if the sequence $p_R(n)$ is polynomially close to log-concave, then the algorithm produces samples of size n with a high acceptance rate. Otherwise, if the process is unsuccessful, we have strong evidence that $p_R(n)$ is not log-concave, which is also a useful outcome.

Regardless of the restricted family of partitions, we may bound R to lie under $y = 2n/x$ when we desire samples of size n . This allows us to represent states using $O(n^{1/2} \log n)$ space by Claim 3.2.1, and allows each transition of \mathcal{M} to be made with probability $\Theta(1/n^{1/2})$. By a similar path coupling argument and Theorem 2.2.2, we see that the mixing time of \mathcal{M} , for any parameter λ , is $O(n^{5/2} \log^2(n) \log(\varepsilon^{-1}))$. The details are omitted here. Note that in any restricted setting it suffices that $\Pr[|\sigma| = k] \leq \Pr[|\sigma| = n] \cdot \text{poly}(n)$, for all $k \geq 0$, to achieve an expected polynomial time acceptance rate, which is a substantially weaker property than log-concavity. Lastly, we reiterate that for approaches where we need uniformly random partitions with size close to n , our algorithm always produces useful samples in polynomial time.

CHAPTER IV

MIXING TIMES OF THE WEIGHTED FORTRESS MODEL

In this chapter, we study the problem of sampling perfect matchings of finite, simply connected regions of the square-octagon lattice, which was introduced in Section 1.2.2. We are interested in a natural local Markov chain that is known to be ergodic and has been used in practice to sample random perfect matchings to discover their properties. However, unlike related Markov chains used for sampling domino and lozenge tilings, this Markov chain on the square-octagon lattice appears to converge slowly. To understand why, we introduce and analyze a weighted Markov chain on *turning paths*, a related path structure on \mathbb{Z}^2 , and prove that this chain can converge in polynomial or exponential time, depending on the settings of the parameters. This provides the first rigorous slow mixing result for this model. Our slow mixing results rely on Peierls arguments, which we use to identify a cut in the state space of a Markov chain that has small conductance [61]. Our rapid mixing result relies on novel bijection between turning paths and 3-colorings of regions of the grid.

4.1 Fortresses

Perfect matchings arise in many natural computational contexts, and have been the cornerstone problem underlying many fundamental complexity questions. They are also of specific interest to the statistical physics community, where they are studied in the context of *dimer models*. Here, edges in a matching represent diatomic molecules, or dimers, and perfect matchings of a lattice region correspond to dimer packings. Physicists study the properties of these physical systems by relating fundamental thermodynamic quantities to weighted sums over the set of all configurations of the system, in our case the set of all perfect matchings of the lattice region.

The seminal work of Edmonds established that the decision and construction problems, i.e. efficiently deciding if a given graph has a perfect matching and finding it if so, were in P [29]. Subsequently, Valiant showed that counting perfect matchings is #P-complete, so it is believed that there is no such polynomial time general solution [79]. As a consequence, there has been a great deal of interest in finding both efficient *approximate* counting algorithms, as well as efficient exact counting algorithms in restricted settings. Jerrum, Sinclair and Vigoda showed how to approximately count and sample perfect matchings in any bipartite graph efficiently, although the complexity remains open on general graphs [48]. Alternatively, in 1969, Kasteleyn et al. developed a robust method to exactly count perfect matchings on any planar graph in polynomial time by calculating a Pfaffian on a directed version of the adjacency matrix [49, 78]. In fact, when the underlying graph is a lattice region, determinant-based methods for counting matchings have been shown to be even more efficient [36, 52].

Matchings on lattices arise naturally as well. For example, on finite regions of the hexagonal lattice, perfect matchings correspond to lozenge tilings of the dual region, and on finite regions of the Cartesian lattice \mathbb{Z}^2 , they correspond to domino tilings of the dual. A common Markov Chain approach for sampling perfect matchings on these lattices that is popular among experimentalists is based on *rotations*. Specifically, on \mathbb{Z}^2 the Markov Chain evolves by choosing a unit face uniformly, and if this face contains two edges of the matching on opposite sides, the Markov chain replaces those edges with the other two opposing edges on the face with some probability. A similar approach on the hexagonal lattice replaces three alternating edges around a hexagonal face with the complement set of alternating edges with some probability.

This Markov chain based on dimer rotations was first studied by Propp and Wilson [65]. They innovated their “coupling-from-the-past” algorithm for this problem, and showed that it could be run on dimer covers of the Cartesian lattice \mathbb{Z}^2 to generate perfectly uniform samples of perfect matchings, although there were no guarantees

that the algorithm would terminate in expected polynomial time. The proof that the expected time to converge is polynomially bounded was provided by Luby et al. [55], Randall and Tetali [68] and further improved by Wilson [82]. Coupling-from-the-past has subsequently been used to study matchings on many other lattices as well, providing perfectly random samples, although not always efficiently. This paradigm gave rise to many conjectures about the convergence times and stationary distributions underlying these chains. A compelling example is perfect matchings on the square-octagon lattice, Λ_{so} , where the dual is a dimer problem on a graph of squares and triangles known as “fortresses” [64]. Many remarkable properties of lozenge and domino tilings, such as the existence of frozen regions at equilibrium, are known to hold for fortresses [64].

There is a natural analogue to the dimer-rotating Markov chain for fortress graphs, that has been used experimentally to study these matchings. This chain is known to connect the state space of perfect matchings [11], but nothing is known rigorously about its convergence time. Although related Markov chains on other lattices are known to converge in polynomial time, including the Cartesian and hexagonal lattices, simulations suggest this chain may in fact require exponential time on the square-octagon lattice to converge. As an informal explanation of the motivation for why the convergence time of this Markov chain is likely exponential, it is useful to interpret these perfect matchings as “contours” [11, 42, 54]. In particular, given any perfect matching on a simply connected region of the square-octagon lattice, we first contract the four vertices of each square into a single vertex, leaving only the edges bordering two octagons. The resulting configuration will be a collection of edges on the Cartesian lattice where every vertex, except possibly those on the boundary, must have even degree, and where each vertex of degree 2 must be incident to one horizontal and one vertical edge (see Figure 3).

If we decompose these sets of edges by pairing non-crossing adjacent edges at the degree 4 vertices, we therefore get a collection of “turning paths,” that terminate at odd-degree boundary vertices, and closed “turning cycles.” The “turning” property refers to

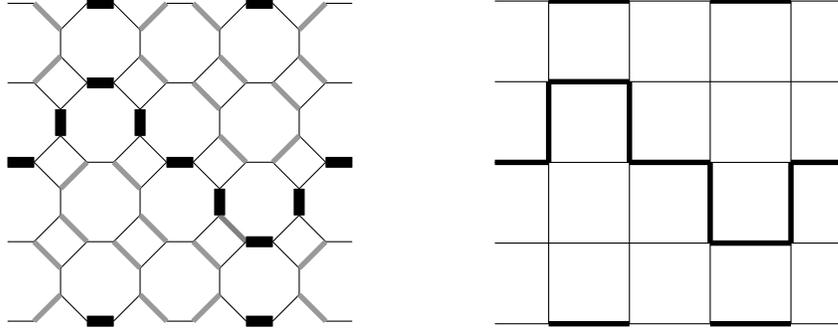


Figure 3: The mapping between (a) perfect matchings of G and (b) turning graphs of G^* .

the fact that traversals of the edges of a path or cycle are required to turn left or right at *every* step.

It is important to note that this map is not bijective, and each turning graph is the image of 2^k perfect matchings on the square-octagon lattice, where k is the number of degree 0, or *free*, vertices in σ (see Figure 4). Each free vertex corresponds to a square on Λ_{so} containing two matched edges, and there are exactly two ways this can occur. Thus each turning graph σ has weight proportional to $2^{V(\sigma)}$, where $V(\sigma)$ is the number of free vertices in σ . This tells us configurations with more free vertices will have greater weight, and this weighting penalizes configurations with long paths and cycles. This is the key insight gained by considering this transformation, as it allows us to use analysis similar to other models in statistical physics, most notably the Ising model of ferromagnetism, that are slowly mixing when long contours are similarly disfavored [54].

As with many statistical physics models, we see a relationship between the rate of convergence of local Markov chains and an underlying *phase transition* in the physical model itself. For the Ising model, a fundamental model of ferro-magnetism, local algorithms are known to converge in polynomial time (in the diameter of the region) at high temperature, but require exponential time at low temperature [46, 54, 69]. On \mathbb{Z}^2 , there is a sharp phase transition: there is a critical temperature below which the chain is slowly mixing (requiring exponential time), and at and above which it is rapidly mixing

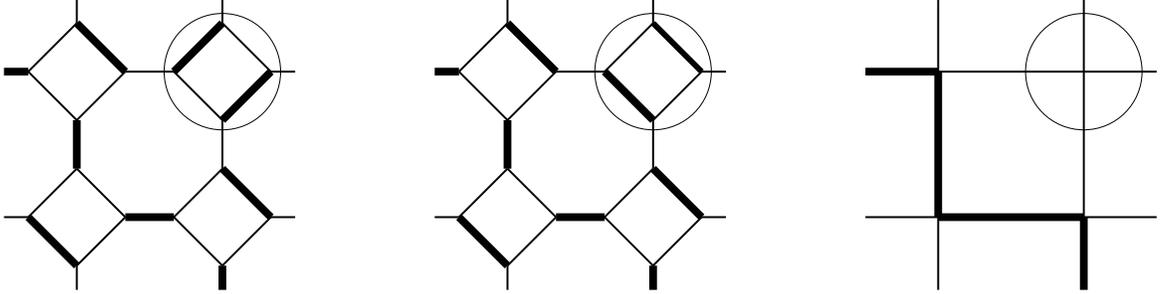


Figure 4: Two possible orientations (a) and (b) for each free vertex in G^* (c).

(converging in polynomial time) [54]. A similar behavior is seen for weighted independent sets on \mathbb{Z}^2 as we change the “activity” (or “fugacity”), a parameter that controls the expected density of an independent set. Local sampling algorithms for independent sets are known to be rapidly mixing when this parameter is small, favoring sparse independent sets [80], and slow to converge when this parameter is high, favoring denser independent sets [14].

To understand the interaction of these turning paths, we introduce a more general weighted version of the model to expose a phase transition in the mixing time. Such an approach was taken recently in the context of triangulations [20] and rectangular dissections and dyadic tilings [18], revealing similar dichotomies. A similar approach was previously considered to study a different, nonlocal Markov chain on sets of perfect and near-perfect matchings on the square-octagon lattice [11], but the behavior of the more natural local dimer-rotating Markov chain studied here remains open.

We focus our attention on a modified version of the Aztec Diamond graph, which we call the “Decorated Aztec Diamond.” The Decorated Aztec Diamond has 4 additional edges attached to the corners of diamond, which cause the corner vertices to have an odd number of incident edges. These four boundary vertices must be connected by a pair of turning paths in one of the two non-crossing ways, and moving between these two classes of configurations requires passing through configurations where the two paths touch. We will show that for this to happen, the paths must be quite long, which is exponentially

less likely at equilibrium for appropriate settings of the parameters. We then argue that because it will take exponential time to reach such a configuration, the Markov chain will require exponential time to converge. Formalizing this type of intuition is often challenging, however, and this particular problem has been open since proposed by Jim Propp in 1997 [63, 64].

We now state our results. For simplicity of notation, our terminology throughout the paper is based on weighted turning graphs rather than matchings. Let $G \subset \mathbb{Z}^2$ be a finite region on the Cartesian lattice, and let T be the set of turning graphs on G . (i.e., all vertices $v \in G \setminus \partial(G)$ have even degree, and any traversal must “turn” at each vertex.) For input parameters $\lambda > 0$ and $\mu > 0$, we define the distribution as follows. Let $\sigma \in T$ be a turning graph. Then

$$\pi_{\lambda, \mu}(\sigma) = \lambda^{|E(\sigma)|} \mu^{|V(\sigma)|} / Z,$$

where $E(\sigma)$ are the edges in σ , $V(\sigma)$ are the “free” vertices in σ , those that are *not* incident with any edge. $Z = \sum_{\tau \in T} \lambda^{|E(\tau)|} \mu^{|V(\tau)|}$, is a normalizing constant known as the *partition function*.

When $\mu = 1$, we weight configurations $\sigma \in T$ by $\lambda^{|E(\sigma)|}$, favoring shorter contours when $\lambda < 1$ and longer ones when $\lambda > 1$. We show that when $\mu = 1$ and $\lambda < 1/(2\sqrt{e})$ or $\lambda > 2\sqrt{e}$, the Markov chain \mathcal{M} mixes slowly. (A duality in the lattice implies that when $\mu = 1$, if the chain is slow for $\lambda = \lambda^*$ then it is also slow for $\lambda = 1/\lambda^*$.) For $\mu > 1$, we show that if $\lambda < \sqrt{\mu}/2\sqrt{e}$ or $\lambda > 2\mu\sqrt{e}$, the Markov chain \mathcal{M} again mixes slowly.

4.2 The Fortress Model

We begin by formalizing our model. Let $\Lambda_{so} = (V, E)$ be the infinite square-octagon lattice, and let G_{so} be a finite, simply connected region of Λ_{so} . We are interested in randomly sampling from the set of perfect matchings on G_{so} , which we denote $\Omega_{SO} = PM(G_{so})$. The local Markov chain \mathcal{M}_{so} , at any initial perfect matching σ_{so} , first chooses a face $\in G_{so}$ uniformly from the interior faces. If the boundary of the

face contains alternate edges of σ_{so} , the Markov chain then attempts to transition to the configuration with the complement alternate edges about the face, and with all other edges unchanged. We call this a “rotation” of the edges about the face.

It will be convenient to introduce an alternate representation of perfect matchings on the square-octagon lattice as *turning contours* on \mathbb{Z}^2 . We distinguish two types of edges of G_{so} , the edges that border both a square and an octagon, which we call *square* edges, and those that have octagons on both sides, which we call *octagon* edges. Consider the map from Λ_{so} to \mathbb{Z}^2 that contracts every square into a single vertex, eliminating self loops. This map eliminates all square edges in Λ_{so} , and leaves the octagon edges in Λ_{so} as edges of \mathbb{Z}^2 . Let G_C be the image of G_{so} under this map, and let $\sigma \subset E(G_C)$ be the image of perfect matching $\sigma_{so} \in PM(G_{so})$. G_C is a simply connected region of \mathbb{Z}^2 , formed from the octagon edges of G_{so} , and σ_C corresponds to the octagon edges of σ_{so} .

We observe that for $\sigma_{so} \in PM(G_{so})$, there must be an even number of octagon edges incident to any interior square, and that if there are only two incident octagon edges, then they are not diametrically opposite. This means that σ_C must have an even number of edges incident to every interior vertex, and if only two edges are incident to a vertex, they must form a right angle, or “turn.” The squares on the boundary of G_{so} can have an even or odd number of outgoing edges, and are therefore mapped to vertices in G_C that retain an “even” or “odd” designation that indicates the parity of incident edges that any configuration σ_C may have. It follows that any valid perfect matching σ_{so} on G_{so} maps to a σ_C that can be decomposed into cycles and paths that begin and end at odd-parity vertices on the boundary. Since our paths and cycles alternate between horizontal and vertical edges of G_C , we call such configurations *turning graphs* [11].

To correspond with the set Ω_{SO} of all perfect matchings on G_{so} , let Ω_C be the corresponding set of all turning graphs on G_C . There is a well-structured many-to-one map between Ω_{SO} and Ω_C . Let a vertex of G_C be called a *free vertex* of σ_C if it is not incident to any edge of σ_C . For every $\sigma_C \in \Omega_C$ with exactly k free vertices, there

are exactly 2^k pre-images $\in G_{so}$, as each “free” vertex in σ corresponds to a square in G_{so} whose edges can freely be matched in exactly two ways, independently of all other vertices (see Figure 4).

4.2.1 Weighted Turning Graphs

It will be useful to consider a generalized, weighted model of turning graphs on G_C .

First, we introduce a natural weighted model on Ω_{SO} . Given a perfect matching $\sigma_{so} \in \Omega_{SO}$, let $\#N(\sigma_{so})$ be the number of octagon edges. For input parameter $\lambda > 0$, let the weight of σ_{so} be defined as $\lambda^{\#N(\sigma_{so})}/Z$, where $Z = \sum_{\sigma_{so} \in PM(G_{so})} \lambda^{\#N(\sigma_{so})}$ is a normalizing constant, also known as the *partition function*.

Projecting this weighting to Ω_C , for a particular turning graph σ_C , it follows that $\pi(\sigma_C)$ is

$$\pi(\sigma_C) = \pi_\lambda(\sigma_C) = \frac{2^{k(\sigma_C)} \lambda^{|\sigma_C|}}{Z},$$

where $k(\sigma_C)$ is the number of free vertices of σ_C . Given a turning contour on G_C sampled according to the prescribed probability distribution, we can easily sample a perfect matching on G_{so} by flipping a bit for each free vertex uniformly at random, and assigning one of the two orientations of the perfect matching at each free vertex accordingly (see Figure 4).

We further generalize this model by introducing a parameter μ , and letting the weight of a configuration

$$\pi(\sigma_C) = \pi_{\lambda,\mu}(\sigma_C) = (\mu^{k(\sigma_C)} \lambda^{|\sigma_C|})/Z,$$

where

$$Z = \sum_{\tau \in \Omega_C} \mu^{k(\tau)} \lambda^{|\tau|}.$$

For convenience, we denote this probability model on Ω_C as $\pi_{\lambda,\mu} : \Omega_C \rightarrow \mathbb{R}$. Note that the case where $\mu = 2$ corresponds to perfect matchings of the square-octagon lattice. By setting $\mu = 1$ in this model, we effectively ignore the effect of the free vertices and the weight of a configuration is more directly influenced by the underlying geometry of

the turning graphs. We show that techniques used to analyze this special case can be extended to the general case of arbitrary μ .

4.2.2 A Markov chain on weighted turning graphs

A natural Markov chain that has been considered in the context of perfect matchings on the square-octagon lattice iteratively took a square or octagon face and rotated all the edges present if this resulted in a valid configuration. Rotations on square faces did not affect the weight of a configuration, while rotations of an octagonal face could increase or decrease the weight of a configuration multiplicatively by λ^4 .

We define the local Markov chain \mathcal{M} on turning graphs Ω_C , starting at any initial configuration σ_0 . The number of steps t required to produce samples sufficiently close to equilibrium will be discussed subsequently.

The Markov chain \mathcal{M}

Repeat for t steps:

- Choose a face x of G_C uniformly at random.
- If no edges of x are in σ_t , let σ be the turning path created by adding the edges of face x .
- If every edge of x is in σ_t , let σ be the turning path created by removing the edges of face x .
- With probability $\min(1, \frac{\pi(\sigma)}{\pi(\sigma_t)})$, let $\sigma_{t+1} = \sigma$, and with the remaining probability, let $\sigma_{t+1} = \sigma_t$.

This Markov chain represents precisely the octagon rotating moves of \mathcal{M}_{so} and ignores the square rotating moves. Note that two configurations that differ by only a square rotation will map to the same turning graph. The fact that \mathcal{M} connects the state space

Ω_C of turning graphs of \mathbb{G}_C and is aperiodic follows from the ergodicity of \mathcal{M}_{so} on perfect matchings of the square octagon lattice [63].

For all $\epsilon > 0$, the *mixing time* $\tau(\epsilon)$ of Markov chain \mathcal{M} is defined as

$$\tau(\epsilon) = \min\{t : \max_{x \in \Omega_C} \frac{1}{2} \sum_{y \in \Omega_C} |P^t(x, y) - \pi(y)| \leq \epsilon, \forall t' \geq t\}.$$

We say that a Markov chain is *rapidly mixing* (or *polynomially mixing*) if the mixing time is bounded above by a polynomial in n and $\log(\epsilon^{-1})$. It is *slowly mixing* if it is bounded below by an exponential function. In Section 4.3, we bound the mixing time of the Markov chain \mathcal{M} on the turning graph model at various input parameters λ when $\mu = 1$. In Section 4.4, we extend these results to the more general model when $\mu > 1$.

4.3 *Mixing of the Markov Chain \mathcal{M} on $\pi_{\lambda,1}$*

The proofs that \mathcal{M} is slowly mixing use so-called ‘‘Peierls arguments,’’ first introduced to study phase transitions in statistical physics models [61]. The Peierls argument identifies an exponentially small (or unlikely) set in the state space by defining a map from this set to the entire state space with exponential gain in weight. When this unlikely set is a cut in the state space, this implies that the chain will take exponential time to move from one side of the cut to the other. This therefore implies that the chain is slowly mixing.

It is fairly simple to show that the chain mixes exponentially slowly when $\lambda > 4$ or $\lambda < 1/4$. We improve this by using a more careful combinatorial analysis, thereby showing slow mixing when $\lambda < 1/(2\sqrt{e})$ or $\lambda > 2\sqrt{e}$. The proof that \mathcal{M} is rapidly mixing when $\mu = 1$ and $\lambda = 1$ relies on a novel bijection between turning graphs and 3-colorings of the grid.

We first focus on the problem of sampling from our weighted distribution over turning graphs in the case when $\mu = 1$. Here, the weight of a contour depends only on the number of edges in the contour. We will show in Sections 4.3.1 and 4.3.2 that \mathcal{M} is slowly mixing when λ is sufficiently small (or sufficiently large) by bounding the chain’s conductance.

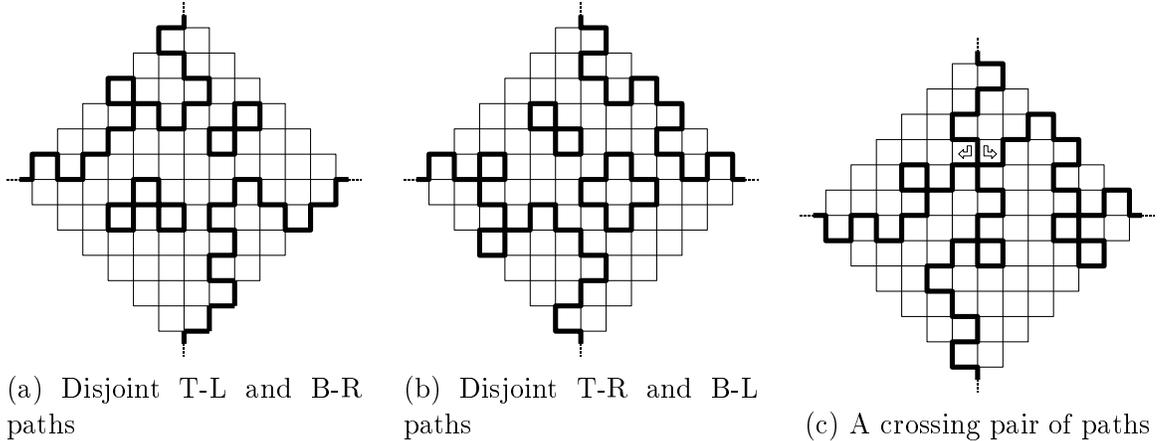


Figure 5: Configurations in (a) Ω_L , (b) Ω_R , and (c) Ω_C respectively.

Conversely, in Section 4.3.3 we show when \mathcal{M} is polynomially mixing when $\lambda = 1$ by reducing to a related chain on 3-colorings of finite regions of the grid.

4.3.1 Slow mixing of \mathcal{M} on $\pi_{\lambda,1}$ for $\lambda < 1/2\sqrt{e}$

We will begin by showing that when $\lambda < 1/2\sqrt{e}$, the Markov Chain \mathcal{M} mixes slowly on a decorated version of a certain diamond graph also known as the “Aztec Diamond” graph G [30]. Starting with the standard Aztec Diamond graph of order n , G_n , we add extra edges E_B to the four corners of the diamond, which we denote as “T, R, B, L” to indicate the top, bottom, left and right corners of the diamond respectively. Let this subgraph of \mathbb{Z}^2 be G_C , and the corresponding region of Λ_{so} be G_{so} . We force a perfect matching in G_{so} to include our added corner edges, since those edges are degree 1. Therefore in G_C , each of those four corners must be the endpoint of some turning path. By construction, these four corners are in fact the *only* possible endpoints of a turning path, and therefore in any turning graph σ_C of G_C , there must either be turning paths from T to L and B to R, or alternately from T to R and B to L. (See Figure 5).

We will use these paths to define our cut, with the cut Ω_C separating the “left” and “right” sets Ω_L and Ω_R . Let Ω_L to be the set of configurations with paths from T to L and B to R but *not* from T to R or B to L. Similarly, let Ω_R be the set of configurations

with paths from T to R and B to L but not from T to L or B to R. Finally, let the cut Ω_C consist of all states where both paths exist. For any state in Ω_C , we can identify a “crossing pair of paths” that are a *connected* set of edges that are the union of a T to L path and a B to R path. This crossing pair of paths has the interesting property that it could alternatively also have been interpreted as the union of a T to R path and a B to L path. In order to pass from configurations in Ω_L to Ω_R , the Markov chain \mathcal{M} must pass through a crossing configuration in Ω_C [11].

When λ is small, we are favoring configurations where the turning paths are short, i.e. have few total edges. We will show that average configurations in Ω_C have many more edges than those of Ω_L or Ω_R , and thereby we will show that Ω_C has small weight relative to Ω_L and Ω_R . Recall that the conductance of a subset $S \subseteq \Omega$ is defined as $\Phi(S) = \sum_{s_1 \in S, s_2 \in \bar{S}} \pi(s_1)P(s_1, s_2)/\pi(S)$. We will show that the conductance of the chain is exponentially small when $\lambda < 1/(2\sqrt{e})$. This suggests that it will take a long time to transition between configurations in Ω_L and Ω_R if we have to pass through Ω_C , and from Theorem 2.3.1, we will therefore conclude that \mathcal{M} mixes exponentially slowly.

Theorem 4.3.1: When $\lambda < 1/(2\sqrt{e})$, the mixing time of the Markov Chain \mathcal{M} on $\pi_{\lambda,1}$, weighted turning graphs of the Aztec Diamond G_n , is at least

$$\tau(\epsilon) \geq n(2\lambda\sqrt{e})^{-4n} \ln \epsilon^{-1}.$$

PROOF: For any $\sigma \in \Omega_C$, we first decompose the edges of σ into its crossing pair of paths, the union of a T to L turning path and a B to R turning path that share a vertex, which we call the *crossing vertex*. For any crossing pair of paths, we can uniquely identify the lexicographically first crossing vertex as a special vertex. These paths define four regions of G , one for each diagonal boundary, that can be viewed as a maximal connected component of faces that do not cross any edges of the crossing pair of paths (through they may cross edges of cycles). We will refer to these regions by the two corners of G that they border, e.g. the “T-L” region shown in Figure 5.

We will show that Ω_C is an exponentially small cut in our state space, thereby bounding the conductance. We first describe a map $\phi_r : \Omega_C \rightarrow \Omega$ such that for any $\sigma \in \Omega_C$, the weight of the image $\pi(\phi_r(\sigma))$ is exponentially larger in n than $\pi(\sigma)$. We construct $\phi_r(\sigma)$ for $\sigma \in \Omega_C$ as follows (see Figure 6). Given a state $\sigma \in \Omega_C$, take a pair of crossing paths in σ with maximal edges, and call the edges in this pair of crossing paths C . We remove the edges of C from G , and then shift all edges in σ from the B-L region up by one edge (we increase the y-coordinate by 1), and all edges from the T-R region down by one edge. Finally, we add in edges along the bottom left and top-right boundaries of G to form a valid turning graph.

It will be convenient to partition Ω_C into sets $\Omega_{C,h,v}$ for $h, v \geq 0$ as follows. Given $\sigma \in \Omega_C$, consider the lexicographically first pair of crossing paths in σ . We separate these into “top-left” and “bottom-right” turning paths that meet at their lexicographically first crossing point x . We define the “horizontal path” as the sub path of the top-left path from the left vertex to x , concatenated with the sub path of the bottom-right path from x to the right vertex. We similarly define the “vertical path” from the top vertex to x to the bottom vertex passing through x . This horizontal path, viewed as a left-to-right path, contains some $h \geq 0$ “backwards” edges from right to left. Similarly the “vertical path” has $v \geq 0$ backwards edges from bottom to top. In this case, we say that $\sigma \in \Omega_{C,h,v}$.

Note that since the horizontal path ends exactly $2n$ edges to the right of its origin, it contains exactly $2n + 2h$ total horizontal edges. Moreover, it is the union of two turning paths, alternating horizontal and vertical edges, so the number of vertical edges in the horizontal path must be $2n + 2h + \delta_h$, where $\delta_h = 0$ if the edges at the point of intersection are vertical, and $\delta_h = -2$ if they are horizontal. Similarly, the vertical path has exactly $2n + 2v$ vertical edges and $2n + 2v + \delta_v$ horizontal edges, with $|\delta_v| \leq 2$. We note that $\delta_v = \delta_h = \delta$, as if the horizontal path crosses the point of intersection with horizontal edges, then the vertical path must cross the point of intersection with vertical

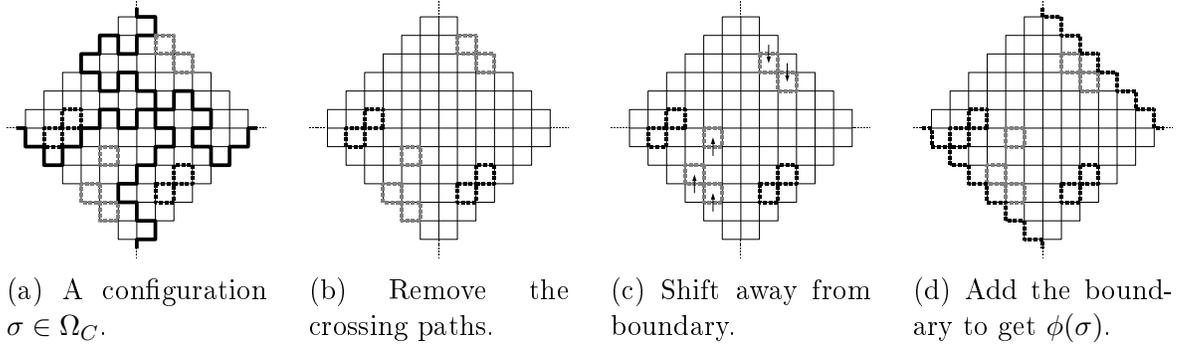


Figure 6: The mapping $\phi : \Omega_C \rightarrow \Omega$.

edges.

If $\delta = 0$, we may encode the horizontal and vertical paths essentially as two separate, equal length interleaved bit sequences, one sequence for the horizontal moves, another for the vertical moves, and a single special symbol x to indicate the location of the crossing. Given the location x , we can also account for the case where $\delta = -2$ by adding a single extra bit.

By this encoding, we see that a bound on the number of pre-images of the map ϕ that have left-right paths of type h is at most

$$n \binom{2n+2h}{h} \binom{2n+2h}{n+h}.$$

Similarly, the number of preimages of ϕ that have top-down paths of type v is at most

$$n \binom{2n+2v}{v} \binom{2n+2v}{n+v}.$$

First, we see that by Sterling's approximation,

$$\begin{aligned} \binom{2n+2h}{n+h} \binom{2n+2v}{n+v} &\leq \frac{2^{2n+2h}}{\sqrt{\pi(n+h)}} \frac{2^{2n+2v}}{\sqrt{\pi(n+v)}} \\ &= \frac{2^{4n+2h+2v}}{\pi \sqrt{n+h} \sqrt{n+v}}. \end{aligned}$$

Similarly by Sterling's approximation and the well known approximation for e ,

$$\begin{aligned}
\binom{2n+2h}{h} \binom{2n+2v}{v} &\leq \frac{(2n+2h)^h}{h!} \frac{(2n+2v)^v}{v!} \\
&\leq \frac{(2e)^{h+v}}{2\pi\sqrt{hv}} \left(1 + \frac{n}{h}\right)^h \left(1 + \frac{n}{v}\right)^v \\
&\leq \frac{e^{2n}(2e)^{h+v}}{2\pi\sqrt{hv}}.
\end{aligned}$$

Finally, it follows that $P(h, v)$, the total number of preimages of type (h, v) , is therefore at most

$$\begin{aligned}
P(h, v) &\leq 2n^2 \binom{2n+2h}{h} \binom{2n+2h}{n+h} \binom{2n+2v}{v} \binom{2n+2v}{n+v} \\
&\leq 2 \frac{n^2 2^{4n+2h+2v} (2e)^{h+v} (e)^{2n}}{2\pi^2 \sqrt{hv} \sqrt{n+h} \sqrt{n+v}}.
\end{aligned}$$

We map from a configuration with $8n + 4v + 4h$ edges in the crossing paths to a configuration with exactly $4n$ new edges, so it follows that for all $\sigma \in \Omega_{C,h,v}$, the gain in weight $\pi(\phi_r(\sigma))/\pi(\sigma) = \lambda^{-(4n+4h+4v)}$. Summing over all possible $0 \leq h, v \leq n^2$, we conclude:

$$\begin{aligned}
\pi(\Omega_C) &= \sum_{h,v} \pi(\Omega_{C,h,v}) \\
&\leq \sum_{h,v} \sum_{\sigma \in \Omega_{C,h,v}} \pi(\phi(\sigma)) \frac{\pi(\sigma)}{\pi(\phi(\sigma))} \\
&\leq \sum_{h,v} \sum_{\sigma \in \Omega_{C,h,v}} \pi(\phi(\sigma)) \lambda^{(4n+4h+4v)} \\
&\leq \sum_{h,v} \lambda^{(4n+4h+4v)} \frac{2n^2 2^{4n+2h+2v} (2e)^{h+v} (e)^{2n}}{\pi \sqrt{hv} \sqrt{n+h} \sqrt{n+v}} \\
&\leq \sum_{h,v} \frac{2n^2}{\pi \sqrt{hv} \sqrt{n+h} \sqrt{n+v}} \frac{(2\lambda\sqrt{e})^{4n+4h+4v}}{(2e)^{h+v}} \\
&\leq 2n(2\lambda\sqrt{e})^{4n}.
\end{aligned}$$

We find that for any constant $\lambda < 1/(2\sqrt{e})$, the probability $\pi(\Omega_C)$ is exponentially small in n . We can conclude that the conductance $\Phi_{\mathcal{M}}$ of the Markov chain \mathcal{M} must be bounded by

$$\begin{aligned}
\Phi_{\mathcal{M}} &\leq \sum_{s_1 \in \Omega_R, s_2 \in \overline{\Omega_R}} \pi(s_1) P(s_1, s_2) / \pi(\Omega_R) \\
&\leq \pi(\Omega_C) / \pi(\Omega_R) \\
&\leq 2 \cdot 2n(2\lambda\sqrt{e})^{4n}.
\end{aligned}$$

By Theorem 2.3.1, it follows that $\tau(\epsilon)$, the mixing time of \mathcal{M} , satisfies

$$\tau(\epsilon) \geq \frac{1}{8n} (2\lambda\sqrt{e})^{-4n} \ln \epsilon^{-1},$$

so we require exponentially many steps to converge when $\lambda < 1/(2\sqrt{e})$. ■

4.3.2 Slow mixing of \mathcal{M} on $\pi_{\lambda,1}$ for $\lambda > 2\sqrt{e}$

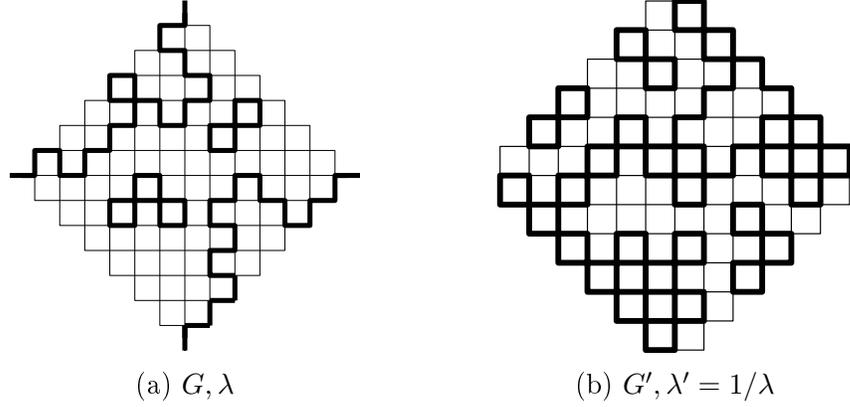


Figure 7: Weight-preserving bijection between $\sigma \subset G$ at parameter λ and $\sigma' \subset G'$ at $\lambda' = 1/\lambda$.

Perhaps surprisingly, we can also show that the chain is slowly mixing when λ is large. This actually follows from a duality between the edges and non-edges in a turning graph on the grid.

Specifically, we show that when each edge present in the turning contour is given weight at least $\lambda > 2\sqrt{e}$, the Markov Chain \mathcal{M} also mixes slowly on a slightly different version of the Aztec Diamond graph. Rather than prove this case directly, we exhibit a bijection between the model on graph G for any $\lambda < 1$, and the complimentary model on $G' = G$ with altered boundary conditions for $\lambda' = 1/\lambda > 1$.

For each vertex v in G with parity $p(v)$, set the parity of that vertex to $\deg(v) - p(v)$ in G' . It follows that the complementary turning graph $C' = E \setminus C$ on G' will be a valid turning graph that, by construction, will satisfy the parity boundary conditions of G' , as a vertex v with k incident edges in G corresponds to a vertex with $\deg(v) - k$ incident edges in E' . Let G be the Aztec diamond graph described in the previous section, and let G' be the Aztec Diamond graph with boundary conditions modified as in Figure 7.

Corollary 4.3.2: When $\lambda > 2\sqrt{e}$, the mixing time of the Markov Chain \mathcal{M} on $\pi_{\lambda,1}$, weighted turning graphs of the Aztec Diamond G_n , is at least

$$\tau(\epsilon) \geq n \left(\frac{\lambda}{2\sqrt{e}} \right)^{-4n} \ln \epsilon^{-1}.$$

PROOF: We show that the missing edges in this model G' behave exactly like the present edges in G , and will form turning paths of missing edges between vertices that have difference in parity between its degree and other parity requirement. It follows then that the unnormalized weight of a turning graph C' with parameter $\lambda' = 1/\lambda$ is exactly

$$G' = \lambda'^{|C'|} = \lambda^{|E|-|C|} = \lambda^{|E|} \lambda'^{|C|} = \lambda^{|E|} \lambda^{|C|}.$$

Since this is exactly the weight of the corresponding turning path C of G multiplied by $\lambda^{|E|}$, it follows that the normalization

$$Z' = \sum_{C' \in G'} \lambda'^{|C'|} = \lambda^{|E|} Z.$$

Thus, the normalized probability $\pi(C') = \pi(C)$.

The Markov chain \mathcal{M} behaves exactly the same on both models, by adding or removing edges with probabilities depending on the relative weights of the current and proposed next state. Thus \mathcal{M} on G with parameter λ behaves exactly the same as \mathcal{M} on G' with parameter $1/\lambda$. The corollary then follows immediately from Theorem 4.3.1. ■

4.3.3 Polynomial mixing of \mathcal{M} on $\pi_{\lambda,1}$ when $\lambda = 1$

On the positive side, we now show that the chain \mathcal{M} does converge to equilibrium efficiently when $\mu = \lambda = 1$. Our proof relies on the fact that a corresponding Markov chain on proper three colorings of finite regions of \mathbb{Z}^2 is known to be polynomially mixing [55, 68]. We will then describe a novel bijection between turning paths of a region of the grid $G = (V, E)$ with three-colorings of the faces of G , both subject to certain boundary conditions. This bijection will allow us to infer that the Markov chain \mathcal{M} is polynomially mixing on any region G of \mathbb{Z}^2 when $\mu = \lambda = 1$.

A proper three coloring of a region G of the grid is a labeling of each vertex v in G with a color chosen from $\{0, 1, 2\}$ such that no edge of G has two ends with the

same color. Three colorings of graphs are a natural combinatorial structure that arise in numerous contexts across mathematics and computer science. They are also studied in statistical physics, notably as the anti-ferromagnetic Potts model for general graphs, and the so called “6 vertex ice model” for regions of \mathbb{Z}^2 . In many of these settings, the following natural, local Markov chain, known as the single-site *Glauber Dynamics*, is of key interest.

Let G be a finite, connected region of \mathbb{Z}^2 , and Ω_3 be the space of three colorings of G . We define \mathcal{M}_3 as follows:

Beginning at any initial coloring σ_0 ,

The Markov chain \mathcal{M}_3

Repeat for t steps:

- Choose a vertex v of G uniformly at random, and $b \in \{0, 1, 2\}$ uniformly at random.
- If vertex v can be colored with b to obtain a valid three-coloring, then let σ_{t+1} be this configuration.
- Otherwise $\sigma_{t+1} = \sigma_t$.

In the context of sampling Eulerian rotations, Luby et. al. proved that the chain \mathcal{M}_3 is rapidly mixing for any region G of \mathbb{Z}^2 with fixed colors on the boundary of G [55]. Later, Goldberg et. al. proved that \mathcal{M}_3 is also rapidly mixing for rectangular regions G with no boundary conditions [37]. Cannon and Randall then generalized these results to more complicated regions G with hybrid fixed and free boundary conditions [19]. Importantly, these results made use of the *height representation* of a 3-coloring of G .

Since we will be construction a coloring of the faces of G , we will describe what follows in the context of face coloring, although these results were originally described in the context of coloring vertices. Given a region G of \mathbb{Z}^2 , a *height function* on the faces $F(G)$ of G is an assignment $h : F(G) \rightarrow \mathbb{Z}$ such that for any two neighboring faces v, w

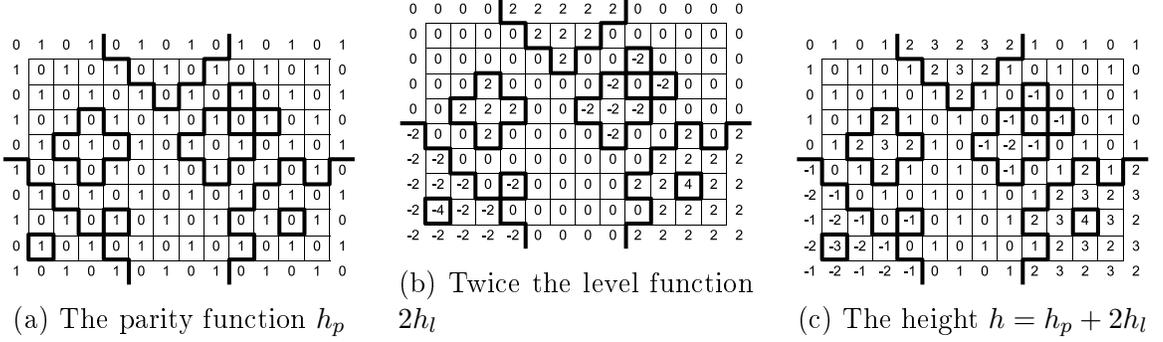


Figure 8: Coloring representations of boundary conditions and turning graphs.

on G ,

$$|h(v) - h(w)| = 1.$$

Every height function is assigned a canonical 3-coloring of G simply by assigning color $h(f) \bmod 3$ to each face f . As each 3-coloring has multiple possible height functions, we may fix an arbitrary face f_0 of G and declare its height to be 0. There is then a bijection between all 3 colorings of G where f_0 is colored 0 and all valid height functions of G where $h(f_0) = 0$ [55]. Note that as the graph of the faces of G is bipartite (and also a region of \mathbb{Z}^2 , all vertices with even height must lie on the same side of the bipartition as f_0 . We call these “even” faces, and all other vertices “odd” faces.

We are now ready to describe a novel bijection between turning graphs of a region of the grid $G = (V, E)$ with three-colorings of the faces of G , both subject to fixed boundary conditions. Essentially, we can think of a turning graph as every other level curve of the height function of some 3-coloring. We will then observe that a face rotation transition of \mathcal{M} in the turning graph G corresponds exactly to changing the color assigned to a single face of G , which is a move of \mathcal{M}_3 . This will allow us to translate the proofs of rapid mixing of \mathcal{M}_3 in the context of three colorings to a proof of rapid mixing of \mathcal{M} in the context of turning graphs.

Theorem 4.3.3: Given a simple, connected region G of \mathbb{Z}^2 , along with boundary conditions representing the parity of incident edges allowed in any turning graph of G at

each boundary vertex v , we can assign colors to the faces on the external boundary of G such that there is a bijection between the set of turning graphs on G satisfying the given vertex boundaries and the set of 3-colorings on the faces of G satisfying the constructed coloring boundaries.

PROOF: It will be convenient to first understand the properties of the mapping from height functions to turning paths. Let f_0 be any face on the exterior boundary, and without loss of generality we consider only height functions that set $h(f_0) = 0$. For any integer k , we say that the faces at level k are those with height value either $2k$ or $2k + 1$. By properties of the height function, faces at level k can only be adjacent to faces at level $k - 1$, k , or $k + 1$.

Consider the (possibly non-simple) path of edges along the boundary between maximally connected regions of faces at height levels k and $k + 1$. By construction, this path must have odd faces on one side, and even faces on the other side. It is easy to see that this can only happen if this path has the turning property. By construction, the paths between levels k and $k + 1$ cannot touch the paths between $k - 1$ and k (or any other levels), as this would imply that there are vertices connected by a path of two edges with at least a height difference of at least three.

We define the map ρ of the height function h to be the set of all edges between levels of h , which by the above argument form a valid turning graph on G . This turning graph can be decomposed into paths and cycles. The paths end at edges that separate boundary faces from two different levels. Since the locations of these endpoints depend only on the values of the height function on the boundaries of σ , it follows that any σ_2 with the same values of the height function on the boundaries must map to a turning graph with paths that end in the same locations as σ .

It is easy to show that this map from height functions to turning paths is injective, as if two height functions differ in their assignment at any face f , the difference in assigned height values must be at least 2 by parity. Therefore that vertex would belong

to different levels in the map from each height function. As we fixed face f_0 to have color 0 in both maps, the number of turning paths crossed in any path from f_0 to f would be different in the two maps, and therefore the two turning graphs must be different as well.

To show that this map is surjective, we consider the following reverse map ρ^{-1} from the set of turning graphs to height functions. Say we are given a simply connected region G of \mathbb{Z}^2 with parities $p(v)$ on each vertex on the boundary. We begin by identifying the vertices for which $p(v)$ is odd; these are the starting points for the turning paths of G . These vertices separate the boundary into disjoint sections of edges, which we identify as B_1, B_2, \dots . We may identify each boundary section as either a set of edges or exterior faces as convenient. For any turning graph σ on G , the edges of σ separate the faces of G into maximal, connected regions R_1, R_2, \dots , and each boundary section connects to exactly one (possibly empty) region.

We begin by fixing a face f_0 of the external boundary of G . We will say that the *height parity* $h_p(f) = 0$ for even faces and $h_p(f) = 1$ for odd faces. We then assign a *height level* to each region as follows. Let R_0 be the region that contains f_0 . We first set $h_l(R_0) = 0$. We require that if regions R_1 and R_2 share a boundary with odd faces on R_1 's side and even faces on R_2 's side, then $h_l(R_1) + 1 = h_l(R_2)$. We can assign height levels to every region by performing a depth-first search beginning at R_0 and assigning each region a height relative to its parent following the above requirement. For any face f in R_1 , we say that $h_l(f) = h_l(R_1)$.

We can then define $h(f) = h_p(f) + 2h_l(f)$. Let f_1, f_2 be two adjacent faces of G . If they are in the same region, $|h(f_1) - h(f_2)| = |h_p(f_1) - h_p(f_2)| = 1$. If they are in differing regions, then without loss of generality let f_1 be the odd face. Then $|h(f_1) - h(f_2)| = |h_p(f_1) - h_p(f_2) + 2(h_l(f_1) - h_l(f_2))| = |(1) + 2(-1)| = 1$. Therefore, h is a valid height function. The corresponding turning graph $\rho(h)$ consists of the edges between every other level of the h , which by construction are exactly the edges of σ .

Therefore $\sigma = \rho(h)$, and we have shown that ρ is a bijection. ■

Corollary 4.3.4: The Markov chain \mathcal{M} when $\lambda = \mu = 1$ is polynomially mixing on finite regions of the grid G .

PROOF: The Markov chain \mathcal{M} that adds or removes a single square of edges around a face f would create or destroy a region that consists of a single vertex. This would either increase or decrease the height function at that vertex by exactly two. This corresponds to the local Markov chain on 3-colorings that changes the color at a single square at a time. This chain was shown to be polynomially mixing on all subsets of \mathbb{Z}^2 with any fixed boundary conditions by a coupling argument on a related chain [55]. ■

4.4 *Mixing of \mathcal{M} on $\pi_{\lambda,\mu}$ for general $\mu > 1$*

We extend our analysis of the special case when $\mu = 1$ to the general model for any $\mu > 1$ by considering an amortized “cost” for each non-free vertex in σ , distributed among its incident edges.

Theorem 4.4.1: When $\lambda < \sqrt{\mu}/(2\sqrt{e})$, the mixing time of the Markov Chain \mathcal{M} on $\pi_{\lambda,\mu}$, weighted turning graphs of the Aztec Diamond G_n , is at least

$$\tau(\epsilon) \geq n(2\lambda\sqrt{e})^{-4n} \ln \epsilon^{-1}.$$

PROOF: To handle the case where $\mu > 1$, we need to consider the change in the number of vertices used by the turning graph. We follow the structure of Theorem 4.3.1, keeping both the structure of the proof and the map ϕ .

Let σ be a configuration in $\Omega_{C,v,h}$. As in Theorem 4.3.1, we see that σ has $8n+4h+4v$ edges in some pair of crossing paths. It follows that the sum of all degrees of all vertices incident to these edges must add to $16n+8h+8v$. This pair of crossing paths includes at least the topmost and bottommost vertex at each x coordinate, and thus must contain at least $4n$ vertices of degree 2. The degrees of the remaining vertices therefore sum

to $8n + 8h + 8v$. Since the maximum degree of any vertex is 4, there must be at least $2n + 2h + 2v$ other vertices used by this pair of crossing paths. The map $\phi(\sigma)$ removes this pair of crossing paths, and adds two paths of exactly $4n$ edges and $4n$ vertices. Thus, we have a net gain of at least $2n + 2h + 2v$ vertices between σ and $\phi(\sigma)$. Thus, the change in weight for any $\sigma \in \Omega_{C,h,v}$ will be

$$\begin{aligned} \pi(\phi_r(\sigma))/\pi(\sigma) &\geq \mu^{2n+2h+2v} \lambda^{-(4n+4h+4v)} \\ &= (\lambda/\sqrt{\mu})^{-(4n+4h+4v)}. \end{aligned}$$

As in Theorem 4.3.1, this implies by Theorem 2.3.1 that \mathcal{M} mixes exponentially slowly when $\lambda/\sqrt{\mu} < 1/2\sqrt{e}$, or more simply $\lambda < \sqrt{\mu}/2\sqrt{e}$. \blacksquare

We now similarly analyze the case where $\lambda > 1$ and obtain a result analogous to Corollary 4.3.2 for this more general case. Following the bijection in Corollary 4.3.2 that maps a turning graph in G with the complementary graph in G' , we could immediately conclude from Theorem 4.4.1 that for any $\mu < 1$, \mathcal{M} is slowly mixing whenever $\lambda > 2\sqrt{e}/\sqrt{\mu}$. However, we are chiefly interested in the case when $\mu > 1$, especially when $\mu = 2$. In this case, we can reason directly from Corollary 4.3.2.

Theorem 4.4.2: When $\lambda > 2\mu\sqrt{e}$, the mixing time of the Markov Chain \mathcal{M} on $\pi_{\lambda,\mu}$, weighted turning graphs of the Aztec Diamond G_n , is at least

$$\tau(\epsilon) \geq n(2\lambda\sqrt{e})^{-4n} \ln \epsilon^{-1}.$$

PROOF: We proceed similarly to the proof of Theorem 4.4.1, but with one important difference. By the nature of the bijection, the mapping ϕ in this context doesn't remove edges and add shorter ones, it removes *non edges*, and adds a shorter path of unchosen edges, potentially increasing the total number of chosen vertices in the process.

However, as in the argument of Theorem 4.4.1, the sum of the degrees of vertices incident to these edges, other than the boundaries, adds to $8n + 8h + 8v$. It follows then

that *at most* $4n + 4h + 4v$ vertices will be added by the map ϕ in the complementary context.

Thus, as before, the change in weight for any $\sigma \in \Omega_{C,h,v}$ is

$$\begin{aligned}\pi(\phi_r(\sigma))/\pi(\sigma) &\geq \mu^{2n+2h+2v} \lambda^{-(4n+4h+4v)} \\ &= (\lambda/\mu)^{-(4n+4h+4v)}.\end{aligned}$$

Treating μ as a constant, by Corollary 4.3.2 we conclude that \mathcal{M} mixes slowly whenever $(\lambda/\mu) > 2\sqrt{e}$, or when $\lambda > 2\mu\sqrt{e}$. ■

CHAPTER V

CLUSTERING AND MIXING TIMES FOR SEGREGATION MODELS ON \mathbb{Z}^2 .

In our final chapter, we will introduce and analyze a generalized version of the Schelling segregation model using some of the same techniques that we adapted for the study of the fortress model in Chapter 4. The Schelling segregation model attempts to explain possible causes of racial segregation in cities. Schelling considered residents of two types, where everyone prefers that the majority of his or her neighbors are of the same type. He showed through simulations that even mild preferences of this type can lead to segregation if residents move whenever they are not happy with their local environments. We generalize the Schelling model to include a broad class of bias functions determining individuals happiness or desire to move, called the General Influence Model. We show that for any influence function in this class, the dynamics will be rapidly mixing and cities will be integrated (i.e., there will not be clustering) if the racial bias is sufficiently low. Next we show complementary results for two broad classes of influence functions: Increasing Bias Functions (IBF), where an individual's likelihood of moving increases each time someone of the same color leaves (this does not include Schelling's threshold models), and Threshold Bias Functions (TBF) with the threshold exceeding one half, reminiscent of the model Schelling originally proposed. For both classes (IBF and TBF), we show that when the bias is sufficiently high, the dynamics take exponential time to mix and we will have segregation and a large "ghetto" will form.

5.1 Introduction

The Schelling Segregation Model was introduced by Thomas Schelling in 1971 to explain how global behavior can arise from small individual preferences [72]. In Schelling’s original model, agents are one of two colors and move if there are too many neighbors of the opposite color within their immediate neighborhood. Simulations show that configurations rapidly become segregated with like colored neighbors clustered together. Schelling used this simple model to argue that “micro-motives” can determine “macro-behavior,” thereby forming the basis for Agent-Based Computational Economics.

Despite extensive interest in the Schelling model and its many variants, almost all research remains non-rigorous. Our goal here is to consider families of Schelling models in an attempt to put them on firmer footing. There are many natural extensions worth considering: How large a neighborhood is relevant to one’s happiness, and do all neighbors within this neighborhood influence us equally? Can residents move away, or are they restricted to remain in the city? Are all houses occupied, or are there empty houses (say, foreclosures) that might be even less desirable to have in one’s proximity? Is one’s happiness determined solely by the color of the majority of one’s neighbors, as Schelling originally proposed, or does one get increasingly happy or unhappy as new people of one type or the other move into the neighborhood? Are decisions to move somewhere based on each person’s relative happiness, or is one less likely to move to a house where he is not wanted if doing so decreases the happiness of his new neighbors?

Economists and social scientists use statistical and non-rigorous computational tools to study the dynamics and limiting distributions, as well as for connecting the model to real world populations [8, 38, 71, 81]. Even the concept of segregation or clustering typically is not formally defined. An exception is the rigorous analysis of the Schelling model in the one-dimensional setting [15, 35, 51, 84]. Additional rigorous work has considered further variations designed to simplify the neighbors’ interactions for some specific, basic models [39, 51, 71, 85].

5.1.1 Relation to spin systems.

The concept of micro-motives effecting macro-behavior is well-studied and far better understood in the statistical physics community, where it is used to explain fundamental concepts such as phase transitions. The Schelling model itself is reminiscent of many physical models, most notably spin systems such as the Ising model which are used to understand ferro-magnetism. In the Ising model, vertices of a graph, say a finite region $G = (V, E)$ of \mathbb{Z}^2 , are assigned + or - spins, and neighboring vertices prefer to have the same spin. Although in the original Schelling model a person's happiness depends only on the color of the majority of his neighbors, in the Ising analogue everyone is incrementally more likely to move as more people of the opposite color move into their neighborhood.

Specifically, in the Ising model we are given a parameter λ that is a function of temperature, and the stationary probability of a configuration $\sigma \in \{\pm 1\}^V$ is

$$\pi(\sigma) = \lambda^{|\{x,y: (x,y) \in E, \sigma(x)=\sigma(y)\}|} / Z,$$

where

$$Z = \sum_{\sigma \in \{\pm 1\}^V} \lambda^{|\{x,y: (x,y) \in E, \sigma(x)=\sigma(y)\}|}$$

is the normalizing constant known as the partition function. Glauber dynamics is a Markov chain on Ising configurations that changes one spin at a time using Metropolis probabilities to force the chain to converge to π . The Ising model on \mathbb{Z}^2 is known to undergo a *phase transition*, i.e., there exists a value λ_c such that when $\lambda < \lambda_c$, the Glauber dynamics for the Ising model mixes in time polynomial in $|V|$ and when $\lambda > \lambda_c$, it mixes in exponential time [46, 69, 54, 77]. Moreover, the phase transition in the mixing time is accompanied by a corresponding transition in the stationary distribution of the Markov chain; at low λ , an average sample from the steady state is “evenly mixed” with regards to the proportions of spins, while at high lambda, an average sample is *clustered*, and has large regions of predominantly one spin type. Indeed, the Ising model has been

studied empirically as an alternative to the Schelling model [71, 75, 76]. In open systems at low temperature (high bias) the population will become predominantly one color or the other, and in closed systems (arising as a fixed magnetization Ising model), large clusters of one color (or spin) will form, indicating segregation [77, 83].

While extensions of the Ising model on \mathbb{Z}^2 have been examined extensively by physicists and mathematicians, the resulting models are typically less-tractable and give little insight into Schelling variants (such as neighborhoods of size larger than 4, unoccupied houses, or bias functions that do not scale geometrically with the number of differently colored neighbors). A lot is known about the Ising model on graphs with more than nearest-neighbor interactions see, e.g., Chapters 2 and 9 of [62] and general spin systems on \mathbb{Z}^d have been shown to have a phase transition whenever there is a phase transition in the associated mean field model for certain classes of interactions [13, 12, 21]. However, while these results apply only to certain classes of interactions, they fail to give insight into more general utility functions which more closely resemble the original Schelling model.

5.1.2 Generalized segregation models.

We consider a generalization of the Schelling model called the General Influence Model (GIM) and give rigorous results demonstrating a dichotomy in mixing times and clustering for two broad classes. The GIM considers *open cities* in a *non-saturated* setting, with neighborhoods of *any radius*, and where moving is based on the *product of everyone's happiness*. *Open cities* allow residents to move away, while *closed cities* require fixed racial demographics. *Unsaturated cities* allow houses to be unoccupied. An individual's happiness is a function depending only on the number of unoccupied, red and blue houses within a certain radius. This function can be a threshold, as suggested by Schelling, a geometric function, similar to the Ising model, or anything else. Moreover, these influence functions are controlled by parameters measuring the strength of these

biases, so for any influence function we can study the effects of large or small racial bias.

First, we consider a natural extension of the Schelling dynamics where people move according to the relative global happiness and we analyze the mixing time, or the time to approach equilibrium. The relevance of bounding the mixing time to understanding Schelling dynamics is indirect and will help us discern properties of the stationary distribution. Second, we formalize a concept of clustering in order to predict when typical configurations are likely to be segregated or integrated. We show that for any influence function, the dynamics will be fast mixing and cities will be integrated (i.e, there will not be clustering) if the racial bias is sufficiently low. Next, we show complementary results for two broad classes of influence functions. The first is for Increasing Bias Functions (IBF), where an individual’s likelihood of moving increases each time someone of the other color moves close or someone of the same color leaves (this does not include Schelling’s threshold model). The second is for Threshold Bias Functions (TBF) when the threshold is more than one half, reminiscent of the model Schelling originally proposed. Here a resident is happy as long as the majority of his neighbors share his color, and is unhappy otherwise, regardless of the actual percentage. For both classes (IBF and TBF) we show that when the bias is sufficiently high, the dynamics take exponential time to mix and we will have segregation. Note that because we are considering open cities, segregation means the city will become predominantly one color, a large ghetto, and slow mixing means that it will take exponentially long for the city to transition from a ghetto of one color to one of the other color. It’s important to note that this does *not* imply that it will take long to see the emergence of ghettos or for the configuration to “stabilize” as one large ghetto; it only means that it will take exponentially long to transition from one essentially stable configuration to another. (We also have initial results showing that these results can be extended to closed cities where our definition of clustering also holds for populations with any fixed racial demographics.)

In Section 5.2 we formalize the General Influence Model, which we subsequently view

as a Markov chain on the set of all housing assignments. We also formalize definitions of mixing times and clustering that we will use to establish dichotomies in the subsequent sections. In Section 5.3 we provide the proofs of fast mixing for all influence functions at low bias and slow mixing for the IBF and TBF classes at high bias. Finally, we conclude with some open problems.

5.2 Preliminaries

We first formalize our generalization of the Schelling model, which we call the *General Influence Model* (GIM), and present some background on the mixing time of Markov chains and clustering.

5.2.1 The General Influence Model.

Let Ω be the set of all 3-colorings of the faces of the $n \times n$ grid G_n , where the colors represent the types of occupants in a housing grid. We label the possible colors B, R and U where B and R represent two types of residents, red and blue, U represents an unoccupied house and we refer to each of these as B, R , or U -faces respectively (see e.g., Figure 10). An *occupied* face refers to a B or R -face. We denote the color of face x in configuration σ as $\sigma(x)$. To simplify our notation, we let $\sigma_{x_1=c_1, x_2=c_2, \dots}$ denote the configuration σ with face x_i colored c_i , for each specified i .

We consider a natural Markov chain \mathcal{M} on Ω whose transitions alter the color of one face at a time. We select a face $x \in G_n$ and a color $c \in \{B, R, U\}$ uniformly at random, then set face x to color c with probability that depends on the total change in “happiness” of the configuration. The happiness of any occupied face is determined by the colors of faces within a radius of r , and the weight of a configuration is the product of the happiness of each occupied face.

Formally, we are given a fixed radius r as a parameter of the model. Each resident (or occupied face) is influenced equally by all $N = 2r^2 + 2r$ neighbors which we define as faces within taxicab distance r . We are also given a utility function $u : \{(s, d) : s, d \in$

$[0, N], s + d \leq N\} \rightarrow [0, 1]$, that relates the coloring of a resident's neighborhood to its happiness with an arbitrary bias (or utility) function. For an occupied face x , let $s(\sigma, x)$ be the number of neighbors of x that have the same color as x in σ and $d(\sigma, x)$ be the number of neighbors of x which have a different, but occupied color. (i.e. R - for B -faces and vice versa) in σ . The happiness of an occupied face x is defined to be $u(s(\sigma, x), d(\sigma, x))$. We also require that for all $d \geq 1$, the utility function u satisfies $u(s+1, d-1) \geq u(s, d) \geq u(s, d-1)$. In other words, one prefers a same colored neighbor to an oppositely colored neighbor to an abandoned house. For our model, we require that $u(0, 0) = 0$ and $u(N, 0) = 1$ for normalization purposes.

We will state our results in terms of bounds on the discrete partial derivatives of the utility function u . In particular, let

$$\begin{aligned} u'_\alpha &= \min_{a,b} \{u(a+1, b) - u(a, b-1)\}, \\ u'_\beta &= \max_{a,b} \{u(a+1, b) - u(a, b-1)\}, \\ u'_\kappa &= \min_{a,b} \{u(a+1, b) - u(a, b)\}, \text{ and} \\ u'_\gamma &= \max_{a,b} \{u(a+1, b) - u(a, b)\}. \end{aligned}$$

The Markov chain \mathcal{M} performs moves using the Metropolis transition probabilities with respect to the distribution π which we will define (see, e.g., Chapter 3 of [50]). The weight π of a configuration σ is defined as

$$\pi(\sigma) = \prod_{x:\sigma(x) \neq U} \lambda^{u(s(\sigma,x), d(\sigma,x))} / Z,$$

where $Z = \sum_{\sigma \in \Omega} \prod_{x:\sigma(x) \neq U} \lambda^{u(s(\sigma,x), d(\sigma,x))}$ is the normalizing constant. We are now ready to formally define \mathcal{M} .

The Markov chain \mathcal{M} :¹

¹We present the results in the unsaturated setting where we allow empty houses. For the saturated model the Markov chain allows houses to move between B and R in one move, indicating that a new resident will move in as soon as one vacates a house. All of the proofs carry over in this case and are in fact simpler.

Starting at any σ_0 , at step t iterate the following:

- Choose a face x of G_n , and a color $c \in \{B, R, U\}$ uniformly at random.
- If $\sigma_t(x) = U$, with probability 1 let $\sigma_{t+1} = \sigma_{t,x=c}$.
- If $\sigma_t(x) = R$ and $c = U$, with probability $\pi(\sigma_{t,x=U})/\pi(\sigma_{t,x=R})$ let $\sigma_{t+1} = \sigma_{t,x=c}$.
- If $\sigma_t(x) = B$ and $c = U$, with probability $\pi(\sigma_{t,x=U})/\pi(\sigma_{t,x=B})$ let $\sigma_{t+1} = \sigma_{t,x=c}$.
- With the remaining probability, let $\sigma_{t+1} = \sigma_t$.

This Markov chain trivially connects the state space since we can always reach the empty configuration from any starting configuration.

The General Influence Model (GIM) is a generalization of many well-studied models on the grid. For example, if we let $r = 1$ (each resident has $N = 4$ neighbors), and $u(s, d) = s/4$, then (after a suitable change of variables), this model is equivalent to the non-saturated Ising model on the grid [40]. Here, B -faces correspond to $+$ spins and R -faces correspond to $-$ spins. The influence on a site is the number of matching neighbors, and the fact that $u(s, d) = s/4$ means that this influence is linearly proportional to the corresponding exponent of λ in the weight of the configuration.

If instead we let $r = 1$ and $u(s, d) = U_0(s - d)$, where U is a step function, then this model corresponds to a reversible version of the original Schelling Model based on thresholds [76, 71]. Here, a site is “happy” if it has at least as many neighbors of the same color as the opposite color. If we let $r = 1$, and $u(s, d) = U_{N/2}(s)$, we have another variant of the Schelling Model where a site is “happy” if at least half of its neighbors are of the same color.

5.2.2 Mixing and clustering.

We give rigorous results demonstrating a dichotomy in mixing times and clustering for two broad classes. Here we formally define both mixing time and clustering. For all

$\epsilon > 0$, the *mixing time* $\tau(\epsilon)$ of \mathcal{M} is defined as

$$\min\{t : \max_{x \in \Omega} \frac{1}{2} \sum_{y \in \Omega} |P^t(x, y) - \pi(y)| \leq \epsilon, \forall t' \geq t\}.$$

We say that a Markov chain is *rapidly mixing* if the mixing time is bounded above by a polynomial in n and $\log(\epsilon^{-1})$ and *slowly mixing* if it is bounded below by an exponential function. In Section 5.3, we bound the mixing time of the Markov chain \mathcal{M} under different conditions.

In order to characterize whether a configuration is segregated or integrated, we determine whether one group of residents has “clustered.” We build on a concept of clustering developed in [58] based on the presence of a large region with small perimeter that is densely filled with either R - or B -faces.

More precisely, we will define a *cluster region* $C = (C_F, C_E)$ where C_F is a set of faces in the grid G_n and C_E is a connected set of edges that contains every edge which is adjacent to a face in C_F and a face in $\overline{C_F} = G_n \setminus C_F$. The perimeter of a region C is $|C_E|$.

Definition 5.2.1: Given a configuration $\sigma \in \Omega$, we say that the X -faces are *c-clustered* if σ contains a cluster region C satisfying:

1. the perimeter of C (i.e. $|C_E|$) is at most cn and
2. the density of X -faces in C_F is at least c and in $\overline{C_F}$ is at most $1 - c$.

This definition is useful to characterize clustering in open and closed cities, but in open cities the region will be the entire grid and a random configuration will be predominantly one color or the other.

5.3 Bounding the Mixing Time

We begin by showing a dichotomy in the mixing time of \mathcal{M} at high and low bias. First, we show that for any IBF and TBF utility function with threshold exceeding one half, \mathcal{M}

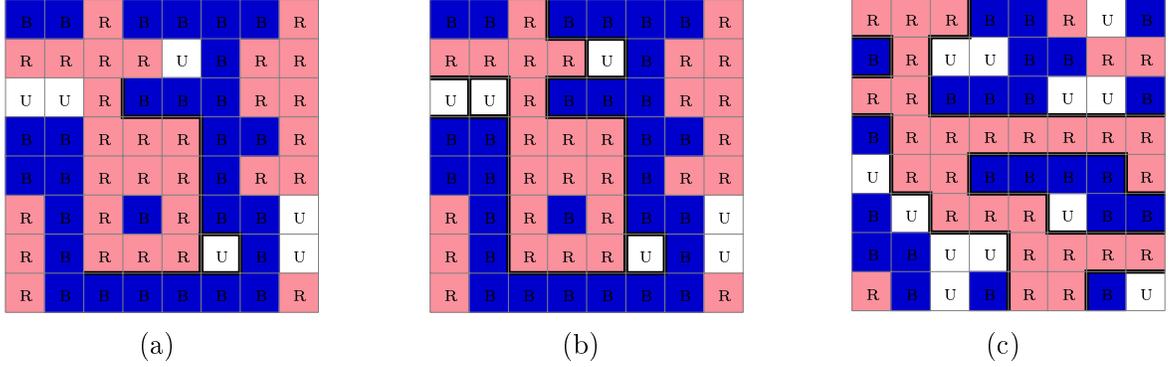


Figure 9: (a) A configuration with a contour, (b) the corresponding fat contour, and (c) an R -cross.

is slowly mixing when λ is sufficiently high. Then we show for all utility functions u , \mathcal{M} is rapidly mixing if λ is sufficiently low.

The proofs of fast mixing and integration at low bias use standard coupling and information-theoretic arguments. The proofs of slow mixing and segregation at high bias are subtle and significantly more challenging. In fact, it is not clear whether the latter results extend to the whole class of GIMs, as our proofs only verify that segregation occurs in the IBF and TBF settings.

The strategy used to show slow mixing of Markov chains and clustering effects is a *Peierls argument*, which originated in physics in order to study Gibbs measures on the infinite lattice. The argument works by showing certain types of configurations are exponentially unlikely by using combinatorial maps and information theory. In the context of Markov chains, Peierls arguments can be used to show that cut sets in the state space are exponentially unlikely, and this is sufficient to show that the Markov chain will require exponential time to converge to equilibrium. Similarly, in the context of clustering, we can use a similar argument to show that configurations that are integrated, or lack large clustered components, also have exponentially small probability at equilibrium.

The proofs of slow mixing build on some techniques established previously, but these

pieces had to be put together in novel ways. We use a strategy introduced in [70] to partition the state space according to topological features, namely monochromatic crosses (similarly colored neighboring houses that connect all four sides of the housing region) and fault lines, or long paths separating houses of different colors. Configurations with fault lines form the cut in the state space, and our objective is then to show that they have exponentially small probability. For the Ising model on \mathbb{Z}^2 , for instance, completing the argument is simple because we can reverse the spins (or flip the colors) of all houses on one side of the fault to move to a new configuration with exponentially larger stationary probability. The introduction of unoccupied houses complicates this approach, but we use a technique used in [41] by characterizing the cut as configurations with “fat faults.” The greater challenge occurs when the radius of influence is larger than 1 and residents are equally influenced by neighbors up to r houses away, for $r > 1$. In this case faults or fat faults are not sufficient and reversing the colors on one side of a fault can actually decrease the probability of a configuration. To address this we introduce the notion of bridges and build a complex of fat faults connecting components that are within distance r .

The arguments are fine tuned to the specific classes, IBF, where everyone gets increasingly happy as more people of their color move into their neighborhood, and TBF, where residents are unhappy unless some threshold over 50% is reached. Either of these conditions give us the leverage to push through the Peierls argument and show that the cutset has exponentially small probability. The significance of 50% is that if we change the color of a resident who is currently happy then he necessarily becomes unhappy, and this only happens in a threshold model when the threshold is beyond one half.

5.3.1 Slow mixing at high λ .

We begin by extending the concept of fat faults introduced in [41] to *fat faults* that are essentially large boundaries that can “jump” up to a distance of r . By showing that

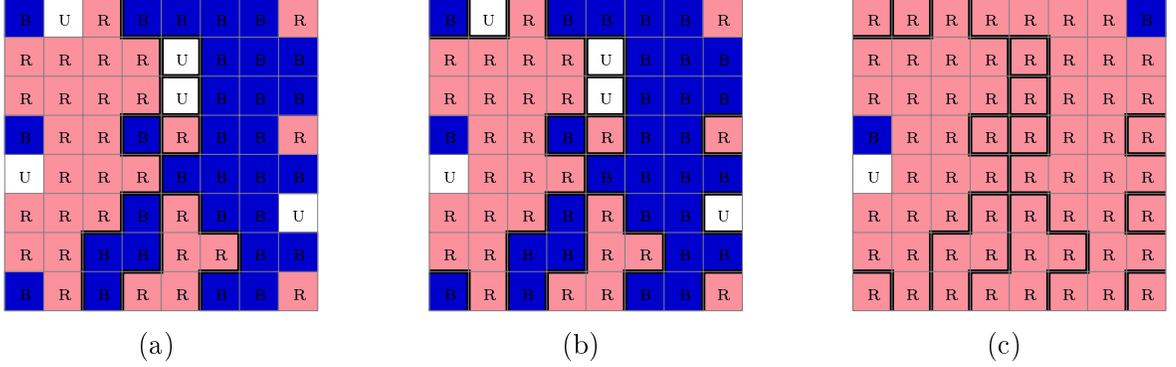


Figure 10: (a) A configuration σ with a fault line, (b) the 1-extended fault, and (c) $\phi(\sigma)$.

these types of faults are unlikely for sufficiently large values of λ , we show that \mathcal{M} mixes exponentially slowly when the utility function is in the IBF or TBF class. We begin by describing the general technique and then give the detailed proofs for the IBF and TBF classes. We make use of the well known relationship between the *conductance* and the mixing time of a Markov chain to show that three sets Ω_B, Ω_R and Ω_F , which we will define shortly, partition the state space with Ω_F being a cutset with exponentially small weight. This lets us show that the conductance of the chain is small, and we can conclude the chain mixes exponentially slowly. (See [47, 73] for details.) The *conductance* of an ergodic Markov chain \mathcal{M} with stationary distribution π and transition matrix P is

$$\Phi_{\mathcal{M}} = \min_{\substack{S \subseteq \Omega \\ \pi(S) \leq 1/2}} \sum_{s_1 \in S, s_2 \in \bar{S}} \pi(s_1) P(s_1, s_2) / \pi(S).$$

The following theorem relates the conductance and mixing time (see [73, 47]).

In order to define the three sets that form our cut we start with some terminology. We call a pair of faces within taxicab distance r to be an *influence*, and refer to this as a *bad influence* if the two faces are colored differently or are both U -faces. Influences at distance 1, adjacent faces, we call *edges* since they correspond to edges of the $n \times n$ grid. We define a *contour* to be a connected set of bad edges and a *fat contour* (see [41] and Figure 9) to be a maximally connected set of bad edges.

A fat contour, or set of fat contours, partitions the faces of the grid into regions whose

border along any single fat contour is monochromatic. With respect to a single contour, we call these R -regions, B -regions, etc. to denote the color along their border. Note that the entire regions are not necessarily monochromatic, as a B -bordered region may fully enclose a set of R faces that do not border the contour. Also note that U -regions are single squares, since all 4 sides of a U -face are bad edges. For example, see Figure 9b where the fat contour partitions the configuration into a B -region, a R -region and 4 U -regions. Given two fat contours c_1 and c_2 , c_1 is within distance r of c_2 if there exists a face adjacent to c_1 that is within taxicab distance r of a face adjacent to c_2 , and these faces are in different regions, where the regions are the unique regions defined by c_1 and c_2 . We can think of all the disjoint fat contours of a configuration to be connected to each other in an auxiliary graph if they are within distance r of each other. We then define an r -extended contour to be the union of all fat contours in a maximally connected component of this auxiliary graph.

We say that a configuration has a *monochromatic cross* if it has a connected monochromatic connected set of B -faces or R -faces that touches all four sides of the grid (see Figure 9c). We will refer to a monochromatic cross as a B -cross or a R -cross depending on the color of the faces. A fat contour that spans from the top to bottom or left to right of the grid is a *fault line*. We use the fact that every configuration falls into one of three disjoint classes: Ω_B (those with a B -cross), Ω_R (those with a R -cross), and Ω_F (those with a fault line). It is known that Ω_B , Ω_R , and Ω_F partition the state space Ω , and moves of the Markov chain \mathcal{M} cannot directly move from Ω_B to Ω_R or vice-versa, and thus must move through Ω_F [41].

Our goal is to show that Ω_F is an exponentially small cut in our state space by exhibiting a mapping $\phi_r : \Omega_F \rightarrow \Omega$ such that for any $\sigma \in \Omega_F$, the image $\phi_r(\sigma)$ “fixes” a fault line by reversing the colors in some of the monochromatic regions that border the r -extended contour containing the fault line. This causes many more same-color interactions, yielding a gain $\pi(\phi_r(\sigma))/\pi(\sigma)$ that is exponentially large in n . This gain is

exponentially larger than the total weight of all potential pre-images $\in \Omega_F$ of any state $\in \Omega$, from which we can conclude that $\pi(\Omega_F)$ is exponentially small.

We construct $\phi_r(\sigma)$ for $\sigma \in \Omega_F$ as described below (see Figure 10).

- Take the lexicographically first fault line in σ .
- Find the r -extended contour (and associated regions) which contains this fault line.
- Finally, for the regions defined by the r -extended contour, map all U -regions to R -faces and within any B -region change all R -faces to B -faces and all B -faces to R -faces.

We note that all faces within distance r of the fat fault line in σ will map to R -faces in $\phi_r(\sigma)$. This map causes all elements within distance r of the fault line to be mapped to R -faces. We also note that no bad influences are created by the map ϕ_r between previously good influences - this can only happen to faces P and Q if they are within r of each other, and also in different fault regions. However, if they are in different fault regions, some fault edge must pass through any shortest path between P and Q , and the r -extended contour would necessarily pick up the borders of the monochromatic regions containing P and Q . Thus, the mapping ϕ_r would cause both P and Q to map to R -faces.

We now bound the number of pre-images of a configuration β such that ϕ_r repairs a r -extended contour of length m (i.e. $\sigma : \phi_r(\sigma) = \beta$). Starting on one of $4n$ points on the border, a r -extended contour can be expressed by a depth first search of m edges, using at most $2m$ steps, and each step travels in up to $2r^2 + 2r$ directions. Each monochromatic region is surrounded by at least four edges, and each edge is on the boundary of two regions. Thus, there are at most $m/2$ distinct regions bordering this contour, each of which can be colored one of 3 ways. Therefore, there are at most $4n3^{m/2}(2r^2 + 2r)^m$ pre-images σ such that $\phi_r(\sigma)$ fixes this contour.

5.3.1.1 Increasing Bias Functions.

We first present result for utility functions u with bounded u'_α .

Theorem 5.3.1: For the Markov chain \mathcal{M} , with radius r and utility function u with $u'_\alpha > 0$, there exists a constant $\lambda_1 = \lambda_1(r, u'_\alpha)$ such that \mathcal{M} mixes exponentially slowly when $\lambda > \lambda_1$.

PROOF: We partition Ω_F into sets $\Omega_{F,m}$ where $\sigma \in \Omega_{F,m}$ if m is the number of bad edges fixed by ϕ_r . We observe that for two adjacent faces I and J with a bad edge, every face that influences both I and J will share a bad influence with at least one of them. Thus each of these $2r^2 - 2$ faces, excluding I, J , gains at least one new neighbor of the same type, which causes an increase of happiness of at least u'_α . Any one influence between any P and Q is counted at most 8 times in this way, once for each potential bad edge bordering P or Q . Also, the happiness of both P and Q improve from is. Thus, we see a gain of at least $u'_\alpha((2r^2 - 2)/4 + 1)$ per face bordering the fault line. Let $\sigma \in \Omega_{F,m}$, then by applying ϕ_r we fix a r -extended contour with m edges and the gain in weight satisfies

$$\frac{\pi(\phi_r(\sigma))}{\pi(\sigma)} \geq (\lambda)^{u'_\alpha \frac{m}{4}(2r^2-1)} \geq (\lambda)^{u'_\alpha \frac{mr^2}{4}}.$$

Next, let

$$\lambda > \lambda_1 = (9(4r^2 + 4r)^4)^{(r^2 u'_\alpha)^{-1}}.$$

Then we have:

$$\begin{aligned} \pi(\Omega_F) &= \sum_{m=n}^{2n^2} \sum_{x \in \Omega_{F,m}} \pi(\phi_r(x)) \frac{\pi(x)}{\pi(\phi_r(x))} \\ &\leq \sum_{m=n}^{2n^2} \sum_{x \in \Omega_{F,m}} \pi(\phi_r(x)) (\lambda^{u'_\alpha})^{-mr^2/4} \end{aligned}$$

$$\begin{aligned}
&\leq \sum_{m=n}^{2n^2} 2n(2r^2 + 2r)^m \cdot 3^{m/2} (\lambda^{-u'_\alpha m r^2/4}) \\
&\leq \sum_{m=n}^{2n^2} 2n 2^{-n/4} \leq 4n^3 2^{-n/4}.
\end{aligned}$$

Next, we will combine this bound on $\pi(\Omega_F)$ with the detailed balance condition which states that for an ergodic reversible Markov chain on Ω with transition matrix P and stationary distribution π , (see e.g. [73])

$$\forall i, j \in \Omega \quad P_{ij} \pi(i) = P_{ji} \pi(j).$$

Thus, we have that

$$\begin{aligned}
\Phi_{\mathcal{M}} &= \sum_{s_1 \in \Omega_R, s_2 \in \bar{\Omega}_R} \pi(s_1) P(s_1, s_2) / \pi(\Omega_R) \\
&\leq \sum_{s_1 \in \Omega_R, s_2 \in \bar{\Omega}_F} \pi(s_2) P(s_2, s_1) / \pi(\Omega_R) \\
&\leq \pi(\Omega_F) / \pi(\Omega_R).
\end{aligned}$$

By symmetry, we know that

$$\pi(\Omega_R) = \pi(\Omega_B) = (1 - \pi(\Omega_F)) / 2.$$

Thus, the conductance of \mathcal{M} is at most

$$\begin{aligned}
\Phi_{\mathcal{M}} &\leq \pi(\Omega_F) / \pi(\Omega_R) \\
&= 2\pi(\Omega_F) / (1 - \pi(\Omega_F)) \\
&\leq 2\pi(\Omega_F) \\
&\leq 8n^3 2^{-n/4}. \quad \blacksquare
\end{aligned}$$

By Theorem 2.3.1, it follows that $\tau(\epsilon)$, the mixing time of \mathcal{M} , satisfies

$$\tau(\epsilon) \geq (n^{-3} 2^{n/4-4} - 1) \ln \epsilon^{-1}.$$

5.3.1.2 *Threshold Bias Functions.*

We now consider the threshold variant where a face needs θ matching neighbors to be happy, so $u(s, d) = U_\theta(s)$, where U is a step function with threshold θ . Here $u'_\alpha = 0$ so we cannot apply the bounds in the previous subsection. However, a key observation allows us to apply our technique to a certain class of threshold utility functions.

Theorem 5.3.2: For the Markov Chain \mathcal{M} , with radius r , neighborhood size $N = 2r^2 + 2r$, threshold $\theta > \frac{1}{2} + \frac{1}{2r+2}N$ and utility function $u(s, o) = U_\theta(s)$, there exists a constant $\lambda_2 = \lambda_2(r)$ such that \mathcal{M} mixes exponentially slow when $\lambda > \lambda_2$.

PROOF: We again partition Ω_F into sets $\Omega_{F,m}$ where $\sigma \in \Omega_{F,m}$ if m is the number of bad edges fixed by ϕ_r . Again, every two adjacent faces I and J with a bad edge shares a neighborhood of $2r^2 - 2$ faces, excluding I and J . Thus if

$$\theta > r^2 + 2r = (2r^2 + 2r) \left(\frac{1}{2} + \frac{1}{2r+2} \right),$$

both I and J cannot be happy. Thus the mapping ϕ_r will cause at least one of I and J to become happy (from unhappy), leading to a gain of 1 per edge of the fault line. This gain is counted at most 4 times, once for each edge bordering the fixed face. Thus, we see a gain of at least $m/4$ by fixing a contour of size m , or an amortized gain of at least $1/4$ per such face. Again, we let

$$\lambda > \lambda_2 = (9(4r^2 + 4r)^4).$$

Then we have:

$$\begin{aligned} \pi(\Omega_F) &\leq \sum_{m=n}^{2n^2} \sum_{x \in \Omega_{F,m}} \pi(\phi_r(x)) (\lambda^{u'_\alpha})^{-m/4} \\ &\leq \sum_{m=n}^{2n^2} 2n(2r^2 + 2r)^m \cdot 3^{m/2} (\lambda^{-m/4}) \\ &\leq 4n^3 2^{-n/4}. \end{aligned} \quad \blacksquare$$

By the same argument as in the case of Increasing Bias Function, it follows that $\tau(\epsilon)$, the mixing time of \mathcal{M} , satisfies

$$\tau(\epsilon) \geq (n^{-3}2^{n/4-4} - 1) \ln \epsilon^{-1}.$$

5.3.2 Rapid mixing at low λ .

In contrast, we show that when λ is sufficiently low, we can guarantee that the chain mixes in polynomial time for all utility functions. Our bound on λ depends on the discrete partial derivative

$$u'_\gamma = \max_{a,b} \{u(a+1, b) - u(a, b)\}.$$

The proof relies on the now standard path coupling technique (see, e.g., [17]). We present the results in the unsaturated setting where we allow empty houses. For the saturated model the Markov chain allows houses to move between B and R in one move, indicating that a new resident will move in as soon as one vacates a house. All of the proofs carry over in this case and are in fact simpler. We prove the following.

Theorem 5.3.3: For the Markov Chain \mathcal{M} , with radius r and utility function u , there exists a constant $\lambda_3 = \lambda_3(r, u'_\gamma)$ such that \mathcal{M} is fast mixing when $1 \leq \lambda < \lambda_3$.

PROOF: We use a path coupling argument with the natural coupling. Notice that a move of \mathcal{M} consists of selecting a face f and a color c . The coupling uses the same face and color for both configurations. The distance metric we use is the minimal number of steps of \mathcal{M} required to change one configuration into another. At any face, it takes at most two steps to change the color at that face to any possible color. Thus, the maximum distance between any two configurations is $2n^2$.

In order to apply the path coupling theorem, we consider pairs of configurations at distance 1, without loss of generality let them be $(\sigma = \sigma_{g=U}, \sigma_{g=R})$. For notational purposes, for a given face y , it will be helpful to use the shorthand $u_y = u(s(\sigma, y), d(\sigma, y))$

to describe the total utility on face y . Since we are interested in the changes to this utility as a function of changing faces near y , we will also use the shorthand $u_y(a, b) = u(s(\sigma, y) + a, d(\sigma, y) + b)$ to mean the utility on face y if a additional same colored tiles and b additional opposite colored tiles are in the neighborhood of y . As the probability of a move depends on the set of neighbors near a tile, it will also be helpful to let $R(y)$ denote an indicator for the event that site y is colored R in σ , $B(y)$ an indicator for the event that y is colored B in σ , $C(y)$ an indicator for the event that $d(y, g) \leq r$, and $F(y)$ an indicator for the event $d(y, g) > r$. Roughly speaking, C and F indicate if y is “close” or “far” from g .

Let f be the face selected by \mathcal{M} . The distance can increase or decrease if $f = g$; here we consider three cases.

- If $f = g$ and $c = R$, then we accept both moves with probability 1, decreasing the distance by 1.
- If $f = g$ and $c = B$, then configuration $\sigma_{g=U}$ will accept the transition with probability 1, while the move is disallowed for $\sigma_{g=R}$; thus increasing the distance by 1.
- If $f = g$ and $c = U$, then the distance decreases by 1 with the probability that $\sigma_{g=R}$ transitions to σ , $\frac{\pi(\sigma_{g=U})}{\pi(\sigma_{g=R})}$. Every occupied face in the neighborhood around g will lose one occupied neighbor, and every R-face will also lose one same colored neighbor. Thus:

$$\begin{aligned} \frac{\pi(\sigma_{g=U})}{\pi(\sigma_{g=R})} &= \frac{1}{\lambda^{u_g}} \prod_{\substack{y: \sigma(y) \neq U, \\ d(g, y) \leq r}} \frac{\lambda^{u_y}}{\lambda^{u_y(A(y), 1)}} \\ &\geq \frac{1}{\lambda^{u_g}} \frac{1}{\lambda^{u'_\gamma s(g) + u'_\beta d(g)}} \end{aligned}$$

We now consider other cases where the distance between configurations can increase, namely whenever $f \neq g$. We again consider three cases:

- If $f = U$, both transitions are accepted with probability 1 and the distance does not change.
- If $f = R$, the probability that we increase the distance by 1 is the difference in the chance that $\sigma_{g=U}$ becomes U at f but $\sigma_{g=R}$ does not. This is exactly $\left| \frac{\sigma_{f=0,g=0}}{\sigma_{f=R,g=0}} - \frac{\sigma_{f=0,g=R}}{\sigma_{f=R,g=R}} \right|$. In the first term, every face within r of f is losing an occupied neighbor, and ever R face is losing a same-colored neighbor. The second term is more complicated. Every face within r of f is still losing an occupied neighbor, but g influences not only f , but also those neighbors that are within r of both g and f . Also, these neighbors are affected differently if the face is an A or B face. In this case,

$$\begin{aligned}
& \left| \frac{\sigma_{f=0,g=0}}{\sigma_{f=R,g=0}} - \frac{\sigma_{f=0,g=R}}{\sigma_{f=R,g=R}} \right| \\
&= \left| \frac{1}{\lambda^{u_f}} \prod_{\substack{y:\sigma(y) \neq U \\ d(y,f) \leq r}} \frac{\lambda^{u_y(-R(y),-B(y))}}{\lambda^{u_y}} - \right. \\
&\quad \left. \frac{1}{\lambda^{u_f(1,1)}} \prod_{\substack{y:\sigma(y) \neq U \\ d(y,f) \leq r}} \frac{\lambda^{u_y(-R(y)F(y),-R(y)F(y))}}{\lambda^{u_y(R(y)C(y),B(y)C(y))}} \right| \\
&\leq \frac{1}{\lambda^{u_f}} \left(\prod_{\substack{y:\sigma(y) \neq U \\ d(y,f) \leq r, d(y,g) > r}} \frac{\lambda^{u_y(-R(y),-B(y))}}{\lambda^{u_y}} \right) \\
&\quad \cdot \left| \frac{1}{\lambda^{u'_\kappa s(g)} \lambda^{u'_\alpha d(g)}} - \frac{1}{\lambda^{u'_\gamma s(g)} \lambda^{u'_\beta d(g)}} \right| \\
&\leq \left| \frac{1}{\lambda^{u'_\kappa s(g)} \lambda^{u'_\alpha d(g)}} - \frac{1}{\lambda^{u'_\gamma} \lambda^{u'_\gamma s(g)} \lambda^{u'_\beta d(g)}} \right| \\
&\leq 1 - \frac{1}{\lambda^{u'_\gamma + (u'_\gamma - u'_\kappa) s(g)} \lambda^{(u'_\beta - u'_\alpha) d(g)}}
\end{aligned}$$

- Similarly, if $f = B$, this is bounded by

$$\leq 1 - \frac{1}{\lambda^{u'_\beta} \lambda^{(u'_\beta - u'_\alpha) s(g)} \lambda^{(u'_\gamma - u'_\kappa) d(g)}}$$

Let $\eta = \max(u'_\gamma - u'_\kappa, u'_\beta - u'_\alpha)$. (Note that for the Ising model, $\eta = 0$.) The expected change in distance is then

$$\begin{aligned}
& \mathbb{E}[\Delta(\sigma_{g=U}, \sigma_{g=R})] \\
& \leq \frac{1}{3n^2} \left(\frac{-1}{\lambda^{u_g}} \frac{1}{\lambda^{u'_\gamma s(g) + u'_\beta d(g)}} \right. \\
& \quad + \quad s(g) \left(1 - \frac{1}{\lambda^{u'_\gamma}} \frac{1}{\lambda^{(u'_\gamma - u'_\kappa)s(g)} \lambda^{(u'_\beta - u'_\alpha)d(g)}} \right) \\
& \quad + \quad d(g) \left(1 - \frac{1}{\lambda^{u'_\beta}} \frac{1}{\lambda^{(u'_\beta - u'_\alpha)s(g)} \lambda^{(u'_\gamma - u'_\kappa)d(g)}} \right) \Big) \\
& \leq \frac{1}{3n^2} \left(\frac{-1}{\lambda^{2u'_\gamma s(g) + 2u'_\beta d(g)}} \right. \\
& \quad + \quad N \left(1 - \frac{1}{\lambda^{u'_\gamma s(g) + u'_\beta d(g)}} \frac{1}{\lambda^{\eta N(u'_\gamma s(g) + u'_\beta d(g))}} \right)^{1/N} \Big) \\
& \leq \frac{-1}{3n^2} \left(\frac{1}{\lambda^{2u'_\gamma s(g) + 2u'_\beta d(g)}} \right. \\
& \quad \left. - \quad (\log(\lambda^{\eta N(2u'_\gamma s(g) + 2u'_\beta d(g))})) \right)
\end{aligned}$$

where the second to last step uses the inequality of arithmetic and geometric means, and the final step uses the fact that

$$\lim_{n \rightarrow \infty} n(1 - x^{1/n}) \rightarrow -\log x \quad \blacksquare$$

from below. Recall that $\eta \leq u'_\gamma \leq u'_\beta$. Thus we see our expected change is negative whenever the value $v = \lambda^{\eta N(u'_\gamma + u'_\beta)}$ satisfies $1/v > \log v$. This occurs if

$$1 \leq \lambda \leq (1.8)^{\eta/(2r^2-1)} = 1 + O(1/r^2)$$

Setting $\lambda = (1.5)^{\eta/(2r^2-1)}$, the expected change in distance is at most $-.2612/3n^2$ per step. At last applying the path coupling theorem [17] gives the bound on the mixing time,

$$\tau(\varepsilon) \leq \frac{3n^2 \log(2n^2 \varepsilon^{-1})}{.2612} = O(n^2 \log(n \varepsilon^{-1})).$$

REFERENCES

- [1] ALDOUS, D., “Random walk on finite groups and rapidly mixing Markov chains,” *In Seminaire de Probabilites XVII*, pp. 243–297, 1983.
- [2] ALDOUS, D., “Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists,” *Euclid*, vol. 5, 1999.
- [3] ALMKVIST, G. and ANDREWS, G., “A Hardy-Ramanujan formula for restricted partitions,” *Journal of Number Theory*, vol. 38, no. 2, pp. 135 – 144, 1991.
- [4] ANDREWS, G., *The theory of partitions*. Addison-Wesley, Reading, MA, 1976.
- [5] ANDREWS, G. and ERIKSSON, K., *Integer partitions*. Integer Partitions, Cambridge University Press, 2004.
- [6] ARRATIA, R. and DESALVO, S., “Probabilistic divide-and-conquer: a new exact simulation method, with integer partitions as an example,” 2015.
- [7] AYYER, A., BOUTTIER, J., CORTEEL, S., and NUNZI, F., “Multivariate juggling probabilities,” *Electronic Journal of Probability*, vol. 20, 2014.
- [8] BARDE, S., “Back to the future: A simple solution to Schelling segregation,” *CoRR*, 2011.
- [9] BENJAMINI, I., BERGER, N., C., H., and MOSSEL, E., “Mixing times of the biased card shuffling and the asymmetric exclusion process,” *Trans. Amer. Math. Soc*, 2005.
- [10] BERESTYCKI, N. and PITMAN, J., “Gibbs distributions for random partitions generated by a fragmentation process,” *Journal of Statistical Physics*, vol. 127, pp. 381–418, 2007.
- [11] BHATNAGAR, N., GREENBERG, S., and RANDALL, D., “The effect of boundary conditions on mixing rates of Markov chains,” *Proceedings of APPROX/RANDOM*, pp. 280–291, 2006.
- [12] BISKUP, M. and CHAYES, L., “Rigorous analysis of discontinuous phase transitions via mean-field bounds,” *Communications in Mathematical Physics*, vol. 238, pp. 53–93, 2003.
- [13] BISKUP, M., CHAYES, L., and CRAWFORD, N., “Mean-field driven first-order phase transitions in systems with long-range interactions,” *Journal of Statistical Physics*, vol. 122, pp. 1139–1193, 2004.

- [14] BLANCA, A., GALVIN, D., RANDALL, D., and TETALI, P., “Phase coexistence and slow mixing for the hard-core model on \mathbb{Z}^2 ,” *Proceedings of APPROX/RANDOM*, pp. 379–394, 2013.
- [15] BRANDT, C., IMMORLICA, N., KAMATH, G., and KLEINBERG, R., “An analysis of one-dimensional Schelling segregation,” in *Proceedings of the 44th Symposium on Theory of Computing (STOC)*, pp. 789–804, 2012.
- [16] BRUINIER, J. and ONO, K., “Algebraic formulas for the coefficients of half-integral weight harmonic weak Maass forms,” *Advances in Mathematics*, vol. 246, pp. 198 – 219, 2013.
- [17] BUBLEY, R. and DYER, M., “Faster random generation of linear extensions,” *Discrete Mathematics*, vol. 201, pp. 81–88, 1999.
- [18] CANNON, S., MIRACLE, S., and RANDALL, D., “Phase transitions in random dyadic tilings and rectangular dissections,” *Proceedings of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1573–1589, 2015.
- [19] CANNON, S. and RANDALL, D., “Sampling on lattices with free boundary conditions using randomized extensions,” *Proceedings of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1952–1971, 2016.
- [20] CAPUTO, P., MARTINELLI, F., SINCLAIR, A. J., and STAUFFER, A., “Random lattice triangulations: Structure and algorithms,” in *Proceedings of the Forty-fifth annual ACM Symposium on Theory of Computing*, pp. 615–624, 2013.
- [21] CHAYES, L., “Mean field analysis of low-dimensional systems,” *Communications in Mathematical Physics*, vol. 292, pp. 303–341, 2009.
- [22] COMTET, A., MAJUMDAR, S., and OUVRY, S., “Integer partitions and exclusion statistics,” *Journal of Physics A: Mathematical and Theoretical*, vol. 40, no. 37, pp. 11255–11269, 2007.
- [23] COUSINS, B. and VEMPALA, S., “A cubic algorithm for computing Gaussian volume,” in *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA ’14*, pp. 1215–1228, 2014.
- [24] DESALVO, S. and PAK, I., “Log-concavity of the partition function,” *The Ramanujan Journal*, vol. 38, no. 1, pp. 61–73, 2014.
- [25] DIACONIS, P. and SALOFF-COSTE, L., “Comparison theorems for reversible Markov chains,” *The Annals of Applied Probability*, vol. 3, pp. 696–730, 1993.
- [26] DUCHON, P., FLAJOLET, P., LOUCHARD, G., and SHAEFFER, G., “Boltzmann samplers for the random generation of combinatorial structures,” *Combinatorics, Probability and Computing*, vol. 13, pp. 577–625, 2004.

- [27] DYER, M., FRIEZE, A., and KANNAN, R., “A random polynomial-time algorithm for approximating the volume of convex bodies,” *J. ACM*, 1991.
- [28] DYER, M. and GREENHILL, C., “A more rapidly mixing Markov chain for graph colorings,” *Random Structures & Algorithms*, vol. 13, pp. 285–317, 1998.
- [29] EDMONDS, J., “Paths, trees, and flowers,” *Canad. J. Math.*, vol. 17, pp. 449–467, 1965.
- [30] ELKIES, N., KUPERBERG, G., LARSEN, M., and PROPP, J. G., “Alternating-sign matrices and domino tilings,” *Journal of Algebraic Combinatorics*, vol. 2, pp. 111–132, 1992.
- [31] FILL, J., “Background on the gap problem,” *Unpublished manuscript*, 2003.
- [32] FILL, J., “An interesting spectral gap problem,” *Unpublished manuscript*, 2003.
- [33] FLAJOLET, P., FUSY, E., and PIVOTEAU, C., “Boltzmann sampling of unlabeled structures,” pp. 201–211, 2007.
- [34] FRISTEDT, B., “The structure of random partitions of large integers,” *Transactions of the American Mathematical Society*, vol. 337, 1993.
- [35] GERHOLD, S., GLEBSKY, L., SCHNEIDER, C., WEISS, H., and ZIMMERMANN, B., “Computing the complexity for Schelling segregation models,” *Communications in Nonlinear Science and Numerical Simulation*, vol. 13, pp. 2236–2245, 2008.
- [36] GESSEL, I. and VIENNOT, X., “Binomials determinants, paths and hook length formulae,” *Advances in Mathematics*, vol. 58, pp. 300–321, 1985.
- [37] GOLDBERG, L. A., MARTIN, R., and PATERSON, M., “Random sampling of 3-colorings in z^2 ,” *Random Structures & Algorithms*, pp. 279–302, 2004.
- [38] GRAUWIN, S., “Effect of local coordination on a Schelling-type segregation model,” *CoRR*, 2008.
- [39] GRAUWIN, S., GOFFETTE-NAGOT, F., and JENSEN, P., “Dynamic models of residential segregation: an analytical solution,” *Journal of Public Economics*, vol. 96, pp. 124–141, 2012.
- [40] GREENBERG, S., PASCOE, A., and RANDALL, D., “Sampling biased lattice configurations using exponential metrics,” in *Proceedings of the 20th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 76–85, 2009.
- [41] GREENBERG, S. and RANDALL, D., “Convergence rates of Markov chains for some self-assembly and non-saturated Ising models,” *Theoretical Computer Science*, vol. 410, pp. 1417–1427, 2009.
- [42] GREENBERG, S. and RANDALL, D., “Slow mixing of Markov chains using fault lines and fat contours,” *Algorithmica*, vol. 58, pp. 911–927, 2010.

- [43] HARDY, G. and RAMANUJAN, S., “Asymptotic formulae in combinatory analysis,” *Proceedings of the London Mathematical Society*, vol. 17, pp. 75–115, 1918.
- [44] HASTINGS, W. K., “Monte Carlo sampling methods using Markov chains and their applications,” *Biometrika*, vol. 57, pp. 97–109, 1970.
- [45] JAMES, G. and KERBER, A., *The representation theory of the symmetric group*. Encyclopedia of Mathematics and its Applications, Cambridge University Press, 1984.
- [46] JERRUM, M. and SINCLAIR, A., “Polynomial-time approximation algorithms for the Ising model,” *Society for Industrial and Applied Mathematics Journal on Computing*, vol. 22, pp. 1087–1116, 1993.
- [47] JERRUM, M. R. and SINCLAIR, A. J., “Approximate counting, uniform generation and rapidly mixing markov chains,” *Information and Computation*, vol. 82, pp. 93–133, 1989.
- [48] JERRUM, M. R., SINCLAIR, A. J., and VIGODA, E., “A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries,” *Journal of the ACM*, vol. 41, pp. 671–697, 2006.
- [49] KASTELEYN, P. W., “The statistics of dimers on a lattice : I. The number of dimer arrangements on a quadratic lattice,” *Physica*, pp. 1209–1225, 1961.
- [50] LEVIN, D., PERES, Y., and WILMER, E., *Markov chains and mixing times*. American Mathematical Society, 2006.
- [51] LEWIS-PYE, E., BARMPALIAS, G., and ELWES, R., “Digital morphogenesis via Schelling segregation,” *CoRR*, 2013.
- [52] LINDSTRÖM, B., “On the vector representations of induced matroids,” *Bulletin of the London Mathematical Society*, vol. 5, pp. 85–90, 1973.
- [53] LOVÁSZ, L. and VEMPALA, S., “Simulated annealing in convex bodies and an $O(n^4)$ volume algorithm,” in *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2003.
- [54] LUBETZY, S. and SLY, A., “Critical Ising on the square lattice mixes in polynomial time,” *Communications in Mathematical Physics*, pp. 815–836, 2012.
- [55] LUBY, M., RANDALL, D., and SINCLAIR, A. J., “Markov chain algorithms for planar lattice structures,” *SIAM Journal on Computing*, vol. 31, pp. 167–192, 2001.
- [56] MADRAS, N. and RANDALL, D., “Markov chain decomposition for convergence rate analysis,” *Annals of Applied Probability*, 2002.
- [57] METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H., and TELLER, E., “Equation of State Calculations by Fast Computing Machines,” *The Journal of Chemical Physics*, vol. 21, pp. 1087–1092.

- [58] MIRACLE, S., RANDALL, D., and STREIB, A., “Clustering in interfering binary mixtures,” in *Proceedings of the 14th international workshop and 15th international conference on Approximation, randomization, and combinatorial optimization: algorithms and techniques*, APPROX’11/RANDOM’11, pp. 652–663, 2011.
- [59] NIJENHUIS, A. and WILF, H., *Combinatorial algorithms*. Academic Press, 1978.
- [60] OKOUNKOV, A., “Symmetric functions and random partitions,” 2002.
- [61] PEIERLS, R., “On Ising’s model of ferromagnetism,” *Proceedings of the Cambridge Philosophical Society*, p. 477, 1936.
- [62] PRESUTTI, E., *Scaling Limits in Statistical Mechanics and Microstructures in Continuum Mechanics*. Theoretical and mathematical physics, Springer Berlin Heidelberg, 2009.
- [63] PROPP, J. G., “Diabolo tilings of fortresses.” <http://www.math.wisc.edu/~propp/diabolo.ps.gz>.
- [64] PROPP, J. G., “Mixing time for dimers on the square-octagon graph.” <http://mathoverflow.net/questions/168637/mixing-time-for-dimers-on-the-square-octagon-graph>.
- [65] PROPP, J. G. and WILSON, D. B., “Exact sampling with coupled Markov chains and applications to statistical mechanics,” *Random Structures and Algorithms*, pp. 223–252, 1996.
- [66] RADEMACHER, H., “On the partition function $p(n)$,” *Proceedings of the London Mathematical Society*, vol. 43, pp. 241–254, 1937.
- [67] RANDALL, D. and SINCLAIR, A., “Self-testing algorithms for self-avoiding walks,” *Journal of Mathematical Physics*, vol. 41, 2000.
- [68] RANDALL, D. and TETALI, P., “Analyzing glauber dynamics by comparison of Markov chains,” *Journal of Mathematical Physics*, vol. 41, pp. 1598–1615, 2000.
- [69] RANDALL, D. and WILSON, D. B., “Sampling spin configurations of an Ising system,” in *Proceedings of the 10th Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA, pp. 959–960, 1999.
- [70] RANDALL, D., “Slow mixing of glauber dynamics via topological obstructions,” in *Proceedings of the 17th Annual ACM-SIAM Symposium on Discrete Algorithm (SODA)*, pp. 870–879, 2006.
- [71] ROGERS, T. and MCKANE, A., “A unified framework for Schelling’s model of segregation,” *Journal of Statistical Mechanics: Theory and Experiments*, vol. P07006, 2011.
- [72] SCHELLING, T., “Dynamic models of segregation,” *Journal of Mathematical Sociology*, vol. 1, pp. 143–186, 1971.

- [73] SINCLAIR, A. J., *Algorithms for random generation and counting*. Progress in theoretical computer science, Birkhäuser, 1993.
- [74] SQUIRE, M., “Efficient generation of integer partitions. unpublished manuscript,” 1993.
- [75] STAUFFER, D., “Social applications of two-dimensional Ising models,” *American Journal of Physics*, vol. 76, pp. 470–473, 2008.
- [76] STAUFFER, D. and SOLOMON, S., “Ising, Schelling, and self-organizing segregation,” *The European Physics Journal B*, vol. 57, pp. 473–479, 2007.
- [77] SWENDSEN, R. and WANG, J., “Nonuniversal critical dynamics in Monte Carlo simulations,” *Phys. Rev. Lett.*, vol. 58, pp. 86–88, 1987.
- [78] TEMPERLEY, H. N. V. and FISHER, M. E., “Dimer problem in statistical mechanics-an exact result,” *Philosophical Magazine*, 1961.
- [79] VALIANT, L., “The complexity of computing the permanent,” *Theoretical Computer Science*, pp. 189 – 201, 1979.
- [80] VERA, J., VIGODA, E., and YANG, L., “Improved bounds on the phase transition for the hard-core model in 2-dimensions,” *Proceedings of APPROX/RANDOMs*, pp. 699–713, 2013.
- [81] VINKOVIK, D. and KIRKMAN, A., “A physical analogue of the Schelling model,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, pp. 19261–19265, 2006.
- [82] WILSON, D. B., “Mixing times of lozenge tiling and card shuffling Markov chains,” *The Annals of Applied Probability*, vol. 1, pp. 274–325, 2004.
- [83] WOLFF, U., “Collective Monte Carlo updating for spin systems,” *Phys. Rev. Lett.*, vol. 62, pp. 361–364, 1989.
- [84] YOUNG, H., *Individual strategy and social structure: an evolutionary theory of institutions*. Princeton paperbacks, 2001.
- [85] ZHANG, J., “A dynamic model of residential segregation,” *Journal of Mathematical Sociology*, vol. 28, pp. 147–170, 2004.