

THE EFFECTS OF BIAS ON SAMPLING ALGORITHMS AND COMBINATORIAL OBJECTS

A Thesis
Presented to
The Academic Faculty

by

Sarah B. Miracle

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

School of Computer Science
Georgia Institute of Technology
May 2015

Copyright © 2015 by Sarah B. Miracle

THE EFFECTS OF BIAS ON SAMPLING ALGORITHMS AND COMBINATORIAL OBJECTS

Approved by:

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

Professor Eric Vigoda
School of Computer Science
Georgia Institute of Technology

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

Professor Tom Trotter
School of Mathematics
Georgia Institute of Technology

Professor Milena Mihail
School of Computer Science
Georgia Institute of Technology

Date Approved: 18 March 2015

*To my parents Betty and Charlie Miracle who have always encouraged
me to learn*

and to my husband Erik who has supported me in so many ways.

ACKNOWLEDGEMENTS

I would first and foremost like to thank my advisor Dana Randall. From my first day in graduate school Dana has been inspiring and encouraging. Not only has she taught me to be how to be a researcher, but she has served as a mentor and friend. Without her support and guidance this thesis would not exist.

During my time at Georgia Tech I have been fortunate to work with many talented researchers including Prateek Bhakta, Sarah Cannon, Amanda Streib and Prasad Tetali. I would especially like to thank Amanda Streib who worked with me to write my first few papers and continues to provide support and advice.

I would also like to thank the members of my committee. Their help in the process of writing and defending this defense has been invaluable. I would especially like to thank Prasad Tetali and Eric Vigoda who both served on my proposal committee and wrote numerous letters of recommendation for me.

Lastly I would like to acknowledge my undergraduate professors at Vanderbilt who helped me become interested in computer science and inspired me to return to graduate school. Specifically, I'd like to thank Douglas Fisher, J. Michael Fitzpatrick and Jeremy Spinrad.

TABLE OF CONTENTS

DEDICATION	iii
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	viii
LIST OF FIGURES	ix
SUMMARY	xi
I INTRODUCTION	1
1.1 Sampling Algorithms	3
1.1.1 Markov Chain Basics	4
1.1.2 Example 1: Card Shuffling Algorithms	5
1.1.3 Example 2: Asymmetric Simple Exclusion Processes	7
1.2 Phase Transitions	8
1.2.1 The Ising Model	9
1.2.2 The Hard-Core Model	10
1.3 Sampling with Bias	11
1.3.1 Biased Permutations	12
1.3.2 Rectangular Dissections	12
1.3.3 Clustering in Colloids	13
1.3.4 Schelling’s Model of Segregation	14
II MARKOV CHAIN ANALYSIS TECHNIQUES	17
2.1 Markov Chain Basics	17
2.2 Coupling and Path Coupling	19
2.3 Comparison	21
2.4 Conductance	22
III SELF-ORGANIZING LISTS AND BIASED PERMUTATIONS	23
3.1 Biased Permutations	23

3.2	Formalizing the Markov Chains	27
3.3	An Example that is Slowly Mixing	29
3.3.1	Slow Mixing of Biased Staircase Walks	31
3.3.2	Extending Slow Mixing of Staircase Walks to Biased Permutations	34
3.4	Choose Your Weapon	42
3.4.1	The Inversion Table Representation.	43
3.4.2	Fast Mixing of the Inversion Chain	45
3.4.3	Fast Mixing of the Nearest-Neighbor Chain	49
3.5	League Hierarchies	52
3.5.1	Fast Mixing of the Binary Tree Chain	54
3.5.2	Fast Mixing of the Nearest-Neighbor Chain	58
IV	DYADIC TILINGS AND RECTANGULAR DISSECTIONS	64
4.1	Rectangular Dissections and Dyadic Tilings	64
4.1.1	Results.	67
4.1.2	Techniques.	70
4.2	Preliminaries	71
4.2.1	Formalizing the Markov Chains	72
4.2.2	Proving Ergodicity	74
4.3	Fast Mixing for Dyadic Tilings when $\lambda < 1$	75
4.4	Slow Mixing for General and Dyadic Tilings	81
4.4.1	Slow Mixing when $\lambda > 1$	83
4.4.2	Slow Mixing for General Tilings when $\lambda < 1$	84
V	COLLOIDS AND INTERFERING BINARY MIXTURES	91
5.1	Clustering in Colloids	91
5.2	Interfering Binary Mixtures	93
5.3	Clustering in Model 1	95
5.3.1	Formalizing the Model	96
5.3.2	The Clustering Property	97

5.3.3	Main Results	98
5.3.4	Clustering at High Density for Model 1	99
5.3.5	No Clustering at Low Density for Model 1	107
5.4	Generalizing from Model 1 to Interfering Binary Mixtures	108
5.4.1	Defining Interfering Binary Mixtures	109
5.4.2	Example Interfering Binary Mixtures	111
5.4.3	Relating the Weight of a Configuration to the Perimeter	116
5.4.4	Clustering at High Density for Interfering Binary Mixtures	117
5.4.5	No Clustering at Low Density for Interfering Binary Mixtures	122
VI	SEGREGATION MODELS ON \mathbb{Z}^2	125
6.1	The Schelling Segregation Model	125
6.1.1	Relation to Spin Systems.	127
6.1.2	Generalized Segregation Models.	128
6.2	Preliminaries	130
6.2.1	The General Influence Model.	130
6.2.2	Clustering.	133
6.3	Bounding the Mixing Time	134
6.3.1	Slow mixing at High λ	136
6.3.2	Rapid mixing at Low λ	142
6.4	Segregation or Integration at Stationarity	146
6.4.1	Segregation at High λ for the IBF and TBF Classes.	146
6.4.2	Integration at Low λ	151
	REFERENCES	155

LIST OF TABLES

1	Parameters for the Example Interfering Binary Mixtures	116
---	--	-----

LIST OF FIGURES

1	A 1-dimensional asymmetric simple exclusion process (a) and the corresponding staircase walk (b).	8
2	An example of fluctuating bias with exponential mixing time and staircase walks in S_1, S_2 , and S_3	31
3	A graphical representation of the relative weights of three different permutations.	36
4	A move that swaps an arbitrary $(1, 0)$ pair.	40
5	The inversion table for a permutation.	44
6	The canonical path for transposing 5 and 7.	50
7	A set \mathbf{P} with tree structure and the corresponding tree-encoding of the permutation 519386742.	53
8	The stages in the canonical path for transposing 5 and 7. Notice that the elements in S are underlined.	59
9	Stages 1 and 2 of the canonical path for transposing 5 and 7.	60
10	(a) An equitable rectangular dissection and (b) a dyadic tiling of the 16×16 square. Shaded rectangles are not dyadic.	66
11	\mathcal{M}_{64} and $\widehat{\mathcal{M}}_{64}$ after 1,000,000 simulated steps for various values of λ , starting with all vertical rectangles of width 1 and height 64.	69
12	Some tilings for $n = 16$. Tilings (a) and (b) differ by an edge flip. Dyadic tilings (c) and (d) differ by an edge flip.	73
13	Rectangle S of area $2n$ in marginal tilings A_t and B_t	76
14	An area $2n$ rectangle S bisected by horizontal edge e in A_t and vertical edge f in B_t . Four “bad” edge flips g, h, i, j exist only if A_t and B_t are tiled in the neighborhood of S as shown.	78
15	A sequence of edge flips from σ_h to σ^*	81
16	The tiling s_{64}	86
17	An example labeling of the bars and regions surrounding bars.	90
18	Model 1, squares and diamonds on the $n \times n$ grid L_n	93
19	Subfigures (a-c) are interfering binary mixtures while (d) is not.	94
20	A configuration $\sigma \in \widehat{\Omega} \setminus \Psi$ and the image $f(\sigma)$ of σ in Ψ	100

21	Before and after one step of the construction of a c -bridge system for a region R ; the solid lighter grey area is exterior to R	101
22	A c -bridge system for $\sigma \in \hat{\Omega} \setminus \Psi$; the image $f_1(\sigma)$; and $f(\sigma) = f_2 \circ f_1(\sigma)$	103
23	Model 2 and the connection with bond percolation	106
24	Example drawings of Model 3 and Model 4	111
25	Computing δ and γ for Model 3	112
26	Computing δ and γ for Model 4	113
27	Examples drawings of Model 5 , Model 6 and Model 7	113
28	n -boxes ($n = 5$) for (a) Model 4 (b) Model 5 (c) Model 7	113
29	A packing of 5-boxes for (a) Model 4 (b) Model 5 (c) Model 7	114
30	Computing δ and γ for Model 5	114
31	Computing δ and γ for Model 6	115
32	Computing δ and γ for Model 7	116
33	(a) A configuration with a contour, (b) the corresponding fat contour, and (c) an R -cross.	134
34	(a) A configuration σ with a fault line, (b) the 1-extended fault, and (c) $\phi(\sigma)$	136

SUMMARY

Markov chains are algorithms that can provide critical information from exponentially large sets efficiently through random sampling. These algorithms are ubiquitous across numerous scientific and engineering disciplines, including statistical physics, biology and operations research. In this thesis we solve sampling problems at the interface of theoretical computer science with applied computer science, discrete mathematics, statistical physics, chemistry and economics. A common theme throughout each of these problems is the use of bias.

The first problem we study is biased permutations which arise in the context of self-organizing lists. Here we are interested in the mixing time of a Markov chain \mathcal{M}_{nn} that performs nearest neighbor transpositions in the non-uniform setting. We are given “positively biased” probabilities $\{p_{i,j} \geq 1/2\}$ for all $i < j$ and let $p_{j,i} = 1 - p_{i,j}$. In each step, the chain \mathcal{M}_{nn} chooses two adjacent elements k , and ℓ and exchanges their positions with probability $p_{\ell,k}$. We define two general classes of bias and give the first proofs that the chain is rapidly mixing for both. We also demonstrate that the chain is not always rapidly mixing by constructing an example requiring exponential time to converge to equilibrium.

Next we study rectangular dissections of an $n \times n$ lattice region into rectangles of area n , where $n = 2^k$ for an even integer k . We consider a weighted version of a natural edge flipping Markov chain where, given a parameter $\lambda > 0$, we would like to generate each rectangular dissection (or dyadic tiling) σ with probability proportional to $\lambda^{|\sigma|}$, where $|\sigma|$ is the total edge length. First we look at the restricted case of *dyadic tilings*, where each rectangle is required to have the form $R = [s2^u, (s+1)2^u] \times [t2^v, (t+1)2^v]$, where s, t, u and v are nonnegative integers. Here we show there is a phase transition:

when $\lambda < 1$, the edge-flipping chain mixes in time $O(n^2 \log n)$, and when $\lambda > 1$, the mixing time is $\exp(\Omega(n^2))$. The behavior for general rectangular dissections is more subtle, and we show the chain requires exponential time when $\lambda > 1$ and when $\lambda < 1$.

The last two problems we study arise directly from applications in chemistry and economics. Colloids are binary mixtures of molecules with one type of molecule suspended in another. It is believed that at low density typical configurations will be well-mixed throughout, while at high density they will separate into clusters. We characterize the high and low density phases for a general family of discrete *interfering colloid models* by showing that they exhibit a “clustering property” at high density and not at low density. The clustering property states that there will be a region that has very high area to perimeter ratio and very high density of one type of molecule. A special case is mixtures of squares and diamonds on \mathbb{Z}^2 which correspond to the Ising model at fixed magnetization.

Subsequently, we expanded techniques developed in the context of colloids to give a new rigorous underpinning to the Schelling model, which was proposed in 1971 by economist Thomas Schelling to understand the causes of racial segregation. 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. We show that for any influence function in this class, the dynamics will be rapidly mixing and cities will be integrated if the racial bias is sufficiently low. However when the bias is sufficiently high, we show the dynamics take exponential time to mix and a large cluster of one type will form.

CHAPTER I

INTRODUCTION

Randomized algorithms have transformed the field of algorithms and become an essential tool in the algorithmic toolbox. In many applications algorithms that use randomness are simpler or faster than known deterministic algorithms for the same problem and in some cases they offer the only efficient solution. Markov chains are a type of randomized algorithm that have enabled a lot of progress by using sampling to gain insight into exponentially large sets. In fact, Markov chains are ubiquitous across numerous scientific and engineering disciplines, particularly statistical physics, biology, operations research, and computer science.

Algorithms based on Markov chains perform a random walk on a set of configurations. A chain is designed so that after enough steps of the random walk, it will converge to a useful distribution over the whole set. For example, if a deck of cards is shuffled long enough the cards will be in random order with each permutation roughly equally likely. The number of steps of the random walk that are necessary to come close to the desired distribution is referred to as the convergence time of the chain. Markov chains that converge quickly provide good tools for sampling, approximate counting and many other applications. Researchers have used Markov chains that converge to a desired distribution in time logarithmic in the size of the set of configurations in order to approximately solve hard problems, most notably estimating the permanent of a matrix [51] and estimating the volume of a convex body [31].

It is necessary to bound the convergence time of Markov chain algorithms to ensure that we know what distribution we are sampling from. In the past couple of decades, computer scientists have created numerous new techniques to bound the convergence

time of Markov chains, including coupling and canonical paths and indirect methods such as decomposition and comparison, among others [73]. Although great strides have been made there remain large classes of Markov chains for which we are unable to determine whether they converge in polynomial time or prove this even when we believe they do.

In this thesis we solve various problems where the convergence rate of a natural Markov chain had eluded analysis or problems that arise from applications where a Markov chain was used in a surprising field and rigorous analysis techniques were needed. A common theme throughout these problems is the use of bias. In each of these problems we are interested in sampling from or studying a distribution where certain configurations are more likely than others depending on the value of a bias parameter or parameters. Bias arises in these problems in three main ways. The first way is settings where the unbiased or uniform distribution is well understood, the biased distribution arises naturally in some application and here the techniques that work in the unbiased case break down. Secondly, biased distributions arise naturally in many diverse application areas where understanding their behavior helps us to answer open questions within the application area. In both of these settings our goal is to develop new analytical tools to address the setting with bias and answer open questions within the respective application areas. Thirdly, in some settings we are really interested in understanding the unbiased setting but use the biased setting to gain insight and develop tools with the goal of eventually applying them in the unbiased setting.

Studying bias is not new and the effect of adding bias to a model is varied and often quite interesting. Adding bias to a model can speed up local Markov chains. For example in card shuffling adding a constant bias speeds up the convergence rate of a local Markov chain which makes nearest-neighbor swaps from $O(n^3 \log n)$ to $O(n^2)$. As we shall see, allowing the bias to vary according to the specific cards being

swapped can result in very different behavior and in the most general case is still open. In contrast, for many models adding bias has the beautiful effect of causing a phase transition. A phase transition occurs when a change to a parameter controlling the microscopic interactions such as temperature causes a macroscopic change to the system. For example, water starts to freeze when the temperature reaches 32° Fahrenheit. Often the phase transition in the stationary distribution is mirrored by a phase transition in the convergence times of local Markov chains. Many of the models we discuss in this thesis have this effect when bias is added.

The remainder of the introduction gives related background for understanding the technical contributions of this thesis and their broader importance. In Section 1.1 we describe sampling algorithms and specifically Markov chains more formally and give several relevant examples. In Section 1.2, we describe the phenomena of phase transitions which may arise in biased settings and give examples in several classical models. Finally, in Section 1.3, we discuss additional ways bias can become relevant and briefly describe the four models which we study in the remainder of the thesis.

1.1 Sampling Algorithms

The Russian mathematician Andrey Markov began studying the stochastic processes which would later be called Markov chains in 1906 [60]. Today Markov chains are used to generate random samples, simulate dynamic processes, approximately count and many other uses in countless different application areas including chemistry, operations research, biology and social science. They are also studied by different disciplines with very different motivations including computer science, mathematics and statistical physics. For example, theoretical computer scientists use Markov chains not only to obtain efficient sampling algorithms but also for applications such as approximate counting. Typically computer scientists are interested in designing and analyzing Markov chains whose convergence times are provably logarithmic in

the size of the set of configurations. We make progress on the problems in this thesis partly by applying rigorous Markov chain analysis techniques from computer science, probability and statistical physics to the study of problems that occur in more applied domains. By looking at these problems from these different perspectives we have not only made breakthroughs by solving open problems from more applied disciplines but also developed new analytical techniques that can be applied in other settings.

1.1.1 Markov Chain Basics

Markov chains perform a random walk on a large set of configurations called the *state space*. The random walk is memoryless so the probability of moving from one configuration to another depends only on the current configuration and not on any previous configurations. Formally, a Markov chain is a sequence of random variables X_0, X_1, X_2, \dots where the X_i 's satisfy the *Markov property*, meaning that the conditional distribution of X_i given X_0, \dots, X_{i-1} depends only on X_{i-1} . The X_i 's take on values from a finite set called the state space Ω . In this thesis we only consider ergodic Markov chains (which we define formally in the next section) with a finite state space. The *transition matrix* $\mathcal{P} = \{p_{j,k}\}$ of a Markov chain specifies the probability with which to move from one state in the state space to another. Specifically,

$$p_{j,k} = \Pr(X_{i+1} = k | X_i = j).$$

The probabilities in the transition matrix \mathcal{P} are chosen so that after a sufficient number of steps the chain converges to the desired distribution (referred to as the stationary distribution) over the state space.

The time a Markov chain takes to converge to its stationary distribution π , which we have previously referred to as the convergence time, is measured in terms of the distance between π and \mathcal{P}^t , the distribution at time t . More formally, the *total*

variation distance at time t , from the worst starting state X_0 , is

$$\|\mathcal{P}^t, \pi\|_{tv} = \max_{x \in \Omega} \frac{1}{2} \sum_{y \in \Omega} |\mathcal{P}^t(x, y) - \pi(y)|,$$

where $\mathcal{P}^t(x, y)$ is the t -step transition probability and Ω is the state space. For all $\epsilon > 0$, the *mixing time* τ of \mathcal{M} is defined as

$$\tau(\epsilon) = \min\{t : \|\mathcal{P}^t, \pi\|_{tv} \leq \epsilon\}.$$

Let n be the size of each configuration in Ω . 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 from below by an exponential in n or more specifically $\exp(\Omega(n))$. Next, we present several relevant examples of Markov chains.

1.1.2 Example 1: Card Shuffling Algorithms

An illustrative example is the classical problem of card shuffling. Card shuffling algorithms are used around the world in situations where users are interested in generating a random ordering of a deck of cards in order to play some sort of card game, or other settings where random permutations are required. Numerous card shuffling algorithms have been studied extensively by mathematicians and computer scientists, including the overhand shuffle, pile shuffle and, most famously, the riffle shuffle which was analyzed rigorously by Bayer and Diaconis [5]. Consider the following simple “nearest neighbor” Markov chain for card shuffling. Start with a deck of cards in any order and pick two neighboring cards uniformly at random. With probability $1/2$ flip the two cards and with probability $1/2$ do nothing. Repeat this process, at each step picking a pair of neighboring cards uniformly at random. Eventually the deck will be shuffled. This simple card shuffling algorithm which we refer to as the nearest-neighbor transposition chain is an example of a Markov chain. Here the mixing time is the number of steps until the deck is approximately randomly shuffled. For

the nearest-neighbor transposition chain, Wilson [89] showed that the mixing time is $\theta(n^3 \log n)$ where n is the number of cards in the deck.

Next, consider a biased version of card shuffling that favors putting cards in sorted order. Starting with a deck of n cards labeled from 1 to n (a permutation of the numbers from 1 to n), at each step instead of flipping adjacent cards with probability $1/2$, put them in order with a probability $p \neq 1/2$ and out of order with probability $1 - p$. This biased version of card shuffling was studied by Benjamini et al [6] and shown to converge in time $O(n^2)$. Interesting, the chain converges faster in this biased setting than the uniform setting.

This can be further generalized so that at each step the cards are put in order with a probability that depends on the two specific cards. This setting arises naturally from the Move-Ahead-One list update algorithm for self-organizing lists [34]. In the Move-Ahead-One protocol, elements are chosen according to some underlying distribution and they move up by one in a linked list after each request is served, if possible. Thus, the most frequently requested elements will move toward the front of the list and will require less access time. By studying the mixing time of the relevant Markov chain we can learn about the effectiveness of the Move-Ahead-One protocol under different bias constraints.

In this biased setting many of the standard techniques for analyzing mixing times which work in the uniform case breakdown. Later in this thesis we define two classes of bias that fit within this more general setting and give the first proofs that the nearest-neighbor Markov chain is rapidly mixing for both classes. Additionally and possibly more interestingly, we also demonstrate that the chain is not always rapidly mixing by constructing an example requiring exponential time to converge to equilibrium.

1.1.3 Example 2: Asymmetric Simple Exclusion Processes

The 1-dimensional “asymmetric simple exclusion process” (ASEP) was introduced in 1970 by Spitzer as a model for interacting particles on a lattice [79]. This model has appeared in the study of a wide array of physical phenomena including the transport of macromolecules through thin vessels, traffic flow and surface growth [39]. Start with a 1-dimension lattice (a line) with n locations for particles to occupy and a fixed number of particles. Each location can be occupied by at most one particle. At each step a particle is selected uniformly at random. With probability p the particle moves to the right (as long as this site isn’t occupied) and with probability $q = 1 - p$ it moves to the left (as long as this site isn’t occupied). We can view an ASEP as a staircase walk on the grid (see Figure 1), where an unoccupied site corresponds to a step to the right and an occupied site corresponds to a step down. Here the equivalent process adds or removes a box along the boundary with the appropriate probability (adding with probability p and removing with probability q). If $p = q$ then this chain samples from the uniform distribution where all staircase walks are equally likely. The biased setting when $p \neq q$ shows up in the area of nanotechnology as a model for DNA-based self assembly. Here the Markov chain simulates the assembly of a substrate so understanding the mixing times gives information about the efficiency of the construction. Interestingly, this setting can also be viewed as a special case of biased card shuffling.

Benjamini et al [6] show that for the bias setting when $p \neq q$, the chain converges in $\Theta(n^2)$ time. Greenberg et. al. [43] match these results and generalize the result on ASEPs to sampling biased surfaces in two and higher dimensions in optimal $\Theta(n^d)$ time. Again, as in the case of biased card shuffling, in the biased setting when $p \neq q$ many of the standard techniques for analyzing mixing times which work in the uniform case breakdown. In the process of proving their result for sampling biased surfaces, Greenberg et. al. develop a new version of the standard path coupling theorem

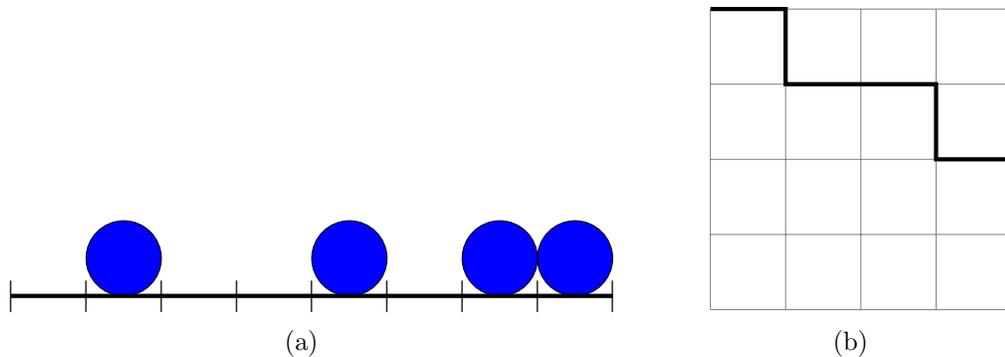


Figure 1: A 1-dimensional asymmetric simple exclusion process (a) and the corresponding staircase walk (b).

which allows exponential distances in order to compensate for the bias. This new technique developed by looking at the chain on staircase walks in the biased setting has subsequently been used for other applications including analyzing a Markov chain for sampling biased dyadic tilings which we discuss in Chapter 4.

1.2 Phase Transitions

Sampling with Markov chains is a common tool used in statistical physics when exact solutions are unavailable. Often sampling is used to observe the existence of a phase transition and in many of these cases the Markov chain algorithms used to sample themselves also undergo a phase transition. A phase transition occurs when a small change to a parameter such as temperature causes a large-scale change to the system. For example, when water is heated and reaches 100°C it begins to boil and the water changes from a liquid to a gas. A phase transition in the stationary distribution corresponding to some physical behavior like the change from a liquid to a gas is often accompanied by a phase transition in the mixing time of local Markov chains. Understanding how fast a simple local Markov chain converges under different settings of a bias parameter can show when a phase transition occurs in the physical model and conversely the physical model can indicate when local Markov chains are efficient.

1.2.1 The Ising Model

A classic model used to study phase transitions is the Ising Model. This model was named after Ernst Ising who began working on the model in the early 1920's [47]. The Ising model was originally introduced as a model of ferromagnetism. When a block of metal such as iron is placed next to a magnet, the metal will become magnetized. The iron's ability to retain its magnetic field varies depending on the temperature and above a critical temperature it undergoes a sudden change and is no longer able to retain its magnetic field at all. The Ising model turns out to be very useful in sampling settings as we will see. In Chapter 5 we will see how it is relevant to models of colloids in chemistry and in Chapter 6 we will see its relevance to modeling segregation.

The Ising model in 2-dimensions is most commonly studied on a $n \times n$ square of the lattice \mathbb{Z}^2 . Each vertex is assigned one of two spins $+$ or $-$. The weight of a particular configuration $\sigma \in \{\pm 1\}^{n^2}$ is

$$\pi(\sigma) = e^{-\beta|E_d(\sigma)|}/Z,$$

where $E_d(\sigma)$ is the set of edges whose endpoints have different spins in σ , $\beta > 0$ is the inverse temperature and $Z = \sum_{\sigma \in \Omega} e^{-\beta|E_d(\sigma)|}$ is the normalizing constant. Glauber dynamics is a Markov chain on Ising configurations that changes one spin at a time using Metropolis probabilities (see Section 2.1) 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 β_c such that when $\beta < \beta_c$, the Glauber dynamics for the Ising model mixes in time polynomial in $|V|$ (the number of vertices in the graph) and when $\beta > \beta_c$, it mixes in exponential time [61, 62]. Moreover, the phase transition in the mixing time is accompanied by a corresponding transition in the stationary distribution of the Markov chain; at low β (high temperature), an average sample from the steady state is “evenly mixed” with regards to the proportions of spins. In contrast, at high β (low temperature), an average sample has long-range order and is likely to have a

large region of predominantly one spin type or *cluster*. In fact, the precise limiting shape of the cluster known as the Wulff shape has been extensively studied using sophisticated techniques (see e.g., [28] and the references therein).

1.2.2 The Hard-Core Model

Another classic model that is believed to exhibit a phase transition is the hard-core lattice gas model. Given a graph G , an *independent set* is a subset S of the vertices such that for any two vertices $x, y \in S$, there are no edges $e = (x, y)$. In graph theory, an independent set is also referred to as a stable set while in statistical physics it is referred to as a *hard-core configuration* and the vertices in the independent set correspond to the placement of particles of gas. The shape of the particles prevent two particles from occupying neighboring vertices. The following natural Markov chain is referred to as the *Glauber dynamics* for independent sets. Start from any independent set; notice that the empty set is an independent set. At each step select a vertex v in the graph uniformly at random. If none of v 's neighbors are in the independent set then put v in the independent set with probability $1/2$ and put v out with probability $1/2$. Here the state space is the set of all independent sets on the graph G and two adjacent configurations differ by adding or removing a vertex from the independent set.

In the statistical physics setting, we are given a bias parameter $\lambda > 0$ that represents the “fugacity” or “activity” of the gas. We are interested in sampling independent sets I from the Gibbs distribution $\pi(I) = \lambda^{|I|}/Z$, where $|I|$ is the size of the independent set and $Z = \sum_{I \in \Omega} \lambda^{|I|}$ is the normalizing constant. If $\lambda > 1$, then larger more dense independent sets are more likely in the stationary distribution, while if $\lambda < 1$, smaller independent sets are favored. This model arises in statistical physics and is referred to as the *hard-core lattice gas model* (see e.g., [37, 76]).

For \mathbb{Z}^2 , it is believed in the statistical physics community that there is a critical

point $\lambda_c \approx 3.79$ where the model undergoes a phase transition [4], but it remains open whether there even is a single critical point. When $\lambda > \lambda_c$ it is believed the model is in the *non-uniqueness regime* and there are multiple equilibrium states while when $\lambda < \lambda_c$ it is in the *uniqueness regime* and there is a single equilibrium state. Informally, this is because for large values of λ , dense independent sets are favored which correspond to configurations where vertices in the independent set lie primarily on either the odd or the even sublattice. Local Markov chains that modify a small number of vertices in each move, including Glauber dynamics, are known to be efficient on \mathbb{Z}^2 at fugacity $\lambda < 2.48$ [85] and inefficient when $\lambda > 5.3646$ [13]. Since local Markov chains only make small changes to move from a configuration primarily on the even sublattice to one primarily on the odd sublattice the chain must go through a configuration with roughly half the vertices on the even sublattice and half on the odd. These configurations must have significantly smaller independent sets making them exponentially unlikely and creating a bottleneck in the state space that results in slow mixing.

1.3 Sampling with Bias

Sampling with bias is a theme connecting the problems in this thesis, focussing on problems at the interface of theoretical computer science with discrete mathematics, applied computer science, statistical physics, chemistry and economics. Our analysis draws us to answer questions that arrive in the specific application domains and develop new analytical techniques that can be applied in the unbiased regime or other settings. A common element underlying all of the solutions is looking at these problems from the perspective of multiple fields and by applying techniques from discrete mathematics, theoretical computer science and statistical physics.

1.3.1 Biased Permutations

The first problem we study in Chapter 3 is biased permutations, or biased card shuffling. In particular, we study the mixing time of the following Markov chain \mathcal{M}_{nn} on permutations of n integers that performs nearest neighbor transpositions in the non-uniform setting, a problem arising in the context of self-organizing lists. We are given “positively biased” probabilities $\{p_{i,j} \geq 1/2\}$ for all $i < j$ and let $p_{j,i} = 1 - p_{i,j}$. Given a permutation of n integers, in each step, the chain \mathcal{M}_{nn} chooses two adjacent elements k , and ℓ and exchanges their positions with probability $p_{\ell,k}$. Here we define two general classes and give the first proofs that the chain is rapidly mixing for both. In the first case we are given input probabilities r_1, \dots, r_{n-1} with $1/2 \leq r_i < 1$ for all i and we set $p_{i,j} = r_i$ for all $i < j$. In the second we are given a binary tree with n leaves labeled $1, \dots, n$ and input probabilities $1/2 \leq q_1, \dots, q_{n-1} < 1$ associated with all of the internal vertices, and we let $p_{i,j} = q_{i \wedge j}$ for all $i < j$. Our bounds on the mixing time of \mathcal{M}_{nn} rely on bijections between permutations, *inversion tables* and *asymmetric simple exclusion processes* (ASEPs) that allow us to express moves of the chain in the context of these other combinatorial families. We also demonstrate that the chain is not always rapidly mixing by constructing an example requiring exponential time to converge to equilibrium. This proof relies on a reduction to biased lattice paths in \mathbb{Z}^2 . This is based on joint work with Bhakta, Streib and Randall which appears in the 2013 ACM-SIAM Symposium on Discrete Algorithms [8].

1.3.2 Rectangular Dissections

The second problem in Chapter 4 has a geometric flavor and involves sampling rectangular dissections. *Rectangular dissections*, or subdivisions of a lattice region into rectangles, arise in VLSI layout, mapping graphs and planning. In Chapter 4 we study equitable dissections, where all rectangles have equal area, in particular the case of partitioning an $n \times n$ square into n rectangles of area n , where $n = 2^k$ for

some even integer k . We consider a biased version of this problem in which we are given a parameter $\lambda > 0$, and would like to generate each rectangular dissection (or dyadic tiling) σ with weight proportional to $\lambda^{|\sigma|}$, where $|\sigma|$ is the total edge length. Varying λ allows us to favor dissections with many long thin or approximately square rectangles. We consider a natural edge-flipping Markov chain and show that there is a phase transition in the case of *dyadic tilings*, where each rectangle is required to have the form $R = [s2^u, (s+1)2^u] \times [t2^v, (t+1)2^v]$, where s, t, u and v are nonnegative integers. When $\lambda < 1$, the edge-flipping chain mixes in time $O(n^2 \log n)$, but when $\lambda > 1$, the mixing time is $\exp(\Omega(n^2))$. Simulations suggest that the chain is fast when $\lambda = 1$, but this case remains open. The behavior for general rectangular dissections is more complicated, including establishing ergodicity of the chain. As in the dyadic case, we show that the edge-flipping Markov chain requires exponential time when $\lambda > 1$. Surprisingly, the chain also requires exponential time when $\lambda < 1$, which we argue using different reasoning. Simulations suggest that the chain is fast at the isolated point $\lambda = 1$. This work with Cannon and Randall appeared in the 2015 ACM-SIAM Symposium on Discrete Algorithms [20].

1.3.3 Clustering in Colloids

The third problem we study in Chapter 5 are colloids or binary mixtures of molecules with one type of molecule suspended in another. Examples include milk and glue. A property observed is that at low density typical configurations will be well-mixed throughout, while at high density they will separate into clusters. Researchers have modeled this with non overlapping shapes in two dimensions and have confirmed this behavior experimentally. However, to even simulate these colloid models required various heuristics, so we cannot conclude they are accurate. A notable exception is a colloid model by Frenkel and Louis [36] who confirmed this behavior on a discrete model. They do this rigorously by mapping non overlapping mixtures of squares and

diamonds to an Ising model with fixed magnetization (meaning the number of positive spins remains fixed), thus inheriting the phase transition from the Ising model [4]. We generalize this special model by defining a class of *interfering binary mixtures*. This class includes well studied models like the Ising model on \mathbb{Z}^2 and independent sets on \mathbb{Z}^2 . Although models in this class do not reduce to the Ising model as does the case studied by Frenkel and Louis, they share enough of the properties that we can extend their results to this class. We define a “clustering property” and prove these models exhibit clustering at high density and not at low density. Informally, the clustering property states that there will be a region that has very high area, very small perimeter, and high density of one type of molecule. Although our proofs use standard techniques they are substantially more complicated because we need to maintain the number of molecules of one type. This work is based on joint work with Streib and Randall for a special case that appeared in the 2012 International Workshop on Randomization and Computation [65], and which we further generalize in this thesis.

1.3.4 Schelling’s Model of Segregation

The final model we study in Chapter 6 is the Schelling segregation model, which attempts to explain how even small racial bias of individuals can cause 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. 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. Although in the original

Schelling model a person is either happy or unhappy while in the Ising analogue, everyone is incrementally more unhappy as more people of the opposite color move into their neighborhood and thus more likely to move. Indeed, the Ising model has been proposed as an alternative to the Schelling model [75, 80, 81]. 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 [82, 90].

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 [69] 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 [12, 11, 22]. 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.

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, closely representing the model Schelling originally proposed.

For both the IBF and TBF classes, 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. This joint work with Bhakta and Randall appeared in the 2014 ACM-SIAM Symposium on Discrete Algorithms [7].

As we explore these examples further we will see how bias can capture features inherent in the models, can be a tool for speeding up algorithms, can cause surprising behavior such as phase transitions, and sometimes can give insight into the unbiased cases. These examples reveal the depth and beauty of randomized algorithms for sampling, but of course that is just our personal bias.

We begin in the next chapter by describing several Markov chain analysis techniques we will use for our proofs. Then, in the remaining four chapters, we explore each of the four examples above in more detail.

CHAPTER II

MARKOV CHAIN ANALYSIS TECHNIQUES

The foundations of Markov chain analysis were developed in probability theory. It was shown that the eigenvalue gap, or spectral gap, of the transition matrix provides a good bound on the mixing time of a Markov chain. Thus, calculating the eigenvalues of the transition matrix is sufficient to determine the mixing time of a Markov chain. However, these techniques miss the idea of scaling and efficiency of analysis. The state spaces and transition matrices of Markov chains studied in theoretical computer science (and in this thesis) are typically exponentially large, so it is infeasible to directly compute eigenvalues and other metrics known to control convergence times. As a consequence, there has been a need to develop indirect methods to allow us to obtain good bounds on the mixing time. We begin by giving some background on Markov chains and then present some of the main tools we use to upper and lower bound the time required by these Markov chain algorithms.

2.1 Markov Chain Basics

We begin by defining some useful properties for Markov chains. A Markov chain is *irreducible* if for any two states σ, τ there exists an integer $t_{\sigma, \tau}$ such that

$$\Pr(X_{t_{\sigma, \tau}} = \tau | X_0 = \sigma) > 0.$$

Informally, this implies that from any state we can reach any other state. A Markov chain is *aperiodic* if for every state σ there exists a t such that for all $t' \geq t$,

$$\Pr(X_{t'} = \sigma | X_0 = \sigma) > 0.$$

Markov chains that are not aperiodic can easily be made aperiodic by adding a small “self-loop” probability at each state. In other words, for each state σ ensure that

$p_{\sigma,\sigma} > 0$. A *lazy* chain has self-loop probabilities of $1/2$ everywhere ($p_{i,i} = 1/2$ for all states i). A Markov chain is *ergodic* if it is both irreducible and aperiodic.

Lemma 2.1.1: Any finite, ergodic Markov chain converges to a unique stationary distribution π . Specifically, for all $i, j \in \Omega$, we have that

$$\lim_{t \rightarrow \infty} \mathcal{P}^t(i, j) = \pi(j).$$

Additionally, for an ergodic Markov chain with transition probabilities \mathcal{P} , if some assignment of probabilities π satisfies the *detailed balance condition*

$$\pi(\sigma)\mathcal{P}(\sigma, \tau) = \pi(\tau)\mathcal{P}(\tau, \sigma)$$

for every $\sigma, \tau \in \Omega$ and $\sum_{i \in \Omega} \pi(i) = 1$, then π is the stationary distribution of the Markov chain (see e.g., [55]). If a Markov chain M satisfies the detailed balance condition for some distribution π then it is *reversible*.

Given a desired stationary distribution π , the Metropolis-Hastings algorithm [64] provides a simple way to define the transition probabilities of a chain in order to guarantee that it converges to the desired distribution π . Given a configuration $\sigma \in \Omega$, define a *neighbor* of σ to be any configuration $\tau \in \Omega$ such that $\Pr(X_{t+1} = \tau | X_t = \sigma) > 0$. Given a state space Ω , let Δ be the maximum number of neighbors of any configuration in Ω .

The Metropolis-Hasting Algorithm

Starting at any configuration σ , repeat:

- Pick a neighbor τ of σ uniformly with probability $\frac{1}{2\Delta}$.
- Move to τ with probability $\min(1, \frac{\pi(\tau)}{\pi(\sigma)})$.
- With the remaining probability, stay at σ .

Next we introduce some of the analysis tools we will use in the remainder of the thesis to give upper and lower bounds on the mixing times of Markov chains.

2.2 Coupling and Path Coupling

One of the most common computer science techniques for showing a Markov chain is fast mixing is coupling. A *coupling* of a Markov chain \mathcal{M} is a joint Markov process on $\Omega \times \Omega$ such that the marginals each agree with \mathcal{M} and, once the two coordinates coalesce, they move in unison. More formally, *coupling* is a stochastic process $(X_t, Y_t)_{t=0}^\infty$ on $\Omega \times \Omega$ with the properties:

1. Each of the processes X_t and Y_t , viewed in isolation, is a faithful copy of \mathcal{M} .

Specifically, for all $x, y \in \Omega$,

$$\Pr[X_{t+1} = y | X_t = x] = \mathcal{P}(x, y) = \Pr[Y_{t+1} = y | Y_t = x].$$

2. If $X_t = Y_t$, then $X_{t+1} = Y_{t+1}$.

The coupling theorem bounds the mixing time in terms of the expected time of coalescence of any coupling. For initial states x, y let $T^{x,y} = \min\{t : X_t = Y_t | X_0 = x, Y_0 = y\}$, and define the *coupling time* to be $T = \max_{x,y} \mathbb{E}[T^{x,y}]$. The following well-known (see e.g. [1]) result relates the mixing time to the coupling time.

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

Path coupling arguments, introduced by Bubley and Dyer [18], are a convenient way of bounding the mixing time of a Markov chain by considering only a subset U of the joint state space $\Omega \times \Omega$ of a coupling. By considering an appropriate metric ϕ on Ω , proving that the two marginal chains, if in a joint configuration in subset U , get no farther away in expectation after one iteration is sufficient to show that \mathcal{M} is rapidly mixing. We will use the following path coupling theorem due to Dyer and Greenhill [32].

Theorem 2.2.2 (Dyer and Greenhill): Let ϕ be an integer-valued metric 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 $\sum_{i=0}^{r-1} \phi(Z_i, Z_{i+1}) = \phi(X_t, Y_t)$. Let \mathcal{M} be a Markov chain on state space Ω and let (X_t, Y_t) be a coupling of \mathcal{M} . Suppose there exists a $\beta \leq 1$ such that

$$\mathbb{E}[\phi(X_{t+1}, Y_{t+1})] \leq \beta \phi(X_t, Y_t),$$

for all $X_t, Y_t \in U$.

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

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

2. If $\beta = 1$, (i.e., $\mathbb{E}[\Delta\phi(X_{t+1}, Y_{t+1})] \leq 0$, for all $X_t, Y_t \in U$), 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 $X_t \neq Y_t$. Then the mixing time of \mathcal{M} satisfies

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

Greenberg et. al. [43] introduced a version of the path coupling theorem which is particularly useful because the values taken by the metric ϕ can be exponential in n , yet as long as the distance between two chains in a coupling decreases by some constant multiplicative factor with each move of the joint Markov process, the Markov chain is provably rapidly mixing.

Theorem 2.2.3 (Greenberg, et. al.): Let $\phi : \Omega \times \Omega \rightarrow \mathbb{R}^+ \cup \{0\}$ be a metric that takes on 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 $\sum_{i=0}^{r-1} \phi(Z_i, Z_{i+1}) = \phi(X_t, Y_t)$.

Let \mathcal{M} be a lazy Markov chain on state space Ω and let (X_t, Y_t) be a coupling of \mathcal{M} . Suppose there exists a $\beta < 1$ such that,

$$\mathbb{E}[\phi(X_{t+1}, Y_{t+1})] \leq \beta \phi(X_t, Y_t),$$

for all $X_t, Y_t \in U$.

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

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

2. If there exists $\kappa, \eta \in (0, 1)$ such that $\Pr[|\phi(X_{t+1}, Y_{t+1}) - \phi(X_t, Y_t)| \geq \eta\phi(X_t, Y_t)] \geq \kappa$ for all t provided that $X_t \neq Y_t$, then the mixing time of \mathcal{M} satisfies

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

2.3 Comparison

The comparison method due to Diaconis and Saloff-Coste [25] can be used to infer the mixing time of one chain given the mixing time of another, similar chain. If \mathcal{P}' and \mathcal{P} are the transition matrices of two reversible Markov chains on the same state space Ω with the same stationary distribution π , the comparison method [25, 71] allows us to relate the mixing times of these two chains. We will assume that we know the mixing time of \mathcal{P}' and are trying to determine the mixing time of \mathcal{P} . Let

$$E(\mathcal{P}) = \{(\sigma, \beta) : \mathcal{P}(\sigma, \beta) > 0\}$$

and

$$E(\mathcal{P}') = \{(\sigma, \beta) : \mathcal{P}'(\sigma, \beta) > 0\}$$

denote the sets of edges of the two graphs, viewed as directed graphs. For each σ, β with $\mathcal{P}'(\sigma, \beta) > 0$, define a path $\gamma_{\sigma\beta}$ using a sequence of states $\sigma = \sigma_0, \sigma_1, \dots, \sigma_k = \beta$ with $\mathcal{P}(\sigma_i, \sigma_{i+1}) > 0$, and let $|\gamma_{\sigma\beta}|$ denote the length of the path. Let

$$\Gamma(v, \omega) = \{(\sigma, \beta) \in E(\mathcal{P}') : (v, \omega) \in \gamma_{\sigma\beta}\}$$

be the set of paths that use the transition (v, ω) of \mathcal{P} . Finally, let $\pi_* = \min_{\rho \in \Omega} \pi(\rho)$ and define

$$A = \max_{(v, \omega) \in E(\mathcal{P})} \frac{1}{\pi(v)\mathcal{P}(v, \omega)} \sum_{\Gamma(v, \omega)} |\gamma_{\sigma\beta}| \pi(\sigma) \mathcal{P}'(\sigma, \beta).$$

We use the following formulation of the comparison method due to Randall and Tetali [71].

Theorem 2.3.1 (Randall and Tetali): Given two reversible and lazy Markov chains each with stationary distribution π , transition matrices \mathcal{P} and \mathcal{P}' and mixing times $\tau(\epsilon)$ and $\tau'(\epsilon)$, respectively. Define A and π_* as above, then for $0 < \epsilon < 1$, we have

$$\tau(\epsilon) \leq \frac{4 \log(1/(\epsilon\pi_*))}{\log(1/2\epsilon)} A\tau'(\epsilon).$$

2.4 Conductance

For several of our results we show that a Markov chain is slow mixing by demonstrating that the state space contains a bottleneck that requires exponential expected time to cross. We use the bottleneck to bound the conductance of the Markov chain. Formally, the *conductance* of an ergodic Markov chain \mathcal{M} with distribution π and transition matrix \mathcal{P} , is

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

We can then use the bound on conductance to bound the mixing time using the following theorem that relates the conductance and mixing time (see, e.g., [49]).

Theorem 2.4.1: For any Markov chain \mathcal{M} with conductance $\Phi_{\mathcal{M}}$ and mixing time $\tau(\epsilon)$, for all $\epsilon > 0$ the mixing time of \mathcal{M} on state space Ω satisfies

$$\tau(\epsilon) \geq \left(\frac{1}{4\Phi_{\mathcal{M}}} - \frac{1}{2} \right) \log \left(\frac{1}{2\epsilon} \right).$$

CHAPTER III

SELF-ORGANIZING LISTS AND BIASED PERMUTATIONS

We begin with work on biased card shuffling which was introduced in Section 1.1.2. Specifically, in this chapter we study the mixing time of a Markov chain \mathcal{M}_{nn} on permutations that performs nearest neighbor transpositions in the non-uniform setting, a problem arising in the context of self-organizing lists. We are given “positively biased” probabilities $\{p_{i,j} \geq 1/2\}$ for all $i < j$ and let $p_{j,i} = 1 - p_{i,j}$. In each step, the chain \mathcal{M}_{nn} chooses two adjacent elements k and ℓ and exchanges their positions with probability $p_{\ell,k}$. Here we define two general classes and give the first proofs that the chain is rapidly mixing for both. In the first case we are given constants r_1, \dots, r_{n-1} with $1/2 \leq r_i \leq 1$ for all i and we set $p_{i,j} = r_i$ for all $i < j$. In the second we are given a binary tree with n leaves labeled $1, \dots, n$ and constants q_1, \dots, q_{n-1} associated with all of the internal vertices, and we let $p_{i,j} = q_{i \wedge j}$ for all $i < j$. Our bounds on the mixing time of \mathcal{M}_{nn} rely on bijections between permutations, *inversion tables* and *asymmetric simple exclusion processes* (ASEPs) that allow us to express moves of the chain in the context of these other combinatorial families. We also demonstrate that the chain is not always rapidly mixing by constructing an example requiring exponential time to converge to equilibrium. This proof relies on a reduction to biased lattice paths in \mathbb{Z}^2 .

3.1 Biased Permutations

Sampling from the permutation group S_n is one of the most fundamental problems in probability theory. A natural Markov chain that has been studied extensively

is a symmetric chain, \mathcal{M}_{nn} , that iteratively makes nearest neighbor transpositions on adjacent elements. We are given a set of input probabilities $\mathbf{P} = \{p_{i,j}\}$ for all $1 \leq i, j \leq n$ with $p_{i,j} = 1 - p_{j,i}$. At each step, the Markov chain \mathcal{M}_{nn} uniformly chooses a pair of adjacent elements, i and j , and puts i ahead of j with probability $p_{i,j}$, and j ahead of i with probability $p_{j,i} = 1 - p_{i,j}$.

The problem of biased permutations arises naturally from the Move-Ahead-One (MA1) list update algorithm and was considered by Fill [34, 35]. In the MA1 protocol, elements are chosen according to some underlying distribution and they move up by one in a linked list after each request is serviced, if possible. Thus, the most frequently requested elements will move toward the front of the list and will require less access time. If we consider a pair of adjacent elements i and j , the probability of performing a transposition that moves i ahead of j is proportional to i 's request frequency, and similarly the probability of moving j ahead of i is proportional to j 's frequency, so the transposition rates vary depending on i and j and we are always more likely to put things in order (of their request frequencies) than out of order. Fill asked for which $\mathbf{P} = \{p_{i,j}\}$ the chain is rapidly mixing.

Despite the simplicity of the model, the mixing times of only a few special cases are known. After a series of papers [26, 24], Wilson [89] showed that in the unbiased case when $p_{i,j} = 1/2$ for all i, j the mixing time is $\Theta(n^3 \log n)$, with upper and lower bounds within a factor of two. Subsequently Benjamini et al. [6] considered a constant bias version of this chain, where we are given a fixed parameter $0 \leq p \leq 1$ such that $p \neq 1/2$ and $p_{i,j} = p$ for all $i < j$ and $p_{i,j} = 1 - p$ for $i > j$. They relate this biased shuffling Markov chain to a chain on an asymmetric simple exclusion process (ASEP) and showed that they both converge in $\Theta(n^2)$ time. These bounds were matched by Greenberg et al. [43] who also generalized the result on ASEP's to sampling biased surfaces in two and higher dimensions in optimal $\Theta(n^d)$ time. Note that when the bias is a constant for all $i < j$ there are other methods for sampling from the stationary

distribution, but studying the Markov chain \mathcal{M}_{nn} is of independent interest, partly because of the connection to ASEPs and other combinatorial structures. Finally, we also have polynomial bounds on the mixing time when each of the $p_{i,j}$ for $i < j$ is equal to $1/2$ or 1 ; in this case we are sampling linear extensions of a partial order over the set $\{1 \dots n\}$, and the chain \mathcal{M}_{nn} was shown by Bublely and Dyer [17] to mix in $O(n^3 \log n)$ time.

It is easy to see that \mathcal{M}_{nn} is not always rapidly mixing. Consider, for example, n elements $1 \dots n$ such that $p_{i,j} = 1$ for all $1 \leq i < j \leq n - 1$, $p_{n,i} = .9$ for $i \leq n/2$ and $p_{i,n} = .9$ for $i > n/2$. Then the first $n - 1$ elements will stay in order once they become ordered. All n places where the last element can be placed have nonzero stationary probability, but the configurations that have this last element at the beginning or end of the permutation will have exponentially larger stationary probability than the configuration that has this last element near the middle of the permutation. This defines an exponentially small cut in the state space and we can conclude that the nearest neighbor transposition chain must be slowly mixing for this choice of \mathbf{P} .

To avoid such situations, we restrict our attention to the *positively biased* setting where for all $i < j$, we have $1/2 \leq p_{i,j} \leq 1$. Note that any transposition that puts elements in the proper order has probability at least $1/2$, so starting at any permutation, we can always perform a series of transpositions to move to the ordered permutation $1, 2, \dots, n$ without ever decreasing the stationary probability. It is also worth noting that the classes for which the chain is known to mix rapidly are all positively biased. Fill [34, 35] conjectured that when \mathbf{P} is positively biased and also satisfies a monotonicity condition where $p_{i,j} \leq p_{i,j+1}$ and $p_{i,j} \geq p_{i+1,j}$ for all $1 \leq i < j \leq n$, then the chain is always rapidly mixing. In fact, he conjectured that the spectral gap is minimized when $p_{i,j} = 1/2$ for all i, j , a problem he refers to as the “gap problem.” Fill verified the conjecture for $n = 4$ and gave experimental evidence for slightly larger n .

In this chapter, we make progress on the question of determining for which values of \mathbf{P} the chain \mathcal{M}_{nn} is rapidly mixing. First, we show that restricting \mathbf{P} to be positively biased is not sufficient to guarantee fast convergence to equilibrium. Our example uses a reduction to ASEPs and biased lattice paths. The construction is motivated by models in statistical physics that exhibit a phase transition arising from a “disordered phase” of high entropy and low energy, an “ordered phase” of high energy and low entropy, and a bad cut separating them that is both low energy and entropy. We note that this example does not satisfy the monotonicity condition of Fill, thus leaving his conjecture open, but does give insight into why bounding the mixing rate of the chain in more general settings has proven quite challenging.

In addition, we identify two new classes of input probabilities \mathbf{P} for which we can prove that the chain is rapidly mixing. It is important to note that these classes are not necessarily monotone. The first, which we refer to as “Choose Your Weapon,” we are given a set of input parameters $1/2 \leq r_1, \dots, r_{n-1} < 1$ representing each player’s ability to win a duel with his or her weapon of choice. When a pair of neighboring players are chosen to compete, the dominant player gets to choose the weapon, thus determining his or her probability of winning the match. In other words, we set $p_{i,j} = r_i$ when $i < j$. We show that the nearest neighbor transposition chain \mathcal{M}_{nn} is rapidly mixing for any choice of $\{r_i\}$. The second class, which we refer to as “League Hierarchies,” is defined by a binary tree with n leaves labeled $1, \dots, n$. We are given q_1, \dots, q_{n-1} with $1/2 \leq q_i < 1$ for all i , each associated with a distinct internal node in the tree. We then set $p_{i,j} = q_{i \wedge j}$ for all $i < j$. We imagine that the two subtrees under the root represent two different leagues, where each player from one league have a fixed advantage over each player from the other. Moreover, each league is subdivided into two sub-leagues, and each player from one has a fixed advantage over a player from the other, and so on recursively. We prove that there is a Markov chain based on transpositions (not necessarily nearest neighbors) that is always rapidly mixing for

positively biased \mathbf{P} defined as League Hierarchies. Moreover, if the $\{q_i\}$ additionally satisfy “weak monotonicity” (i.e., $p_{i,j} \leq p_{i,j+1}$ if $j > i$) then the nearest neighbor chain \mathcal{M}_{nn} is also rapidly mixing. Note that both the choose-your-weapon and the tree-hierarchy classes are generalizations of the constant bias setting, which can be seen by taking all parameters r_i or q_i to be constant.

Our proofs rely on various combinatorial representations of permutations, including Inversion Tables and families of ASEPs. In each case there is a natural Markov chain based on (non necessarily adjacent) transpositions for which we can more easily bound the mixing time in the new context. We then interpret these new moves in terms of the original permutations in order to derive bounds on the mixing rate of the nearest neighbor transposition via comparison methods. These new chains that allow additional, but not necessarily all, transpositions are also interesting in the context of permutations and these related combinatorial families. Finally, we note that the choose-your-weapon class is actually a special case of the league-hierarchy class, but the proofs bounding the mixing rate of \mathcal{M}_{nn} are simpler and yield faster mixing times, so we present these proofs separately in Sections 3.4 and 3.5.

3.2 Formalizing the Markov Chains

We begin by formalizing the nearest neighbor and transposition Markov chains. Let $\Omega = S_n$ be the set of all permutations $\sigma = (\sigma(1), \dots, \sigma(n))$ of n integers. We consider Markov chains on Ω whose transitions transpose two elements of the permutation. Recall we are given a set \mathbf{P} , consisting of $p_{i,j} \in [0, 1]$ for each $1 \leq i \neq j \leq n$, where $p_{j,i} = 1 - p_{i,j}$. In this chapter we only consider sets \mathbf{P} which are positively biased and bounded away from 1. Specifically, for any $i < j$, $1/2 \leq p_{i,j} < 1$. The Markov chain \mathcal{M}_{nn} will sample elements from Ω as follows.

The Nearest Neighbor Markov chain \mathcal{M}_{nn}

Starting at any permutation σ_0 , repeat:

- At time t , select index $i \in [n - 1]$ uniformly at random (u.a.r.).
 - Exchange the elements $\sigma_t(i)$ and $\sigma_t(i + 1)$ with probability $p_{\sigma_t(i+1),\sigma_t(i)}$ to obtain σ_{t+1} .
 - With probability $p_{\sigma_t(i),\sigma_t(i+1)}$ do nothing so that $\sigma_{t+1} = \sigma_t$.

The chain \mathcal{M}_{nn} connects the state space, since every permutation σ can move to the ordered permutation $(1, 2, \dots, n)$ (and back) using the bubble sort algorithm. Since \mathcal{M}_{nn} is also aperiodic, this implies that \mathcal{M}_{nn} is ergodic. It is easy to see that for \mathcal{M}_{nn} , the distribution

$$\pi(\sigma) = \left(\prod_{i < j: \sigma(i) < \sigma(j)} \frac{p_{i,j}}{p_{j,i}} \right) Z^{-1},$$

where Z is the normalizing constant $\sum_{\sigma \in \Omega} \left(\prod_{i < j: \sigma(i) < \sigma(j)} \frac{p_{i,j}}{p_{j,i}} \right)$, satisfies detailed balance, and is thus the stationary distribution (see, e.g., [78]).

Next, we define the Markov chain \mathcal{M}_{tr} which can make any transposition at each step, while maintaining the stationary distribution π . The transition probabilities of \mathcal{M}_{tr} can be quite complicated, since swapping two distant elements in the permutation consists of many transitions of \mathcal{M}_{nn} , each with different probabilities.

The All-Transposition Markov chain \mathcal{M}_{tr}

Starting at any permutation σ_0 , repeat:

- At time t , select indices $i, j \in [n - 1]$ u.a.r.
 - Swap the elements $\sigma_t(i), \sigma_t(j)$ with probability $p_{\sigma_t(j),\sigma_t(i)} \prod_{i < k < j} \frac{p_{\sigma_t(j,k)}}{p_{\sigma_t(i,k)}}$ to obtain σ_{t+1} .
 - Otherwise, do nothing so that $\sigma_{t+1} = \sigma_t$.

The transition probabilities of \mathcal{M}_{tr} are chosen so that the distribution π will satisfy detailed balance, and therefore be the stationary distribution of \mathcal{M}_{tr} . In the

following sections, we will introduce two other Markov chains whose transitions are a subset of those of \mathcal{M}_{tr} , but for which we can describe the transition probabilities succinctly.

3.3 An Example that is Slowly Mixing

We begin by presenting an example that is positively biased yet takes exponential time to mix. In particular, we show that there are positively biased \mathbf{P} for which the chains \mathcal{M}_{nn} and even \mathcal{M}_{tr} require exponential time to converge to equilibrium. The key component used in the construction of these \mathbf{P} values is an example of slow mixing which was discovered by Pascoe and Randall [68] in the context of tile-based self-assembly models and is of independent interest in this setting.¹ We use a mapping from biased permutations to multiple particle ASEP configurations with n zeros and n ones. The resulting ASEPs are in bijection with staircase walks [43], which are sequences of n ones and n zeros, that correspond to paths on the Cartesian lattice from $(0, n)$ to $(n, 0)$, where each 1 represents a step to the right and each 0 represents a step down (see Figure 2b). In [43], Greenberg et al. examined the Markov chain which attempts to swap a neighboring $(0, 1)$ pair, which equivalently adds or removes a unit square from the region below the walk, with probability depending on the position of that unit square. Each unit square (x, y) is assigned a bias $\lambda_{x,y}$ and the probability of each staircase walk w is proportional to the product of the bias assigned to each square below the walk. More formally, the stationary weight of a walk w is $\pi(w) = Z^{-1} \prod_{xy < w} \lambda_{x,y}$, where $xy < w$ whenever the square at (x, y) lies underneath the walk w and Z is the normalizing constant. For example, the walk w

¹In [8] we give a simpler example also based on work by Pascoe and Randall in the context of tile-based self-assembly models. However, this example requires $p_{i,j} = 1$ for $i < j \leq n$ or $n < i < j$ (for permutations on $2n$ numbers), ensuring that once the elements $1, 2, \dots, n$ get in order, they stay in order (and similarly for the elements $n + 1, n + 2, \dots, 2n$). The example presented here, although more complex, allows all $p_{i,j}$ values to be bounded away from 1.

in Figure 3(a) has weight

$$\pi(w) = Z^{-1} \lambda_{1,5} \lambda_{1,6} \lambda_{2,6} \lambda_{3,6} \lambda_{1,7} \lambda_{2,7} \lambda_{3,7} \lambda_{4,7} \lambda_{1,8} \lambda_{2,8} \lambda_{3,8} \lambda_{4,8}$$

We show that there are settings of the $\{\lambda_{x,y}\}$ which cause the chain to be slowly mixing from any starting configuration (or walk). Consider the $\{\lambda_{x,y}\}$ given in Figure 2a where roughly all squares have bias $1/2$ except those in the upper right corner which have bias 1. The idea is that if you start at a configuration that never touches the upper right corner (for example, the walk from $(0, n)$ to $(0, 0)$ to $(n, 0)$) you can expect to stay within \sqrt{n} of the diagonal from $(0, n)$ to $(n, 0)$ and never discover the walks with high weight that go through the upper right corner. More specifically, with these bias values, we show that at stationarity the most likely configurations will be concentrated near the diagonal from $(0, n)$ to $(n, 0)$ (the high entropy, low energy states) or they will extend close to the point (n, n) (the high energy, low entropy states) but it will be unlikely to move between these sets of states because there is a bottleneck that has both low energy and low entropy. Finally, we give a map from biased permutations to biased lattice paths to produce a positively biased set of probabilities \mathbf{P} for which \mathcal{M}_{nn} also requires exponential time to mix.

Suppose, for ease of notation, that we are sampling permutations with $2n$ entries (having an odd number of elements will not cause qualitatively different behavior). We begin by setting $p_{i,j} = 1$ when $i < j \leq n$ or $n < i < j$, ensuring that once the elements $1, 2, \dots, n$ get in order, they stay in order (and similarly for the elements $n+1, n+2, \dots, 2n$). Since the smallest (largest) n elements of the biased permutation never change order once they get put in increasing order, permutations with these elements out of order have zero stationary probability. Hence, we can represent the smallest n numbers as ones and the largest n numbers as zeros, assuming that within each class the elements are in increasing order. Thus, in this setting, we have a bijection between staircase walks and permutations and it suffices to show a set of $\{\lambda_{x,y}\}$, staircase walk bias values, for which the Markov chain on staircase walks is

slow. We do this first and then show how we can extend this to the more interesting case where we set $p_{i,j} = 1 - \beta$ when $i < j \leq n$ or $n < i < j$, and we have a more complex map between permutations and staircase walks.

3.3.1 Slow Mixing of Biased Staircase Walks

First we define a set of $\{\lambda_{x,y}\}$, staircase walk bias values, for which the Markov chain on staircase walks is slow. Let $M = 2n^{2/3}$, $0 < \delta < \frac{1}{2}$ be a constant, $\epsilon = \beta = 1/n^2$. We will define the $\{\lambda_{x,y}\}$ as follows (see Figure 2a):

$$\lambda_{x,y} = \begin{cases} 1 - \delta & \text{if } x + y > n + M; \\ \frac{1}{2} + \epsilon & \text{otherwise.} \end{cases} \quad (3.3.1)$$

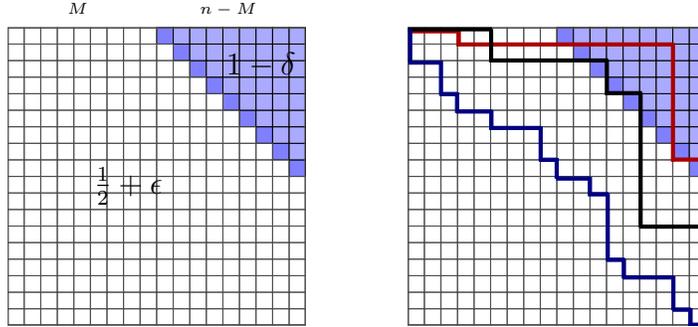


Figure 2: An example of fluctuating bias with exponential mixing time and staircase walks in S_1, S_2 , and S_3 .

We identify sets W_1, W_2, W_3 such that $\pi(W_2)$ is exponentially smaller than both $\pi(W_1)$ and $\pi(W_3)$, but to get between W_1 and W_3 , the Markov chain on staircase walks must pass through W_2 , the cut. We prove that for the set $\{\lambda_{x,y}\}$ defined above, the Markov chain on staircase walks has a bad cut. Then we use the *conductance* (see Section 2.4) to prove it is slowly mixing.

For a staircase walk w , define the *height of w_i* as $\sum_{j \leq i} w_j$, and let $\max(w)$ be the maximum height of w_i over all $1 \leq i \leq 2n$. Let W_1 be the set of walks w such that $\max(w) < n + M$, W_2 the set of walks such that $\max(w) = n + M$, and W_3 the set of walks such that $\max(w) > n + M$. That is, W_1 is the set of walks that never reach the

dark blue diagonal in Figure 2b, W_2 is the set whose maximum peak is on the dark blue line, and W_3 is the set which crosses that line and contains squares in the light blue triangle. Define $\gamma = (1/2 + \epsilon)/(1/2 - \epsilon)$, which is the ratio of two configurations that differ by swapping a $(0, 1)$ pair with probability $\frac{1}{2} + \epsilon$. First we notice that since the maximal staircase walk is in W_3 ,

$$\pi(W_3) \geq \frac{1}{Z} \gamma^{n^2 - \frac{(n-M)^2}{2}} (\delta^{-1} - 1)^{\frac{(n-M)^2}{2}}.$$

Also, $\pi(W_1) = \frac{1}{Z} \sum_{w \in W_1} \gamma^{A(w)}$, where $A(w)$ is the number of unit squares below w . We have that

$$\begin{aligned} \pi(W_1) &= \frac{1}{Z} \sum_{w \in W_1} \gamma^{A(w)} \\ &\leq \frac{1}{Z} \sum_{w \in W_1} \gamma^{n^2 - \frac{(n-M)^2}{2}} \\ &\leq \frac{1}{Z} \binom{2n}{n} \gamma^{n^2 - \frac{(n-M)^2}{2}} \\ &\leq \frac{1}{Z} (2e)^n \gamma^{n^2 - \frac{(n-M)^2}{2}} \\ &\leq \frac{1}{Z} \gamma^{n^2 - \frac{(n-M)^2}{2}} (\delta^{-1} - 1)^{\frac{(n-M)^2}{2}} \\ &\leq \pi(W_3) \end{aligned}$$

for large enough n , since $1/\delta > 2$ is a constant. Hence $\pi(W_1) \leq \pi(W_3)$. We will show that $\pi(W_2)$ is exponentially small in comparison to $\pi(W_1)$ (and hence also to $\pi(W_3)$).

$$\pi(W_2) = \frac{1}{Z} \sum_{\sigma \in W_2} \gamma^{A(\sigma)} \leq \frac{\gamma^{n^2} |W_2|}{Z}.$$

We bound $|W_2|$ as follows. The unbiased Markov chain is equivalent to a simple random walk $w_{2n} = X_1 + X_2 + \dots + X_{2n} = 0$, where $X_i \in \{+1, -1\}$ and where a $+1$ represents a step to the right and a -1 represents a step down. We call this random walk *tethered* since it is required to end at 0 after $2n$ steps. Compare walk w_{2n} with

the untethered simple random walk $w'_{2n} = X'_1 + X'_2 + \dots + X'_{2n}$.

$$\begin{aligned}
P\left(\max_{1 \leq t \leq 2n} w_t \geq M\right) &= P\left(\max_{1 \leq t \leq 2n} w'_t \geq M \mid w'_{2n} = 0\right) \\
&= \frac{P(\max_{1 \leq t \leq 2n} w'_t \geq M)}{P(w'_{2n} = 0)} \\
&= \frac{2^{2n}}{\binom{2n}{n}} P\left(\max_{1 \leq t \leq 2n} w'_t \geq M\right) \\
&\approx \sqrt{\pi n} P\left(\max_{1 \leq t \leq 2n} w'_t \geq M\right).
\end{aligned}$$

Since the $\{X'_i\}$ are independent, we can use Chernoff bounds to see that

$$P\left(\max_{1 \leq t \leq 2n} w'_t \geq M\right) \leq 2nP(w'_{2n} \geq M) \leq 2ne^{-\frac{M^2}{2n}}.$$

Together these show that

$$P\left(\max_{1 \leq t \leq 2n} w_t \geq M\right) < e^{-n^{1/3}},$$

by definition of M . Therefore we have

$$\begin{aligned}
\pi(W_2) &\leq \frac{1}{Z} \gamma^{n^2} |W_2| \leq \frac{\gamma^{n^2}}{Z} \binom{2n}{n} e^{-n^{1/3}} \\
&\leq \frac{1}{Z} \binom{2n}{n} e^{-n^{1/3}+1} (1 - e^{-n^{1/3}}) \\
&\leq \frac{1}{Z} |W_1| e^{-n^{1/3}+1} \\
&\leq e^{-n^{1/3}+1} \pi(W_1),
\end{aligned}$$

as desired. Thus, $\pi(W_2)$ is exponentially smaller than $\pi(W_1)$ for every value of δ and the conductance satisfies

$$\begin{aligned}
\Phi &\leq \sum_{x \in W_1} \frac{\pi(x)}{\pi(W_1)} \sum_{y \in W_2} P(x, y) \\
&\leq \sum_{x \in S_1} \frac{\pi(x)}{\pi(W_1)} \pi(W_2) \\
&\leq e^{-n^{1/3}+1} \pi(W_1) \leq \frac{e^{-n^{1/3}+1}}{2}.
\end{aligned}$$

Using Theorem 2.4.1, we can use the bound on conductance to bound the mixing time.

3.3.2 Extending Slow Mixing of Staircase Walks to Biased Permutations

Next, we extend the above proof to the more general case where we set $p_{i,j} = 1 - \beta$ when $i < j \leq n$ or $n < i < j$, and we have a more complex map between permutations and staircase walks. Again, let $M = 2n^{2/3}$, $0 < \delta < \frac{1}{2}$ be a constant, $\epsilon = \beta = 1/n^2$. For $i < j \leq n$ or $n < i < j$, $p_{i,j} = 1 - \beta$, ensuring that the elements $1, 2, \dots, n$ are likely to be in order (and similarly for the elements $n + 1, n + 2, \dots, 2n$). The remaining $p_{i,j}$ values are defined as follows:

$$p_{i,j} = \begin{cases} 1 - \beta & i < j \leq n \text{ or } n < i < j; \\ 1 - \delta & \text{if } i + 2n - j + 1 \geq n + M; \\ \frac{1}{2} + \epsilon & \text{otherwise.} \end{cases} \quad (3.3.2)$$

We again identify sets S_1, S_2, S_3 such that $\pi(S_2)$ is exponentially smaller than both $\pi(S_1)$ and $\pi(S_3)$, but to get between S_1 and S_3 , \mathcal{M}_{nn} and \mathcal{M}_{tr} must pass through S_2 , the cut. In order to do this, we will define a map from permutations to staircase walks by representing the smallest n numbers as ones and the largest n numbers as zeros. More precisely, given a permutation σ , let $f(\sigma)$ be a sequence of ones and zeros, where $f(\sigma)_i = 1$ if $i \leq n$ and 0 otherwise. For example, the permutation $\sigma = (5, 1, 7, 8, 4, 3, 6, 2)$ maps to $f(\sigma) = (0, 1, 0, 0, 1, 1, 0, 1)$. If the first n and last n elements were always in order then, the probability that an adjacent 1 and a 0 swap in \mathcal{M}_{nn} depends on how many ones and zeros occur before that point in the permutation. Specifically, if element i is a 0 and element $i + 1$ is a 1 then we swap them with probability $\frac{1}{2} + \epsilon$ if the number of ones occurring before position x plus the number of zeros occurring after $i + 1$ is less than $n + M - 1$. Otherwise, they swap with probability $1 - \delta$. Equivalently, the probability of adding a unit square at position $v = (x, y)$ is $\frac{1}{2} + \epsilon$ if $x + y \leq n + M$, and $1 - \delta$ otherwise; see Figure 2b. We will show that in this case, the Markov chain is slow. The idea is that in the

stationary distribution, there is a good chance that the ones and zeros will be well-mixed, since this is a high entropy situation. However, the identity permutation also has high weight, and the parameters are chosen so that the entropy of the well-mixed permutations balances with the energy of the maximum (identity) permutation, and that to get between them is not very likely (low entropy and low energy). We prove that for the set \mathbf{P} defined above, \mathcal{M}_{nn} and \mathcal{M}_{tr} have a bad cut. Then we use the *conductance* (see Section 2.4) to prove \mathcal{M}_{nn} and \mathcal{M}_{tr} are slowly mixing.

Given a staircase walk w , define σ_w to be the highest weight permutation σ such that $f(\sigma) = w$. Notice that σ_w is the permutation where elements $1, 2, \dots, n$ and elements $n+1, n+2, \dots, 2n$ are each in order (for example, $\sigma_{10110010} = (1, 5, 2, 3, 6, 7, 4, 8)$). First, we will show how the combined weight of all permutations that map to w relates to $\pi(\sigma_w)$. We will assign bias to the squares as follows. Each square (i, j) is given weight $\lambda_{i,j} = p_{i,j}/p_{j,i}$, where the squares are numbered as shown in Figure 3(a). Thus the weight of σ_w satisfies

$$\pi(\sigma_w) = \frac{(\beta^{-1} - 1)^{n^2-n} \prod_{xy < w} \lambda_{x,y}}{Z}.$$

The factor $(\beta^{-1} - 1)^{n^2-n}$ comes from having the first and last n elements in order. For any staircase walk w , define $\Pi(w) = \sum_{\sigma: f(\sigma)=w} \pi(\sigma)$, the sum of the weights of the permutations which map to w . We will show that with our choice of β , we have $\pi(w)$ within a factor of 2 of the weight of σ_w .

Lemma 3.3.1: Given the set of probabilities \mathbf{P} defined in Equation 3.3.2, for all staircase walks w , the weight $\Pi(w)$ satisfies the following,

$$\pi(\sigma_w) < \Pi(w) < 2\pi(\sigma_w).$$

PROOF: First, notice that since $f(\sigma_w) = w$ and $\Pi(w) = \sum_{\sigma: f(\sigma)=w} \pi(\sigma)$, we trivially have that $\Pi(w) > \pi(\sigma_w)$. Given any permutation σ , let $h_1(\sigma)$ be the number of inversions between the first n numbers, specifically, pairs $(i, j) : 1 \leq i, j \leq n, i <$

$j, \sigma_i > \sigma_j$. Similarly, let $h_2(\sigma)$ be the number of inversions in the second n numbers, specifically, pairs $(i, j) : n < i, j \leq 2n, i < j, \sigma_i > \sigma_j$. We will start by showing that

$$\frac{\pi(\sigma)}{\pi(\sigma_w)} \leq \left(\frac{\delta^{-1} - 1}{\gamma(\beta^{-1} - 1)} \right)^{h_1(\sigma) + h_2(\sigma)}. \quad (3.3.3)$$

$\lambda_{1,5}$	$\lambda_{2,5}$	$\lambda_{3,5}$	$\lambda_{4,5}$
$\lambda_{1,6}$	$\lambda_{2,6}$	$\lambda_{3,6}$	$\lambda_{4,6}$
$\lambda_{1,7}$	$\lambda_{2,7}$	$\lambda_{3,7}$	$\lambda_{4,7}$
$\lambda_{1,8}$	$\lambda_{2,8}$	$\lambda_{3,8}$	$\lambda_{4,8}$

(a) Permutation 15236478

$\lambda_{1,5}$	$\lambda_{2,5}$	$\lambda_{3,5}$	$\lambda_{4,5}$
$\lambda_{1,6}$	$\lambda_{2,6}$	$\lambda_{3,6}$	$\lambda_{4,6}$
$\lambda_{1,7}$	$\lambda_{2,7}$	$\lambda_{3,7}$	$\lambda_{4,7}$
$\lambda_{1,8}$	$\lambda_{2,8}$	$\lambda_{3,8}$	$\lambda_{4,8}$

(b) Permutation 35416278

$\lambda_{1,5}$	$\lambda_{2,5}$	$\lambda_{3,5}$	$\lambda_{4,5}$
$\lambda_{1,6}$	$\lambda_{2,6}$	$\lambda_{3,6}$	$\lambda_{4,6}$
$\lambda_{1,7}$	$\lambda_{2,7}$	$\lambda_{3,7}$	$\lambda_{4,7}$
$\lambda_{1,8}$	$\lambda_{2,8}$	$\lambda_{3,8}$	$\lambda_{4,8}$

(c) Permutation 38415276

Figure 3: A graphical representation of the relative weights of three different permutations.

Given a walk w , consider any permutation $\sigma : f(\sigma) = w$ and let σ_1 be the sub-permutation corresponding to the first n integers. Similarly let σ_2 be the sub-permutation corresponding to the last n integers. We start by studying the effect of inversions within σ_1 on the weight of σ relative to the weight of σ_w . If σ_1 and σ_2 contain no inversions (i.e. $h_1(\sigma) = h_2(\sigma) = 0$) then $\pi(\sigma) = \pi(\sigma_w)$. First, assume $h_2(\sigma) = 0$, the effect of inversions between the first n elements is to reorder the columns of the walk. For example in Figure 3(b), $\sigma_1(1) = 3$ so each inversion $(1, i)$ is replaced by a $(3, i)$ inversion and $\lambda_{1,i}$ is replaced with $\lambda_{3,i}$ in the weight. This can be visualized as moving the shaded squares (those included in the weight) from column i to column $\sigma(i)$ (see Figure 3(b)). If $h_1(\sigma) = k, h_2(\sigma) = 0$ the weight is now the product of the bias of the shaded squares (after rearranging the columns as specified due to σ_1) times $(\beta^{-1} - 1)^{n^2 - n - k} / Z$. We want to determine if σ_1 has a inversions how much can this change the weight of the σ . Notice that if we shift a column by i since the boundary between the region where p_{ij} is $1 - \delta$ and where it is $1/2 + \epsilon$ is a diagonal (see Figure 2), we can increase the weight of the permutation by at

most $\left(\frac{\delta^{-1}-1}{\gamma}\right)^i$. Each column gets shifted by $\sigma(i) - i$. We are only interested in the case where $\sigma(i) - i > 0$, otherwise, the weight decreases. Let $I(i)$ be the number of inversions associated with i , specifically $j > i : \sigma(j) < \sigma(i)$. We will show that $\sigma(i) - i \leq I(i)$. To see this consider any permutation σ . Consider the permutations $\sigma' = \sigma(1)\sigma(2)\sigma(3)\dots\sigma(i-1)r_1, r_2, r_3, \dots$ where r_1, \dots, r_k are the remaining integers not included in $\sigma(1)\dots\sigma(i-1)$. Assume $\sigma(i) = r_j$ then $I(i) = j$ and since every number less than $\sigma(i)$ occurs before $\sigma(i)$ in σ' , it follows that $\sigma_i \leq i + I(i)$ implying that $\sigma(i) - i \leq I(i)$ as desired. If $h_2(\sigma) \neq 0$, we can use the exact same argument to bound the increase in weight due to inversion between the last n elements. These inversions correspond to switching rows of the staircase walk instead of columns. For example, see Figure 3(c). Similarly, each row i moves a distance of $\sigma(i) - i$ so we can bound these in the exact same way. Combining these gives us the following,

$$\begin{aligned}
\frac{\pi(\sigma)}{\pi(\sigma_w)} &\leq (\beta^{-1} - 1)^{-(h_1(\sigma)+h_2(\sigma))} \left(\frac{\delta^{-1} - 1}{\gamma}\right)^{\sum_{i:1 \leq i \leq n, \sigma(i) > i} \sigma(i) - i + \sum_{i:n < i \leq 2n, \sigma(i) > i} \sigma(i) - i} \\
&\leq (\beta^{-1} - 1)^{-(h_1(\sigma)+h_2(\sigma))} \left(\frac{\delta^{-1} - 1}{\gamma}\right)^{\sum_{i=0}^n I(i) + \sum_{i=n+1}^{2n} I(i)} \\
&\leq (\beta^{-1} - 1)^{-(h_1(\sigma)+h_2(\sigma))} \left(\frac{\delta^{-1} - 1}{\gamma}\right)^{h_1(\sigma)+h_2(\sigma)} \\
&\leq \left(\frac{\delta^{-1} - 1}{\gamma(\beta^{-1} - 1)}\right)^{h_1(\sigma)+h_2(\sigma)}.
\end{aligned}$$

Next, notice that there are most n^{i+j} permutations $\sigma : f(\sigma) = w, h_1(\sigma) = i, h_2(\sigma) = j$. This is because we can think of this as first choosing a permutation of the first n elements with i inversions and then choosing a permutation of the next n elements with j inversions. The number of permutations of n elements with i inversions is upper bounded by n^i . This is straightforward to see in the context of the bijection with inversion tables discussed in Section 3.4.1. Combining this with

Equation 3.3.3 gives the following:

$$\begin{aligned}
\Pi(w) &= \sum_{i=0}^{\binom{n}{2}} \sum_{j=0}^{\binom{n}{2}} \sum_{\sigma: f(\sigma)=w, h_1(\sigma)=i, h_2(\sigma)=j} \pi(\sigma) \\
&\leq \sum_{i=0}^{\binom{n}{2}} \sum_{j=0}^{\binom{n}{2}} \sum_{\sigma: f(\sigma)=w, h_1(\sigma)=i, h_2(\sigma)=j} \pi(\sigma_w) \left(\frac{\delta^{-1} - 1}{\gamma(\beta^{-1} - 1)} \right)^{i+j} \\
&\leq \pi(\sigma_w) \sum_{i=0}^{\binom{n}{2}} \sum_{j=0}^{\binom{n}{2}} n^{i+j} \left(\frac{\delta^{-1} - 1}{\gamma(\beta^{-1} - 1)} \right)^{i+j} \\
&< 2\pi(\sigma_w) \quad \blacksquare
\end{aligned}$$

We are now ready to prove the main theorem of the section.

Theorem 3.3.2: There exists a positively biased preference set \mathbf{P} for which the mixing time $\tau(\epsilon)$ of the Markov chain \mathcal{M}_{nn} with preference set \mathbf{P} satisfies

$$\tau(\epsilon) = \Omega \left(e^{n^{1/3}} \log(\epsilon^{-1}) \right).$$

PROOF: As in the earlier example, for a staircase walk w , define the *height of w_i* as $\sum_{j \leq i} w_j$, and let $\max(w)$ be the maximum height of w_i over all $1 \leq i \leq 2n$. Again let W_1 be the set of walks w such that $\max(w) < n + M$, W_2 the set of walks such that $\max(w) = n + M$, and W_3 the set of walks such that $\max(w) > n + M$. Let S_1 be the set of permutations σ such that $f(\sigma) \in W_1$, S_2 the permutations such that $f(\sigma) \in W_2$ and S_3 the permutations such that $f(\sigma) \in W_3$. That is, W_1 is the set of walks that never reach the dark blue diagonal in Figure 2b, W_2 is the set whose maximum peak is on the dark blue line, and W_3 is the set which crosses that line and contains squares in the light blue triangle. Define $\gamma = (1/2 + \epsilon)/(1/2 - \epsilon)$, which is the ratio of two configurations that differ by swapping a $(0, 1)$ pair with probability $\frac{1}{2} + \epsilon$. First we notice that since the maximum weight permutation (which maps to the maximal tiling) is in S_3 ,

$$\pi(S_3) \geq \frac{1}{Z} \gamma^{n^2 - \frac{(n-M)^2}{2}} (\delta^{-1} - 1)^{\frac{(n-M)^2}{2}} (\beta^{-1} - 1)^{n^2 - n}.$$

Using Lemma 3.3.1, $\pi(S_1) \leq \frac{2}{Z} \sum_{w \in W_1} \gamma^{A(w)} (\beta^{-1} - 1)^{n^2 - n}$, where $A(w)$ is the number of unit squares below w . We have that

$$\begin{aligned}
\pi(S_1) &\leq \frac{2}{Z} \sum_{w \in W_1} \gamma^{A(w)} (\beta^{-1} - 1)^{n^2 - n} \\
&\leq \frac{2}{Z} \sum_{w \in W_1} \gamma^{n^2 - \frac{(n-M)^2}{2}} (\beta^{-1} - 1)^{n^2 - n} \\
&\leq \frac{2}{Z} \binom{2n}{n} \gamma^{n^2 - \frac{(n-M)^2}{2}} (\beta^{-1} - 1)^{n^2 - n} \\
&\leq \frac{1}{Z} (2e)^{n+1} \gamma^{n^2 - \frac{(n-M)^2}{2}} (\beta^{-1} - 1)^{n^2 - n} \\
&\leq \frac{1}{Z} \gamma^{n^2 - \frac{(n-M)^2}{2}} (\delta^{-1} - 1)^{\frac{(n-M)^2}{2}} (\beta^{-1} - 1)^{n^2 - n} \\
&\leq \pi(S_3)
\end{aligned}$$

for large enough n , since $1/\delta > 2$ is a constant. Hence $\pi(S_1) \leq \pi(S_3)$. We will show that $\pi(S_2)$ is exponentially small in comparison to $\pi(S_1)$ (and hence also to $\pi(S_3)$).

$$\pi(S_2) \leq \frac{2}{Z} \sum_{\sigma \in W_2} \gamma^{A(w)} (\beta^{-1} - 1)^{n^2 - n} \leq \frac{2\gamma^{n^2} (\beta^{-1} - 1)^{n^2 - n} |W_2|}{Z}.$$

We bound $|W_2|$ in exactly the same way as before. The unbiased Markov chain is equivalent to a simple random walk $w_{2n} = X_1 + X_2 + \dots + X_{2n} = 0$, where $X_i \in \{+1, -1\}$ and where a $+1$ represents a step to the right and a -1 represents a step down. We call this random walk *tethered* since it is required to end at 0 after $2n$ steps. Compare walk w_{2n} with the untethered simple random walk $w'_{2n} = X'_1 + X'_2 + \dots + X'_{2n}$.

$$\begin{aligned}
P\left(\max_{1 \leq t \leq 2n} w_t \geq M\right) &= P\left(\max_{1 \leq t \leq 2n} w'_t \geq M \mid w'_{2n} = 0\right) \\
&= \frac{P(\max_{1 \leq t \leq 2n} w'_t \geq M)}{P(w'_{2n} = 0)} \\
&= \frac{2^{2n}}{\binom{2n}{n}} P\left(\max_{1 \leq t \leq 2n} w'_t \geq M\right) \\
&\approx \sqrt{\pi n} P\left(\max_{1 \leq t \leq 2n} w'_t \geq M\right).
\end{aligned}$$

Since the $\{X'_i\}$ are independent, we can use Chernoff bounds to see that

$$P\left(\max_{1 \leq t \leq 2n} w'_t \geq M\right) \leq 2nP(w'_{2n} \geq M) \leq 2ne^{-\frac{M^2}{2n}}.$$

Together these show that

$$P\left(\max_{1 \leq t \leq 2n} W_t \geq M\right) < e^{-n^{1/3}},$$

by definition of M . Therefore we have

$$\begin{aligned} \pi(S_2) &\leq \frac{2}{Z} \gamma^{n^2} (\beta^{-1} - 1)^{n^2-n} |W_2| \leq \frac{2\gamma^{n^2} (\beta^{-1} - 1)^{n^2-n}}{Z} \binom{2n}{n} e^{-n^{1/3}} \\ &\leq \frac{(\beta^{-1} - 1)^{n^2-n}}{Z} \binom{2n}{n} e^{-n^{1/3}+1} (1 - e^{-n^{1/3}}) \\ &\leq \frac{(\beta^{-1} - 1)^{n^2-n}}{Z} |S_1| e^{-n^{1/3}+1} \\ &\leq e^{-n^{1/3}+1} \pi(S_1), \end{aligned}$$

as desired. Thus, $\pi(S_2)$ is exponentially smaller than $\pi(S_1)$ for every value of δ and the conductance satisfies

$$\begin{aligned} \Phi &\leq \sum_{x \in S_1} \frac{\pi(x)}{\pi(S_1)} \sum_{y \in S_2} P(x, y) \\ &\leq \sum_{x \in S_1} \frac{\pi(x)}{\pi(S_1)} \pi(S_2) \\ &\leq e^{-n^{1/3}+1} \pi(S_1) \leq \frac{e^{-n^{1/3}+1}}{2}. \quad \blacksquare \end{aligned}$$

Hence, by Theorem 2.4.1, $\tau(\epsilon)$, the mixing time of \mathcal{M}_{nn} satisfies

$$\tau(\epsilon) \geq \frac{1}{2} \left(e^{n^{1/3}-1} - 1 \right) \log \left(\frac{1}{2\epsilon} \right).$$

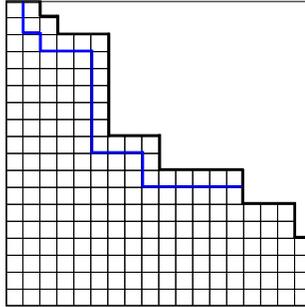


Figure 4: A move that swaps an arbitrary $(1, 0)$ pair.

In fact, this proof can be extended to the more general Markov chain where we can swap any 1 with any 0, as long as we maintain the correct stationary distribution. This is easy to see, because any move that swaps a single 1 with a single 0 can only change the maximum height by at most 2 (see Figure 4). If we expand S_2 to include all configurations with maximum height $n + M$ or $n + M + 1$, $\pi(S_2)$ is still exponentially smaller than $\pi(S_1) \leq \pi(S_3)$. Hence the Markov chain that swaps an arbitrary $(1, 0)$ pair still takes exponential time to converge.

Next, we show that there exists a value of δ for which $\pi(S_3) = \pi(S_1)$, which will imply that $\pi(S_2)$ is also exponentially smaller than $\pi(S_3)$, and hence the set S_2 forms a bad cut, regardless of which state the Markov chain begins in.

Lemma 3.3.3: There exist a constant δ , $\frac{1}{65} < \delta < \frac{1}{2}$, such that for this choice of δ , $\pi(S_3) = \pi(S_1)$.

PROOF: To find this value of δ , we will rely on the continuity of the function $f(\xi) = Z\pi(S_3) - Z\pi(S_1)$ with respect to $\xi = (1 - \delta)/\delta$. Let $a(\sigma)$ be the number of non-inversions in σ between i and j , $i < j$ such that $p_{ij} = 1/2 + \epsilon$ (for any highest weight configuration σ_w this corresponds to the number of tiles above the diagonal M in w) and let $b(\sigma)$ be the number of non-inversions in σ between i and j such that $p_{ij} = 1 - \delta$ (the number of tiles below the diagonal M). Notice that $Z\pi(S_1)$ is constant with respect to ξ and $Z\pi(S_3) = \sum_{\sigma \in S_3} \gamma^{b(\sigma)} \xi^{a(\sigma)} (\beta^{-1} - 1)^{n^2 - n - h_1(\sigma) - h_2(\sigma)}$ is just a polynomial in ξ . Therefore $Z\pi(S_3)$ is continuous in ξ and hence $f(\xi)$ is also continuous with respect to ξ . Moreover, when $\xi = \gamma$, clearly $Z\pi(S_3) < Z\pi(S_1)$, so $f(\gamma) < 0$. We will show that $f(4e^2) > 0$, and so by continuity we will conclude that there exists a value of ξ satisfying $\gamma < \xi < 4e^2$ for which $f(\xi) = 0$ and $Z\pi(S_3) = Z\pi(S_1)$. Clearly this implies that for this choice of ξ , $\pi(S_3) = \pi(S_1)$, as desired. To obtain the corresponding value of δ , we notice that $\delta = 1/(\xi + 1)$. In particular, δ is a constant satisfying $\frac{1}{65} < \delta < \frac{1}{2}$.

Thus it remains to show that $f(4e^2) > 0$. First we notice that since the maximal

tiling is in S_3 , $\pi(S_3) \geq Z^{-1} \gamma^{n^2 - \frac{(n-M)^2}{2}} \xi^{\frac{(n-M)^2}{2}} (\beta^{-1} - 1)^{n^2 - n}$. Also,

$$\pi(S_1) = Z^{-1} \sum_{\sigma \in S_1} \gamma^{a(\sigma)} (\beta^{-1} - 1)^{n^2 - n - h_1(\sigma) - h_2(\sigma)} < Z^{-1} \binom{2n}{n} \gamma^{n^2 - \frac{(n-M)^2}{2}} 2(\beta^{-1} - 1)^{n^2 - n}.$$

Therefore

$$\pi(S_1)/\pi(S_3) < \frac{2 \binom{2n}{n}}{\xi^{\frac{(n-M)^2}{2}}} \leq (2e)^n \xi^{-n/2} = 1$$

since $\xi = 4e^2$. Hence $f(4e^2) = Z\pi(S_3) - Z\pi(S_1) > Z\pi(S_3) - Z\pi(S_3) = 0$, as desired. ■

Remark: In the setting of biased staircase walks, if the bias on each unit square (x, y) satisfies $\lambda_{x,y} \geq 2$, Pascoe and Randall give polynomial bounds on the mixing time of the Markov chain which adds or removes a unit square from under the walk [68]. Using a similar reduction to the one used in the proof of Theorem 3.3.2, in the biased permutations setting, these results can provide a class of positively biased \mathbf{P} for which there are $O(n^2)$ input parameters and \mathcal{M}_{nn} is rapidly mixing.

3.4 Choose Your Weapon

Despite the slow mixing example outlined in the previous section, there are many cases for which the chain will be rapidly mixing. We define two new classes for which we can rigorously demonstrate this and we provide the proofs in the next two sections.

For the first class, imagine a community of n people, each with a unique combative talent. Each member has his or her weapon of choice, and a competition with any other member of the community using this weapon affords that person a fixed advantage. When two people are chosen to compete, they each prefer using their own weapon of choice, so we resolve this by letting the person with the higher rank (e.g., age, seniority, etc.) choose the weapon they both will use. At any point in time our competitors are ordered and nearest neighbors are randomly selected to compete, where upon the winner is moved in front of the loser in the ordering.

To formalize the ‘‘Choose Your Weapon’’ scenario, we are given

$$1/2 \leq r_1, r_2, \dots, r_{n-1} < 1$$

and the set \mathbf{P} satisfies $p_{i,j} = r_i$, if $i < j$ and $p_{i,j} = 1 - p_{j,i}$ if $j < i$. The moves of the nearest neighbor Markov chain \mathcal{M}_{nn} formalize the competitions, and our goal is to bound the mixing rate of this chain. Notice that this class includes the constant bias case studied by Benjamini et al. as a special case, and indeed our analysis yields an independent and simpler proof that the nearest neighbor Markov chain \mathcal{M}_{nn} is rapidly mixing in that context.

We shall show that the chain \mathcal{M}_{nn} is always rapidly mixing for probabilities \mathbf{P} defined in this way. Our proof relies on a bijection between permutations and *Inversion Tables* [54, 83] that, for each element i , record how many elements $j > i$ come before i in the permutation. We consider a Markov chain \mathcal{M}_{inv} that simply increments or decrements a single element of the inversion table in each step; using the bijection with permutations this corresponds to transpositions of elements that are not necessarily nearest neighbors to the Markov chain \mathcal{M}_{nn} . Remarkably, this allows \mathcal{M}_{inv} to decompose into a product of simple one-dimensional random walks and bounding the convergence time is very straightforward. Finally, we use comparison techniques [25, 71] to bound the mixing time of the nearest neighbor chain \mathcal{M}_{nn} .

3.4.1 The Inversion Table Representation.

The Markov chain \mathcal{M}_{inv} acts on the *inversion table* for the permutation [54, 83], which has an entry for each $i \in [n]$ counting the number of inversions involving i ; that is, the number of values $j > i$ where j comes before i in the permutation (see Figure 5). It is easy to see that the i th element of the inversion table is an integer between 0 and $n - i$. In fact, the function I is a bijection between the set of permutations and the set \mathcal{I} of all possible inversion tables (all sequences $X = (x_1, x_2, \dots, x_n)$ where $0 \leq x_i \leq n - i$ for all $i \in [n]$). To see this, we will construct a permutation from any inversion table $X \in \mathcal{I}$. Place the element 1 in the $(x_1 + 1)$ st position of the permutation. Next, there are $n - 1$ slots remaining. Among these, place the element 2 in the $(x_2 + 1)$ st

position remaining (ignoring the slot already filled by 1). Continuing, after placing $i - 1$ elements into the permutation, there are $n - i + 1$ slots remaining, and we place the element i into the $(x_i + 1)$ st position among the remaining slots. This proves that I is a bijection from S_n to \mathcal{I} .

Given this bijection, a natural algorithm for sampling permutations is the following local Markov chain on inversion tables: select a position $i \in [n]$ and attempt to either add one or subtract one from x_i , according to the appropriate probabilities. In terms of permutations, this amounts to adding or removing an inversion involving i without affecting the number of inversions involving any other integer, and is achieved by swapping the element i with an element $j > i$ such that every element in between is smaller than both i and j . If i moves ahead of j , this move happens with probability $p_{i,j}$ because for each k originally between i and j , $p_{k,i} = r_k = p_{k,j}$ (since $k < i$ and $k < j$), so the net effect of the move is neutral. The detailed balance condition ensures that π is the correct stationary distribution. Formally, the Markov chain \mathcal{M}_{nn} is defined as follows.

The Inversion Markov chain \mathcal{M}_{inv}

Starting at any permutation σ_0 , repeat:

- With probability $1/2$ let $\sigma_{t+1} = \sigma_t$.
- Otherwise, select $(i, b) \in [n] \times \{-1, +1\}$ u.a.r.
 - If $b = +1$ let j be the first element after i in σ_t such that $j > i$ (if such a j does not exist let $\sigma_{t+1} = \sigma_t$). With probability $p_{j,i}$, obtain σ_{t+1} from σ_t by swapping i and j .
 - If $b = -1$ let j be the last element before i in σ_t such that

$$\begin{array}{rcl} \sigma & = & 8 \ 1 \ 5 \ 3 \ 7 \ 4 \ 6 \ 2 \\ I(\sigma) & = & 1 \ 7 \ 2 \ 3 \ 1 \ 2 \ 1 \ 0 \end{array}$$

Figure 5: The inversion table for a permutation.

$j > i$ (if such a j does not exist let $\sigma_{t+1} = \sigma_t$). With probability $p_{i,j}$, obtain σ_{t+1} from σ_t by swapping i and j .

This Markov chain contains the moves of \mathcal{M}_{nn} (and therefore also connects the state space). Although elements can jump across several elements, it is still fairly local compared with the general transposition chain \mathcal{M}_{tr} which has $\binom{n}{2}$ choices at every step, since \mathcal{M}_{inv} has at most $2n$.

3.4.2 Fast Mixing of the Inversion Chain

The inversion Markov chain \mathcal{M}_{inv} can be viewed as a product of n independent processes. The i th process is a one-dimensional random walk bounded between 0 and $n - i$ that moves up by one with probability r_i and down by one with probability $1 - r_i$; its mixing time is $O(n^2 \log n)$, unless r_i is bounded away from $1/2$, in which case its mixing time is $O(n)$. We make moves in each chain with probability $1/n$, since we update one random walk at a time. The main tool we use for proving rapid mixing of \mathcal{M}_{inv} is coupling which was introduced in Section 2.2.

For the general case where the r_i 's are not bounded about from $1/2$, we will first bound the mixing time of each one-dimensional walk by using the following Lemma due to Luby, Randall and Sinclair [59] to bound the coupling time.

Lemma 3.4.1: Let ϕ be an integer valued metric defined on $\Omega \times \Omega$ which takes values in $[0, B]$, and $\phi(x, y) = 0$ iff $x = y$. Let \mathcal{M} be a Markov chain on Ω and let (X_t, Y_t) be a coupling of \mathcal{M} , with $\phi_t = \phi(X_t, Y_t)$. Suppose the coupling satisfies $\mathbb{E}[\Delta\phi_t | X_t, Y_t] \leq 0$ and, whenever $\phi_t > 0$, $\mathbb{E}[(\Delta\phi_t)^2 | X_t, Y_t] \geq V$. Then the expected coupling time from initial states x, y satisfies

$$\mathbb{E}T^{x,y} \leq \frac{\phi_0(2B - \phi_0)}{V}.$$

Next, we will use the following theorem, which relates the mixing time of a product of independent Markov chains to the mixing time of each component to bound the

mixing time of \mathcal{M}_{inv} . Similar results have been proved before in other settings (i.e., see [6, 9] and Corollary 12.12 of [55]). The proof can be found in [8].

Theorem 3.4.2: Suppose the Markov chain \mathcal{M} is a product of N independent Markov chains $\{\mathcal{M}_i\}$, where \mathcal{M} updates each \mathcal{M}_i with probability p_i . If $\tau_i(\epsilon)$ is the mixing time for \mathcal{M}_i and $\tau_i(\epsilon) \geq 4 \ln \epsilon$ for each i , then

$$\tau(\epsilon) \leq \max_{i=1,2,\dots,N} \frac{2}{p_i} \tau_i \left(\frac{\epsilon}{2N} \right).$$

Now we are ready to prove the following theorem, bounding the mixing time of \mathcal{M}_{inv} .

Theorem 3.4.3: Given input probabilities $1/2 \leq r_1, r_2, \dots, r_{n-1} < 1$, let $\mathbf{P} = \{p_{i,j} = r_{\min\{i,j\}}\}$. The mixing time of \mathcal{M}_{inv} with preference set \mathbf{P} satisfies $\tau(\epsilon) = O(n^3 \log(n\epsilon^{-1}))$.

PROOF: To prove our theorem, we need to analyze the one-dimensional process $\mathcal{M}(r, k)$, bounded between 0 and k , which chooses to move up with probability $r/2 \geq 1/4$, down with probability $(1-r)/2$ and does nothing with probability $1/2$ at each step, if possible. This simple random walk is well-studied; we include the proof for completeness. We use Lemma 3.4.1. First, we define a natural distance metric $\phi(X_t, Y_t) = \phi_t$ on pairs X_t, Y_t of walks where ϕ_t is the distance between the two walks at time t (i.e., if $X_t = i$ and $Y_t = j$ then $\phi_t = |i - j|$). We construct a coupling on the two lazy walks where with probability $r/2$ the walk X_t moves up (Y_t does nothing), with probability $(1-r)/2$ the walk X_t moves down (Y_t does nothing), similarly with probability $r/2$ the walk Y_t moves up (X_t does nothing) and with probability $(1-r)/2$ the walk Y_t moves down (X_t does nothing). The directions for Y_t and X_t are chosen independently. Once the walks collide, they make the exact same moves. Since the two walks never move at the same time, they will never jump over each other. Without loss of generality, assume that X_t is above Y_t then there are at most two cases where

the distance increases. Namely, if X_t moves up which happens with probability $r/2$ or if Y_t moves down which happens with probability $(1-r)/2$. The distance is decreased if X_t moves down or Y_t moves up which happens with probabilities $(1-r)/2$ and $r/2$ respectively. Thus, $\mathbb{E}[\Delta\phi_t|X_t, Y_t] \leq r/2 + (1-r)/2 - r/2 - (1-r)/2 = 0$. Similarly, whenever $\phi_t > 0$, $\mathbb{E}[(\Delta\phi_t)^2|X_t, Y_t] \geq 1(r/2) + 1((1-r)/2) + 1((1-r)/2) \geq 1/2 = V$. Notice that $0 \leq \phi_0 \leq n = B$. Applying Lemma 3.4.1 with these choices of B , V and ϕ_0 gives the following:

$$\mathbb{E}T^{x,y} \leq \frac{\phi_0(2B - d_0)}{V} \leq \frac{n(2n - 0)}{1/2} = 4n^2.$$

To bound the mixing time using our bound on the coupling time we apply Theorem 2.2.1 as follows:

$$\tau(\epsilon) \leq \lceil Te \ln \epsilon^{-1} \rceil = \lceil 4n^2 e \ln \epsilon^{-1} \rceil.$$

Next we use Theorem 3.4.2 to bound the mixing time of \mathcal{M}_{inv} as follows:

$$\tau(\epsilon) \leq \frac{2}{1/n} (\lceil 4n^2 e \ln \epsilon^{-1} \rceil) = O(n^3 \log(n\epsilon^{-1})).$$

■

When each r_i is bounded away from $1/2$ and 1 , by using the path coupling theorem we obtain a stronger result. We use Theorem 2.2.3 due to Greenberg, Pascoe and Randall [43].

Theorem 3.4.4: Given input probabilities $1/2 \leq r_1, r_2, \dots, r_{n-1} < 1$ and a positive constant c such that $c + 1/2 < r_i < 1 - c$ for $1 \leq i \leq n - 1$, let $\mathbf{P} = \{p_{i,j} = r_{\min\{i,j\}}\}$. The mixing time of \mathcal{M}_{inv} with preference set \mathbf{P} satisfies

$$\tau(\epsilon) = O(n^2 \ln(n\epsilon^{-1})).$$

PROOF: Here we use path coupling. We use the natural coupling on inversion tables where we choose the same element i for both $X = X_t = (x_1, \dots, x_n)$ and $Y = Y_t =$

(y_1, \dots, y_n) for each step t . A pair of inversion tables (X, Y) is in $U \subset I \times I$ if Y can be obtained from X by adding or subtracting 1 from a single x_i . We define a distance metric $\phi(X, Y)$ on pairs X, Y of inversion tables as a sum over the distances between the entries x_i and y_i . Let $\alpha_i = 1/(2(1 - r_i))$ and define

$$\phi(X, Y) = \sum_{i=1}^n \sum_{j=\min\{x_i, y_i\}}^{\max\{x_i, y_i\}-1} \alpha_i^j.$$

Let $(X, Y) \in U$ and suppose Y is obtained from X by adding 1 to x_i . Then as before, any move of \mathcal{M}_{inv} whose smaller index is not i succeeds or fails with the same probability in X and Y . There are two moves that decrease the distance: adding 1 to x_i , which happens with probability $(1 - r_i)/(4n)$, or subtracting 1 from y_i , which happens with probability $r_i/(4n)$. Both of these moves decrease the distance by $\alpha_i^{x_i}$. On the other hand, \mathcal{M}_{inv} proposes adding 1 to y_i with probability $(1 - r_i)/(4n)$, which increases the distance by $\alpha_i^{x_i+1}$, and \mathcal{M}_{inv} proposes subtracting 1 from x_i and succeeds with probability $r_i/(4n)$, increasing the distance by $\alpha_i^{x_i-1}$. Thus the expected change in distance is

$$\begin{aligned} E[(\phi_{t+1} - \phi_t) | X_t, Y_t] &= \frac{1}{4n} (-\alpha_i^{x_i} + (1 - r_i)\alpha_i^{x_i+1} + r_i\alpha_i^{x_i-1}) \\ &= \frac{\alpha_i^{x_i-1}}{4n} (-\alpha_i + (1 - r_i)\alpha_i^2 + r_i) \\ &= \frac{\alpha_i^{x_i-1}}{4n} \left(r_i - \frac{1}{4(1 - r_i)} \right) \\ &= \frac{\alpha_i^{x_i}}{2n} \cdot \frac{-(2r_i - 1)^2}{4} \\ &= \frac{-\phi_t(2r_i - 1)^2}{8n}, \end{aligned}$$

■

since $\phi_t = \alpha_i^{x_i}$. Hence $E[\phi_{t+1}] \leq \phi_t(1 - (2r_i - 1)^2/(8n))$. Moreover, the maximum distance between any two inversion tables is

$$B = \sum_{i=1}^{n-1} 1 + \alpha_i + \dots + \alpha_i^{n-i} = \sum_{i=1}^{n-1} \frac{\alpha_i^n - 1}{\alpha_i - 1} = O(n\alpha_{\max}^n),$$

where $\alpha_{\max} = \max_i \{\alpha_i\} = \frac{1}{2(1-r_{\max})}$. Hence

$$\ln(B\epsilon^{-1}) = O(n \log(n\epsilon^{-1})),$$

so by Theorem 2.2.2, we have

$$\tau(\epsilon) = O(n^2 \log(n\epsilon^{-1})).$$

Remark: The proofs of Theorem 3.4.3 and Theorem 3.4.4 also apply to the case where the probability of swapping i and j depends on the object with lower rank (i.e., we are given r_2, \dots, r_n and we let $p_{i,j} = r_j$ for all $i < j$). This case is related to a variant of the MA1 list update algorithm, where if a record is requested, we try to move the associated record x ahead of its immediate predecessor in the list, if it exists. If it has higher rank than its predecessor, then it always succeeds, while if its rank is lower we move it ahead with probability $f_x = r_x / (1 + r_x) \leq 1$.

3.4.3 Fast Mixing of the Nearest-Neighbor Chain

First, we show that the two Markov chains \mathcal{M}_{inv} and \mathcal{M}_{nn} have the same stationary distribution. Then we will use the comparison method, Theorem 2.3.1, to infer a bound on the mixing time of \mathcal{M}_{nn} from the bounds on the mixing time of \mathcal{M}_{inv} shown in Theorem 3.4.3 and Theorem 3.4.4.

Theorem 3.4.5: Given input probabilities $1/2 \leq r_1, r_2, \dots, r_{n-1} < 1$ and a positive constant c such that $r_i < 1 - c$ for $1 \leq i \leq n - 1$, let $\mathbf{P} = \{p_{i,j} = r_{\min\{i,j\}}\}$.

1. If $\forall 1 \leq i \leq n - 1, c + 1/2 < r_i$, the mixing time of \mathcal{M}_{nn} with preference set \mathbf{P} satisfies

$$\tau(\epsilon) = O(n^7 \log(n\epsilon^{-1}) \log(\epsilon^{-1})).$$

2. Otherwise, the mixing time of \mathcal{M}_{nn} with preference set \mathbf{P} satisfies

$$\tau(\epsilon) = O(n^8 \log(n\epsilon^{-1}) \log(\epsilon^{-1})).$$

PROOF: In order to apply Theorem 2.3.1, we need to define, for any transition $e = (\sigma, \beta)$ of the Markov chain \mathcal{M}_{inv} , a sequence of transitions of \mathcal{M}_{nn} . Let e be a transition of \mathcal{M}_{inv} which performs a transposition on elements $\sigma(i)$ and $\sigma(j)$, where $i < j$. Recall \mathcal{M}_{inv} can only swap $\sigma(i)$ and $\sigma(j)$ if all the elements between them are smaller than both $\sigma(i)$ and $\sigma(j)$. To obtain a sufficient bound on the congestion along each edge, we ensure that in each step of the path, we do not decrease the weight of the configuration. This is easy to do; in the first stage, move $\sigma(i)$ to the right, one step at a time, until it swaps with $\sigma(j)$. This removes an inversion of the type $(\sigma(i), \sigma(k))$ for every $i < k < j$, so clearly we have not decreased the weight of the configuration at any step. Next, move $\sigma(j)$ to the left, one step at a time, until it reaches position i . This completes the move e , and at each step, we are adding back an inversion of the type $(\sigma(j), \sigma(k))$ for some $i < k < j$. Since $\sigma(k) = \min\{\sigma(j), \sigma(k)\} = \min\{\sigma(i), \sigma(k)\}$, we have $p_{\sigma(i), \sigma(k)} = p_{\sigma(j), \sigma(k)}$ for every $i < k < j$, so in this stage we restore all the inversions destroyed in the first stage, for a net change of $p_{\sigma(i), \sigma(j)}$. See Figure 6.

Given a transition (v, ω) of \mathcal{M}_{nn} we must upper bound the number of canonical paths $\gamma_{\sigma\beta}$ that use this edge, which we do by bounding the amount of information needed in addition to (v, ω) to determine σ and β uniquely. For moves in the first stage, all we need to remember is $\sigma(i)$, because we know $\sigma(j)$ (it is the element moving forward). We also need to remember where $\sigma(j)$ came from. Given this information along with v and ω we can uniquely recover (σ, β) . Thus there are at most n^2 paths which use any edge (v, ω) . Also, notice that the maximum length of any path is $2n$.

5	2	3	7
2	5	3	7
2	3	5	7
2	3	7	5
2	7	3	5
7	2	3	5

Figure 6: The canonical path for transposing 5 and 7.

Next we bound the quantity A which is needed to apply Theorem 2.3.1. Let $\lambda = \max_{i < j} p_{i,j}/p_{j,i}$. Recall that we have guaranteed that $\pi(\sigma) \leq \max\{\pi(v), \pi(\omega)\}$. Assume first that $\pi(\sigma) \leq \pi(v)$. Then

$$\begin{aligned} A &= \max_{(v,\omega) \in E(P)} \left\{ \frac{1}{\pi(v)P(v,\omega)} \sum_{\Gamma(v,\omega)} |\gamma_{\sigma\beta}| \pi(\sigma) P'(\sigma, \beta) \right\} \\ &\leq \max_{(v,\omega) \in E(P)} \sum_{\Gamma(v,\omega)} 2n \frac{P'(\sigma, \beta)}{P(v,\omega)} \\ &\leq \max_{(v,\omega) \in E(P)} \sum_{\Gamma(v,\omega)} 2n \frac{1/(2n)}{\frac{1}{(1+\lambda)(n-1)}} = O(n^3). \end{aligned}$$

If, on the other hand, $\pi(\sigma) \leq \pi(\omega)$, then we use detailed balance to obtain:

$$\begin{aligned} A &= \max_{(v,\omega) \in E(P)} \left\{ \frac{1}{\pi(v)P(v,\omega)} \sum_{\Gamma(v,\omega)} |\gamma_{\sigma\beta}| \pi(\sigma) P'(\sigma, \beta) \right\} \\ &= \max_{(v,\omega) \in E(P)} \left\{ \frac{1}{\pi(\omega)P(\omega, v)} \sum_{\Gamma(v,\omega)} |\gamma_{\sigma\beta}| \pi(\sigma) P'(\sigma, \beta) \right\} \\ &\leq \max_{(v,\omega) \in E(P)} \sum_{\Gamma(v,\omega)} 2n \frac{P'(\sigma, \beta)}{P(\omega, v)} \\ &\leq \max_{(v,\omega) \in E(P)} \sum_{\Gamma(v,\omega)} 2n \frac{1/(2n)}{\frac{1}{(1+\lambda)(n-1)}} = O(n^3). \quad \blacksquare \end{aligned}$$

In either case, we have $A = O(n^3)$. Then $\pi_* = \min_{\rho \in \Omega} \pi(\rho) \geq (\lambda^{\binom{n}{2}} n!)^{-1}$ where λ is defined as above, so $\log(1/(\epsilon \pi_*)) = O(n^2 \log \epsilon^{-1})$, since λ is bounded from above by a positive constant. Appealing to Theorem 2.3.1 and Theorem 3.4.4, if the r_i 's are bounded away from $1/2$, we have that the mixing time of \mathcal{M}_{nn} satisfies

$$\tau(\epsilon) = O(n^7 \log(n\epsilon^{-1}) \log(\epsilon^{-1})).$$

Similarly, appealing to Theorem 2.3.1 and Theorem 3.4.3 we have that the mixing time of \mathcal{M}_{nn} satisfies

$$\tau(\epsilon) = O(n^8 \log(n\epsilon^{-1}) \log(\epsilon^{-1})).$$

Remark: If the input probabilities are not bounded away from 1 by a constant but instead by some function of n , then using the same proof as Theorem 3.4.5 we can obtain a bound on the mixing time. Specifically, given input probabilities $1/2 \leq r_1, r_2, \dots, r_{n-1} < 1 - 1/f(n)$ let $\mathbf{P} = \{p_{i,j} = r_{\min\{i,j\}}\}$. Then the mixing time of \mathcal{M}_{nn} with preference set \mathbf{P} satisfies

$$\tau(\epsilon) = O(n^8 f(n) \log(n\epsilon^{-1}) \log(f(n)\epsilon^{-1})).$$

If additionally, there exists a positive constant c such that $r_i > 1/2 + c$ for all $1 \leq i \leq n-1$, then the mixing time of \mathcal{M}_{nn} with preference set \mathbf{P} satisfies

$$\tau(\epsilon) = O(n^7 f(n) \log(nf(n)\epsilon^{-1}) \log(f(n)\epsilon^{-1})).$$

Note that in this case it is necessary to also modify the proof of Theorem 3.4.4.

3.5 League Hierarchies

We now introduce a second general class of input probabilities \mathbf{P} for which we show \mathcal{M}_{nn} is always rapidly mixing. Imagine a sporting franchise consisting of an A-league with stronger players and a B-league with weaker players. We assume that any player from the A-league has a fixed advantage over any player from the B-league, representing his or her probability of winning in a matchup. Within each of these leagues we have tier-1 and tier-2 players, where again a player from the stronger tier has a fixed probability of winning a competition against a tier-2 player. Likewise for the tiers in the other league, but of course the fixed advantage there can be different. This partition of each tier into stronger and weaker players continues recursively. To formalize the class of “League Hierarchies,” let T be a proper rooted binary tree with n leaf nodes, labeled $1, \dots, n$ in sorted order. Each non-leaf node v of this tree is labeled with a value $\frac{1}{2} \leq q_v < 1$. For $i, j \in [n]$, let $i \wedge j$ be the lowest common ancestor of the leaves labeled i and j . We say that \mathbf{P} has *league structure* T if $p_{i,j} = q_{i \wedge j}$. For example, Figure 7(a) shows a set \mathbf{P} such that $p_{1,4} = .8$, $p_{4,9} = .9$, and $p_{5,8} = .7$.

We define matches by pairing up adjacent players in the current ranking and then we promote the winners, thus simulating \mathcal{M}_{nn} .

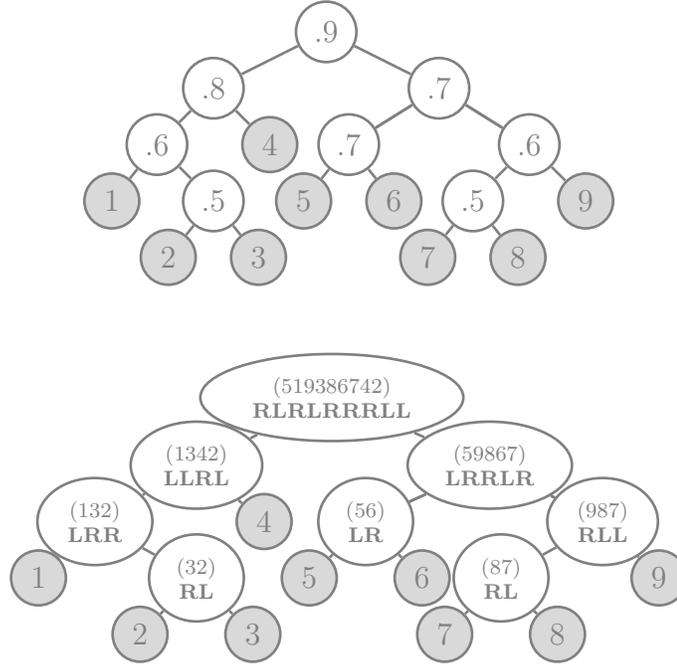


Figure 7: A set \mathbf{P} with tree structure and the corresponding tree-encoding of the permutation 519386742.

To show \mathcal{M}_{nn} is rapidly mixing for any input probabilities in the League Hierarchy class, we introduce a new combinatorial representation of each permutation that will be useful for the proofs. This representation associates a bit string b_v to each node v of a binary tree with n leaves. Specifically, $b_v \in \{L, R\}^{\ell_v}$ where ℓ_v is the number of leaves in t_v , the subtree rooted at v , and for each element i of the sub-permutation corresponding to the leaves of t_v , $b_v(i)$ records whether i lies under the left or the right branch of v (see Figure 7(b)). The set of these bit strings is in bijection with the permutations. We consider a chain $\mathcal{M}_{tree}(T)$ that allows transpositions when they correspond to a nearest neighbor transposition in exactly one of the bit strings. Thus, the mixing time of $\mathcal{M}_{tree}(T)$ decomposes into a product of $n - 1$ ASEP chains and we can conclude that the chain $\mathcal{M}_{tree}(T)$ is rapidly mixing using results in the constant bias case [6, 43]. Again, we use comparison techniques to conclude that \mathcal{M}_{nn} is also

rapidly mixing when we have weak monotonicity, although we show that $\mathcal{M}_{tree}(T)$ is always rapidly mixing.

3.5.1 Fast Mixing of the Binary Tree Chain

We define the Markov chain $\mathcal{M}_{tree}(T)$ over permutations, given set \mathbf{P} with league structure T .

The Markov chain $\mathcal{M}_{tree}(T)$

Starting at any permutation σ_0 , repeat:

- Select distinct $a, b \in [n]$ with $a < b$ u.a.r.
- If every number between a and b in the permutation σ_t is not a descendant in T of $a \wedge b$, obtain σ_{t+1} from σ_t by placing a, b in order with probability $p_{a,b}$, and out of order with probability $1 - p_{a,b}$, leaving all elements between them fixed.
- Otherwise, $\sigma_{t+1} = \sigma_t$.

First, we show that this Markov chain samples from the same distribution as \mathcal{M}_{nn} .

Lemma 3.5.1: The Markov chain $\mathcal{M}_{tree}(T)$ has the same stationary distribution as \mathcal{M}_{nn} .

PROOF: Let π be the stationary distribution of \mathcal{M}_{nn} , and let (σ_1, σ_2) be any transition in $\mathcal{M}_{tree}(T)$ such that $P_{tree}(\sigma_1, \sigma_2) > 0$ where P_{tree} is the transition matrix of $\mathcal{M}_{tree}(T)$. It suffices to show that the detailed balance condition holds for this transition with the stationary distribution π . Specifically we will show that

$$\pi(\sigma_1)P_{tree}(\sigma_1, \sigma_2) = \pi(\sigma_2)P_{tree}(\sigma_2, \sigma_1).$$

Recall that we may express $\pi(\sigma) = \left(\prod_{i < j: \sigma(i) < \sigma(j)} \frac{p_{i,j}}{p_{j,i}} \right) Z^{-1}$ where $Z = \sum_{\sigma \in \Omega} \pi(\sigma)$. The transition (σ_1, σ_2) transposes some two elements $a <_{\sigma_1} b$, where every element

between a and b in σ_1 (note that these are the same as the elements between a and b in σ_2) is not a descendant of $a \wedge b$ in T . Although swapping arbitrary non-adjacent elements could potentially change the weight of the permutation dramatically, for any element c that is not a descendant in T of $a \wedge b$ the relationship between a and c is the same as the relationship between b and c . Thus, the league structure ensure that swapping a and b only changes the weight by a multiplicative factor of $p_{a,b}/p_{b,a}$. Let $X = \{x_1, \dots, x_k\}$ be the elements between a and b in σ_1 . Thus, the path from a or b to x_i in T must pass through $a \wedge b$ and go to another part of the tree. For every such element x_i , $a \wedge x_i = (a \wedge b) \wedge x_i = b \wedge x_i$.

From the observation, we see from the league structure that $p_{a,x_i} = p_{b,x_i}$ for every x_i between a and b . Also, we see that either both $a < x_i, b < x_i$ or $a > x_i, b > x_i$, since all numbers between a, b are necessarily descendants of $a \wedge b$. Define $S = \{x \in X : x < a, b\}$ and $B = \{x \in X : x > a, b\}$.

Therefore,

$$\frac{\pi(\sigma_1)}{\pi(\sigma_2)} = \frac{p_{a,b} \prod_{x \in S} (p_{x,b}/p_{b,x}) \prod_{x \in B} (p_{a,x}/p_{x,a})}{p_{b,a} \prod_{x \in S} (p_{x,a}/p_{a,x}) \prod_{x \in B} (p_{b,x}/p_{x,b})} = \frac{p_{a,b}}{p_{b,a}}.$$

This is exactly the ratio of the transition probabilities in $\mathcal{M}_{tree}(T)$

$$\frac{\pi(\sigma_1)}{\pi(\sigma_2)} = \frac{p_{a,b}}{p_{b,a}} = \frac{P_{tree}(\sigma_2, \sigma_1)}{P_{tree}(\sigma_1, \sigma_2)},$$

thus the detailed balance condition is satisfied and $\mathcal{M}_{tree}(T)$ also has stationary distribution π . ■

The key to the proof that $\mathcal{M}_{tree}(T)$ is rapidly mixing is again to decompose the chain into $n - 1$ independent Markov chains, $\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_{n-1}$, one for each non-leaf node of the tree T . We introduce an alternate representation of a permutation as a set of binary strings arranged like the tree T . We use the characters L and R for our binary representation instead of 0 and 1 for convenience. For each non-leaf node v in the tree T , let $L(v)$ be its left descendants, and $R(v)$ be its right descendants. We now do the following:

- For each non-leaf node v do the following:
 - List each descendant x of v in the order we encounter them in the permutation σ . These are parenthesized in Figure 7(b).
 - For each listed element x , write a L if $x \in L(v)$ and a R if $x \in R(v)$. This is the final encoding in Figure 7(b).

We see that any σ will lead to an assignment of binary strings at each non-leaf node v with $L(v)$ L 's and $R(v)$ R 's. This is a bijection between the set of permutations and the set of assignments of binary strings to the tree T . Given any such assignment of binary strings, we can recursively reconstruct the permutation σ as follows:

- For each leaf node i , let its string be the string “ i ”.
- For any node n with binary string b ,
 - Determine the strings of its two children. Call these s_L, s_R .
 - Interleave the elements of s_L with s_R , choosing an element of s_L for each L in b , and an element of s_R for each R .

With this bijection, we first analyze $\mathcal{M}_{tree}(T)$'s behavior over tree representations and later extend this analysis to permutations. The Markov chain $\mathcal{M}_{tree}(T)$, when proposing a swap of the elements a and b , will only attempt to swap them if a, b correspond to some *adjacent* L and R in the string associated with $a \wedge b$. Swapping a and b does not affect any other string, so each non-leaf node v represents an independent exclusion process with $L(v)$ L 's and $R(v)$ R 's. These exclusion processes have been well-studied [17, 89, 6, 43]. We use the following bounds on the mixing times of the symmetric and asymmetric simple exclusion processes.

Theorem 3.5.2: Let \mathcal{M} be the exclusion process with parameter $1/2 \leq p < 1$ on a binary string of length k .

1. If $p > 1/2 + c$ for some positive constant c , then $\tau(\epsilon) = O(k^2 \log(\epsilon^{-1}))$. [6]
2. Otherwise, $\tau(\epsilon) = O(k^3 \log(k/\epsilon))$. [17]

The bounds in Theorem 3.5.2 refer to the exclusion process which selects a position at random and swaps the two elements in that position with the appropriate probability.

Since each exclusion process \mathcal{M}_i operates independently, the overall mixing time will be roughly n times the mixing time of each piece, slowed down by the inverse probability of selecting that process. Each \mathcal{M}_i has a different size, and a different mixing time relative to its size. To employ the bounds from Theorem 3.5.2, we will use Theorem 3.4.2, which relates the mixing time of a product of independent Markov chains to the mixing time of each component. Finally, we can prove that $\mathcal{M}_{tree}(T)$ is rapidly mixing.

Theorem 3.5.3: Given input probabilities $1/2 \leq q_1, q_2, \dots, q_{n-1} < 1$ let $\mathbf{P} = \{p_{i,j} = q_{i \wedge j}\}$.

1. If $\forall 1 \leq i \leq n-1, c + 1/2 < q_i$ for some positive constant c , the mixing time of $\mathcal{M}_{tree}(T)$ with preference set \mathbf{P} satisfies

$$\tau(\epsilon) = O(n^3 \log(n\epsilon^{-1})).$$

2. Otherwise, the mixing time of $\mathcal{M}_{tree}(T)$ with preference set \mathbf{P} satisfies

$$\tau(\epsilon) = O(n^4 \log(n\epsilon^{-1})).$$

PROOF: In order to apply Theorem 3.4.2 to the Markov chain $\mathcal{M}_{tree}(T)$, we note that for a node whose associated bit string has length k , the probability of selecting a move that corresponds to two neighboring bits in the string is $\frac{\binom{k-1}{2}}{\binom{n}{2}} = \frac{k-1}{2n(n-1)}$. Combining the first bound from Theorem 3.5.2 with Theorem 3.4.2 where $N = n-1$ gives the following result,

$$\tau(\epsilon) = O\left(\frac{n(n-1)}{k-1} k^3 \log(2(n-1)k/\epsilon)\right) = O(n^4 \log(n\epsilon^{-1})).$$

If all of the chains have probabilities that are bounded away from $1/2$, then we can use the second bound from Theorem 3.5.2 to obtain

$$\tau(\epsilon) = O\left(\frac{n(n-1)}{k-1}k^2 \log(2(n-1)/\epsilon)\right) = O(n^3 \log(n\epsilon^{-1})).$$

3.5.2 Fast Mixing of the Nearest-Neighbor Chain

Next, we show that \mathcal{M}_{nn} is rapidly mixing when \mathbf{P} has league structure and is *weakly monotone*:

Definition 3.5.1: The set \mathbf{P} is *weakly monotone* if properties 1 and *either* 2 or 3 are satisfied.

1. $p_{i,j} \geq 1/2$ for all $1 \leq i < j \leq n$, and
2. $p_{i,j+1} \geq p_{i,j}$ for all $1 \leq i < j \leq n-1$ or
3. $p_{i-1,j} \geq p_{i,j}$ for all $2 \leq i < j \leq n$.

We note that if \mathbf{P} satisfies all three properties then it is *monotone*, as defined by Jim Fill [35].

The comparison proof in this setting is similar to the comparison proof in Section 2.3, except we allow elements between $\sigma(i)$ and $\sigma(j)$ that are larger or smaller than both i and j . This poses a problem, because we may not be able to move $\sigma(j)$ towards $\sigma(i)$ without greatly decreasing the weight. However, we can resolve this if \mathbf{P} is weakly monotone. Specifically, we are now ready to prove the following theorem.

Theorem 3.5.4: Given input probabilities $1/2 \leq q_1, q_2, \dots, q_{n-1} < 1$ and a positive constant c such that $r_i < 1 - c$ for $1 \leq i \leq n-1$ and $\mathbf{P} = \{p_{i,j} = q_{i \wedge j}\}$ is weakly monotone.

1. If $\forall 1 \leq i \leq n-1, c + 1/2 < r_i$ the mixing time of \mathcal{M}_{nn} with preference set \mathbf{P} satisfies

$$\tau(\epsilon) = O(n^7 \log(n\epsilon^{-1}) \log(\epsilon^{-1})).$$

2. Otherwise, the mixing time of \mathcal{M}_{nn} with preference set \mathbf{P} satisfies

$$\tau(\epsilon) = O(n^8 \log(n\epsilon^{-1}) \log(\epsilon^{-1})).$$

PROOF: Throughout this proof we assume that \mathbf{P} satisfies properties 1 and 2 of the weakly monotone definition. If instead \mathbf{P} satisfies property 3, then the proof is almost identical. In order to apply Theorem 2.3.1 to relate the mixing time of \mathcal{M}_{nn} to the mixing time of $\mathcal{M}_{tree}(T)$ we need to define for each transition of $\mathcal{M}_{tree}(T)$ a canonical path using transitions of \mathcal{M}_{nn} . Let $e = (\sigma, \beta)$ be a transition of $\mathcal{M}_{tree}(T)$ which performs a transposition on elements $\sigma(i)$ and $\sigma(j)$. If there are no elements between $\sigma(i)$ and $\sigma(j)$ then e is already a transition of \mathcal{M}_{nn} and we are done. Otherwise, σ contains the string $\sigma(i), \sigma(i+1), \dots, \sigma(j-1), \sigma(j)$ and β contains $\sigma(j), \sigma(i+1), \dots, \sigma(j-1), \sigma(i)$. From the definition of $\mathcal{M}_{tree}(T)$ we know that for each $\sigma(k)$, $k \in [i+1, j-1]$, either $\sigma(k) > \sigma(i), \sigma(j)$ or $\sigma(k) < \sigma(i), \sigma(j)$. Define $S = \{\sigma(k) : \sigma_k < \sigma(i), \sigma(j)\}$ and $B = \{\sigma(k) : \sigma_k > \sigma(i), \sigma(j)\}$. To obtain a good bound on the congestion along each edge we must ensure that the weight of the configurations on the path are not smaller than the weight of σ . Thus, we define three stages in our path from σ to β . In the first, we shift the elements of S to the left, removing an inversion with each element of B . In the second stage we move $\sigma(i)$ next to $\sigma(j)$ and in the third stage we move $\sigma(j)$ to $\sigma(i)$'s original location. Finally, we shift the elements of S to the right to return them to their original locations. See Figure 8.

Stage 1: At a high-level in this stage we are shifting the elements in S to the left

Stage 1	5	8	9	<u>2</u>	10	<u>3</u>	<u>4</u>	<u>1</u>	7
Stage 2	5	<u>2</u>	8	9	<u>3</u>	10	<u>4</u>	<u>1</u>	7
	<u>2</u>	8	9	<u>3</u>	10	<u>4</u>	<u>1</u>	5	7
Stage 3	<u>2</u>	8	9	<u>3</u>	10	<u>4</u>	<u>1</u>	7	5
Stage 4	7	<u>2</u>	8	9	<u>3</u>	10	<u>4</u>	<u>1</u>	5
	7	8	9	<u>2</u>	10	<u>3</u>	<u>4</u>	<u>1</u>	5

Figure 8: The stages in the canonical path for transposing 5 and 7. Notice that the elements in S are underlined.

$\sigma(l) \in B$.

Stage 4: Finally we want to return the elements in S and B to their original position. Starting with the left-most element in S that was moved in Stage 1, perform the nearest neighbor swaps to the right necessary to return it to its original position. Finally, if originally $\sigma(j-1) \in B$, then move $\sigma(i)$ to the original location of $\sigma(j)$. Note that if originally $\sigma(j-1) \notin B$, then $\sigma(i)$ was placed in the original location of $\sigma(j)$ at the end of Stage 2. It is clear from the definition of the stages that the weight of a configuration never decreases below the weight of $\min(\pi(\sigma), \pi(\beta))$.

Given a transition (v, ω) of \mathcal{M}_{mn} we must upper bound the number of canonical paths $\gamma_{\sigma\beta}$ that use this edge. Thus, we analyze the amount of information needed in addition to (z, w) to determine σ and β uniquely. First we record whether (σ, β) is already a nearest neighbor transition or which stage we are in. Next for any of the 4 stages we record the original location of $\sigma(i)$ and $\sigma(j)$. Given this information, along with v and ω , we can uniquely recover (σ, β) . Hence, there are at most $4n^2$ paths through any edge (v, ω) . Also, note that the maximum length of any path is $4n$.

Next we bound the quantity A which is needed to apply Theorem 2.3.1. Recall that we have guaranteed that $\pi(\sigma) \leq \max\{\pi(v), \pi(\omega)\}$. Assume that $\pi(\sigma) \leq \pi(v)$. Let $\lambda = \max_{i < j} p_{i,j}/p_{j,i}$. Then

$$\begin{aligned} A &= \max_{(v,\omega) \in E(P)} \left\{ \frac{1}{\pi(v)P(v,\omega)} \sum_{\Gamma(v,\omega)} |\gamma_{\sigma\beta}| \pi(\sigma) P'(\sigma, \beta) \right\} \\ &\leq \max_{(v,\omega) \in E(P)} \sum_{\Gamma(v,\omega)} 2n \frac{P'(\sigma, \beta)}{P(v, \omega)} \\ &\leq \max_{(v,\omega) \in E(P)} \sum_{\Gamma(v,\omega)} 2n \frac{1/\binom{n}{2}}{(1+\lambda)(n-1)} = O(n^2). \end{aligned}$$

If, on the other hand, $\pi(\sigma) \leq \pi(\omega)$, then we use detailed balance to obtain:

$$\begin{aligned}
A &= \max_{(v,\omega) \in E(P)} \left\{ \frac{1}{\pi(v)P(v,\omega)} \sum_{\Gamma(v,\omega)} |\gamma_{\sigma\beta}| \pi(\sigma) P'(\sigma, \beta) \right\} \\
&= \max_{(v,\omega) \in E(P)} \left\{ \frac{1}{\pi(\omega)P(\omega, v)} \sum_{\Gamma(v,\omega)} |\gamma_{\sigma\beta}| \pi(\sigma) P'(\sigma, \beta) \right\} \\
&\leq \max_{(v,\omega) \in E(P)} \sum_{\Gamma(v,\omega)} 2n \frac{P'(\sigma, \beta)}{P(\omega, v)} \\
&\leq \max_{(v,\omega) \in E(P)} \sum_{\Gamma(v,\omega)} 2n \frac{1/\binom{n}{2}}{\frac{1}{(1+\lambda)(n-1)}} = O(n^2). \quad \blacksquare
\end{aligned}$$

In either case, we have $A = O(n^2)$. Then $\pi_* = \min_{\rho \in \Omega} \pi(\rho) \geq (\lambda \binom{n}{2} n!)^{-1}$ where λ is defined as above, so $\log(1/(\epsilon \pi_*)) = O(n^2 \log \epsilon^{-1})$, as above.. Appealing to Theorem 2.3.1 and the first bound from Theorem 3.5.3 if the q_i 's are bounded away from $1/2$, we have that the mixing time of \mathcal{M}_{nn} satisfies

$$\tau(\epsilon) = O(n^7 \log(n\epsilon^{-1}) \log(\epsilon^{-1})).$$

Similarly, appealing to Theorem 2.3.1 and the second bound from Theorem 3.5.3 we have that the mixing time of \mathcal{M}_{nn} satisfies

$$\tau(\epsilon) = O(n^8 \log(n\epsilon^{-1}) \log(\epsilon^{-1})).$$

Remark: By repeating Stage 1 of the path a constant number of times, it is possible to relax the weakly monotone condition slightly if we are satisfied with a polynomial bound on the mixing time.

Remark: If the input probabilities are not bounded away from 1 by a constant but instead by some function of n , then using the same proof as Theorem 3.5.4 we can obtain a bound on the mixing time. Specifically, given input probabilities $1/2 \leq r_1, r_2, \dots, r_{n-1} < 1 - 1/f(n)$ let $\mathbf{P} = \{p_{i,j} = r_{\min\{i,j\}}\}$. Then the mixing time of \mathcal{M}_{nn} with preference set \mathbf{P} satisfies

$$\tau(\epsilon) = O(n^8 f(n) \log(n\epsilon^{-1}) \log(f(n)\epsilon^{-1})).$$

If additionally, there exists a positive constant c such that $r_i > 1/2 + c$ for all $1 \leq i \leq n - 1$, then the mixing time of \mathcal{M}_{nn} with preference set \mathbf{P} satisfies

$$\tau(\epsilon) = O(n^7 f(n) \log(n\epsilon^{-1}) \log(f(n)\epsilon^{-1})).$$

CHAPTER IV

DYADIC TILINGS AND RECTANGULAR DISSECTIONS

In this chapter, we study *rectangular dissections* of an $n \times n$ lattice region into rectangles of area n , where $n = 2^k$ for an even integer k . We consider a natural edge-flipping Markov chain whose mixing time is open both for rectangular dissections and when the state space is restricted to *dyadic tilings*, where each rectangle is required to have the form $R = [s2^u, (s+1)2^u] \times [t2^v, (t+1)2^v]$, where s, t, u and v are nonnegative integers.

We introduce a biased version of these Markov chains where, given a parameter $\lambda > 0$, we would like to generate each rectangular dissection (or dyadic tiling) σ with probability proportional to $\lambda^{|\sigma|}$, where $|\sigma|$ is the total edge length. We show there is a phase transition in the dyadic setting: when $\lambda < 1$, the edge-flipping chain mixes in time $O(n^2 \log n)$, and when $\lambda > 1$, the mixing time is $\exp(\Omega(n^2))$. Simulations suggest that the chain converges quickly when $\lambda = 1$, but this case remains open. The behavior for general rectangular dissections is more subtle, and even establishing ergodicity of the chain requires a careful inductive argument. As in the dyadic case, we show that the edge-flipping Markov chain for rectangular dissections requires exponential time when $\lambda > 1$. Surprisingly, the chain also requires exponential time when $\lambda < 1$, which we show using a different argument. Simulations suggest that the chain converges quickly at the isolated point $\lambda = 1$.

4.1 *Rectangular Dissections and Dyadic Tilings*

Rectangular dissections arise in the study of VLSI layout [23], mapping graphs for floor layouts [67, 87], and routings and placements [93] and have long been of interest to combinatorialists [15, 84]. In each of these applications, a lattice region needs to be

partitioned into rectangles whose corners lie on lattice points such that the dissection satisfies some appropriate additional constraints. For example, *equitable rectangular dissections* require that all rectangles in the partition have the same area [45] (see Figure 10). We are interested in understanding what random equitable rectangular dissections look like as well as finding efficient methods for sampling these dissections.

There has also been interest in the special case of *dyadic tilings*, or equitable rectangular dissections into dyadic rectangles. A *dyadic rectangle* is a set of the form

$$R = [s2^u, (s + 1)2^u] \times [t2^v, (t + 1)2^v]$$

where s, t, u and v are nonnegative integers. A dyadic tiling of the $2^k \times 2^k$ square is a set of 2^k dyadic rectangles, each of area 2^k , whose union is the full square. See Figure 10(b). Janson et al. [48] studied the asymptotics A_k , the number of dyadic tilings of the $2^k \times 2^k$ square where $k \in \mathbf{Z}^+$. They show that every dyadic tiling must have a fault line, that is, a line bisecting the square in the vertical or horizontal direction which avoids non-trivial intersection with all rectangles in the tilings. This allows them to derive the recurrence $A_k = 2A_{k-1}^2 - A_{k-2}^4$ and show that asymptotically $A_k \sim \phi^{-1}\omega^{2^k}$, where $\phi = (1 + \sqrt{5})/2 = 1.6180\dots$ is the golden ratio and $\omega = 1.84454757$ is a constant.

Although equitable partitions of lattice regions into rectangles or triangles have been extensively studied, many fundamental questions remain open. A notable exception is dissections into rectangles with area 2, commonly known as *domino tilings* or the *dimer model* from statistical physics. Researchers have discovered remarkable properties of these tilings, including striking underlying combinatorial structures [52], statistical properties of random tilings [53], and analysis showing various Markov chains for generating them are efficient [33, 59, 70].

Triangular dissections have been explored extensively as well, both when the vertices are in general position and when they are vertices of a planar lattice. On the

Cartesian lattice \mathbf{Z}^2 , the problem becomes finding equitable (or *unimodular*) triangulations of a lattice region, where each triangle has area $1/2$. See [57] for an extensive history of work on triangulations.

Interestingly, in each of these cases, a certain “edge-flip” Markov chain has been identified that connects the state space of allowable dissections. For example, for domino tilings, the Markov chain iteratively removes a length 2 edge bordering two dominoes and replaces it with a length 2 edge in the orthogonal direction, effectively replacing two vertical dominoes with two horizontal ones, or vice versa. This chain is known to be rapidly mixing [59, 71, 89]. In the case of dyadic tilings, there is again a natural edge-flip chain that connects the set of possible configurations – if there are two neighboring rectangles in the tiling that share an edge, we can remove that edge and retiling the larger composite rectangle with the edge that bisects it in the orthogonal direction, provided the new tiling is still dyadic (see Figure 12(c),(d)). The mixing rate of this edge-flip chain was left open in [48], although the authors argue that a different, nonlocal, Markov chain containing additional moves does converge quickly to equilibrium.

Another edge-flip chain also connects the state space of triangulations by replacing an edge bordering two triangles with the edge connecting the other two vertices if the quadrilateral formed by their union is convex. The edge-flip chain on triangulations of general point sets has been the subject of much interest in the computational geometry

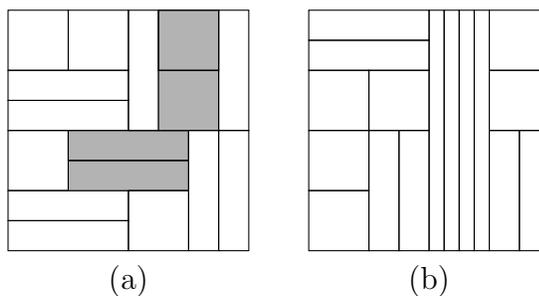


Figure 10: (a) An equitable rectangular dissection and (b) a dyadic tiling of the 16×16 square. Shaded rectangles are not dyadic.

community (see, e.g., [88]). In the unweighted case the chain has only been analyzed when the points are in convex position [63, 66], in which case the triangulations are enumerated by the Catalan numbers.

Recently, Caputo et al. [21] introduced a weighted version of the lattice triangulation dissection problem and discovered remarkable behavior. Each triangulation σ on a finite region of \mathbf{Z}^2 is assigned a weight $\lambda^{|\sigma|}$, where $\lambda > 0$ is some input parameter. They conjecture there is a phase transition at $\lambda = 1$ and that when $\lambda < 1$ there are no long-range correlations of the triangles and Markov chains based on local edge flips converge in polynomial time, while when $\lambda > 1$ there will be large regions of aligned long-thin triangles and local Markov chains will require exponential convergence time. They verify this conjecture when $\lambda > 1$ and when $\lambda < \lambda_0 < 1$ for some suitably small constant λ_0 . Their conjecture is supported by the intuition that when λ is large, triangulations with many long-thin triangles will be favored, and the geometry will force these triangles to align in the same direction. In contrast, when $\lambda < 1$, triangles with large aspect ratio will be preferred, the chain will be rapidly mixing, and there will not be any long-range order.

4.1.1 Results.

In this chapter, we study a weighted version of the equitable rectangular dissection problem and explore the mixing time of an appropriate edge-flip Markov chain. Let $n = 2^k$, for k an even integer, and let Λ_n be the $n \times n$ lattice region. We will be considering rectangular dissections of Λ_n into rectangles of area n in the dyadic and general cases. Let Ω_n be the set of dyadic tilings of Λ_n and let $\widehat{\Omega}_n$ be the set of rectangular dissections of Λ_n into rectangles of area n that are not necessarily dyadic. In the weighted setting, we are given an input parameter $\lambda > 0$ and the weight of a dyadic tiling $\sigma \in \Omega_n$ is $\pi(\sigma) = \lambda^{|\sigma|}/Z$, where $|\sigma|$ is the total length of edges in σ and $Z = \sum_{\sigma \in \Omega_n} \lambda^{|\sigma|}$ is the normalizing constant known as the partition function.

Likewise, in the general dissection setting, for $\alpha \in \widehat{\Omega}_n$, we define $\widehat{\pi}(\alpha) = \lambda^{|\alpha|} / \widehat{Z}$, where $|\alpha|$ is the total length of α and \widehat{Z} is again the normalizing constant.

Let \mathcal{M}_n be the edge-flip Markov chain on Ω_n that replaces an edge bordering two rectangles with the perpendicular bisector of the combined area $2n$ rectangle, provided the resulting tiling remains dyadic (details are given in Section 2.). It is easy to generalize this chain to the weighted setting by modifying the transition probabilities so that the chain converges to distribution π . Likewise, we can define the natural generalization of the edge-flip chain $\widehat{\mathcal{M}}_n$ on $\widehat{\Omega}_n$ by connecting two dissections if they differ by the the removal and addition of one edge.

The remainder of the chapter will be concerned with the mixing times of \mathcal{M}_n and $\widehat{\mathcal{M}}_n$ as we vary the parameter λ . One might expect the same behavior for weighted rectangular dissections as in the triangulation case, namely that when λ is small we favor balanced rectangles and we might expect the chain to be rapidly mixing, while for λ large we favor long thin rectangles, and we should expect they will mostly align vertically or horizontally. This picture is actually much more complicated in the general case, but precisely what we find in the dyadic setting. In addition, in the dyadic case we have succeeded in closing the gap between the regimes for fast and slow mixing, and prove that there is a phase transition at $\lambda = 1$. The analogous result was only conjectured for triangulations in [21]. Specifically, we prove the following two theorems that establish that the phase transition occurs at $\lambda = 1$ for dyadic tilings.

Theorem 4.1.1: For any constant $\lambda < 1$, the edge-flip chain \mathcal{M}_n on Ω_n converges in time $O(n^2 \log n)$.

Theorem 4.1.2: For any constant $\lambda > 1$, the edge-flip chain \mathcal{M}_n on Ω_n requires time $\exp(\Omega(n^2))$.

Simulations suggest that the chain \mathcal{M}_n is also fast when $\lambda = 1$. See the left column

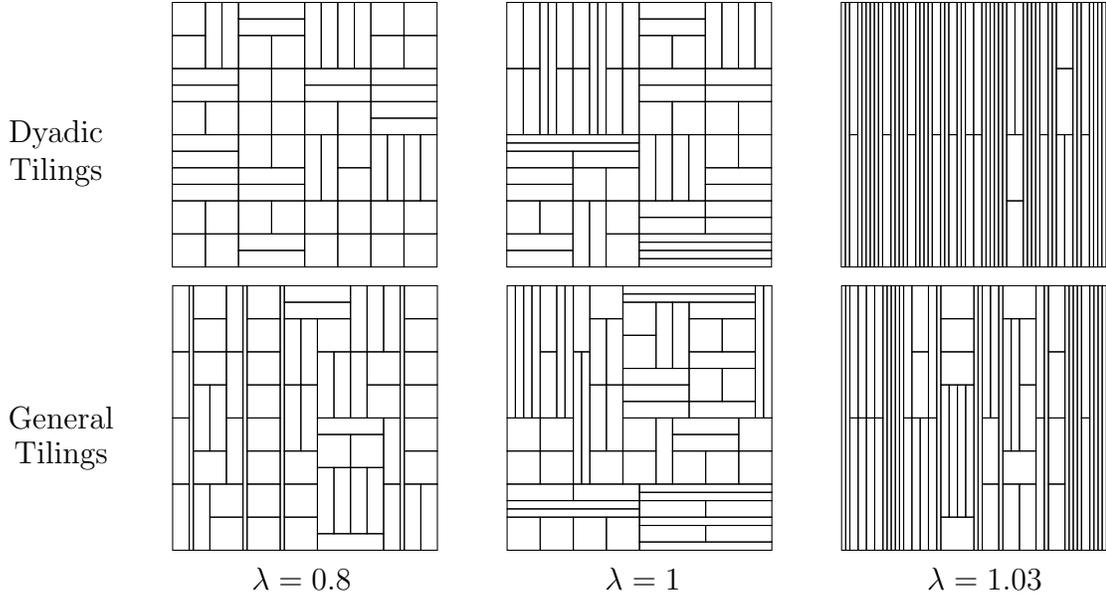


Figure 11: \mathcal{M}_{64} and $\widehat{\mathcal{M}}_{64}$ after 1,000,000 simulated steps for various values of λ , starting with all vertical rectangles of width 1 and height 64.

of Figure 11 for samples generated with various values of λ for \mathcal{M}_{64} .

In the general setting the picture is more surprising. When λ is large, we get the expected results confirming that the Markov chain $\widehat{\mathcal{M}}_n$ requires exponential time. However, we show that the chain also requires exponential time to converge to equilibrium when λ is small, as the following two theorems state.

Theorem 4.1.3: For any constant $\lambda > 1$, the edge-flip chain $\widehat{\mathcal{M}}_n$ on $\widehat{\Omega}_n$ requires time $\exp(\Omega(n^2))$.

Theorem 4.1.4: For any constant $\lambda < 1$, the edge-flip chain $\widehat{\mathcal{M}}_n$ on $\widehat{\Omega}_n$ requires time $\exp(\Omega(n \log n))$.

Even though together these results seem to suggest that the chain will always be slow, the proofs in these two regimes (i.e., $\lambda < 1$ and $\lambda > 1$) show that the reasons underlying the slow mixing results are quite different. When $\lambda > 1$ long thin rectangles are favored, and it will take exponential time to move from a configuration that is predominantly horizontal to one that is vertical. When $\lambda < 1$ “balanced”

rectangles that are close to square are favored. This is enough to dramatically speed up the mixing time in the dyadic case, but in the general setting it causes an obstacle because long thin rectangles that are well separated by many squares (or near squares) will take exponential time to disappear since their removal requires the creation of more long-thin rectangles, and their creation is exponentially unlikely. Both slow mixing proofs when $\lambda > 1$ show that there is a bad cut in an equitable partition of the state space into two equal sized pieces, but the proof in the general setting when $\lambda < 1$ relies critically on a careful choice of the starting configuration. It may indeed be the case that the chain is fast if we start from the most favorable configuration consisting entirely of squares. As before, the convergence time is unknown when $\lambda = 1$, but based on simulations we conjecture that the chain $\widehat{\mathcal{M}}_n$ converges quickly to equilibrium at this isolated point (see the right column of Figure 11).

We note that these results for dyadic tilings are complementary to other phase transitions discovered in the unweighted setting. Angel et al. [2] affirmatively answered a question of Joel Spencer regarding the probability that there is a dyadic tiling if each dyadic rectangle is present with probability p , independent of the others. They show that there is a phase transition for some $p < 1$, at which point the likelihood of there not being such a tiling becomes exponentially small.

4.1.2 Techniques.

Dyadic tilings have rich combinatorial properties that allow us to establish the presence of a phase transition in the convergence times. The proof of fast mixing of \mathcal{M}_n on dyadic tilings when $\lambda < 1$ is based on the method of exponential metrics for path coupling. Similar techniques have been used by Greenberg et al. [43] for lattice paths and Caputo et al. [21] for weighted triangulations, but both of these proofs rely on analysis of lattice paths. Here our proof uses a more traditional analysis based on path coupling by directly analyzing configurations of rectangles. It is worth noting

that the analysis is self-contained and does not rely on computational tools to optimize the weights used in the calculations. We show that \mathcal{M}_n will be rapidly mixing for all $\lambda < 3^{-1/\sqrt{n}}$, which is sufficient to prove fast mixing for any $\lambda < 1$, when n is sufficiently large.

To show slow mixing of the Markov chains \mathcal{M}_n and $\widehat{\mathcal{M}}_n$ in the dyadic and general cases when $\lambda > 1$, we apply a standard Peierls argument. Here, a straightforward analysis suffices to show that configurations without horizontal or vertical long thin rectangles must have exponentially small weight, even after summing over all such configurations. Since we must pass through these very unlikely configurations to move from a mostly horizontal configuration to a mostly vertical one, we can conclude that the mixing time is exponential using a basic flow argument.

The proof of slow mixing for general rectangular dissections when $\lambda < 1$ is considerably more delicate. In this regime, rectangles that are close to square are preferred. We show that it will take exponential time to move from a configuration that has two well-separated long thin rectangles to one that does not have any long thin rectangles by very carefully analyzing required features of these tilings. If the total width of the region being filled with rectangles is $n = 2^k$, and there are at least two rectangles with width 1, then there must be many other thin rectangles in the rectangular dissection. We define the cut set to consist of rectangular dissections that are forced to have significantly more thin rectangles in order to show that there is a bad cut in the state space.

4.2 Preliminaries

We start by formalizing the problems. In the remainder of this chapter, we will refer to equitable rectangular dissections instead as *tilings* in analogy to the widely used designation *dyadic tilings* to provide a uniformity of language.

Let $n = 2^k$ for some even integer k . An n -tiling is a tiling of the $[0, n] \times [0, n]$

lattice Λ_n by n axis-aligned rectangles, each of area n ; see Figure 10. We assume all rectangles are the Cartesian product of two closed intervals, $R = [x_1, x_2] \times [y_1, y_2]$, and are of dimension $2^a \times 2^b$, where $a, b \in \{0, 1, 2, \dots, k\}$ and $a + b = k$. That k is even implies n is a perfect square and there exists a “ground state” tiling consisting entirely of $\sqrt{n} \times \sqrt{n}$ squares; this is critical to the proof of Theorem 4.1.1. A tiling is *dyadic* if all rectangles are of the form $[s2^u, (s + 1)2^u] \times [t2^v, (t + 1)2^v]$ for some nonnegative integers s, t, u, v . We will use the following lemma.

Lemma 4.2.1: For any $a \in \{0, \dots, n - 1\}$ and $b \in \{1, \dots, k - 2\}$, at most one of $[a, a + 2 \cdot 2^b]$ and $[a + 2^b, a + 3 \cdot 2^b]$ can be written in the form $[s2^u, (s + 1)2^u]$ for some nonnegative integers s and u .

PROOF: Suppose $[a, a + 2 \cdot 2^b] = [s2^u, (s + 1)2^u]$ and $[a + 2^b, a + 3 \cdot 2^b] = [t2^v, (t + 1)2^v]$ for some nonnegative integers s, t, u, v . Looking at the first equation, $u = b + 1$ and $a = s2^u = s2^{b+1}$. From the second equation, $v = b + 1$ and $a + 2^b = t2^v = t2^{b+1}$. It then follows that

$$2^b = (a + 2^b) - a = t2^{b+1} - s2^{b+1} = (t - s)2^{b+1}.$$

This is impossible as $t - s$ is integer. ■

4.2.1 Formalizing the Markov Chains

We study two related Markov chains \mathcal{M}_n and $\widehat{\mathcal{M}}_n$ whose state spaces Ω_n and $\widehat{\Omega}_n$, respectively, are all dyadic n -tilings and all n -tilings. Moves in these Markov chains consist of edge flips, which we now define. By an *edge*, we mean a boundary between two adjacent rectangles in a tiling. Two tilings σ_1, σ_2 differ by exactly one *edge flip* if it is possible to remove an edge in σ_1 that bisects a rectangle of area $2n$ and replace it with the bisecting edge in the perpendicular orientation to form σ_2 . For example, in Figure 12, tilings (a) and (b) differ by a single edge flip, as do tilings (c) and (d). We say an edge e is *flippable* if it bisects a rectangle of area $2n$.

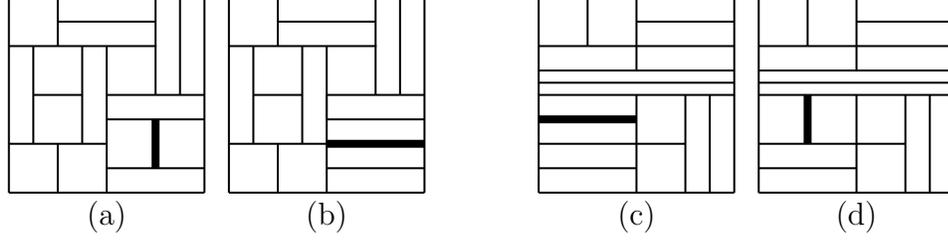


Figure 12: Some tilings for $n = 16$. Tilings (a) and (b) differ by an edge flip. Dyadic tilings (c) and (d) differ by an edge flip.

We consider biased Markov chains with a bias $\lambda \in (0, \infty)$, analogous to [21]. For a tiling σ , let $|\sigma|$ denote the sum of the lengths of all the edges in σ . First, we define the Markov chain $\widehat{\mathcal{M}}_n$ with bias λ . Note all logarithms are assumed to be base 2. Starting at any tiling σ_0 , iterate:

- Choose, uniformly at random, $(x, y, d, o, p) \in$

$$\left\{ \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \dots, \frac{2n-1}{2} \right\} \times \left\{ \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \dots, \frac{2n-1}{2} \right\} \\ \times \{t, l, b, r\} \times \{0, 1\} \times (0, 1).$$

Let R be the rectangle in σ_t containing (x, y) . If $d = t$, let e be the top boundary of R ; if $d = l, b$, or r , let e be the left, bottom, or right boundary of R , respectively.

- If e is a flippable edge and $\log |e| \equiv o \pmod{2}$, let σ' be the tiling obtained by flipping e to new edge e' . If $p < \lambda^{|\sigma'| - |\sigma_t|} = \lambda^{|e'| - |e|}$, then $\sigma_{t+1} = \sigma'$.
- Else, $\sigma_{t+1} = \sigma_t$.

The Markov chain \mathcal{M}_n for dyadic tilings is defined in the same way, interpreting “flippable” to mean flippable into another dyadic tiling; Figure 12 (c) and (d) shows an edge flip between two dyadic tilings that are adjacent in Ω_n .

We note that each rectangle R of any tiling σ is of area n and so contains exactly n points in $\{\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \dots, \frac{2n-1}{2}\} \times \{\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \dots, \frac{2n-1}{2}\}$. A given flippable edge e in σ is thus selected by $2n$ different values of (x, y, d, o) , specifically, the $2n$ points (x, y) in the two

rectangles e separates, each with the appropriate value of d and o . Consequently, a given flippable edge e is selected by (x, y, d, o) with probability $2n \cdot \frac{1}{n^2} \cdot \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{4n} =: q$. This flip then occurs with probability $\min\{1, \lambda^{|\sigma'| - |\sigma|} = \lambda^{|e'| - |e|}\}$, according to the random value of p . These transition probabilities favor long, thin rectangles when $\lambda > 1$ and favor squares or rectangles close to square when $\lambda < 1$. At most one of $(x, y, d, 0, p)$ and $(x, y, d, 1, p)$ results in an edge flip; (x, y, d) selects a potentially flippable edge e in σ_t , and then an edge flip can only occur if the length of e satisfies $\log |e| = o(\text{mod } 2)$. This implies both \mathcal{M}_n and $\widehat{\mathcal{M}}_n$ are lazy and thus aperiodic.

4.2.2 Proving Ergodicity

It remains to be shown that the moves described above connect state spaces Ω_n and $\widehat{\Omega}_n$. Connectivity for Ω_n follows from work on dyadic tilings in [48], specifically from their tree representation of a dyadic tiling. Dyadic constraints ensure rectangles exist in pairs; an edge flip is always possible for every rectangle. In particular, all $1 \times n$ and $n \times 1$ rectangles are adjacent to at least one other rectangle of the same dimensions, so can be eliminated with a single edge flip.

However, connectivity of $\widehat{\Omega}_n$ is much less straightforward, and an interesting result in its own right. Intuitively, issues arise because rectangles in a general n -tiling do not exist in pairs and there may be many rectangles for which no edge flip is possible; it is not even immediately evident that there is a single valid edge flip. Rectangles of height n , or alternately, rectangles of height h where there are no rectangles of larger height, may be well separated by complicated arrangements of tiles. It is not clear how to introduce another rectangle of height h next to an existing rectangle of height h so that both may be eliminated, a necessary step for obtaining a tiling with no rectangles of height h or larger, for instance. A proof showing connectivity of $\widehat{\Omega}_n$, can be found in [20].

From connectivity, it follows that $\widehat{\mathcal{M}}_n$ and \mathcal{M}_n are irreducible and thus ergodic,

so they converge to unique stationary distributions $\widehat{\pi}$ and π , respectively. By detailed balance, the distribution $\widehat{\pi}$ can be given by $\widehat{\pi}(\sigma) = \lambda^{|\sigma|}/\widehat{Z}$, where \widehat{Z} is the normalizing constant. Similarly, $\pi(\sigma) = \lambda^{|\sigma|}/Z$, where Z is the normalizing constant.

4.3 *Fast Mixing for Dyadic Tilings when $\lambda < 1$.*

We prove \mathcal{M}_n is rapidly mixing for all $\lambda < 3^{-1/\sqrt{n}}$. This bound approaches 1 as n grows, so for any $\lambda < 1$ there is sufficiently large n such that the Markov chain \mathcal{M}_n is rapidly mixing. To give some perspective, we note that for all $n \geq 4$, we have fast mixing for all $\lambda < 0.577$, a much better constant than obtained in [21]. Already for $n \geq 1024$ we have fast mixing for all $\lambda < 0.966$. We use a path coupling argument and an exponential metric, as in [43], to prove Theorem 4.1.1.

We now seek to apply the exponential metric theorem of [43], described in full in Section 2.2. Intuitively, we consider the subset U of the joint state space $\Omega_n \times \Omega_n$ of tilings that differ by one edge flip. The main result we need to show is that for any coupling whose joint state is two configurations in U , after one iteration of the Markov chain, the expected distance between the two coupled chains decreases by a constant factor of their original distance. It is crucial to define the appropriate notion of “distance” between two tilings.

Suppose $\lambda < 3^{-1/\sqrt{n}}$. Consider any dyadic tilings σ_1 and σ_2 that differ by one flip between edge e and edge f , both bisecting a common area $2n$ rectangle S . Without loss of generality, suppose that $|e| \geq |f|$. We define the distance between σ_1 and σ_2 to be

$$\phi(\sigma_1, \sigma_2) = \phi(\sigma_2, \sigma_1) := \lambda^{|f|-|e|} \geq 1,$$

and similarly for all other adjacent tilings in Ω_n . We note that the distance between any two adjacent pairs is at least one. For any two tilings σ and σ' that are not adjacent in Ω_n , the distance between them is the minimum over all paths in Ω_n from σ to σ' of the sum of the distances between adjacent tilings along the path, also at

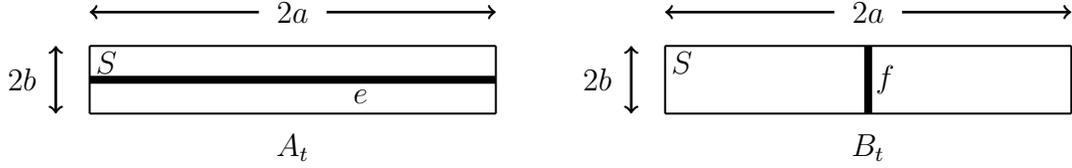


Figure 13: Rectangle S of area $2n$ in marginal tilings A_t and B_t .

least one.

Formally, let $(\mathcal{A}, \mathcal{B})$ denote a coupling of \mathcal{M}_n , where A_t and B_t are the states of the two chains, respectively, after t iterations. Let $\phi_t = \phi_t(A_t, B_t)$ be the distance between the two chains in the coupling $(\mathcal{A}, \mathcal{B})$ after t iterations. Suppose, without loss of generality, A_t and B_t differ by a single flip between edge e and edge f , where $|e| \geq |f|$, e is horizontal in A_t of length $2a$, f is vertical in B_t of length $2b$, and both bisect a rectangle S of area $2n$; see Figure 13.

We wish to bound $e[\phi_{t+1} - \phi_t]$ in terms of ϕ_t . Any potential moves (x, y, d, o, p) that select an edge not in S or on the boundary of S have the same effect on both A_t and B_t and thus, in these cases, $\phi_{t+1} = \phi_t$, as A_{t+1} and B_{t+1} still differ by the same single edge flip. We next note there is a rectangle in valid dyadic tiling A_t of dimension $2a \times b$, implying that $2ab = n = 2^k$. As a and b are powers of 2, $a \geq b$ by assumption, and k is even, then $a = 2^i b$ where i is odd. We now consider two cases, $a \geq 8b$ and $a = 2b$.

Case $a \geq 8b$. We first examine the moves that decrease the distance between the two coupled chains. There are exactly two edge flips decrease the distance between the coupled chains, namely flipping e to f in A_t or flipping f to e in B_t . There are $2n$ values of (x, y, d, o) that select edge e in A_t . Precisely, these are each of the $2n$ points (x, y) in S together with the appropriate direction from among t, b that selects e and the appropriate parity o such that $\log |e| = o \pmod{2}$. Invoking Lemma 4.2.1 and examining the parity o shows these same choices do not yield a flippable edge in B_t ; this is where the value of o plays a critical role, as no edges within or on the

boundary of S in $B_t = \sigma_2$ are of the same length as e . As each such selection occurs with probability $1/(8n^2)$, potential edge flip e is selected with probability $q = 1/(4n)$. In this case the condition for flipping edge e is $p < \lambda^{2b-2a}$, which always occurs as $2b - 2a \leq 0$. After such a flip, $A_{t+1} = B_t$ while $B_{t+1} = B_t$. Thus $\phi_{t+1} = 0$ and the change in distance between the two chains is $-\phi_t = -\lambda^{2b-2a}$. The total contribution to the expected change in $\phi(\mathcal{A}, \mathcal{B})$ from this move is $-q \cdot \lambda^{2b-2a}$.

Similarly, the probability (x, y, d, o) selects edge f in B_t is also $q = 1/(4n)$, and these values do not yield a flippable edge in A_t . Edge f flips only if $p < \lambda^{2a-2b}$, which occurs with probability $\lambda^{2a-2b} < 1$. If this move occurs, then $B_{t+1} = A_t = A_{t+1}$, and the change in distance between A and B is again $-\lambda^{2b-2a}$. The total contribution to the expected change in $\phi(\mathcal{A}, \mathcal{B})$ from this move is

$$-q \cdot \lambda^{2a-2b} \cdot \lambda^{2b-2a} = -q.$$

While the two potential moves above decrease the distance between the chains according to metric ϕ , there are also moves that increase it. For A_t , the top and bottom edges of S are not flippable by Lemma 4.2.1. At first glance there are four other potential edge flips for A_t involving S , specifically flips of the top and bottom halves of S 's left and right boundaries. However, again by Lemma 4.2.1, at most one of the left boundary and the right boundary of S contains flippable edges. Without loss of generality, assume it is the right boundary of S , and label the two potentially flippable edges as g and h . Similarly, for B_t , at first glance there exist four other potential edge flips involving S , specifically the left and right halves of S 's top and bottom boundaries. By Lemma 4.2.1, we assume without loss of generality that only portions of S 's bottom boundary are potentially flippable, and label the two potentially flippable edges as i and j .

Such potential flips only occur if A_t and B_t are tiled in the neighborhood of S as in Figure 14. We suppose this worse case neighborhood tiling exists. Edges g and h are each selected by values (x, y, d, o) in A_t with probability q ; both are then flipped

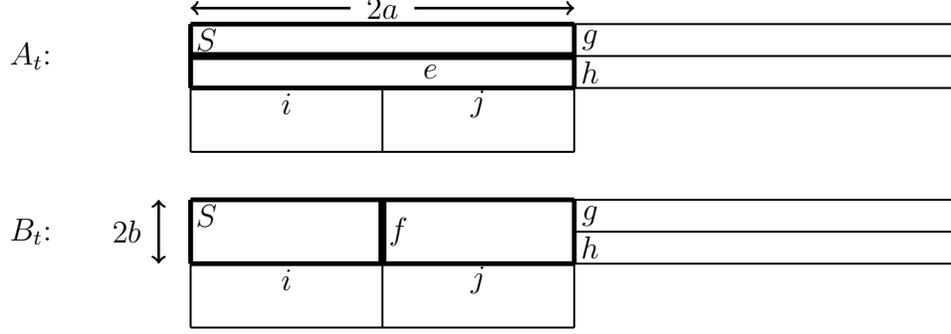


Figure 14: An area $2n$ rectangle S bisected by horizontal edge e in A_t and vertical edge f in B_t . Four “bad” edge flips g, h, i, j exist only if A_t and B_t are tiled in the neighborhood of S as shown.

with probability λ^{4a-b} . The tiling A_{t+1} resulting from this flip is at distance λ^{b-4a} from configuration A_t . The same selection (x, y, d, o) does not result in any flip in B_t , so $B_{t+1} = B_t$. The change in distance between \mathcal{A} and \mathcal{B} for these two moves is at most λ^{b-4a} . In all, the contribution by these moves to the expected change in distance between the coupled chains is at most

$$2 \cdot q \lambda^{4a-b} \cdot \lambda^{b-4a} = 2q.$$

Similarly, edges i and j are selected to be flipped in B_t by values (x, y, d, o) with probability q , and once selected, these edge flips occur if $p < \lambda^{4b-a}$, a bound which is at least 1 for $a \geq 8b$. The tiling B_{t+1} resulting from either flip is at distance at most λ^{4b-a} from configuration B_t . These same values mean that $A_{t+1} = A_t$. Thus the change in distance between \mathcal{A} and \mathcal{B} for these two moves is at most λ^{4b-a} . In all, the contribution by these moves to the expected change in distance between the two chains in the coupling is at most $2 \cdot q \cdot \lambda^{4b-a}$.

In total, we have shown

$$\begin{aligned} e[\phi_{t+1} - \phi_t] &\leq -q - q\lambda^{2b-2a} + 2q + 2q\lambda^{4b-a} \\ &= -q\lambda^{2b-2a}(\lambda^{2a-2b} + 1 - 2\lambda^{2a-2b} - 2\lambda^{2b+a}) \\ &= -q\phi_t(1 - \lambda^{2a-2b} - 2\lambda^{2b+a}) \end{aligned}$$

We first note that as $a \geq 8b$, and in particular, as $a \geq \sqrt{n}$,

$$2a - 2b \geq 2\left(a - \frac{1}{8}a\right) \geq a \geq \sqrt{n}.$$

Additionally, $2b + a \geq a \geq \sqrt{n}$. Thus,

$$\lambda^{2a-2b} + 2\lambda^{2b+a} \leq \lambda^{\sqrt{n}} + 2\lambda^{\sqrt{n}} = 3\lambda^{\sqrt{n}}$$

Provided $\lambda < 3^{-1/\sqrt{n}}$, as we assumed at the start of this section, we have that

$$\lambda^{2a-2b} + 2\lambda^{2b+a} \leq 3\lambda^{\sqrt{n}} < 1$$

Then,

$$e[\phi_{t+1}] \leq (1 - qc)\phi_t,$$

where c is some positive constant, depending on how close λ is to the bound given above. This satisfies the requirement to apply the exponential metric theorem for the case $a \geq 8b$.

Case a = 2b. The analysis of potential good moves and bad moves remains the same as the first case above, though certain probabilities and distances change. Initially, $\phi(A_t, B_t) = \lambda^{2b-2a} = \lambda^{-2b}$, as in the previous case. We note that the contribution to the expected change in distance from good moves flipping edges e and f is still $-q(1 + \lambda^{2b-2a}) = -q(1 + \lambda^{-2b})$. The contributions to the expected change in distance from flipping edges g and h is still $2q$. We note now, however, that for the edges i and j , once selected by (x, y, d, o) , flips now occur with probability $q\lambda^{4b-a} = q\lambda^{2b}$ rather than probability q . Such a move results in a change in distance between the chains in the coupling of $\lambda^{a-4b} = \lambda^{-2b}$. The expected contribution to the change in distance from these moves is now $2q\lambda^{2b}\lambda^{-2b} = 2q$.

In total, we see that in this case,

$$\begin{aligned}
e[\phi_{t+1} - \phi_t] &\leq -q(1 + \lambda^{-2b}) + 4q \\
&= -q\lambda^{-2b}(\lambda^{2b} + 1 - 4\lambda^{2b}) \\
&= -q\phi_t(1 - 3\lambda^{-2b}).
\end{aligned}$$

We note that in this case, $2ab = n$ so $a = 2b = \sqrt{n}$. Provided $\lambda < 3^{-1/\sqrt{n}}$, it follows that $3\lambda^{\sqrt{n}} < 1$, the required condition holds and we get the same bound on $e[\phi_{t+1}]$ as in the previous case, though with a different constant c , also depending on λ .

Theorem 4.3.1: The mixing time of Markov chain \mathcal{M}_n is $O(n^2 \log n)$ for all $\lambda < 3^{-1/\sqrt{n}}$.

PROOF: We apply the exponential metric theorem from [43] (Theorem 2.2.3), using the coupling $(\mathcal{A}, \mathcal{B})$ and metric ϕ defined above.

We first must find an exponential upper bound B on the values ϕ may take. If we let σ^* denote the ground state tiling, tiling the $n \times n$ square with n smaller squares of dimension $\sqrt{n} \times \sqrt{n}$, careful consideration shows that the two dyadic tilings at farthest distance ϕ from σ^* are the tiling consisting of all $n \times 1$ horizontal rectangles σ_h and the tiling consisting of all $1 \times n$ vertical rectangles σ_v . We note that one path in Ω_n from σ_h to σ^* consists of $(\log n)/2 = k/2$ stages, where in each stage $n/2$ edge flips are performed, reducing the length of each of the n rectangles by half; see Figure 15.

The contribution to $\phi(\sigma_h, \sigma^*)$ from each of these edge flips is at most λ^{-n} , and there are $nk/4$ such moves in this particular path in Ω_n from σ_h to σ^* , giving $\phi(\sigma_h, \sigma^*) \leq (nk/4)\lambda^{-n}$. The same holds for σ_v . There is thus a path between any two tilings, through the ground state σ^* , yielding the bound

$$\phi(\sigma_1, \sigma_2) \leq (nk/2)\lambda^{-n} \leq n \log(n)\lambda^{-n}.$$

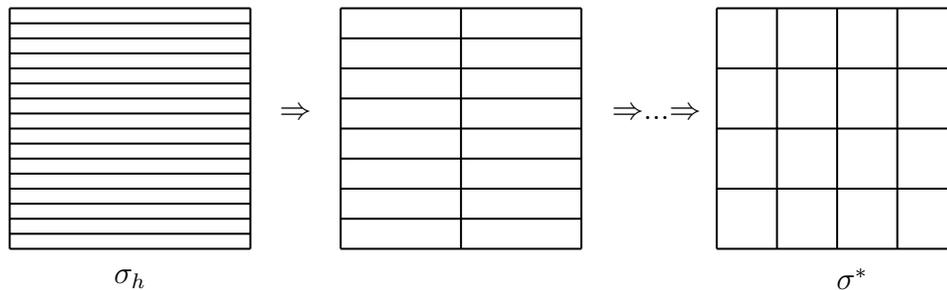


Figure 15: A sequence of edge flips from σ_h to σ^* .

Thus ϕ takes on values in the range

$$\{0\} \cup [1, n \log(n) \lambda^{-n}].$$

We now apply Theorem 2.2.3 with metric ϕ as defined above. We note that ϕ satisfies the path requirement with U being the set of all pairs of tilings that are adjacent in Ω_n , and that ϕ takes on values in $\{0\} \cup [1, B]$ for $B = n \log(n) \lambda^{-n}$. Additionally \mathcal{M}_n is lazy, as discussed in Section 2. For the coupling above, we have demonstrated that $e[\phi_{t+1}] \leq (1 - qc)\phi_t$ whenever $\lambda < 3^{-1/\sqrt{n}}$. Finally, by Theorem 2.2.3, we conclude that

$$\begin{aligned} \tau(\varepsilon) &\leq \frac{\ln(n \log n \lambda^{-n} \varepsilon^{-1})}{qc} \\ &\leq \frac{4n^2}{c} \ln(n \log n \lambda^{-1} \varepsilon^{-1}) = O(n^2 \log(n/\varepsilon)). \end{aligned}$$

When we assume $\varepsilon = 1/4$, as is standard practice, we see $\tau = \tau(1/4) = O(n^2 \log(n))$. ■

This implies for all $\lambda < 1$, \mathcal{M}_n mixes in time $O(n^2 \log n)$, as claimed in Theorem 4.1.1, where the constant in the $O(\cdot)$ notation depends on λ .

4.4 *Slow Mixing for General and Dyadic Tilings*

In this section, we prove that for certain values of λ both chains can require exponential time to converge. We begin by proving that in both the dyadic and general settings, the Markov chains \mathcal{M}_n and $\widehat{\mathcal{M}}_n$ mix slowly when $\lambda > 1$. Next, we show that

for general tilings, unlike in the dyadic case, when $\lambda < 1$, the Markov chain $\widehat{\mathcal{M}}_n$ mixes slowly. In each of these cases we prove that the Markov chain requires exponential time by demonstrating that the state space contains a bottleneck that requires exponential expected time to cross. We use the bottleneck to bound the conductance and then the mixing time of the Markov chain as described in Section 2.4.

A change in terminology will be convenient for the remainder of this section whereby we let $|\sigma|$ be the sum of the perimeters of the rectangles in the dissection (or tiling) σ , rather than the total edge length. This will simplify the analysis. Using detailed balance, we reformulate stationary distributions π and $\widehat{\pi}$ for \mathcal{M}_n and $\widehat{\mathcal{M}}_n$ as follows. Let $w(R)$ be the width of rectangle R and $l(R)$ be the length (height) of R . For convenience, we now let $|\sigma|$ denote the total *perimeter* of σ , that is, $|\sigma| = \sum_{R \in \sigma} 2w(R) + 2l(R)$. We note this total perimeter divided by 2 differs from the total edge length of σ by exactly $2n$. By detailed balance, we rewrite $\pi(\sigma) = \lambda^{|\sigma|/2}/Z = (\prod_{R \in \sigma} \lambda^{w(R)+l(R)})/Z$ and $\widehat{\pi}(\sigma/2) = (\prod_{R \in \sigma} \lambda^{w(R)+l(R)})/\widehat{Z}$; here \widehat{Z} and Z are new normalizing constants, differing from those in Section 4.2.2 by a multiplicative factor of λ^{2n} .

First, we prove the following lemma bounding the number of n -tilings in the general setting which we use in both slow mixing proofs.

Lemma 4.4.1: The number of general tilings of Λ_n satisfies $|\widehat{\Omega}_n| \leq (\log n)^n$.

PROOF: Consider any rectangle R in an n -tiling. By assumption R has dimensions $2^w \times 2^h$ for integers $w, h \in \{0, 1, \dots, k = \log n\}$ and thus has $\log n$ possible heights. Given the height of R , the width is uniquely determined since R has area n . To bound the total number of tilings, there are $\log n$ choices for the height of the rectangle that covers the lowest leftmost unit square of Λ_n . Next, consider the rectangle that covers the lowest leftmost unit square not yet tiled. Given the height of all rectangles ordered in this way the rectangle tiling is uniquely determined. There are n different rectangles with $\log n$ possible heights therefore $|\widehat{\Omega}_n| \leq (\log n)^n$. ■

4.4.1 Slow Mixing when $\lambda > 1$

We start by showing that for both dyadic and general rectangle tilings when $\lambda > 1$, the Markov chains \mathcal{M}_n and $\widehat{\mathcal{M}}_n$ both take exponential time to converge. Informally, consider the tilings with at least one $n \times 1$ rectangle and those with at least one $1 \times n$ rectangle. In order to go between these sets we must go through a tiling where all rectangles have width and length at least 2 and thus perimeter at most $n + 4$. We show these tilings are exponentially unlikely and thus our state space forms a bottleneck.

Proof of Theorem 4.1.2 and Theorem 4.1.3. We note that the proofs are identical for \mathcal{M}_n and $\widehat{\mathcal{M}}_n$; here we show for $\widehat{\mathcal{M}}_n$. We first partition the state space into two sets, B , the set of tilings with no rectangles of dimension $1 \times n$, and \overline{B} , the remainder. Notice that B contains the tiling σ_h where all rectangles are $n \times 1$ and \overline{B} contains the tiling σ_v where all rectangles are $1 \times n$. Both of these tilings have weight $\pi(\sigma_h) = \pi(\sigma_v) = \widehat{Z}^{-1} \lambda^{n(1+n)}$. Therefore,

$$\pi(B) \geq \pi(\sigma_h) \geq \widehat{Z}^{-1} \lambda^{n(1+n)},$$

$$\pi(\overline{B}) \geq \pi(\sigma_v) \geq \widehat{Z}^{-1} \lambda^{n(1+n)}.$$

Let $B_c \subset B$ be the set of tilings containing no $(1 \times n)$ or $(n \times 1)$ rectangles. Every rectangle in every tiling in B_c has perimeter at most $n + 4$ and thus has weight at most $\widehat{Z}^{-1} \lambda^{n(n+4)/2}$. By Lemma 4.4.1, $|B_c| \leq |\widehat{\Omega}_n| \leq (\log n)^n$; we briefly note this is true in the dyadic case as well although tighter bounds exist. Combining these, we see

$$\pi(B_c) \leq (\log n)^n \widehat{Z}^{-1} \lambda^{n(n+4)/2},$$

which is exponentially smaller than the weight of B and \overline{B} .

Using these bounds, we next bound the conductance of the Markov chain and then the mixing time using Theorem 2.4.1. If $\pi(B) \leq 1/2$, then combining the definition

of conductance with the bounds on $\pi(B)$ and $\pi(B_c)$ yields

$$\begin{aligned}
\Phi_{\widehat{\mathcal{M}}_n} &\leq \frac{1}{\pi(B)} \sum_{b_1 \in B, b_2 \in \overline{B}} \pi(b_1) \mathcal{P}(b_1, b_2) \\
&= \frac{1}{\pi(B)} \sum_{b_1 \in B_c, b_2 \in \overline{B}} \pi(b_1) \mathcal{P}(b_1, b_2) \\
&\leq \frac{1}{\pi(B)} \sum_{b_1 \in B_c} \pi(b_1) = \frac{\pi(B_c)}{\pi(B)} \\
&\leq \frac{(\log n)^n Z^{-1} \lambda^{n(n+4)/2}}{Z^{-1} \lambda^{n(1+n)}} = \frac{(\log n)^n}{\lambda^{n^2/2-n}} = \lambda^{-c_1 n^2},
\end{aligned}$$

for constant c_1 and n sufficiently large when λ is a constant greater than 1. Alternatively, if $\pi(B) > 1/2$ then $\pi(\overline{B}) \leq 1/2$ and so by detailed balance and the bounds on $\pi(\overline{B})$ and $\pi(B_c)$,

$$\begin{aligned}
\Phi_{\widehat{\mathcal{M}}_n} &\leq \frac{1}{\pi(\overline{B})} \sum_{b_1 \in B, b_2 \in \overline{B}} \pi(b_2) \mathcal{P}(b_2, b_1) \\
&= \frac{1}{\pi(\overline{B})} \sum_{b_1 \in B, b_2 \in \overline{B}} \pi(b_1) \mathcal{P}(b_1, b_2) \\
&= \frac{1}{\pi(\overline{B})} \sum_{b_1 \in B_c, b_2 \in \overline{B}} \pi(b_1) \mathcal{P}(b_1, b_2) \leq \frac{\pi(B_c)}{\pi(\overline{B})} \\
&\leq \frac{(\log n)^n Z^{-1} \lambda^{n(n+4)/2}}{Z^{-1} \lambda^{n(1+n)}} = \frac{(\log n)^n}{\lambda^{n^2/2-n}} = \lambda^{-c_1 n^2},
\end{aligned}$$

for constant c_1 defined above and n sufficiently large, whenever λ is a constant greater than 1.

In both cases, $\Phi_{\widehat{\mathcal{M}}_n} \leq \lambda^{-c_1 n^2}$. Applying Theorem 2.4.1 proves that for all $\epsilon > 0$, the mixing time of $\widehat{\mathcal{M}}_n$ satisfies

$$\tau(\epsilon) \geq \left(\lambda^{c_1 n^2} / 4 - \frac{1}{2} \right) \log \left(\frac{1}{2\epsilon} \right) = \Omega(\lambda^{c_1 n^2} \ln \epsilon^{-1}).$$

Letting $\epsilon = 1/4$ we have that $\tau = \Omega(\lambda^{c_1 n^2})$, as desired.

4.4.2 Slow Mixing for General Tilings when $\lambda < 1$

Next, we consider general tilings when $\lambda < 1$ and show that in this setting $\widehat{\mathcal{M}}_n$ takes exponential time to converge by again demonstrating a bottleneck in the state space.

In this case however the bottleneck is much more complex. Define a *bar* to be a rectangle of width 1 and length (height) n . The bottleneck in $\widehat{\Omega}_n$ is based on the *separation* of a tiling which measures the distance between the bars in the tiling. More formally, define the *distance* between two bars to be the difference in their x -coordinates plus one. For example, two adjacent bars are at distance 2 and two bars separated by a rectangle of size $2 \times n/2$ are at distance 4. Given an n -tiling, pair the bars in order from left to right (there must be an even number of bars since $n = 2^k$). The *separation* of a tiling is the sum of the distances between each pair of bars. Let S be the set of tilings with separation greater than or equal to $n/2 + 2$ and \bar{S} be the remaining tilings, namely those with separation less than $n/2 + 2$. We show all moves from S to \bar{S} involve a tiling with at least 4 bars and separation $n/2 + 2$, and the total weight of this set of tilings is exponentially smaller than the weight of both S and \bar{S} .

Proof of Theorem 4.1.4. We begin by proving a lower bound on $\pi(S)$ and $\pi(\bar{S})$. Let g_n be the “ground state” tiling consisting entirely of rectangles of size $\sqrt{n} \times \sqrt{n}$. This tiling has perimeter $|g_n| = 4n\sqrt{n}$. Since $g_n \in \bar{S}$ because g_n has no bars and thus separation 0, this implies that $\pi(\bar{S}) > \pi(g_n) = \widehat{Z}^{-1}\lambda^{2n\sqrt{n}}$. Next we will define a special tiling $s_n \in S$. Let s_n have one bar on the far left side of Λ_n and one bar on the far right side of Λ_n . Next to the leftmost bar there is a column with two rectangles of width 2 followed by a column with four rectangles of width 4 and so forth until there is a column with only rectangles of width $2^{k/2-1}$. The remainder of the tiling is filled with rectangles of size $\sqrt{n} \times \sqrt{n}$. Note that the combined width of these columns is $2 + \sum_{i=1}^{k/2-1} 2^i = \sqrt{n}$ so the remainder of the tiling has width $n - \sqrt{n}$ and can be tiled with $n - \sqrt{n}$ rectangles of size $\sqrt{n} \times \sqrt{n}$. Figure 16 shows s_{64} . Configuration s_n has perimeter

$$\begin{aligned} |s_n| &= 4(1 + 2^k) + \sum_{i=1}^{k/2-1} (2^i 2(2^i + 2^{k-i})) + (n - \sqrt{n})4\sqrt{n} \\ &= 4n^{3/2} + n \log n - (4/3)n + (4/3). \end{aligned}$$

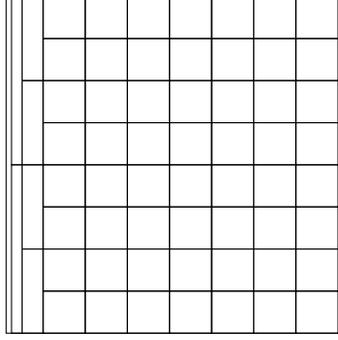


Figure 16: The tiling s_{64} .

As $s_n \in S$ because it has separation n , this implies $\pi(S) > \pi(s_n) = \widehat{Z}^{-1} \lambda^{|s_n|/2}$.

Let S_C be the set of tilings in S from which it is possible to transition to \bar{S} . We will prove that every tiling in S_C has at least four bars and separation exactly $n/2+2$. We use the following lemma.

Lemma 4.4.2: One move of the chain $\widehat{\mathcal{M}}_n$ changes the separation of a tiling by 0, +2 or -2.

PROOF: The only moves of the Markov chain that change the separation are when two bars are added or removed. Let's consider adding two bars first. Let P be the pairing of the bars in the configuration before the two bars are added. There are two cases; either the two bars are added between two bars that were paired in P , or between two pairs of bars. If they are added between two pairs, then they will be paired up in the new pairing and add 2 to the separation. If they are added between two bars b_l and b_r paired in P with distance d , then the new bars will be paired with b_l and b_r . The sum of the distances will remain unchanged. Next, consider the case where two bars are removed. Again, there are two cases. If the two bars are paired, then the separation decreases by two however if the two bars are paired with two other bars the distance remains unchanged. ■

Configuration g_n has separation 0. Since all tilings are connected by the Markov chain $\widehat{\mathcal{M}}_n$ which by Lemma 4.4.2 changes the separation by an even number at each

step, this implies that the separation of all tilings is even. Additionally, to go from S to \overline{S} we must go through a tiling with separation exactly $n/2 + 2$. Given a tiling with two bars and separation $n/2 + 2$ there is no way to decrease the separation and thus no way to transition to \overline{S} . Thus, every tiling in S_C has separation $n/2 + 2$ and at least four bars. Next, we will upper bound the weight of each tiling σ in S_C . To do this, we lower bound the perimeter of any tiling of a lattice region of size $(n/2 - 2) \times n$ and then show that every tiling in S_C has two such regions.

Lemma 4.4.3: Any tiling σ' of an $(n/2 - 2) \times n$ region has perimeter $|\sigma'| \geq 2n^{3/2} + n \log n - (16/3)n - (8/3)$.

PROOF: We will assign each unit square in the lattice region a weight based on the perimeter of the rectangle the square is contained in so that the combined weight of all n squares within a rectangle is equal to the perimeter of the rectangle. Assume the unit square at location (i, j) is contained in a rectangle of size $2^a \times 2^{k-a}$ then the weight $w_{i,j} = 2(2^a + 2^{k-a})/2^k$. Since each rectangle has area 2^k , the sum of all weights $\sum_{j=1}^{n/2-2} \sum_{i=1}^n w_{i,j} = |\sigma'|$. Consider the binary representation of the width $n/2 - 2$ of the region, $011 \dots 110$. Since each rectangle has width 2^a for some integer a this implies that in each row, for each integer $\ell = 1$ to $k/2 - 1$ there must be either a rectangle of width 2^ℓ or multiple rectangles of width smaller than 2^ℓ whose widths add up to 2^ℓ . If there is a single rectangle of width 2^ℓ then the 2^ℓ unit squares in this row contained in this rectangle each have weight $2(2^\ell + 2^{k-\ell})/2^k$. If there is instead multiple smaller rectangles then they will have larger perimeter and thus larger weight. Thus, the combined weight of these unit squares in each row is at least $\sum_{\ell=1}^{k/2-1} 2^\ell (2(2^\ell + 2^{k-\ell})/2^k) = \log n - (4/3) - 8/(3n)$. Since the minimum perimeter rectangle is the $2^{k/2} \times 2^{k/2}$ square, $w_{i,j} \geq 4/2^{k/2}$. Thus the remaining $2^{k-1} - 2^{k/2}$ unit squares in each row have combined weight at least $4(2^{k-1} - 2^{k/2})/2^{k/2} = 2\sqrt{n} - 4$.

This implies that the total perimeter satisfies

$$\begin{aligned}
|\sigma'| &= \sum_{i=1}^n \sum_{j=1}^{n/2-2} w_{i,j} \\
&\geq \sum_{i=1}^n (\log n - (4/3) - 8/(3n) + 2\sqrt{n} - 4) \\
&= 2n^{3/2} + n \log n - (16/3)n - 8/3.
\end{aligned}$$

This is the desired result. ■

Consider any tiling σ with separation $n/2 + 2$ and at least four bars. Label the bars b_1, b_2, \dots, b_B from left to right. Next, label the regions between the pairs of bars $p_1, p_2, \dots, p_{B/2}$ and the gaps between the pairs $g_0, g_1, g_2, \dots, g_{B/2}$, as shown in Figure 17. Let $w(p_i), w(g_i)$ denote the widths of the regions between the bars.

Now, since σ has separation $n/2 + 2$, this implies that $\sum_{i=1}^{B/2} (w(p_i) + 2) = n/2 + 2$. Reorder the tiling so it is ordered $g_0, \dots, g_{B/2}, b_1, b_2, b_3, \dots, b_{B-1}, p_1, p_2, \dots, p_{B/2}, b_B$. Notice that the region $b_4 \cup \dots \cup b_{B-1} \cup p_1 \cup \dots \cup p_{B/2}$ has width $n/2 - 2$ as does the region $g_0 \cup g_1 \cup \dots \cup g_{B/2}$. Thus we can apply Lemma 4.4.3 to show that the total perimeter of tiling σ must be at least

$$\begin{aligned}
|\sigma| &\geq 4(2 + 2^{k+1}) + 2 \left(2n^{3/2} + n \log n - \frac{16}{3}n - \frac{8}{3} \right) \\
&= 4n^{3/2} + 2n \log n - (8/3)n + 8/3.
\end{aligned}$$

Combining this bound with the bound on the number of tilings from Lemma 4.4.1 gives $\pi(S_C) \leq \widehat{Z}^{-1}(\log n)^n \lambda^{2n^{3/2} + n \log n - (4/3)n + (4/3)}$, which is exponentially smaller than the above bound on $\pi(S)$, as desired. Using these bounds we bound the conductance of the Markov chain and then the mixing time using Theorem 2.4.1. If $\pi(S) \leq 1/2$, then combining the definition of conductance with the bounds on $\pi(S)$ and $\pi(S_C)$

yields

$$\begin{aligned}
\Phi_{\widehat{\mathcal{M}}_n} &\leq \frac{1}{\pi(S)} \sum_{s_1 \in S, s_2 \in \bar{S}} \pi(s_1) \mathcal{P}(s_1, s_2) \\
&= \frac{1}{\pi(S)} \sum_{s_1 \in S_C, s_2 \in \bar{S}} \pi(s_1) \mathcal{P}(s_1, s_2) \\
&\leq \frac{1}{\pi(S)} \sum_{s_1 \in S_C} \pi(s_1) = \frac{\pi(S_C)}{\pi(S)} \\
&\leq \frac{(\log n)^n \lambda^{2n^{3/2} + n \log n - (4/3)n + (4/3)}}{\lambda^{2n^{3/2} + (n \log n)/2 - (2/3)n + (2/3)}} \\
&= (\log n)^n \lambda^{(n \log n - (4/3)n + (4/3))/2} \\
&= 2^{n \log \log n} \lambda^{(n \log n - (4/3)n + (4/3))/2} = \lambda^{c_2 n \log n},
\end{aligned}$$

for constant c_2 and n sufficiently large when $\lambda < 1$ is a constant.

If $\pi(S) > 1/2$, then $\pi(\bar{S}) \leq 1/2$, and by detailed balance and bounds on $\pi(S)$, $\pi(\bar{S})$ and $\pi(S_C)$,

$$\begin{aligned}
\Phi_{\widehat{\mathcal{M}}_n} &\leq \frac{1}{\pi(\bar{S})} \sum_{s_1 \in \bar{S}, s_2 \in S} \pi(s_1) \mathcal{P}(s_1, s_2) \\
&= \frac{1}{\pi(\bar{S})} \sum_{s_1 \in \bar{S}, s_2 \in S} \pi(s_2) \mathcal{P}(s_2, s_1) \\
&= \frac{1}{\pi(\bar{S})} \sum_{s_1 \in S_C, s_2 \in \bar{S}} \pi(s_1) \mathcal{P}(s_1, s_2) \leq \frac{\pi(S_C)}{\pi(\bar{S})} \\
&\leq \frac{(\log n)^n \lambda^{4n^{3/2} + 2n \log n - 2n}}{\lambda^{4n^{3/2} + n \log n - n}} = \lambda^{c_2 n \log n},
\end{aligned}$$

for constant c_2 defined above and n sufficiently large when $\lambda < 1$ is a constant. In both cases,

$$\Phi_{\widehat{\mathcal{M}}_n} \leq \lambda^{c_2 n \log n}.$$

Applying Theorem 2.4.1 proves that the mixing time of $\widehat{\mathcal{M}}_n$ satisfies

$$\begin{aligned}
\tau(\epsilon) &\geq \left(\lambda^{-c_2 n \log n} / 4 - \frac{1}{2} \right) \log \left(\frac{1}{2\epsilon} \right) \\
&= \Omega(\lambda^{-c_2 n \log n} \ln \epsilon^{-1}).
\end{aligned}$$

Letting $\epsilon = 1/4$, we have that $\tau = \Omega(\lambda^{-c_2 n \log n})$.

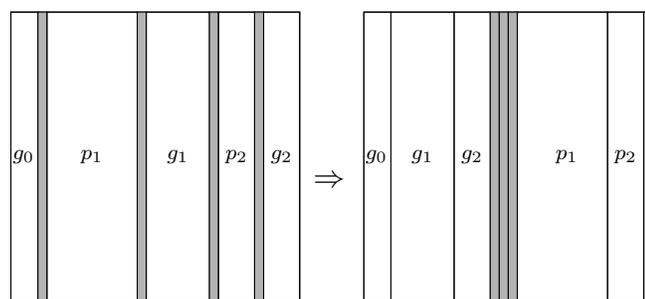


Figure 17: An example labeling of the bars and regions surrounding bars.

CHAPTER V

COLLOIDS AND INTERFERING BINARY MIXTURES

Colloids are binary mixtures of molecules with one type of molecule suspended in another. It is believed that at low density typical configurations will be well-mixed throughout, while at high density they will separate into clusters. In this chapter, we characterize the high and low density phases for a general family of discrete *interfering binary mixtures* by showing that they exhibit a “clustering property” at high density and not at low density. The clustering property states that there will be a region that has very high area, very small perimeter, and high density of one type of molecule. A special case of interfering binary mixtures are mixtures of squares and diamonds on \mathbb{Z}^2 that correspond to the Ising model at fixed magnetization which was introduced in Section 1.2.1.

5.1 Clustering in Colloids

Colloids are mixtures of two types of molecules in suspension where all non-overlapping arrangements are equally likely. When the density of each type of molecule is low, the mixtures are homogeneous and consequently exhibit properties that make them suitable for many industrial applications, including fogs, gels, foods, paints, and photographic emulsions (see, e.g., [10], [46]). In contrast, when the density is high, the two types of molecules separate whereby one type appears to cluster together. Although this behavior is similar to phase transitions that occur in other discrete models, such as the Ising and Potts models, here the two types of molecules do not possess any *enthalpic* forces causing like particles to attract or disparate particles to repel. In contrast, the behavior of colloids is purely *entropic* — the only restriction is a “hard-core” constraint requiring objects to remain in non-overlapping positions,

and clustering occurs at high density because the overwhelming majority of configurations in the stationary distribution are believed to exhibit such a separation. While the experimental study of colloids is pervasive in surface chemistry, material science, physics, and nanotechnology, there has been little rigorous work explaining their behavior. Even running simulations has been challenging because local algorithms will be slow to converge at high density. Dress and Krauth [30] introduced an algorithm to try to overcome this obstacle, but this too was shown to require time exponential in the number of molecules in some cases [66]. Nonetheless, their algorithm seems to be well-behaved in practice, and Buhot and Krauth [19] provided simulations showing strong heuristic evidence of the presence of two distinct phases in colloid models consisting of different sized squares.

Frenkel and Louis [36] studied an interesting discrete model of colloids whose behavior can be related to the Ising model, a standard model of ferromagnetism. Their model consists of mixtures of unit squares in a region of \mathbb{Z}^2 and diamonds of area $1/2$ that sit on lattice edges (see Figure 18). They show that this colloid model, which we call **Model 1**, corresponds to an Ising model, where the density of squares fixes the magnetization and the density of diamonds determines the temperature (see Section 2.1). The Ising model at low temperature is known to exhibit clustering of positive spins. In fact the precise limiting shape of the cluster known as the Wulff shape has been extensively studied using sophisticated techniques (see, e.g. [28], or the references therein). **Model 1** then inherits the phase transition arising in the Ising model which shows there will be clustering (of the squares) at high densities [62]. In this chapter we study clustering using elementary methods that apply to a large class of natural colloid models. We characterize clustering directly in terms of the parameters arising from the model to distinguish between the high and low phases and understand the role the density of each type of molecule plays.

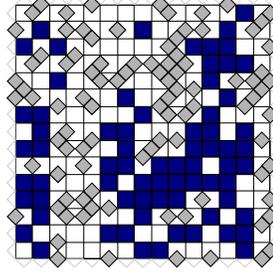


Figure 18: Model 1, squares and diamonds on the $n \times n$ grid L_n

5.2 *Interfering Binary Mixtures*

We consider a class of *interfering binary mixtures*. Define Λ_f as a planar graph where all faces are isomorphic to f . Let (Λ_A, Λ_B) be a pair of such graphs, and consider the intersection of these graphs with some finite region L , where $L_A = \Lambda_A \cap L$ and $L_B = \Lambda_B \cap L$. While we do not require any symmetry or special alignment of the faces in Λ_A and Λ_B , we do impose a few extra conditions (*divisibility* and *bridgeability*) on Λ_A to ensure that the model is reasonably well-behaved (see Section 5.4.1 for details); most natural models satisfy these conditions. A *binary mixture* is a non-overlapping packing of tiles on L , where A -tiles lie on the faces of L_A and B -tiles lie on the faces of L_B . A binary mixture is *interfering* if there exist constants $0 < \delta \leq \gamma$ such that for any face x in L_B and any set S of faces in L_A for which the outer perimeter of S has a nontrivial intersection with x , δ and γ provide upper and lower bounds on the ratio of the area of $S \cap x$ to the length of the perimeter of S contained within x ; that is,

$$0 < \delta \leq \frac{a(S \cap x)}{|\kappa(S) \cap x|} \leq \gamma,$$

where $\kappa(S)$ is the perimeter of S . For example, in **Model 1**, Λ_A is the Cartesian lattice \mathbb{Z}^2 and Λ_B is the set of diamonds bisected by edges in \mathbb{Z}^2 ; then $\delta = \gamma = 1/4$ (Figure 18). Notice that this definition forbids pairs (Λ_A, Λ_B) where an A -tile y and a B -tile x can intersect in an edge, such as in **Model 2** where A -tiles are unit squares on L_n and B -tiles are squares of side length $1/2$ on the half-integer lattice (see Figure

19(d)). This is because in this case we can choose $S = \{y\}$, and the area of $S \cap x$ is 0, even though the perimeter of y intersects x nontrivially, forcing $\delta = 0$. See Figure 19(a-c) for additional examples of interfering binary mixtures. In Section 5.4, we

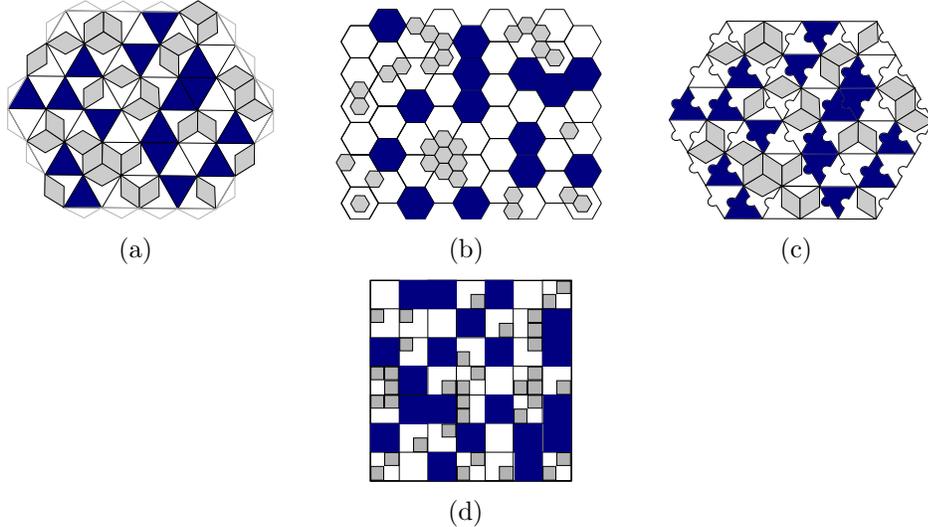


Figure 19: Subfigures (a-c) are interfering binary mixtures while (d) is not.

will give examples of other interfering binary mixtures, including independent sets, that arise naturally in combinatorics and statistical physics and contrast these with a non-interfering binary mixture that provably does not exhibit clustering.

It is often useful to switch from a model where the number of tiles of each type are fixed to a so-called *grand-canonical ensemble* where these are allowed to vary. Here, however, typical configurations would have a preponderance of only one type of tile at most high densities and the balanced configurations we are interested in would be exponentially unlikely. Instead, we fix the number of A -tiles and allow the B -tiles to vary stochastically. Each configuration σ has weight proportional to $\lambda^{d(\sigma)}$, where $d(\sigma)$ is the number of B -tiles in σ . The choice of λ controls the expected density of B -tiles.

Our goal now is to understand when the A -tiles will exhibit clustering in terms of the (expected) density of A -tiles and B -tiles. First we define a clustering property for configurations of tiles. Informally we have clustering if there exists a dense region R

in Λ_A with $\Omega(n^2)$ area and $O(n)$ perimeter. Our main theorems demonstrate that at high density interfering binary mixtures exhibit the clustering property while at low densities they do not.

The key tools in our proofs are careful *Peierls arguments*, used in statistical physics to study uniqueness of the Gibbs state and phase transitions (see, e.g., [27], [29]), and in computer science to study slow mixing of Markov chains (see, e.g., [14], [40], [74]). Peierls arguments allow contours to be added and removed by complementing the interiors of those contours. The main challenge here is maintaining the number of A -tiles, making the arguments considerably more difficult. We introduce the concept of bridge systems, to handle multiple contours by connecting components and to make it possible to efficiently encode the boundaries of all contours removed. The encoding is necessary to account for the entropy/energy tradeoffs in these maps.

We give precise definitions of the clustering property and state the main theorems in Section 5.3. In Sections 5.3.4 and 5.3.5 we prove the two main theorems in the context of **Model 1** and in Section 5.4 we extend our proofs to all interfering binary mixtures.

5.3 *Clustering in Model 1*

Frenkel and Lewis introduced **Model 1** which consists of non-overlapping mixtures of unit squares in a region of \mathbb{Z}^2 and diamonds of area $1/2$ that sit on lattice edges. **Model 1** is equivalent to the Ising model with fixed magnetization and thus inherits the phase transition arising in the Ising model which has been extensively studied. Although our results are weaker than what is already known for the Ising model we begin by proving our results for **Model 1** in order to introduce our techniques and ideas in this simpler context. Later in Section 5.4, we extend these techniques to other models of interfering binary mixtures which do not inherit the results from the Ising model. We begin by formally introducing the model, rigorously defining clustering

and stating our main theorems.

5.3.1 Formalizing the Model

Given constants $\lambda > 1$, and $0 < b < 1/2$, where $bn^2 \in \mathbb{Z}$, define $\Omega = \Omega(b, \lambda)$ as the set of non-overlapping packings of L with bn^2 A-tiles and any number of B-tiles (where a tile can only be placed on a face of its type). We wish to study the distribution $\pi(\rho) = \lambda^{d(\rho)}/Z$, where $d(\rho)$ is the number of B-tiles in ρ and $Z = \sum_{\rho \in \Omega} \lambda^{d(\rho)}$ is a normalizing constant. Our goal is to determine whether a configuration chosen according to π is likely to have clusters of A-tiles.

As mentioned earlier, **Model 1** is equivalent to the Ising model of ferromagnetism with a fixed magnetization, which we will see presently. First, we will define the Ising model on the $n \times n$ grid L_n . Let $\bar{G} = (\bar{V}, \bar{E})$ be the dual lattice region to L_n and let $\rho \in \{+, -\}^{\bar{V}}$ be an assignment of spins to each of the vertices in \bar{V} (i.e., the faces in V). The weight of a configuration is $\bar{\pi}(\rho) = e^{\beta|\bar{E}_d(\rho)|}/\bar{Z}$, where $\bar{E}_d(\rho) \subseteq \bar{E}$ is the set of edges in \bar{G} whose endpoints have different spins in ρ , β is the inverse temperature and \bar{Z} is a normalizing constant.

Returning to **Model 1**, let the *A-structure* $\Gamma(\rho)$ of a configuration ρ in Ω be the configuration σ obtained from ρ by removing all of its B-tiles (diamonds). The set $\hat{\Omega}$ of all such A-structures with bn^2 A-tiles (squares) is called the *projection of Model 1*. Let $\hat{\pi}$ be the induced distribution on $\hat{\Omega}$; that is, for $\sigma \in \hat{\Omega}$, let $\hat{\pi}(\sigma) = \sum_{\rho \in \Gamma^{-1}(\sigma)} \pi(\rho)$. Then the function $f : \hat{\Omega} \rightarrow \{+, -\}^{\bar{V}}$, which replaces each square by a positive spin and each empty face by a negative spin, is a bijection which maps the projection of **Model 1** onto the Ising model. To see this, define the *perimeter* of σ (in Ω or $\hat{\Omega}$) to be the edges that belong to exactly one A-tile in σ , and define $\kappa(\sigma)$ as the length of the perimeter of σ . Let $e(\sigma)$ be the number of edges that are not incident to any A-tile in σ . We find that

$$\hat{\pi}(\sigma) = \sum_{k=0}^{e(\sigma)} \frac{\lambda^k}{Z} \binom{e(\sigma)}{k} = \frac{1}{Z} (1 + \lambda)^{e(\sigma)} = (1 + \lambda)^{2n^2 - 2bn^2} \frac{\mu^{\kappa(\sigma)}}{Z}, \quad (5.3.1)$$

where $\mu = (1 + \lambda)^{-\frac{1}{2}}$. Thus, the total perimeter of the A -structure completely determines the probability that it will show up in Ω . Since the weight of a configuration is determined exactly by the number of edges with opposite spins in L_n , this is the Ising model with a fixed number of positive spins for some λ that is a function of β , known as *fixed magnetization*.

While the perimeter of the A -structure does not exactly determine its probability for the other models of interfering binary mixtures, we will see that they are closely related, and we can still use arguments about the perimeter to bound the weight of configurations. Thus it makes sense to define the clustering property in terms of the perimeter to area ratio, which we do next.

5.3.2 The Clustering Property

The goal of this chapter is to show that when the density of B -tiles is high, interfering binary mixtures cluster, while at low density they do not. First, we characterize clustering in the context of **Model 1**. Intuitively, a configuration has the clustering property if there is a large region densely filled with A -tiles. More precisely, we define a *region* $R = (R_F, R_E)$ where R_F is a set of faces in L_n and R_E is a set of edges where R_E is connected and any edge e which is adjacent to a face in R_F and a face in $\bar{R}_F = L_n \setminus R$ satisfies $e \in R_E$. The length of the perimeter $\kappa(R)$ of a region R is $|R_E|$. Let $c = \min \left\{ \frac{b}{2}, \frac{1}{100} \right\}$.

Definition 5.3.1: A configuration $\sigma \in \Omega$ (or $\Gamma(\sigma) \in \widehat{\Omega}$) has the *clustering property* if it contains a region R which satisfies the following properties:

1. R contains at least $(b - c)n^2$ A -tiles,
2. the perimeter of R is at most $8\sqrt{b}n$, and
3. the density of A -tiles is at least $1 - c$ in R and at most c in \bar{R} .

If a configuration has the clustering property, we show that it contains an $n^{1/3} \times n^{1/3}$ window with high density and one with low density, demonstrating the heterogeneity of the configuration. At the end of Section 5.3.4 we contrast this with **Model 2**, related to bond percolation, which remains homogeneous at all densities.

5.3.3 Main Results

We show that at high density interfering binary mixtures have the clustering property while at low densities they do not. Specifically, we prove the following theorems in the context of **Model 1** on the $n \times n$ region L_n with bn^2 A -tiles and the density of B -tiles determined by λ . In Section 5.4, we show they also hold for other interfering binary mixtures.

Theorem 5.3.1: For $0 < b < 1/2$, there exist constants $\lambda^* = \lambda^*(b) > 1$, $\gamma_1 < 1$ and $n_1 = n_1(b)$ such that for all $n > n_1$, $\lambda \geq \lambda^*$ a random sample from Ω will have the clustering property with probability at least $(1 - \gamma_1^n)$.

Theorem 5.3.2: For $0 < b < 1/2$, there exist constants $\lambda_* = \lambda_*(b) > 0$, $\gamma_2 < 1$ and $n_2 = n_2(b)$ such that for all $n > n_2$, $\lambda \leq \lambda_*$ a random sample from Ω will not have the clustering property with probability at least $(1 - \gamma_2^n)$.

Furthermore, it follows from the proofs that at low density if a dense region R' has area $\Omega(n^2)$ then it must have perimeter $\Omega(n^2)$. Notice that in the case $b > 1/2$ we can obtain comparable results by the symmetry of the A -tiles to the empty space. Indeed, in this case if λ is sufficiently high we will see empty faces clustering within a sea of A -tiles and for low density the empty faces will be well-distributed.

Note that since clustering is just a property of the A -tiles, it suffices to prove Theorems 5.3.1 and 5.3.2 for weighted A -structures $\widehat{\Omega}$, involving just the A -tiles. From this point we focus on $\widehat{\Omega}$, and we refer to A -tiles just as *tiles*.

5.3.4 Clustering at High Density for Model 1

We concentrate first on interfering binary mixtures at high density to prove Theorem 5.3.1. Define $\Psi \subset \hat{\Omega}$ to be the set of configurations that have the clustering property; then we show that $\hat{\pi}(\hat{\Omega} \setminus \Psi) \leq \gamma_1^n$ for some constant $\gamma_1 < 1$. To achieve this, we apply a Peierls argument, in which we define a map $f : \hat{\Omega} \setminus \Psi \rightarrow \hat{\Omega}$ and show that for all $\tau \in \hat{\Omega}$,

$$\sum_{\sigma \in f^{-1}(\tau)} \hat{\pi}(\sigma) \leq \gamma_1^n \hat{\pi}(\tau). \quad (5.3.2)$$

Given a configuration $\sigma \in \hat{\Omega} \setminus \Psi$, the map f removes a large set T of tiles in σ and reassembles them in a single large component in $f(\sigma)$. This decreases the total perimeter of the configuration significantly, and therefore $\hat{\pi}(f(\sigma))$ is exponentially larger than $\hat{\pi}(\sigma)$. The challenge is to bound the number of configurations that map to a given $\tau \in \Psi$ by carefully encoding the preimages of τ .

Some definitions will be helpful. We say two tiles are *adjacent* if their borders share an edge. A *component* is a maximal connected set of tiles, and maximal connected segments of the perimeter of σ are *contours*. The set T of tiles we remove will be a union of components, which we identify using a system of “bridges” connecting these components (Figure 20). The key is that the number of edges in the bridges is at most a constant times the total perimeter of the components bridged. Then if E is the set of all edges in bridges or along contours bridged, we can bound $|f^{-1}(\tau)|$ by the number of ways that those E edges could be distributed in σ . Finally, we show that there is a sparse, roughly square region in the resulting configuration where we can add the T tiles. We complement that region to obtain $f(\sigma)$, which allows us to remember the locations of any components that were not bridged (see Figure 20). Notice that the resulting configuration has much higher weight (much smaller perimeter), as desired.

Building Bridges. Given a region R , let $C(R)$ be the set of contours fully contained within the interior of R and define the *outer contours* to be those in $C(R)$ that are

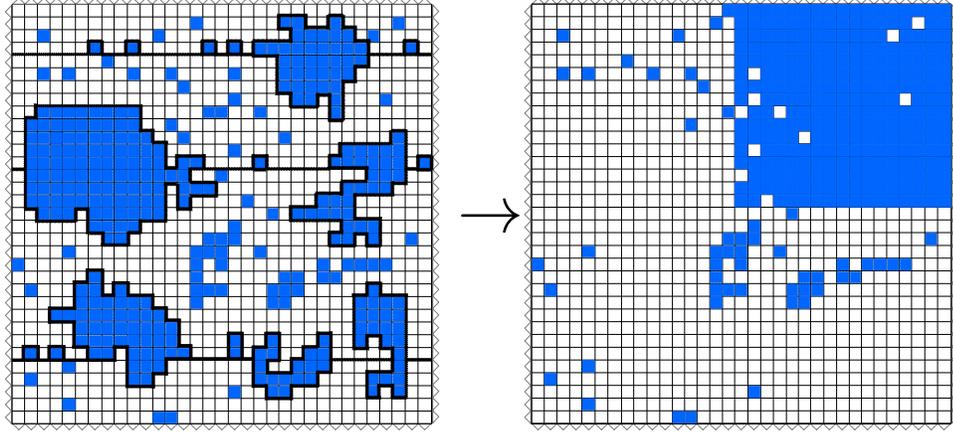


Figure 20: A configuration $\sigma \in \widehat{\Omega} \setminus \Psi$ and the image $f(\sigma)$ of σ in Ψ

not contained in the interior of other contours in $C(R)$. The interior of the outer contours of components are called *holes* and the interior of the outer contours of holes are called *islands*.

Consider first the case in which there are no components with holes. Suppose B is a set of edges of L_n connecting some subset S of the contours to the boundary of L_n . We call B a set of *bridges* and S a set of *bridged contours*. A face in L_n or a tile is called *unbridged* if it is not bounded by a bridged contour. Then (B, S) is a **c -bridge system** for $\sigma \in \widehat{\Omega}$ if the number of unbridged tiles is at most c times the number of unbridged faces, and $|B| \leq \kappa(S)(1 - c)/(2c)$. If σ has components with holes, then first construct a c -bridge system (B, S) for σ' , obtained from σ by filling all the holes. Next for each bridged contour X in σ , construct a c -bridge system for the region in σ bounded by X (treating tiles as empty faces and empty faces as tiles). Recurse until a c -bridge system for each bridged contour at every level of the recursion is obtained. We call this a c -bridge system of σ .

Lemma 5.3.3: There exists a c -bridge system for any configuration $\sigma \in \widehat{\Omega}$.

PROOF: If any components of σ have holes, we may need to recurse as described above. We may assume that we are given a region R in σ with no holes, since otherwise we recurse as described above. Now we use induction on the number of

contours in R . If there are no contours, then clearly (\emptyset, \emptyset) is a c -bridge system for R . Otherwise, define $t(R)$ to be the set of tiles in R and $x(R)$ to be the number of empty faces in R . Let \mathbb{H} be the set of horizontal lines through R . If, for every $H \in \mathbb{H}$, $|t(R) \cap H| < c|R \cap H|$ then we are done, since then (\emptyset, \emptyset) is a c -bridge system for R . Otherwise there exists a horizontal line H such that $|t(R) \cap H| \geq c|R \cap H|$. Then let B be the set of bottom edges of every outer face in $H \cap R$. See Figure 21, where the dark black edges along the line H are the new bridges. Let S be the set of contours connected in this step. We know that $\kappa(S) \geq 2|t(R) \cap H| \geq 2c|R \cap H| \geq 2c/(1-c)|x(R) \cap H|$, so $|B| \leq (1-c)/(2c)\kappa(S)$. We obtain R' from R by removing the faces bounded by a contour in S , as in Figure 21. Then by induction, there exists a c -bridge system (B', S') of R' . Then $\widehat{B} := B \cup B'$ is a set of bridges connecting the contours in $\widehat{S} = S \cup S'$ to each other and to the boundary of R . Moreover, $|\widehat{B}| \leq \frac{1-c}{2c} \kappa(\widehat{S})$ and the number of unbridged tiles is at most c times the number of unbridged faces. Hence $(\widehat{B}, \widehat{S})$ is a c -bridge system for R . ■

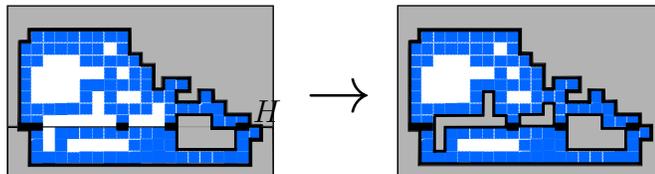


Figure 21: Before and after one step of the construction of a c -bridge system for a region R ; the solid lighter grey area is exterior to R

Once we have a c -bridge system, we can apply a map in which we *complement* an entire region of faces, making tiled faces empty and vice versa. This map significantly reduces the perimeter, but can dramatically change the total number of tiles. Recall we must maintain the total number of tiles, so we may need to supplement by adding extra tiles from another region or we may have extra tiles, which we will put in our “bank” for later. At the end of the process we will find a roughly square region that we can again complement using the bank of extra tiles so that the total number of tiles is restored to bn^2 at minimal cost.

Finding a Sparse Box. We now show that after removing all but cn^2 tiles, there exists a roughly square region of low density where we can place the tiles in our bank.

Lemma 5.3.4: For $(b - c)n^2 \leq a < bn^2$, there exists a constant $n_3 = n_3(b)$ such that for all $n \geq n_3$, if ρ is a configuration with at most cn^2 tiles then ρ contains a region R' such that complementing R' requires a additional tiles and the change in total perimeter is at most $5\sqrt{a}$.

PROOF: Given a region R , let $d(R)$ denote the number of tiles needed to complement R ; this is exactly the area of R minus twice the number of tiles in R . Let $l = \lceil \sqrt{8a/7} \rceil$. First we show that there exists a square $l \times l$ region R such that $d(R) \geq a$. Assume that such a $l \times l$ region does not exist. Divide the grid into $\lfloor \frac{n}{l} \rfloor^2$ disjoint squares with side length l and consider any square. Let t be the number of tiles in the square. The empty volume is at least $l^2 - t$. By assumption each square satisfies $l^2 - t < t + a$, and so $t > \frac{l^2 - a}{2}$. In particular, $8a/7 \leq l^2 < a + 2t \leq a + 2cn^2$, so we know $a < 14cn^2$. This implies that $l \leq \sqrt{8a/7} + 1 \leq 1 + 4\sqrt{cn}$. However, if T is the total number of tiles,

$$cn^2 \geq T > \left\lfloor \frac{n}{l} \right\rfloor^2 \frac{l^2 - a}{2} \geq \frac{n^2}{2} \left(1 - \frac{l}{n}\right)^2 \left(1 - \frac{a}{l^2}\right) > \frac{n^2(1 - \frac{1}{n} - 4\sqrt{c})^2}{16} \geq cn^2,$$

since $c \leq \frac{1}{65}$ and $n \geq n_3$, a contradiction. Therefore there exists an $l \times l$ square R such that $d(R) \geq a$. Remove faces from R one at a time, starting with the bottom row of R and moving across, until we obtain a region $R' \subseteq R$ with $d(R') = a$. This can be done because removing one face at a time changes d by at most 1. This region R' is roughly square and has perimeter at most $4\sqrt{8a/7} < 5\sqrt{a}$.

The Proof of Theorem 5.3.1. Finally we can prove Theorem 5.3.1, showing that for large λ a typical configuration will have the clustering property.

Proof of Theorem 5.3.1. Let $\sigma \in \hat{\Omega} \setminus \Psi$. Construct a c -bridge system (B, S) for L_n as described in Lemma 5.3.3. That is, (B, S) is a set of bridges in L_n connecting some

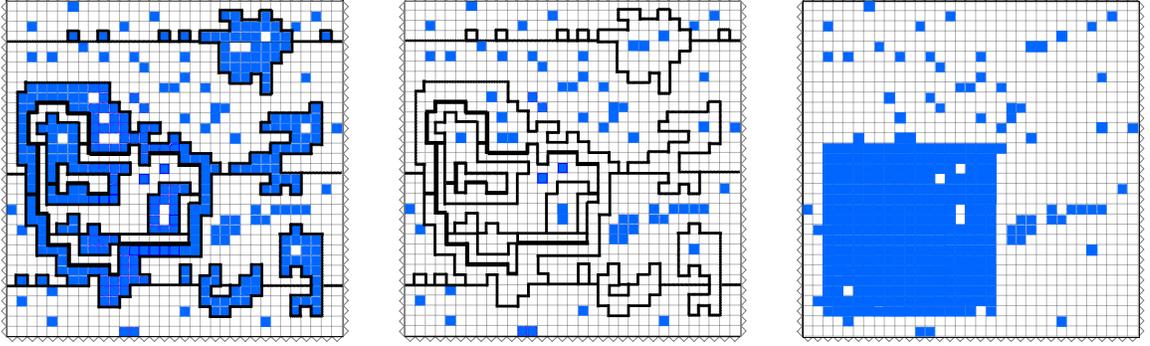


Figure 22: A c -bridge system for $\sigma \in \hat{\Omega} \setminus \Psi$; the image $f_1(\sigma)$; and $f(\sigma) = f_2 \circ f_1(\sigma)$

of the components, some of the holes within those components, some of the islands within those holes, etc. For any bridged contour X , let $r(X)$ be the region bounded by X . If $r(X)$ is a component with holes, then we remove all outer tiles of $r(X)$ and complement all unbridged holes in X , using a subset of the tiles removed to fill in the holes. If $r(X)$ is a hole with islands, then we leave all of the unbridged islands alone. At this point, after complementing some number of regions, we have a bank of extra tiles; let a be the number of tiles in the bank. Notice that by the definition of a c -bridge system, the density of tiles remaining is at most c , so $a \geq (b - c)n^2$.

Let $f_1(\sigma)$ be obtained from σ by removing the bridged components and complementing as described above. Let \mathcal{F}_1 be the image of f_1 on $\hat{\Omega} \setminus \Psi$; note that $\mathcal{F}_1 \not\subset \hat{\Omega}$ since the configurations in \mathcal{F}_1 have too few tiles. Let κ be the total perimeter of all contours bridged. Then for any $\rho \in \mathcal{F}_1$, we claim that the number of preimages of ρ whose bridged contours have total perimeter κ is at most 5^{c_3} for $c_3 = (1 + \frac{1-c}{2c} + \frac{1}{8\sqrt{a}})\kappa$. Consider the c -bridge system obtained above for L_n . Let V denote the leftmost vertical edges of the region. Let $S' = S \cup V$. We perform what is essentially a depth-first-search traversal of the bridge system on S' , starting at the top left corner of L_n . As we traverse an edge we record what type of edge it was using. Then we ‘encode’ the location of the bridges and contours using five bits that represent forward, left, right, bridge east, or bridge west; note that all bridges are horizontal edges, so all edges

in B fall into one of these 5 categories. Whenever we encounter a new bridge B_i , we “process” that bridge by traversing it from the previous contour C_i to the next contour C_{i+1} , then traversing the edges of C_{i+1} . If we encounter another bridge, we process it before continuing. We finish processing B_i when we return to the intersection of C_{i+1} with B_i . Finally, we jump back to the intersection of B_i with C_i and continue traversing C_i . Given the encoded information, there is a unique way to distribute the contours. Hence for all perimeters $\kappa \geq 8\sqrt{a}n$ the number of preimages of ρ whose bridged contours have total perimeter κ is at most $5^{|B|+\kappa+n} \leq 5^{c_3}$. Therefore $|f_1^{-1}(\rho)| \leq \sum_{\kappa \geq 8\sqrt{a}n} 5^{c_3}$.

Let $\rho \in \mathcal{F}_1$ with $bn^2 - a$ tiles. Lemma 5.3.4 shows how to find a region S' in ρ to complement using the a tiles from the bank to obtain τ in such a way that $\kappa(\tau) - \kappa(\rho) \leq 5\sqrt{a}$. Let $f_2(\rho) = \tau$ and $f = f_2 \circ f_1$. We can encode the boundary of S' with $n^2 3^{\kappa(S')} \leq n^2 3^{5\sqrt{a}}$ information. Hence for any $\tau \in \Psi$,

$$|f^{-1}(\tau)| \leq n^2 3^{5\sqrt{a}} \max_{\rho \in f_2^{-1}(\tau)} |f_1^{-1}(\rho)|.$$

Let $\sigma \in \hat{\Omega} \setminus \Psi$, and as above let κ be the total perimeter of components bridged in σ (recall $\kappa(\sigma)$ is the total perimeter of all contours in σ). If $\kappa \leq 8\sqrt{a}$, then $\sigma \in \Psi$, a contradiction. To see this, define the parity of a face to be 1 if it is contained within an odd number of bridged contours and 0 otherwise, and let R be the set of faces with parity 1. Then R has density at least $1 - c$, perimeter at most $8\sqrt{a}$ and $a \geq (b - c)n^2$ tiles. Moreover, \bar{R} has density at most c . Thus R is the region we require, and so $\sigma \in \Psi$. This implies $\kappa > 8\sqrt{a}$. We have shown that $\kappa(\sigma) - \kappa(f(\sigma)) > \kappa - 5\sqrt{a} > \kappa/4$. Let $\tau \in \hat{\Omega}$ and define $f_\kappa^{-1}(\tau)$ to be the set of configurations with perimeter κ that map to τ . Then $|f_\kappa^{-1}(\tau)| \leq n^2 \left(3^{\frac{1}{2\sqrt{7}}} 5^{1 + \frac{1-c}{2c} + \frac{1}{16\sqrt{b}}} \right)^\kappa$ and so

$$\pi(\tau)^{-1} \sum_{\sigma \in f^{-1}(\tau)} \pi(\sigma) \leq \sum_{\sigma \in f^{-1}(\tau)} \mu^{\kappa(\sigma) - \kappa(f(\sigma))} \leq \sum_{\kappa=8\sqrt{a}}^{2n^2} \mu^{\kappa/4} |f_\kappa^{-1}(\tau)| \leq \gamma_1^n,$$

for some $\gamma_1 < 1$, if $\mu \leq \mu^* < \left(3^{\frac{1}{2\sqrt{7}}} 5^{1 + \frac{1-c}{2c} + \frac{1}{16\sqrt{b}}} \right)^{-4}$. Thus the theorem holds if

$$\lambda \geq \lambda^* = \mu^{*-2} - 1.$$

Comparing Model 1 with Model 2. As a corollary to Theorem 5.3.1, we find that for **Model 1** if a configuration has the clustering property then there exists an $n^{1/3} \times n^{1/3}$ window with high density and one with low density. However, in contrast, we will see that for **Model 2**, regardless of λ , the probability that any $n^{1/3} \times n^{1/3}$ box has high density d such that $d > 1.5b$ or low density, $d < 0.5b$, is exponentially small.

Corollary 5.3.5: For $0 < b \leq 1/2$ there exists a constant $n_4 = n_4(b)$ such that for all $n > n_4$, if σ satisfies the clustering property then σ contains square $n^{1/3} \times n^{1/3}$ windows W_1 and W_2 such that the density of tiles in W_1 is at least $.99(1 - c)$ and the density of tiles in W_2 is at most $2.1c$.

PROOF: Let $\sigma \in \Psi$ and let R be the active region given by the clustering property. Consider the set of $n^{4/3}$ windows of side length $n^{1/3}$ that tile L_n . Since R has $a \geq (b - c)n^2$ tiles and density at least $1 - c$, we know that at most $an^{-2/3}/(1 - c)$ windows are contained completely within R . Similarly, since R has perimeter at most $8\sqrt{b}n$ we know that at most $8\sqrt{b}n$ windows intersect the boundary of R . This means that there exists a constant $n_5 = n_5(b)$ such that for $n > n_5$, there exists a window with density d satisfying,

$$d \geq \frac{a}{n^{2/3} \left(\left(\frac{a}{1-c} \right) n^{-2/3} + 8\sqrt{b}n \right)} \geq \frac{99}{100}(1 - c).$$

Next, consider the region $\bar{R} = L_n - R$. From the clustering property we know that \bar{R} has area at least $n^2 - a/(1 - c)$ and contains at most cn^2 tiles. This implies that there are at least $n^{-2/3}(n^2 - a/(1 - c))$ windows intersecting \bar{R} . At most $8\sqrt{b}n$ of these windows can intersect R and these contain at most $8\sqrt{b}n^{5/3}$ tiles from R . Combining these observations implies that there exists a constant $n_6 = n_6(b)$ such that for $n > n_5$,

there exists a window with density d satisfying,

$$d \leq \frac{cn^2 + 8\sqrt{bn}^{5/3}}{n^{2/3}n^{-2/3} \left(n^2 - \frac{a}{1-c}\right)} \leq \frac{99}{49}c.$$

Model 2: A -tiles are unit squares on L_n and B -tiles are squares of side length $1/2$ on the half-integer lattice, (see Figure 24(c)). This model is qualitatively different from the previous models since *the placement* of the A -tiles does not influence *the number of places* in which we can put the B -tiles. In fact, this model is just bond percolation on a rotated grid with a fixed number of edges, where we do not expect clustering at any density. To see the bijection, label a unit square with a Northwest-Southeast diagonal if it lies on an even face and label it with a Northeast-Southwest diagonal otherwise, as in Figure 24(d). Notice that these lines form a subset of the edges of a rotated grid. If we have bn^2 A -tiles then each edge in the rotated grid is present with probability b .

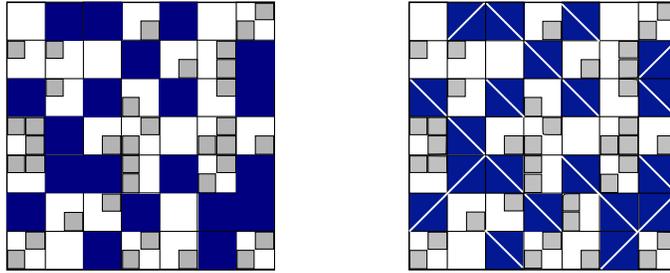


Figure 23: Model 2 and the connection with bond percolation

To illustrate the difference between the behavior of **Model 2** and the interfering binary mixtures, consider an $n^{1/3} \times n^{1/3}$ window in each. In **Model 2**, the probability that any $n^{1/3} \times n^{1/3}$ box has density d such that $d > 1.5b$ or $d < 0.5b$ is less than γ_3^n for some constant $\gamma_3 < 1$. This is straightforward to show since each configuration of bn^2 tiles in n^2 locations has equal likelihood. Thus, the probability that a fixed window has density d is exactly $\binom{n^{2/3}}{dn^{2/3}} \binom{n^2 - n^{2/3}}{bn^2 - dn^{2/3}} / \binom{n^2}{bn^2}$. Using standard approximations and union bounds we can obtain the desired result. In contrast, by Corollary 5.3.5, a configuration with the clustering property has a window with density $d \geq .99(1 - c)$

and a window with density $d \leq 2.1c$. Hence we see markedly different behavior between interfering and non-interfering binary mixtures.

5.3.5 No Clustering at Low Density for Model 1

We now examine the low density case and prove, in the context of **Model 1**, Theorem 5.3.2, stating that at sufficiently low density, typical configurations will not have the clustering property. For small enough λ , the A -tiles will be well-distributed throughout L_n , in the sense that any large dense region must have perimeter on the order of n^2 .

Proof of Theorem 5.3.2. Let $t = \frac{1-2c}{1-c}(b-c)$ and $\delta = (\frac{1-b+t}{t})^t$. Define $\Psi' \subset \widehat{\Omega}$ to be the set of configurations with a region R that has at least $(b-c)n^2$ tiles, perimeter less than αn^2 and density at least $1-c$, where α satisfies $0 < \alpha < \ln 3(\ln(\delta) - b \ln 2)/2$. We will show $\widehat{\pi}(\Psi')$ is exponentially small. From the definition of clustering (Definition 5.4.3), it is straightforward to see that Ψ , the set of configurations that have the clustering property is contained in Ψ' . Thus, if $\widehat{\pi}(\Psi')$ is exponentially small then clustering is exponentially unlikely to occur.

For each $\sigma \in \Psi'$, let R be the lexicographically first region which meets the conditions given above. We will “flip” each face in R (tiles become empty faces and vice versa) to obtain $f(\sigma)$. Since R has density at least $1-c$ and at least $(b-c)n^2$ tiles, this means that there are at most $\frac{c}{1-c}(b-c)n^2$ empty faces in R . So by flipping R we are left with a bank of a_σ tiles such that $a_\sigma \geq \frac{1-2c}{1-c}(b-c)n^2 = tn^2$. Next, we define $N(\sigma)$ to be the set of all configurations obtained from $f(\sigma)$ by adding a_σ tiles back in any a_σ empty locations; then $|N(\sigma)| = \binom{n^2 - (bn^2 - a_\sigma)}{a_\sigma}$. For each $\tau \in \widehat{\Omega}$, we need to bound the number of configurations σ such that $\tau \in N(\sigma)$. For any configuration τ , there are at most 2^{bn^2} configurations β such that $\beta = f(\sigma)$ and $\tau \in N(\sigma)$ for some $\sigma \in \Psi'$. This is due to the fact that there are 2^{bn^2} ways to choose which bn^2 tiles were in their original location and which were removed by f . For each such configuration β

we bound the number of regions R that could have been removed in order to recover the original σ . There are at most bn^2 ways to select an A -tile on the border of R and $3^{\alpha n^2}$ possible perimeters for R , since R has perimeter less than αn^2 . Thus for any configuration τ there are at most $2^{bn^2}(bn^2 3^{\alpha n^2}) \leq (2^b \delta)^{n^2/2}$ configurations σ such that $\tau \in N(\sigma)$.

Finally, we define a weighted bipartite graph $G(\Psi', \widehat{\Omega}, E)$ with an edge of weight $\pi(\sigma)$ between $\sigma \in \Psi'$ and $\tau \in \widehat{\Omega}$ if $\tau \in N(\sigma)$. The total weight of edges is

$$\sum_{\sigma \in \Psi'} \pi(\sigma) |N(\sigma)| \geq \sum_{\sigma \in \Psi'} \pi(\sigma) \binom{n^2 - (bn^2 - a_\sigma)}{a_\sigma} \geq \pi(\Psi') \delta^{n^2}.$$

However, the weight of the edges is at most $\sum_{\tau \in \widehat{\Omega}} \pi(\tau) \mu^{-4bn^2} (2^b \delta)^{n^2/2}$. Let $\mu^* = (2^b/\delta)^{1/(8b)}$ and $\lambda^* = (\mu^*)^{-2} - 1$. Thus for all $\mu < \mu^*$,

$$\pi(\Psi') < \mu^{-4bn^2} (2^b \delta)^{n^2/2} \delta^{-n^2} < \gamma_2^n,$$

for some $\gamma_2 < 1$, completing the proof.

5.4 *Generalizing from Model 1 to Interfering Binary Mixtures*

One may expect behavior similar to **Model 1** as long as our binary mixtures combine tiles that “interfere” in that there has to be some space between them when they are non-overlapping. Indeed this is the case for the class of “interfering binary mixtures” which we define in order to make this intuition rigorous. We will now show how to extend the analysis from **Model 1** to prove that interfering binary mixtures exhibit clustering at sufficiently high density, whereas at sufficiently low density the A -tiles will be well-distributed. Conceptually all of the ideas presented in this section were already introduced in the context of **Model 1**. However, the definitions are quite technical because we must define the gain from a Peierls argument in terms of perimeter, area and overlaps of tiles. In addition, we need to generalize the concept of bridges that were critical to the previous arguments. We start by a series of

definitions to capture these concepts, and then move on to the more technical proofs.

5.4.1 Defining Interfering Binary Mixtures

We begin by formally defining interfering binary mixtures. We require that Λ_A and Λ_B be divisible and bridgeable to ensure that the model is reasonably well-behaved. We begin by defining these restrictions. See Section 5.4.2 for examples of bridgeable and divisible graphs. Let Λ_f be a planar graph whose faces are all isomorphic to f .

Definition 5.4.1: We say Λ_f is *divisible* with parameter ν' if for every integer $n > 0$, there exists a subgraph R_n of Λ_f , which contains n^2 faces, has outer perimeter $\nu'n$, and for each integer $0 < k < n$ there exists a packing of $\lfloor \frac{n}{k} \rfloor^2$ non-overlapping copies of R_k in R_n . We will refer to R_n as an n -box.

For example, **Model 1** is divisible. To see this, consider a $n \times n$ subgraph of \mathbb{Z}^2 and let this be the n -box, R_n . The subgraph R_n has area n^2 and perimeter $4n$. There exists an $n \times n$ subgraph R_n for every integer $n > 0$. For each integer $0 < k < n$ there exist $\lfloor \frac{n}{k} \rfloor^2$ non-overlapping $k \times k$ regions contained in a $n \times n$ region (i.e., a packing of $\lfloor \frac{n}{k} \rfloor^2$ k -boxes). Thus, **Model 1** is divisible with parameter $\nu' = 4$. See Figure 28 and Figure 29 for other examples of boxes and packings.

Definition 5.4.2: Let Λ_f be a divisible planar graph with n -box R_n and \mathcal{H} be the set of horizontal lines which intersect at least one vertex of R_n . The graph Λ_f is *bridgeable* with parameters h_1, h_2 if

- for every horizontal line $H \in \mathcal{H}$, if H has length L (where L is the length of H contained within R_n), then the maximum number of vertices or edges intersecting H , discounting any edges which have at least one vertex intersecting H , is at most h_1L ,
- and the number of horizontal lines $|\mathcal{H}|$ satisfies $|\mathcal{H}| \leq h_2n$.

Let Λ_A be a bridgeable, divisible planar graph with all faces isomorphic to A and n -box R_n . Let Λ_B be a planar graph where all faces are isomorphic to B . We will restrict our attention to the intersection of Λ_A and Λ_B with R_n ; specifically, $L_A = \Lambda_A \cap R_n$ and $L_B = \Lambda_B \cap R_n$. A *binary mixture* is a (nonoverlapping) packing of the region R_n with A -tiles (lying on faces of L_A) and B -tiles (lying on faces of L_B). For any set S of tiles, define $per(S)$ to be the set of edges adjacent to a tile in S and a tile not in S and for any contour C , let $|C|$ be the total length of the contour C . Define $a(S)$ to be the area of the tiles in S . A binary mixture is *interfering* if there exist constants δ and γ such that for any face x in L_B and any set S of faces in L_A for which $|per(S) \cap x| \neq 0$,

$$0 < \delta \leq \frac{a(S \cap x)}{|per(S) \cap x|} \leq \gamma. \quad (5.4.1)$$

Given constants $\lambda > 1$, and $0 < b < 1/2$, where $bn^2 \in \mathbb{Z}$, define $\Omega = \Omega(b, \lambda)$ as the set of non-overlapping packings of L with bn^2 A -tiles and any number of B -tiles (where a tile can only be placed on a face of its type). We wish to study the distribution $\pi(\rho) = \lambda^{d(\rho)}/Z$, where $d(\rho)$ is the number of B -tiles in ρ and $Z = \sum_{\rho \in \Omega} \lambda^{d(\rho)}$ is a normalizing constant. Our goal is to determine whether a configuration chosen according to π is likely to have clusters of A -tiles. Notice that these models can all be seen as weighted independent sets on some graph (including the Ising model), namely, the graph whose vertex set is the set of centers of all faces in L_A and L_B and where two vertices are adjacent if their corresponding faces intersect nontrivially.

Next, we will define several parameters of interfering binary mixtures that will be used in our proofs. These parameters will be vary depending on the specific model. Define α to be the area of a A -tile and β to be the area of a B -tile. Let ν be the perimeter of a A -tile. Let Δ be the maximum degree of any vertex in L_A and ν_e be the length of the shortest edge in L_A . Recall from the definition of a binary mixture that each model also has parameters ν' , h_1 , h_2 , δ and γ because it is divisible, bridgeable and interfering. Table 1 shows the values of these parameters

for the example interfering binary mixtures given in Section 5.4.2.

5.4.2 Example Interfering Binary Mixtures

First, we will see a few other examples of interfering binary mixtures, which help to illuminate the changes necessary to generalize the proof of Theorem 5.3.1.

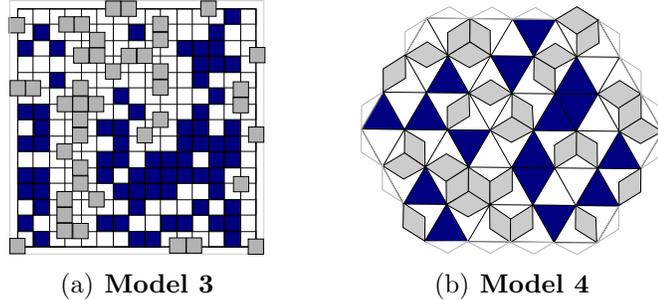


Figure 24: Example drawings of **Model 3** and **Model 4**.

Model 3: A -tiles are unit squares on the grid, $L_A = L_n$ and B -tiles are unit squares centered on the vertices of L_A , see Figure 24(a). It is not hard to see that this model corresponds exactly to an independent set model on the rotated grid where vertices correspond to the centers of A -tiles and B -tiles, and the number of even vertices (A -tiles) is fixed. The number of odd vertices (B -tiles) varies according to λ . To see that **Model 3** is divisible, consider a square of size n by n which contains n^2 faces of the grid. This has perimeter $4n$ and clearly meets the divisible conditions with $\nu' = 4$. For this model the set \mathcal{H} is simply the set of horizontal grid lines. Since there are exactly $n + 1$ of these, we can let $h_2 = 2$ thus $|\mathcal{H}| = n + 1 \leq 2n = h_2n$ as required. For each horizontal line $H \in \mathcal{H}$, the number of vertices or edges intersecting H , discounting any edges which have at least one vertex intersecting H , is exactly the number of vertices intersecting H which is $n + 1$. Since each horizontal line H has length n we let $h_1 = 2$ and thus $n + 1 \leq 2n = h_1(n)$ as required. To determine appropriate values for δ and γ , consider any B -tile x and all choices of S (sets of tiles) that affect x where $|per(S) \cap x| \neq 0$. There are 4 cases as shown in Figure 25. Note that each case displays a single face of L_B . The dark blue regions corresponds

to fragments of A -tiles that overlap the face, each of which is only $1/4$ of the actual A -tile which is contained in S . In order to determine δ and γ we need to calculate the ratio $r = \frac{a(S \cap x)}{|per(S) \cap x|}$ in each case, where $|per(S) \cap x|$ is the portion of the perimeter of S that overlaps with x which is highlighted by the yellow line (in Figure 25) and $a(S \cap x)$ is the area of the region in dark blue. Specifically, in the first case $|per(S) \cap x| = 1$ and $a(S \cap x) = 3/4$ so $r = \frac{a(S \cap x)}{|per(S) \cap x|} = 3/4$. By calculating the ratio r in each case we determine that $\delta = 1/4$, $\gamma = 3/4$ suffice and conclude that **Model 3** is an interfering binary mixture.

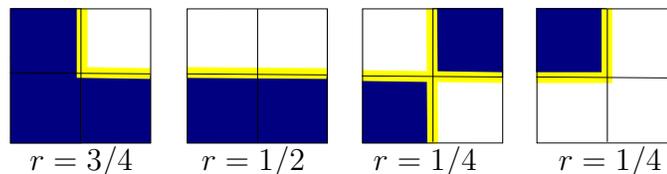


Figure 25: Computing δ and γ for **Model 3**.

Model 4: A -tiles are triangles with perimeter three on the triangular lattice L_A and B -tiles are lozenges bisected by edges of L_A , (see Figure 24(b)). **Model 4** maps bijectively onto an Ising Model with fixed magnetization on L_A . To see that **Model 4** is divisible, consider a box of size n by n as shown in Figure 28(a) which contains n^2 faces of the triangular lattice. This has perimeter $3n$ and these boxes can always be packed tightly by flipping every other box vertically as shown in Figure 29(a). Thus this definition of a n -box clearly meets the divisible conditions with $\nu' = 3$. For this model the set \mathcal{H} is simply the set of horizontal grid lines. Since there are exactly $n + 1$ of these, we can let $h_2 = 2$ thus $|\mathcal{H}| = n + 1 \leq 2n = h_2 n$ as required. For each horizontal line $H \in \mathcal{H}$, the number of vertices or edges intersecting H , discounting any edges which have at least one vertex intersecting H is the number of vertices intersecting H . If n is odd, then the number of vertices on a line is either $\lceil n/2 \rceil + 1$ or $\lceil n/2 \rceil$ and the lines have length $\lceil n/2 \rceil$ or $\lfloor n/2 \rfloor$ respectively. If n is even, then all lines have $n/2 + 1$ vertices and length $n/2$. Thus if $h_1 = 2$ suffices. To determine

appropriate values for δ and γ , consider any B -tile x and all choices of S that affect x where $|per(S) \cap x| \neq 0$. There is only one case with $r = \sqrt{3}/12$ as shown in Figure 26. Thus $\delta = \gamma = \sqrt{3}/12$ suffice and we can conclude that **Model 4** is an interfering binary mixture.

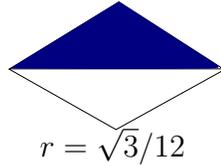


Figure 26: Computing δ and γ for **Model 4**.

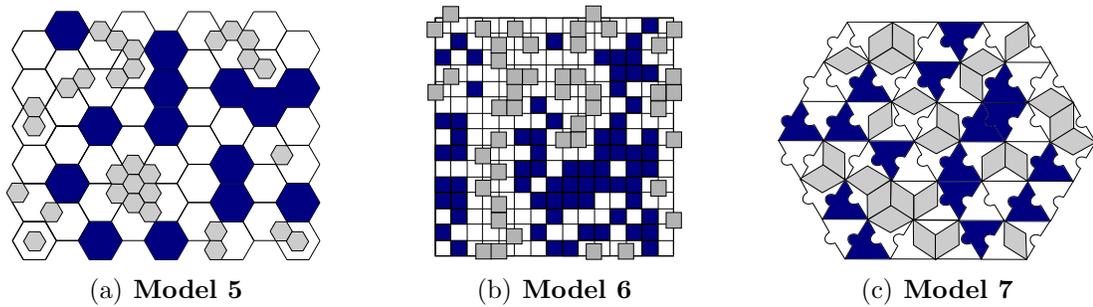


Figure 27: Examples drawings of **Model 5**, **Model 6** and **Model 7**.

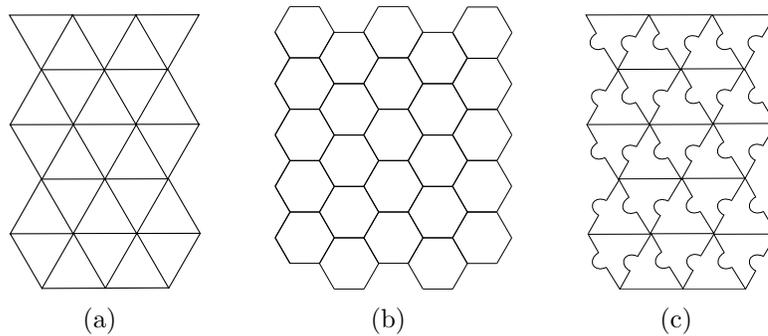


Figure 28: n -boxes ($n = 5$) for (a) **Model 4** (b) **Model 5** (c) **Model 7**

Model 5: A -tiles are hexagons with perimeter 6 on the hexagonal lattice L_A and B -tiles are also hexagons on a smaller hexagonal lattice L_B where each face of L_A contains one face of L_B and bisects 6 others as shown in Figure 27(a). To see that **Model 5** is divisible, consider a box of size n by n as shown in Figure 28(b) with

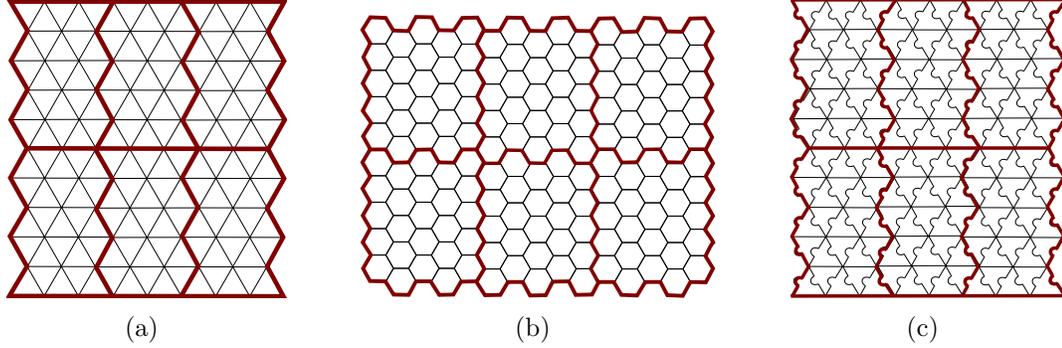


Figure 29: A packing of 5-boxes for (a) **Model 4** (b) **Model 5** (c) **Model 7**

perimeter $8n$ which contains n^2 faces of the hexagonal lattice. This has perimeter $4n$ and, similar to the triangular lattice, these boxes can be packed tightly by flipping every other box vertically as shown in Figure 29(b). Thus this definition of a n -box clearly meets the divisible conditions with $\nu' = 8$. For this model the set \mathcal{H} has size $2n + 2$, so we can set $h_2 = 3$ and thus $|\mathcal{H}| = 2n + 2 \leq 3n = h_2n$ as required. For each horizontal line $H \in \mathcal{H}$, the number of vertices or edges intersecting H , discounting any edges which have at least one vertex intersecting H , is exactly the number of vertices intersecting H which is at most $n + 1$. Since each horizontal line H has length l greater than n , set $h_1 = 2$ and thus $n + 1 \leq 2n = h_1n \leq h_1(l)$ as required. To determine appropriate values for δ and γ , there is one case with $r = 3\sqrt{3}/16$ as shown in Figure 30. Thus, $\delta = \gamma = 3\sqrt{3}/16$ suffice and we can conclude that **Model 5** is an interfering binary mixture.

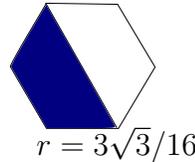


Figure 30: Computing δ and γ for **Model 5**.

Model 6: A -tiles are unit squares on $\Lambda_A = L_n$ and B -tiles are unit squares on Λ_B , where Λ_B is L_n shifted up vertically by $2/3$ and horizontally by $1/2$ as shown in Figure 27(b). Since Λ_A is the same as in **Model 2** we have already shown that

it is divisible with $\nu' = 4$ and bridgeable with $h_1 = 2$ and $h_2 = 2$. To determine appropriate values for δ and γ , we calculate the ratio r for all seven cases given in Figure 31. Thus, we can let $\delta = 1/5$, $\gamma = 1$ and conclude that **Model 6** is an interfering binary mixture.

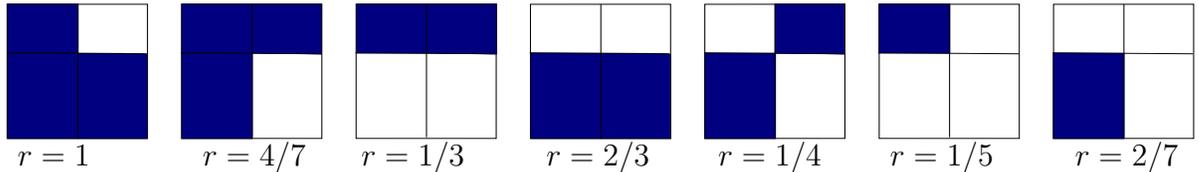


Figure 31: Computing δ and γ for **Model 6**.

Model 7: A -tiles are modified triangles sitting on L_A , a modified version of the triangular lattice where on a subset of edges a semicircle with radius $1/8$ is added or removed as shown in Figure 27(c). B -tiles are lozenges bisected by the edges of the triangular lattice. To see that **Model 7** is divisible, consider a box of size n by n as shown in Figure 28(c) which contains n^2 faces of the modified triangular lattice. This has perimeter $\frac{(10+\pi)n}{4}$ and these boxes can be packed tightly by flipping every other box vertically as shown in Figure 29(c). Thus this definition of a n -box clearly meets the divisible conditions with $\nu' = \frac{10+\pi}{4}$. For this model the set \mathcal{H} is simply the set of horizontal grid lines. Since there are exactly $n + 1$ of these, we can let $h_2 = 2$ thus $|\mathcal{H}| = n + 1 \leq 2n = h_2n$ as required. For each horizontal line $H \in \mathcal{H}$, the number of vertices or edges intersecting H , discounting any edges which have at least one vertex intersecting H is the number of vertices intersecting H which is $n + 1$. Since each horizontal line H intersects $2n$ faces of L_A we let $h_1 = 1$ and thus $n + 1 \leq 2n = h_1(2n)$ as required. To determine appropriate values for δ and γ , there are 3 cases as shown in Figure 32. Thus, $\delta = \frac{32\sqrt{3}-3\pi}{48(6+\pi)}$ and $\gamma = \frac{32\sqrt{3}+3\pi}{48(6+\pi)}$ suffice and we can conclude that **Model 7** is an interfering binary mixture.

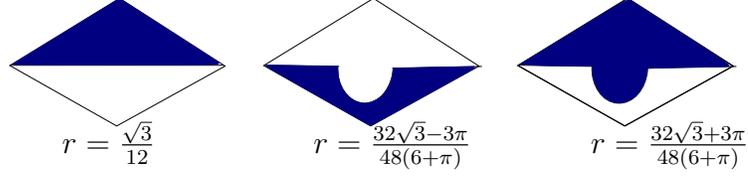


Figure 32: Computing δ and γ for **Model 7**.

Table 1: Parameters for the Example Interfering Binary Mixtures

Parameters		Models					
		1	3	4	5	6	7
α	area of an A -tile	1	1	$\frac{\sqrt{3}}{4}$	$\frac{3\sqrt{3}}{2}$	1	$\frac{\sqrt{3}}{4}$
β	area of an B -tile	$\frac{1}{2}$	1	$\frac{\sqrt{3}}{6}$	$\frac{3\sqrt{3}}{8}$	1	$\frac{\sqrt{3}}{6}$
Δ	max degree of vertex in L_A	4	4	6	3	4	6
γ	upper bound, area to per. ratio	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{\sqrt{3}}{12}$	$\frac{3\sqrt{3}}{16}$	1	$\frac{32\sqrt{3}+3\pi}{48(6+\pi)}$
δ	lower bound, area to per. ratio	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{\sqrt{3}}{12}$	$\frac{3\sqrt{3}}{16}$	1/5	$\frac{32\sqrt{3}-3\pi}{48(6+\pi)}$
ν	perimeter of an A -tile	4	4	3	6	4	$\frac{10+\pi}{4}$
ν_e	length of shortest edge in L_A	1	1	1	1	1	1
ν'	perimeter of a “ n -box”	4	4	3	8	4	$\frac{10+\pi}{4}$
h_1	upper bound, intersections along a line in L_A	2	2	2	2	2	2
h_2	number of horizontal lines in L_A	2	2	2	3	2	2

5.4.3 Relating the Weight of a Configuration to the Perimeter

Recall that the A -structure $\Gamma(\rho)$ of a configuration ρ in Ω is the configuration σ obtained from ρ by removing all of its B -tiles. The set $\widehat{\Omega}$ of all such A -structures with bn^2 A -tiles is called the *projection* of the model. Let $\widehat{\pi}$ be the induced distribution on $\widehat{\Omega}$; that is, for $\sigma \in \widehat{\Omega}$, let $\widehat{\pi}(\sigma) = \sum_{\rho \in \Gamma^{-1}(\sigma)} \pi(\rho)$. For $\sigma \in \widehat{\Omega}$, let $\mathcal{O}(\sigma)$ be the number of *open B -vertices* in σ ; that is, B -vertices which do not intersect any A -tiles in σ . This is the number of places where a B -tile can go. We find that

$$\widehat{\pi}(\sigma) = \sum_{k=0}^{\mathcal{O}(\sigma)} \frac{\lambda^k}{Z} \binom{\mathcal{O}(\sigma)}{k} = \frac{1}{Z} (1 + \lambda)^{\mathcal{O}(\sigma)}. \quad (5.4.2)$$

Thus, $\mathcal{O}(\sigma)$ completely determines the probability that it will show up as an A -structure of a configuration in Ω .

Our goal is to show that our process outlined in Section 5.3.4 for changing the perimeter of a configuration does in fact change the number of open B -vertices as

well (which in turn increases the weight of the configuration). This would allow us to infer that all Interfering Binary Mixtures have the clustering property at high density. Define a component of A -tiles as a maximal connected subset, where x and y are adjacent if they share a vertex (and/or an edge). The number, $\mathcal{O}(\sigma)$, of open vertices in σ is equal to the total number of vertices minus the number of blocked vertices in σ . Thus for any configuration σ , we want to bound the size of $B(\sigma)$, defined as the set of B -faces blocked by A -tiles in σ , in terms of the length of the perimeter of σ , $\kappa(\sigma)$. Define $h(\sigma) := a(B(\sigma)) - a(\sigma)$ to be the area of B -tiles hanging off of X where $a(\sigma)$ is total areas of the A -tiles in σ and $a(B(\sigma))$ is the total area of the B -faces in $B(\sigma)$. Then we obtain the number of blocked B -faces by noting that $|B(\sigma)| = a(B(\sigma))/\beta = \frac{a(\sigma)+h(\sigma)}{\beta}$.

For the Ising models, there is a direct correspondence between perimeter and the number of open B -vertices. However, in general, the shape of a component determines the exact relationship between its perimeter and the area of B -tiles hanging off of the component. Still, for any interfering binary mixture, we can provide constant bounds on their ratio using the bounds given in Equation 5.4.1. Recall that for any B -tile x , if $A(x)$ is the set of A -tiles intersecting x , then for any subset $S \subset A(x)$ such that $|per(S) \cap x| > 0$,

$$0 < \delta \leq \frac{a(x \cap S)}{|per(S) \cap x|} \leq \gamma.$$

Therefore, for any component X , we have $0 < \delta\kappa(X) \leq h(X) \leq \gamma\kappa(X)$, and so

$$\frac{\alpha|\sigma| + \delta\kappa(\sigma)}{\beta} \leq B(\sigma) \leq \frac{\alpha|\sigma| + \gamma\kappa(\sigma)}{\beta}. \quad (5.4.3)$$

5.4.4 Clustering at High Density for Interfering Binary Mixtures

We concentrate first on interfering binary mixtures at high density. The goal of this section is to extend Theorem 5.3.1 for the general class of interfering binary mixtures. First we show how to prove the analogs of Lemmas 5.3.3 and 5.3.4 for all interfering binary mixtures. These imply that a c -bridge system exists and there is an appropriate

region to flip in order to place the remaining tiles in our bank. Finally we use these lemmas to prove that clustering occurs at sufficiently high density for all interfering binary mixtures.

Building Bridges. First we modify the process for building bridge systems to work in the general setting of interfering binary mixtures. Since the edges of Λ_A can take many shapes and orientations, we will use straight horizontal lines as the bridges rather than edges of Λ_A . As before, we will identify a set B of horizontal lines bridging a set S of contours. We will define a flip operation f_1 which will again remove a tiles and all contours in S , leaving a configuration with at most cn^2 tiles. Specifically, $f_1(\sigma, S)$ will do the following: for each face f in σ bounded by an odd number of contours in S , it flips f (tiles become empty faces and empty faces become tiles), whereas for each face bounded by an even number of contours in S it does nothing. If B is a set of horizontal lines in \mathcal{H} and S is a set of contours bridged by B , then we say (B, S) is a c -**bridge system** if after applying f_1 to S , there are at most cn^2 tiles and $|B| \leq \frac{\kappa(S)}{2c}$ where $|H|$ is the length of H and $|B| = \sum_{H \in B} |H|$

Lemma 5.4.1: There exists a c -bridge system for any configuration $\sigma \in \hat{\Omega}$.

PROOF: We begin by introducing some terminology. For any horizontal line H , let $s(H)$ be the contours that intersect H and let $t(H)$ be the combined length of the segments of H that are either contained within or adjacent to a tile. In order to prove the lemma we will first describe an algorithm for finding a c -bridge system for σ which at each step maintains a bridge system (B_i, S_i) with $|B_i| \leq \frac{\kappa(S_i)}{2c}$ (note that this is not a c -bridge system since there may be more than cn^2 tiles left after applying f_1). Initially we start with $B_0 = S_0 = \emptyset$. At each step i of the algorithm we construct $\sigma_i = f_1(\sigma, S_{i-1})$. If there are at most cn^2 tiles in σ_i then we are done. Otherwise, we find a line H for which $t(H) \geq c|H|$ where $|H|$ is the length of the line H , and let $B_i = B_{i-1} \cup H$ and $S_i = S_{i-1} \cup s(H)$. We know at least one such line must exist since

the density of σ_i is strictly greater than c (we assumed there are more than cn^2 tiles in σ_i) and because all faces have the same size and shape. Since $t(H) \geq c|H|$ this implies that $\kappa(s(H)) \geq 2t(H) \geq 2c|H|$. We will repeat this procedure until we reach a step i for which σ_i has at most cn^2 tiles left. Notice that at each step we remove at least one contour and add no new contours. If we remove all contours then there will be zero tiles left and we start with a finite number of contours so this algorithm will terminate. At each step in the algorithm the bridge H that we add satisfies $\kappa(s(H)) \geq 2t(H) \geq 2c|H|$ where $s(H)$ are the contours we add at the same step. Since this is satisfied for each horizontal line and associated contours it is satisfied for the entire set so at the end $|B| \leq \frac{\kappa(S)}{2c}$ as desired and we have a c -bridge system. ■

Finding a Sparse Box. We now show that after removing all but cn^2 tiles, there exists a region with small perimeter and low density where we can place the tiles in our bank. The region will be very similar to a k -box for some k as given by the divisibility condition. For example for **Model 5** the region will look like Figure 28(b).

Lemma 5.4.2: For $(b - c)n^2 \leq a < bn^2$, there exists a constant $n_3 = n_3(b)$ such that for all $n \geq n_3$, if ρ is a configuration with at most cn^2 tiles then ρ contains a region R' such that complementing R' requires a additional tiles and the change in total perimeter is at most $4\nu'\sqrt{a}$.

PROOF: Given a region R , let $d(R)$ denote the number of tiles needed to complement R ; this is exactly the number of faces in R minus twice the number of tiles in R . Let $l = \lceil \sqrt{8a/7} \rceil$. First we show that there exists an l -box such that $d(R) \geq a$. Assume that such a box does not exist. Divide the A -lattice into $\lfloor \frac{n}{l} \rfloor^2$ disjoint l^2 boxes as given by the divisibility condition, and consider any such box. Let t be the number of tiles in the box. The number of empty faces is at least $l^2 - t$. By assumption each l -box satisfies $l^2 - t < t + a$, and so $t > \frac{l^2 - a}{2}$. In particular, $8a/7 \leq l^2 < a + 2t \leq a + 2cn^2$,

because $t < cn^2$ since there are only cn^2 tiles remaining, so we know $a < 14cn^2$. This implies that $l \leq \sqrt{8a/7} + 1 \leq 1 + 4\sqrt{cn}$. However, if T is the total number of tiles,

$$cn^2 \geq T > \left\lfloor \frac{n}{l} \right\rfloor^2 \frac{l^2 - a}{2} \geq \frac{n^2}{2} \left(1 - \frac{l}{n}\right)^2 \left(1 - \frac{a}{l^2}\right) > \frac{n^2 \left(1 - \frac{1}{n} - 4\sqrt{c}\right)^2}{16} \geq cn^2,$$

since $c \leq \frac{1}{65}$ and $n \geq n_3$, a contradiction. Therefore there exists an l -box R_l such that $d(R_l) \geq a$. From the divisibility condition we know that there exists a $(l-1)$ -box R_{l-1} contained in R which in turn contains a $(l-2)$ -box R_{l-2} and so forth. This implies that there exists an i for which $d(R_{l-i}) \geq a \geq d(R_{l-i-1})$ since $d(R_1) < a \leq d(R_l)$. Given this i , start from R_{l-i} and remove faces one at a time, only removing faces that aren't in $d(R_{l-i-1})$ and ensuring that the region remains connected until we obtain a region $R' \subseteq R_{l-i}$ with $d(R') = a$. This can be done because removing one face at a time changes d by at most 1 and we know that $d(R_{l-i}) \geq a \geq d(R_{l-i-1})$. This region has perimeter at most the perimeter of the R_{l-i-1} -box plus the perimeter from the additional faces. Since there are at most $(l-i)^2 - (l-i-1)^2 \leq 2(l-i)$ additional tiles, this additional perimeter is at most $\nu 2(l-i)$. Since $\nu \leq \nu'$ and $i \geq 0$, the total perimeter is at most $\nu'(l-i-1) + \nu 2(l-i) \leq 3\nu' \lceil \sqrt{8a/7} \rceil < 4\nu' \sqrt{a}$.

Proof of the Clustering Theorem. Recall that our definition of the clustering property gives a precise bound on the maximum allowable perimeter for an active region that a configuration with the clustering property can have. However, the precise constant in that bound is not as essential as the fact that the perimeter is of order a square root of the order of the area of the region R . To extend the proof to general interfering binary mixtures, we must modify our definition of clustering slightly to accommodate the bounds we are using to relate the change in open vertices with the change in perimeter. Hence our definition of clustering becomes:

Definition 5.4.3: We say that a configuration $\sigma \in \Omega$ (or $\Gamma(\sigma) \in \widehat{\Omega}$) has the *clustering property* if it contains a region R which satisfies the following properties:

1. R contains at least $(b - c)n^2$ A -tiles,
2. the perimeter of R is at most x_1n , where $x_1 = 8\gamma\nu'\sqrt{b}/\delta$, and
3. the density of A -tiles is at least $1 - c$ in R and at most c in \bar{R} .

Theorem 5.4.3: For any interfering binary mixture and for any $0 < b \leq 1/2$, there exist constants $\lambda^* = \lambda^*(b) > 1$, $\gamma_1 < 1$ and $n_1 = n_1(b)$ such that for all $n > n_1$, $\lambda \geq \lambda^*$ a random sample from Ω will have the clustering property with probability at least $(1 - \gamma_1^n)$.

PROOF: The proof proceeds as in the proof of Theorem 5.3.1. Let $\sigma \in \hat{\Omega} \setminus \Psi$. Construct a c -bridge system (B, S) for L_A as described in Lemma 5.4.1. Next, we apply the flip operation f_1 , defined in the context of building bridges, which flips each face bounded by an odd number of contours. At this point, after complementing some number of regions, we have a bank of a extra tiles. Again, by the definition of a c -bridge system, the density of tiles remaining is at most c , so $a \geq (b - c)n^2$.

Let \mathcal{F}_1 be the image of f_1 on $\hat{\Omega} \setminus \Psi$. Let k be the total perimeter of all contours bridged. Then as before, for any $\rho \in \mathcal{F}_1$, the number of preimages of ρ whose bridged contours have total perimeter k is at most $2^{h_2n}2^{|B|h_1}\Delta^{k/\nu_e}$, since we can encode which horizontal lines were used, which edges of each of those horizontal lines were added as bridges, and the contours themselves (by choosing which of the edges adjacent to a given vertex is next on the contour) using this much information. Since $\Delta \geq 2$, $k > x_1n$ and $|B| \leq k/(2c)$, we have

$$2^{h_2n}2^{|B|h_1}\Delta^{k/\nu_e} \leq \Delta^{\frac{h_2k}{x_1} + \frac{kh_1}{2c} + \frac{k}{\nu_e}} = \Delta^{c_3k},$$

where $c_3 = \frac{h_2}{x_1} + \frac{h_1}{2c} + \frac{1}{\nu_e}$. Therefore $|f_1^{-1}(\rho)| \leq \sum_{k \geq x_1n} \Delta^{c_3k}$.

Let $\rho \in \mathcal{F}_1$ with $bn^2 - a$ tiles. Lemma 5.4.2 shows how to find a region R' in ρ to complement using the a tiles from the bank to obtain τ in such a way that

$\kappa(\tau) - \kappa(\rho) \leq 4\nu'\sqrt{a}$. Let $f_2(\rho) = \tau$ and $f = f_2 \circ f_1$. We can encode the boundary of R' with $n^2\Delta^{\kappa(R')/\nu_e} \leq n^2\Delta^{4\nu'\sqrt{a}/\nu_e}$ information. Hence for any $\tau \in \Psi$,

$$\begin{aligned}
|f^{-1}(\tau)| &\leq n^2\Delta^{4\nu'\sqrt{a}/\nu_e} \max_{\rho \in f_2^{-1}(\tau)} |f_1^{-1}(\rho)| \\
&\leq n^2\Delta^{4\nu'\sqrt{a}/\nu_e} \sum_{k \geq x_1 n} \Delta^{c_3 k} \\
&\leq n^2\Delta^{4\nu'(\frac{k\delta}{8\gamma\nu'})/\nu_e} \sum_{k \geq x_1 n} \Delta^{c_3 k} \\
&= n^2\Delta^{\frac{\delta}{2\gamma\nu_e}} \sum_{k \geq x_1 n} \Delta^{c_3 k} \quad \blacksquare
\end{aligned}$$

Where the second to last step comes from the fact that $a < bn^2$ and $k > x_1 n = (8\gamma\nu'\sqrt{b}/\delta)n$. Let $\sigma \in \hat{\Omega} \setminus \Psi$, and as above let k be the total perimeter of components bridged in σ . Since $\sigma \notin \Psi$, we have $k > x_1 n$. Now, by equation 5.4.3, the change in the number of blocked vertices after applying the function f is at most

$$\left(\frac{\alpha a}{\beta} + \frac{\gamma\kappa(R')}{\beta}\right) - \left(\frac{\alpha a}{\beta} + \frac{\delta k}{\beta}\right) < \frac{-\delta k}{2\beta},$$

since

$$\delta k/2 > \delta x_1 n/2 = 4\gamma\nu'\sqrt{bn} \geq 4\gamma\nu'\sqrt{a} \geq \gamma\kappa(R').$$

Therefore the change $\mathcal{O}(\sigma) - \mathcal{O}(f(\sigma))$ in the number of open vertices is at least $k\frac{\delta}{2\beta}$.

Let $\tau \in \Psi$ and define $f_k^{-1}(\tau)$ to be the set of configurations with perimeter k that map to τ . As shown previously, $|f_k^{-1}(\tau)| \leq n^2\Delta^{\frac{\delta}{2\gamma\nu_e}} \Delta^{c_3 k} = n^2(\Delta^{\frac{\delta}{2\gamma\nu_e} + c_3})^k = n^2(\Delta^{c_4})^k$, where $c_4 = \frac{\delta}{2\gamma\nu_e} + c_3$, and so

$$\pi(\tau)^{-1} \sum_{\sigma \in f^{-1}(\tau)} \pi(\sigma) \leq \sum_{\sigma \in f^{-1}(\tau)} \mu^{\mathcal{O}(\sigma) - \mathcal{O}(f(\sigma))} \leq \sum_{k=x_1 n}^{(\nu'/2)n^2} \mu^{\kappa\delta/(2\beta)} |f_k^{-1}(\tau)| \leq \gamma_1^n,$$

for some $\gamma_1 < 1$, if $\mu \leq \mu^* < (\Delta^{c_4})^{-2\beta/\delta}$. Thus the theorem holds if $\lambda \geq \lambda^* = \mu^{*-2} - 1$.

5.4.5 No Clustering at Low Density for Interfering Binary Mixtures

We now examine the low density case and extend the proof of Theorem 5.3.2, stating that typical configurations will not have the clustering property, to interfering binary

mixtures. For small enough λ , the A -tiles will be well-distributed throughout L_A , in the following sense. Any large dense region must have perimeter on the order of n^2 . Although extending Theorem 5.3.2 to interfering binary mixtures is relatively straightforward, we include the entire proof for completeness.

Theorem 5.4.4: For $0 < b < 1/2$, there exist constants $\lambda_* = \lambda_*(b) > 0, \gamma_2 < 1$ and $n_2 = n_2(b)$ such that for all $n > n_2, \lambda \leq \lambda_*$ a random sample from Ω will not have the clustering property with probability at least $(1 - \gamma_2^n)$.

PROOF: Define $t = \frac{1-2c}{1-c}(b-c)$ and $\delta = (\frac{1-b+t}{t})^t$. Let $\Psi' \subset \widehat{\Omega}$ be the set of configurations with a region R that has at least $(b-c)n^2$ tiles, perimeter less than αn^2 and density at least $1-c$, where α satisfies $0 < \alpha < \ln \Delta(\ln(\delta) - b \ln 2)\nu_e/2$. We will show $\widehat{\pi}(\Psi')$ is exponentially small. From the definition of clustering (Definition 5.4.3), it is straightforward to see that Ψ , the set of configurations that have the clustering property is contained in Ψ' . Thus, if $\widehat{\pi}(\Psi')$ is exponentially small then clustering is exponentially unlikely to occur.

For each $\sigma \in \Psi'$, let R be the lexicographically first region which meets the conditions given above. We will flip each face in R (tiles become empty faces and vice versa) to obtain $f(\sigma)$. Since R has density at least $1-c$ and at least $(b-c)n^2$ tiles this means that there are at most $\frac{c}{1-c}(b-c)n^2$ empty faces in R . So by flipping R we are left with a bank of a_σ tiles such that $a_\sigma \geq \frac{1-2c}{1-c}(b-c)n^2 = tn^2$. Next we define $N(\sigma)$ to be the set of all configurations obtained from $f(\sigma)$ by adding a_σ tiles back in any a_σ empty locations; then $|N(\sigma)| = \binom{n^2 - (bn^2 - a_\sigma)}{a_\sigma}$. For each $\tau \in \widehat{\Omega}$, we need to bound the number of configurations σ such that $\tau \in N(\sigma)$. For any configuration τ there are at most 2^{bn^2} configurations β such that $\beta = f(\sigma)$ and $\tau \in N(\sigma)$ for some $\sigma \in \Psi'$. This is due to the fact that there are 2^{bn^2} ways to choose which bn^2 tiles were in their original location and which were removed by f . For each such configuration β we bound the number of regions R that could have been removed in order to recover the original σ . There are at most bn^2 ways to select an A -tile on the border of R and

$\Delta^{\alpha n^2/\nu_e}$ possible perimeters for R , since R has perimeter less than αn^2 . Thus for any configuration τ there are at most $2^{bn^2} (bn^2 \Delta^{\alpha n^2/\nu_e}) \leq (2^b \delta)^{n^2/2}$ configurations σ such that $\tau \in N(\sigma)$.

Finally, we define a weighted bipartite graph $G(\Psi', \widehat{\Omega}, E)$ with an edge of weight $\pi(\sigma)$ between $\sigma \in \Psi'$ and $\tau \in \widehat{\Omega}$ if $\tau \in N(\sigma)$. The total weight of edges is

$$\sum_{\sigma \in \Psi'} \pi(\sigma) |N(\sigma)| \geq \sum_{\sigma \in \Psi'} \pi(\sigma) \binom{n^2 - (bn^2 - a_\sigma)}{a_\sigma} \geq \pi(\Psi') \delta^{n^2}.$$

However, the weight of the edges is at most $\sum_{\tau \in \widehat{\Omega}} \pi(\tau) \mu^{-\nu bn^2} (2^b \delta)^{n^2/2}$. Let $\mu^* = (2^b/\delta)^{1/(2\nu b)}$ and $\lambda^* = (\mu^*)^{-2} - 1$. Thus for all $\mu < \mu^*$,

$$\pi(\Psi') < \mu^{-\nu bn^2} (2^b \delta)^{n^2/2} \delta^{-n^2} < \gamma_2^n,$$

for some $\gamma_2 < 1$, completing the proof. ■

CHAPTER VI

SEGREGATION MODELS ON \mathbb{Z}^2

In this chapter we use techniques developed in the context of colloids (Chapter 5) to prove results about the Schelling segregation model. 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. In this chapter, 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.

6.1 The Schelling Segregation Model

The Schelling Segregation Model was introduced by Thomas Schelling in 1971 to explain how global behavior can arise from small individual preferences [77]. 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 [3, 41, 75, 86]. 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 [16, 38, 56, 91]. Additional rigorous work has considered further variations designed to simplify the neighbors’ interactions for some specific, basic models [42, 56, 75, 92].

6.1.1 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 [50, 72, 58, 82]. 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 [75, 80, 81]. 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 [82, 90].

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 [69] 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 [12, 11, 22]. 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.

6.1.2 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 is 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 6.2 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 6.3 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, in Section 6.4 we give the corresponding proofs for integration at low bias and segregation at high bias, which will build on the proof ideas established in Section 6.3. Finally, we conclude with some open problems.

6.2 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.

6.2.1 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 34). 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 [55]). 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

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 [43]. 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 [81, 75]. Here, a site is “happy” if it has at least as many neighbors of

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.

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.

6.2.2 Clustering.

We give rigorous results demonstrating a dichotomy in mixing times and clustering for two broad classes. Here we formally define clustering. 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 [66] based on the presence of a large region with small perimeter that is densely filled with either R - or B -faces. In Section 6.4, we will show that a random sample from our model will be exponentially likely to be clustered when the bias is high, and exponentially unlikely to be clustered when the bias is sufficiently low.

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 6.2.1: 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.

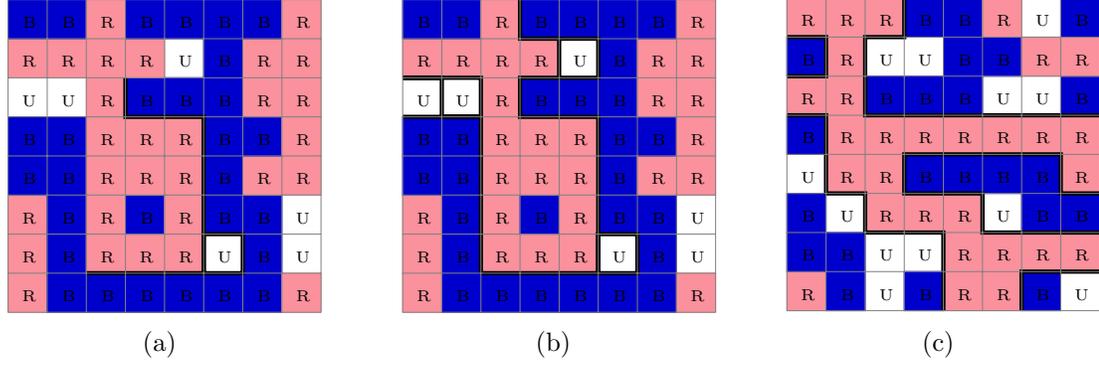


Figure 33: (a) A configuration with a contour, (b) the corresponding fat contour, and (c) an R -cross.

6.3 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} 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 [74] 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 [44] 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.

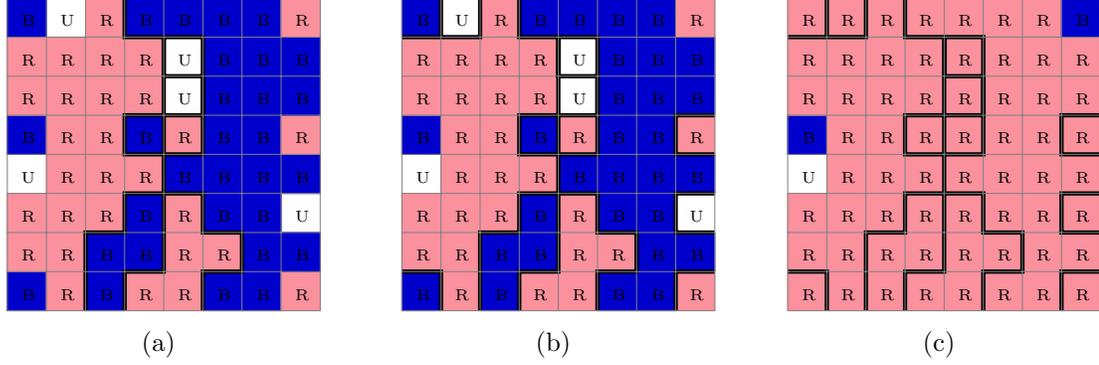


Figure 34: (a) A configuration σ with a fault line, (b) the 1-extended fault, and (c) $\phi(\sigma)$.

6.3.1 Slow mixing at High λ .

We begin by extending the concept of fat faults introduced in [44] to *fat faults* that are essentially large boundaries that can “jump” up to a distance of r . By showing that 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 [49, 78] for details.)

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 [44] and Figure 33) 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 33b 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 33c). 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 [44].

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 34).

- 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.

6.3.1.1 Increasing Bias Functions.

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

Theorem 6.3.1: 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:

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

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

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. [78])

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

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

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

6.3.1.2 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 6.3.2: 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

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

6.3.2 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 6.3.3: 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 every 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,

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

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

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

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

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

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,

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

6.4 Segregation or Integration at Stationarity

We now return to the original motivation behind the Schelling model, namely determining how racial biases can influence segregation in a community. To address this question, we need to formalize how biases contribute to the limiting distributions for the Schelling processes. We consider the Markov chains arising from the Generalized Influence Model and we characterize properties of the stationary distributions. Using insights from Section 3 on mixing times we establish a similar dichotomy indicating integration and segregation at low and high values of λ , respectively. When λ is large, ghettos will form, and configurations will be predominantly one color. However, when λ is small, there will be no clustering of one type and cities will remain integrated.

Our proofs build on combinatorial insights developed in Section 6.3.1 and in [66] to establish clustering (i.e., segregation) for the IBF and TBF models when the bias is high. We characterize clustering by the existence of a region R that has large (quadratic) area, small (linear) perimeter, and whose interior is dense with one of the two colors. A similar notion of clustering was used in [66], but the proofs required the introduction of r -bridges and fat contours to handle unoccupied houses and large radii of influence.

6.4.1 Segregation at High λ for the IBF and TBF Classes.

First, we use the combinatorial techniques developed in Section 6.3.1 to argue that at high λ , configurations will be segregated. In open cities we expect a single ghetto of predominantly R - or B -faces. Specifically, we prove that at high values of λ , a typical configuration will have no large contours and will have high density of either R - or B -faces. We combine techniques used to show clustering [66] with the slow mixing techniques used in Section 6.3.1. Let ρ_R be the density of R -faces and ρ_B be the density of B -faces. We prove the following theorem showing ghettos will form.

Theorem 6.4.1: Assume a valid utility function u with radius r such that $u'_\alpha > 0$

or u is a threshold utility function with $\theta > (\frac{1}{2} + \frac{1}{2r+2})N$, where $N = 2r^2 + 2r$. Given a constant density $d_1 > 1/2$, there exist constants $\gamma_1 = \gamma_1(d_1) < 1$ and $\lambda_1 = \lambda_1(u'_\alpha, r, d_1)$ such that for all $\lambda \geq \lambda_1$ a random sample from Ω will have no contours with more than $d_1 n$ edges and either the density $\rho_R > d_1$ or $\rho_B > d_1$ with probability at least $(1 - \gamma_1^n)$.

PROOF: Using an extension of the techniques from 6.3.1 we show that it is exponentially unlikely for a configuration to have any contour with size greater than $d_1 n$ and that it is exponentially unlikely for $\rho_R, \rho_B < d_1$. The union bound lets us combine these two results.

Let Ω_{d_1} be the set of configuration in Ω which contain a contour longer than $d_1 n$ edges. To show that such configurations are unlikely, we construct a map $\phi_{d_1} : \Omega_{d_1} \rightarrow \Omega$ from configurations with contours of size greater than $d_1 n$ to configurations which have at least one less contour of size greater than $d_1 n$. As in Section 6.3.1, ϕ_{d_1} takes the lexicographically first contour of size greater than $d_1 n$, finds the r -extended contour which contains this contour, changes all U -faces bordering the r -extended contour to R -faces and flips all B -bordered regions adjacent to the contour. Unlike Section 6.3.1 where the contour is a fault line and thus adjacent to the border, our contour is not necessarily anchored to the border.

Next, we bound the number of pre-images of a configuration under ϕ_{d_1} , using a combinatorial argument similar to Section 6.3.1. In Section 6.3.1 the number of configurations with an r -extended contour with m edges which intersect the border is at most $4n3^{m/2}(2r^2 + 2r)^m$. However, with our new function ϕ_{d_1} , the contour might not be connected to the border so the number of configurations with an r -extended contour with m edges is now $2n^23^{m/2}(2r^2 + 2r)^m$, since the number of possible starting points is increased from $4n$ to $2n^2$ (the number of edges in the grid). Additionally, we only guarantee that the r -extended contour has at least $d_1 n$ edges instead of n edges. Let $\Omega_{F,m}$ be defined as in Section 6.3.1 where a configuration $\sigma \in \Omega_{F,m}$ if m

is the number of bad edges fixed by ϕ_{d_1} . The remainder of the proof is the same as in Theorem 6.3.1. If our utility function u satisfies $u'_\alpha > 0$, then we have a gain of at least $\lambda^{u'_\alpha r^2/4}$ per edge of the r -extended contour. Assume

$$\lambda \geq \lambda_1 = (3(2r^2 + 2r))^{4/u'_\alpha r^2},$$

and let $\gamma_1 = 3^{-3d_1/4}$. We then find

$$\begin{aligned} \pi(\Omega_{d_1}) &\leq \sum_{m=d_1 n}^{2n^2} \pi(\Omega_{F,m}) \\ &\leq \sum_m \sum_{x \in \Omega_{F,m}} \pi(\phi_{d_1}(x)) \lambda^{-mu'_\alpha r^2/4} \\ &\leq \sum_m 2n^2 3^{d_1 n/2} (2r^2 + 2r)^{d_1 n} (\lambda)^{-d_1 nu'_\alpha r^2/4} \\ &\leq 4n^4 3^{-d_1 n/2} \leq \gamma_1^n. \end{aligned}$$

Otherwise, if u is a threshold utility function with

$$\theta > \left(\frac{1}{2} + \frac{1}{2r + 12} \right) N,$$

then we have a gain of at least $\lambda^{1/4}$ per edge of the r -extended contour. Assume

$$\lambda \geq \lambda_1 = (3(2r^2 + 2r))^4,$$

and let $\gamma_1 = 3^{-3d_1/4}$. Then, we have that

$$\begin{aligned} \pi(\Omega_{d_1}) &\leq \sum_{m=d_1 n}^{2n^2} \pi(\Omega_{F,m}) \\ &\leq \sum_m \sum_{x \in \Omega_{F,m}} \pi(\phi_{d_1}(x)) \lambda^{-m/4} \\ &\leq \sum_m 2n^2 3^{d_1 n/2} (2r^2 + 2r)^{d_1 n} (\lambda)^{-d_1 n/4} \\ &\leq 4n^4 3^{-d_1 n/2} \leq \gamma_1^n. \end{aligned}$$

To show that it is exponentially unlikely for $\rho_R, \rho_B < d_1$ we construct a map ϕ_S which locates a sufficiently large set of r -extended contours and removes them. Given a set S of r -extended contours, the *size* of the set which we denote as $|S|$ is the sum of the sizes of the distinct r -extended contours contained in S . We show there exists a row P in the grid and a set S of r -extended contours with $|S| \geq (\frac{1-d_1}{2})n$ such that each r -extended contour in S contains at least one vertical edge along P .

Next, we bound the number of pre-images of a configuration under ϕ_S , using an argument similar to Section 6.3.1. There are n possible rows P and for each choice of P there are 2^n different sets of starting points for our depth first search. Given the set of starting points, a depth first search of m edges takes at most $2m$ steps and each move travels in up to $2r^2 + 2r$ directions. Thus there are now $n2^n 3^{m/2} (2r^2 + 2r)^m$ configurations with a row P and set S with $|S| = m$ and each r -extended contour in S intersecting P . Unlike Section 6.3.1, we only guarantee that we are “fixing” at least $\frac{1-d_1}{2}n$ bad edges instead of n edges since $|S| \geq (\frac{1-d_1}{2})n$. If our utility function u satisfies $u'_\alpha > 0$, then we have a gain of at least $\lambda^{u'_\alpha r^2/4}$ per bad edge. In this case, assume

$$\lambda \geq \lambda_2 = (2^{2/(1-d_1)} 3 (2r^2 + 2r))^{4/u'_\alpha r^2},$$

and let $\gamma_1 = 3^{-3(1-d_1)/8}$. Let $\Omega_{F,m}$ be defined as in Section 6.3.1. Combining these results we find,

$$\begin{aligned} \pi(\Omega_S) &\leq \sum_{m=\frac{(1-d_1)n}{2}}^{2n^2} \sum_{x \in \Omega_{F,m}} \pi(\phi_S(x)) \lambda^{-m u'_\alpha r^2/4} \\ &\leq \sum_m n 2^n 3^{\frac{1-d_1}{4}n} (2r^2 + 2r)^{\frac{1-d_1}{2}n} \lambda^{-\frac{(1-d_1)}{2} \frac{n u'_\alpha r^2}{4}} \\ &\leq \gamma_1^n. \end{aligned}$$

Otherwise, if u is a threshold utility function with

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

then we have a gain of at least $\lambda^{1/4}$ per bad edge. In this case, assume

$$\lambda \geq \lambda_2 = (2^{2/(1-d_1)} 3(2r^2 + 2r))^4,$$

and let $\gamma_1 = 3^{-3(1-d_1)/8}$.

$$\begin{aligned} \pi(\Omega_S) &\leq \sum_{m=\frac{(1-d_1)n}{2}}^{2n^2} \sum_{x \in \Omega_{F,m}} \pi(\phi_S(x)) \lambda^{-m/4} \\ &\leq \sum_m n 2^n 3^{\frac{1-d_1}{4}n} (2r^2 + 2r)^{\frac{1-d_1}{2}n} \lambda^{-\frac{(1-d_1)n}{4}} \\ &\leq \gamma_1^n. \end{aligned}$$

It remains to show that there exists a row P and a set S of r -extended contours with $|S| \geq (\frac{1-d_1}{2})n$ such that each r -extended contour in S contains at least one vertical edge along P . First consider the case where the density of B - and R -faces along any row P is low specifically, $\rho_R + \rho_B < \frac{1+d_1}{2}$. This implies that along this row there are at least $(1 - \frac{1+d_1}{2})n = (\frac{1-d_1}{2})n$ U -faces this implies that the maximum set S of r -extended contours which intersect P satisfies $|S| \geq (\frac{1-d_1}{2})n$ (for each U -faces either the edge above or the edge below must be included in S). Next, we can assume the density of B - and R -faces along each row is at least $\frac{1+d_1}{2}$. Let γ_R be the number of R -faces along the left and right boundaries of the grid and similarly let γ_B be the number of B -faces. Since $\gamma_R + \gamma_B \leq 2n$, either $\gamma_R < n$ or $\gamma_B < n$. We assume $\gamma_R < n$. Next, assume there is a row P with at least $(\frac{1-d_1}{2})n$ R -faces. Consider the maximum set S of r -extended contours which intersect P . This set S divides the grid into regions. Now for each R -face t along P , this face is contained within some region which implies that there is an edge of S in the same column as t or the region containing t spans the entire column. If there are no such regions that span entire columns then the size of S is at least as large as the number of R -faces along P implying, $|S| \geq (\frac{1-d_1}{2})n$ as desired. Otherwise we have a region with boundary ψ that is bordered by R -faces and spans an entire column. Since ψ spans an entire

column, each row of the grid contains 2 edges of ψ . Since there are at most n R -faces along the boundary, there are at most n boundary edges contained in ψ implying ψ contains at least n non-boundary edges which implies $|S| \geq n \geq (\frac{1-d_1}{2})n$, as desired.

Finally, if there is no row P with at least $(\frac{1-d_1}{2})n$ R -faces then, since every row has at least $(\frac{1+d_1}{2})n$ B - and R -faces, there must be at least $d_1 n^2 B$ -faces implying $\rho_B \geq d_1$, a contradiction. \blacksquare

6.4.2 Integration at Low λ .

Finally, we provide complementary results showing that at low λ an average sample from the steady state is integrated; there will not be a high density of R or B -faces so we are not likely to have clustering. We prove the following theorem which shows that at low bias, ghettos are unlikely to form.

Theorem 6.4.2: Given a valid utility function u with radius r and constant $c_2 > 10/11$, there exist constants $\gamma_2 = \gamma_2(c_2) < 1$ and $\lambda_2 = \lambda_2(u', r, c_2)$ such that for $\lambda \leq \lambda_2$ a random sample from Ω will be c_2 -clustered or the density ρ_R or $\rho_B > c_2$ with probability at most γ_2^n .

PROOF: First we show that a configuration is exponentially unlikely to be c_2 -clustered. We use a similar technique to show that $\rho_R, \rho_B < c_2$. It is straightforward to combine the two results using a union bound.

Let $\Omega_C \subset \Omega$ be the set of configurations that are c_2 -clustered. We will show that under the conditions stated in the theorem, $\pi(\Omega_C)$ is exponentially small. To show Ω_C is exponentially small, we construct a map $\phi_C : \Omega_C \rightarrow \Omega$, which maps a configuration $\sigma \in \Omega_C$ to the set of all configurations which correspond to removing a c_2 -cluster region C and then selecting $(1 - c_2)n^2$ B -faces or U -faces and changing them to R -faces. Given $\sigma \in \Omega_C$ whose R -faces are c_2 -clustered, define $N(\sigma)$ to be the set of all configurations obtained from σ by removing a c_2 -cluster region C and changing exactly $(1 - c_2)n^2$ B -faces or U -faces to R -faces. If instead the B -faces in σ

are c_2 -clustered the proof is essentially the same and so we omit it. To remove C , we change (or flip) all R -faces to B -faces within C . Once we flip the R -faces and B -faces in C there are at most $(1 - c_2)n^2$ R -faces remaining so $|N(\sigma)| \geq \binom{c_2 n^2}{(1 - c_2)n^2}$. For each configuration $\tau \in \Omega$ we bound the number of configurations σ such that $\tau \in N(\sigma)$. If there exists σ such that $\tau \in N(\sigma)$, then the number of R -faces in τ is at most $2(1 - c_2)n^2$. Since C is a c_2 -cluster region with perimeter at most $c_2 n$, there are at most $2n^2 3^{c_2 n} 3^{2(1 - c_2)n^2}$ possible pre-images of any configuration τ . The factor of 2 is because the configuration could have been R or B -clustered.

Next, given configurations σ, τ such that $\tau \in N(\sigma)$ we derive an upper bound on the ratio $\pi(\sigma)/\pi(\tau)$. Recall the map ϕ_C first removes a c_2 -cluster region C by flipping the R - and B -faces within C . This procedure only changes the ‘‘happiness’’ of faces within distance r of the border. Since there are at most $(2r^2 + 2r + 1)c_2 n$ of these, removing C decreases the weight by at most a factor of $\lambda^{c_2 n(2r^2 + 2r + 1)}$. Changing the color of a single face can decrease the weight of a configuration by at most a factor of $\lambda^{2u'_\beta(2r^2 + 2r)}$. Thus, changing $(1 - c_2)n^2$ B -faces or U -faces to R -faces decreases the weight by at most a factor of $\lambda^{2u'_\beta(1 - c_2)n^2(2r^2 + 2r)}$. Combining these shows that

$$\pi(\sigma)/\pi(\tau) \leq \lambda^\Delta,$$

where

$$\Delta = c_2 n(2r^2 + 2r + 1) + 2u'_\beta(1 - c_2)n^2(2r^2 + 2r).$$

We define a weighted bipartite graph $G(\Omega_D, \Omega, E)$ with an edge weight $\pi(\sigma)$ between $\sigma \in \Omega_D$ and $\tau \in \Omega$ if $\tau \in N(\sigma)$. The total weight of edges W is

$$\begin{aligned} W &= \sum_{\sigma \in \Omega_D} \pi(\sigma) |N(\sigma)| \\ &\geq \sum_{\sigma \in \Omega_D} \pi(\sigma) \binom{c_2 n^2}{(1 - c_2)n^2} \\ &\geq \pi(\Omega_D) \left(\frac{c_2}{(1 - c_2)} \right)^{(1 - c_2)n^2}. \end{aligned}$$

Also, the weight of edges is at most

$$\begin{aligned} W &= \sum_{\tau \in \Omega} \pi(\tau) 2n^2 3^{c_2 n} 3^{2(1-c_2)n^2} \lambda^\Delta \\ &\leq 2n^2 3^{c_2 n} 3^{2(1-c_2)n^2} \lambda^{2(1-\mu)\Delta}. \end{aligned}$$

Combining these equations, assuming

$$\lambda \leq \lambda_2 = \left(\frac{c_2}{10(1-c_2)} \right)^{(4u'_\beta(r^2+r))^{-1}},$$

and letting $\gamma_2 = (10/11)^{1-c_2}$ gives

$$\begin{aligned} \pi(\Omega_D) &\leq 2n^2 3^{c_2 n} 3^{2(1-c_2)n^2} \lambda^{2(1-\mu)\Delta} \left(\frac{1-c_2}{c_2} \right)^{(1-c_2)n^2} \\ &\leq \gamma_2^n. \end{aligned}$$

Next, we show that at low λ we will have $\rho_R, \rho_B < d_2$. Let Ω_D be the set of configuration in Ω for which $\rho_R \geq d_2$ or $\rho_B \geq d_2$. We will show that under the conditions stated in the theorem, $\pi(\Omega_D)$ is exponentially small. Throughout this proof we will assume that $\rho_R \geq d_2$. To show this we will construct a map $\phi_D : \Omega_D \rightarrow \Omega$, which maps a configuration σ to the set of all configurations which correspond to selecting $(1-d_2)n^2$ R -faces and changing them to B -faces. Define $N(\sigma)$ to be the set of all configurations obtained from σ by changing exactly $(1-d_2)n^2$ R -faces to B -faces. Since there are least $d_2 n^2$ R -faces, $|N(\sigma)| \geq \binom{d_2 n^2}{(1-d_2)n^2}$. For each configuration $\tau \in \Omega$ we need to bound the number of configuration σ such that $\tau \in N(\sigma)$. If there exists a σ such that $\tau \in N(\sigma)$ then this implies that the number of B -faces in σ is at most $2(1-d_2)n^2$ and since our map only changes R -faces to B -faces, there are at most $2^{2(1-d_2)n^2+1}$ possible pre-images for σ . The additional factor of 2 is due to the fact that originally either $\rho_R \geq d_2$ or $\rho_B \geq d_2$. We define a weighted bipartite graph $G(\Omega_D, \Omega, E)$ with an edge weight $\pi(\sigma)$ between $\sigma \in \Omega_D$ and $\tau \in \Omega$ if $\tau \in N(\sigma)$. The

total weight W of edges is

$$\begin{aligned}
W &= \sum_{\sigma \in \Omega_D} \pi(\sigma) |N(\sigma)| \\
&\geq \sum_{\sigma \in \Omega_D} \pi(\sigma) \binom{d_2 n^2}{(1-d_2)n^2} \\
&\geq \pi(\Omega_D) \left(\frac{d_2}{1-d_2} \right)^{(1-d_2)n^2}.
\end{aligned}$$

However the weight of the edges is at most

$$\begin{aligned}
W &= \sum_{\tau \in \Omega} \pi(\tau) 2^{2(1-d_2)n^2+1} \\
&\quad (\lambda^{2(1-\mu)(2r^2-1)})^{(1-d_2)n^2} \\
&\leq 2^{2(1-d_2)n^2+1} (\lambda^{2(1-\mu)(2r^2-1)})^{(1-d_2)n^2}.
\end{aligned}$$

Combining these equations, assuming

$$\lambda^{1-\mu} \leq \lambda_2 = \left(\frac{d_2}{5(1-d_2)} \right)^{1/(4r^2-2)},$$

and letting $\gamma_2 = (5/6)^{1-d_2}$ and $d'_2 = (1-d_2)/d_2$ gives the following result

$$\begin{aligned}
\pi(\Omega_D) &\leq \left(d_2'^{(1-d_2)n^2} 2^{2(1-d_2)n^2+1} \right) \\
&\quad \left(\lambda^{2(1-\mu)(2r^2-1)} \right)^{(1-d_2)n^2} \\
&\leq \gamma_2^n.
\end{aligned}$$

■

REFERENCES

- [1] ALDOUS, D., “Random walk on finite groups and rapidly mixing Markov chains,” *In Seminaire de Probabilites XVII*, pp. 243–297, 1983.
- [2] ANGEL, O., HOLROYD, A., KOZMA, G., WÄSTLUND, J., and WINKLER, P., “The phase transition for dyadic tilings,” *Transactions of the AMS* (to appear), 2014.
- [3] BARDE, S., “Back to the future: A simple solution to Schelling segregation,” *CoRR*, 2011.
- [4] BAXTER, R., ENTIG, I., and TSANG, S., “Hard-square lattice gas,” *Journal of Statistical Physics*, vol. 22, pp. 465–489, 1989.
- [5] BAYER, D. and DIACONIS, P., “Trailing the dovetail shuffle to its lair,” *The Annals of Applied Probability*, vol. 2, pp. 295–313, 1992.
- [6] BENJAMINI, I., BERGER, N., HOFFMAN, C., and MOSSEL, E., “Mixing times of the biased card shuffling and the asymmetric exclusion process,” *Trans. Amer. Math. Soc.*, vol. 357, pp. 3013–3029, 2005.
- [7] BHAKTA, P., MIRACLE, S., and RANDALL, D., “Clustering and mixing times for segregation models on \mathbb{Z}^2 ,” in *Proceedings of the 24rd Annual ACM-SIAM Symposium on Discrete Algorithms*, 2014.
- [8] BHAKTA, P., MIRACLE, S., STREIB, A., and RANDALL, D., “Mixing times of Markov chains for self-organizing lists and biased permutations,” in *Proceedings of the 24th ACM/SIAM Symposium on Discrete Algorithms*, SODA ’13, pp. 1–15, 2013.
- [9] BHATNAGAR, N. and RANDALL, D., “Torpid mixing of simulated tempering on the Potts model,” in *Proceedings of the 15th ACM/SIAM Symposium on Discrete Algorithms*, SODA ’04, pp. 478–487, 2004.
- [10] BIRDI, K., *Handbook of Surface and Colloid Chemistry*. CRC Press, 2008.
- [11] 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.
- [12] 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.

- [13] BLANCA, A., GALVIN, D., RANDALL, D., and TETALI, P., “Phase co-existence and slow mixing for the hard-core model on \mathbb{Z}^2 ,” in *Proc. 17th Workshop on Randomization and Computation*, pp. 379–394, 2013.
- [14] BORGS, C., CHAYES, J., FRIEZE, A., KIM, J., TETALI, P., VIGODA, E., and VU, V., “Torpид mixing of some MCMC algorithms in statistical physics,” in *40th IEEE Symp. on Foundations of Computer Science*, FOCS 1999, pp. 218–229, 1999.
- [15] BOROS, E. and FÜREDI, Z., “Rectangular dissections of a square,” *European Journal of Combinatorics*, vol. 9, pp. 271–280, 1988.
- [16] 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.
- [17] BUBLEY, R. and DYER, M., “Faster random generation of linear extensions,” in *Proceedings of the ninth annual ACM-SIAM symposium on Discrete algorithms*, SODA ’98, 1998.
- [18] BUBLEY, R. and DYER, M., “Faster random generation of linear extensions,” *Discrete Math.*, vol. 201, pp. 81–88, 1999.
- [19] BUHOT, A. and KRAUTH, W., “Phase separation in two-dimensional additive mixture,” *Phys. Rev.*, vol. E 59, pp. 2939–2941, 1999.
- [20] CANNON, S., MIRACLE, S., and RANDALL, D., “Phase transitions in random dyadic tilings and rectangular dissections,” *In submission*, 1981.
- [21] CAPUTO, P., MARTINELLI, F., SINCLAIR, A., 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.
- [22] CHAYES, L., “Mean field analysis of low-dimensional systems,” *Communications in Mathematical Physics*, vol. 292, pp. 303–341, 2009.
- [23] DEMICHELI, G., ANTOGNETTI, P., and SANGIOVANNI-VINCENTELLI, A., eds., *Design systems for VLSI circuits: logic synthesis and silicon compilation*. Springer, 1987.
- [24] DIACONIS, P. and SALOFF-COSTE, L., “Comparison techniques for random walks on finite groups,” *The Annals of Applied Probability*, vol. 21, pp. 2131–2156, 1993.
- [25] DIACONIS, P. and SALOFF-COSTE, L., “Comparison theorems for reversible Markov chains,” *Annals of Applied Probability*, vol. 3, pp. 696–730, 1993.
- [26] DIACONIS, P. and SHAHSHAHANI, M., “Generating a random permutation with random transpositions,” *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, vol. 57, pp. 159–179, 1981.

- [27] DOBRUSHIN, R., “The problem of uniqueness of a Gibbs random fields and the problem of phase transition,” *Func. anal. and appl.*, vol. 2, pp. 302–312, 1968.
- [28] DOBRUSHIN, R., KOTECKÝ, R., and SHLOSMA, S., *The Wulff Construction: A Global Shape from Local Interactions*. American Mathematical Society, 1992.
- [29] DOBRUSHIN, R. and SHLOSMA, S., “Constructive criterion for the uniqueness of Gibbs fields,” in *Statistical Physics and Dynamical Systems*, pp. 347–370, Birkhäuser, 1985.
- [30] DRESS, C. and KRAUTH, W., “Cluster algorithm for hard spheres and related systems,” *J. Phys. A: Math. Gen.*, vol. 28, 1995.
- [31] DYER, M., FRIEZE, A., and KANNAN, R., “A random polynomial-time algorithm for approximating the volume of convex bodies,” *J. ACM*, 1991.
- [32] DYER, M. and GREENHILL, C., “A more rapidly mixing Markov chain for graph colorings,” *Random Structures & Algorithms*, vol. 13, pp. 285–317, 1998.
- [33] ELKIES, N., KUPERBERG, G., LARSEN, M., and PROPP, J., “Alternating-sign matrices and domino tilings,” *Journal of Algebraic Combinatorics*, vol. 1, pp. 111–132, 1992.
- [34] FILL, J., “Background on the gap problem,” *Unpublished manuscript*, 2003.
- [35] FILL, J., “An interesting spectral gap problem,” *Unpublished manuscript*, 2003.
- [36] FRENKEL, D. and LOUIS, A., “Phase separation in binary hard-core mixtures: An exact result,” *Physical Review Letters*, vol. 68, pp. 3363–3365, 1992.
- [37] GAUNT, D. and FISHER, M., “Hard-sphere lattice gases,” *The Journal of Chemical Physics*, vol. 45, pp. 2840–2863, 1965.
- [38] 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.
- [39] GOLINELLI, O. and MALLICK, K., “The asymmetric simple exclusion process: an integrable model for non-equilibrium statistical mechanics,” *Journal of Physics A: Mathematical and General*, vol. 39, p. 1267912705, 2006.
- [40] GORE, V. and JERRUM, M., “The Swendsen-Wang algorithm does not always mix rapidly,” *J. Statistical Phys.*, vol. 97, pp. 67–86, 1999.
- [41] GRAUWIN, S., “Effect of local coordination on a Schelling-type segregation model,” *CoRR*, 2008.

- [42] 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.
- [43] GREENBERG, S., PASCOE, A., and RANDALL, D., “Sampling biased lattice configurations using exponential metrics,” in *Proceedings of the twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA ’09, 2009.
- [44] 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.
- [45] HAGGKVIST, R., LINDBERG, P., and LINDSTROM, B., “Dissecting a square into rectangles of equal area,” *Discrete Mathematics*, vol. 47, pp. 321–323, 1983.
- [46] HIEMENZ, P. and RAJAGOPALAN, R., *Principles of Colloid and Surface Chemistry*. CRC Press, 1997.
- [47] ISING, E., “Contribution to the theory of ferromagnetism,” *Z. Phys.*, vol. 31, pp. 253–258, 1925.
- [48] JANSON, S., RANDALL, D., and SPENCER, J., “Random dyadic tilings of the unit square,” in *Random Structures and Algorithms*, vol. 21, pp. 225–251, 2002.
- [49] JERRUM, M. and SINCLAIR, A., “Approximate counting, uniform generation and rapidly mixing Markov chains,” *Information and Computation*, vol. 82, pp. 93–133, 1989.
- [50] JERRUM, M. and SINCLAIR, A., “Polynomial-time approximation algorithms for the Ising model,” *SIAM Journal on Computing*, vol. 22, pp. 1087–1116, 1993.
- [51] JERRUM, M., SINCLAIR, A., and VIGODA, E., “A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries,” *Journal of the ACM*, vol. 51, pp. 671–697, 2004.
- [52] JOCKUSCH, W., PROPP, J., and SHOR, P., “Random domino tilings and the arctic circle theorem,” *arXiv:math/9801068v1*, 1998.
- [53] KENYON, R., “Conformal invariance of domino tilings,” *The Annals of Probability*, vol. 28, pp. 759–795, 2000.
- [54] KNUTH, D., *The Art of Computer Programming*, vol. 3: Sorting and Searching. Addison Wesley, 1973.
- [55] LEVIN, D. A., PERES, Y., and WILMER, E. L., *Markov chains and mixing times*. American Mathematical Society, 2006.
- [56] LEWIS-PYE, E., BARMPALIAS, G., and ELWES, R., “Digital morphogenesis via schelling segregation,” *CoRR*, 2013.

- [57] LOERA, J. D., RAMBAU, J., and SANTOS, F., *Triangulations: Structures for Algorithms and Applications*. Springer, 2010.
- [58] LUBETZY, S. and SLY, A., “Critical Ising on the square lattice mixes in polynomial time,” *Communications in Mathematical Physics*, pp. 815–836, 2012.
- [59] LUBY, M., RANDALL, D., and SINCLAIR, A., “Markov chain algorithms for planar lattice structures,” *SIAM Journal on Computing*, vol. 31, pp. 167–192, 2001.
- [60] MARKOV, A., “Extension of the law of large numbers to dependent quantities [in russian],” *Izv. Fiz-Matem. Obsch. Kazan Univ. (2nd Ser.)*, vol. 15, pp. 135–157, 1906.
- [61] MARTINELLI, E. and OLIVIERI, E., “Approach to equilibrium of Glauber dynamics in the one phase region i: The attractive case,” *Communications in Mathematical Physics*, vol. 161, pp. 447–486, 1994.
- [62] MARTINELLI, F., “Lectures on Glauber dynamics for discrete spin models,” *Lectures Notes in Math*, vol. 1717, pp. 93–191, 1999.
- [63] MCSHINE, L. and TETALI, P., “On the mixing time of the triangulation walk and other catalan structures,” *Randomization Methods in Algorithm Design*, vol. 43, pp. 147–160, 1999.
- [64] METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M., TELLER, A., and TELLER, E., “Equation of state calculations by fast computing machines,” *Journal of Chemical Physics*, vol. 21, pp. 1087–1092, 1953.
- [65] 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 (RANDOM)*, pp. 652–663, 2011.
- [66] MIRACLE, S., RANDALL, D., and STREIB, A. P., “Cluster algorithms for discrete models of colloids with bars,” in *8th Workshop on Analytic Algorithmics and Combinatorics*, ANALCO, pp. 135–150, 2011.
- [67] MURATA, H., FIJIYOSHI, K., WATANABE, T., and KAJITANI, Y., “A mapping from sequence-pair to rectangular dissection,” in *Design Automation Conference*, pp. 625–633, 1997.
- [68] PASCOE, A. and RANDALL, D., “Self-assembly and convergence rates of heterogeneous reversible growth processes,” in *Foundations of Nanoscience*, 2009.
- [69] PRESUTTI, E., *Scaling Limits in Statistical Mechanics and Microstructures in Continuum Mechanics*. Theoretical and mathematical physics, Springer Berlin Heidelberg, 2009.

- [70] PROPP, J. and WILSON, D., “Exact sampling with coupled Markov chains and applications to statistical mechanics,” *Random Structures and Algorithms*, vol. 9, pp. 223–252, 1996.
- [71] RANDALL, D. and TETALI, P., “Analyzing Glauber dynamics by comparison of Markov chains,” *Journal of Mathematical Physics*, vol. 41, pp. 1598–1615, 2000.
- [72] RANDALL, D. and WILSON, D., “Sampling spin configurations of an Ising system,” in *Proceedings of the 10th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 959–960, 1999.
- [73] RANDALL, D., “Mixing,” in *Proceedings of the 44th annual IEEE symposium on Foundations of Computer Science*, 2003.
- [74] 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.
- [75] ROGERS, T. and MCKANE, A., “A unified framework for Schelling’s model of segregation,” *Journal of Statistical Mechanics: Theory and Experiments*, vol. P07006, 2011.
- [76] RUNNELS, L. and COMBS, L., “Exact finite method of lattice statistics. i. square and triangular lattice gases of hard molecules,” *The Journal of Chemical Physics*, vol. 45, pp. 2482–2492, 1966.
- [77] SCHELLING, T., “Dynamic models of segregation,” *Journal of Mathematical Sociology*, vol. 1, pp. 143–186, 1971.
- [78] SINCLAIR, A., *Algorithms for Random Generation & Counting: A Markov Chain Approach*. Birkhäuser, 1993.
- [79] SPITZER, F., “Interaction of Markov processes,” *Advances in Mathematics*, vol. 5, pp. 246–290, 1970.
- [80] STAUFFER, D., “Social applications of two-dimensional Ising models,” *American Journal of Physics*, vol. 76, pp. 470–473, 2008.
- [81] STAUFFER, D. and SOLOMON, S., “Ising, Schelling, and self-organizing segregation,” *The European Physics Journal B*, vol. 57, pp. 473–479, 2007.
- [82] SWENDSEN, R. and WANG, J., “Nonuniversal critical dynamics in Monte Carlo simulations,” *Phys. Rev. Lett.*, vol. 58, pp. 86–88, 1987.
- [83] TURRINI, S., “Optimization in permutation spaces,” *Western Research Laboratory Research Report*, 1996.
- [84] TUTTE, W., “A census of planar maps,” *Canad. J. Math.*, vol. 15, pp. 249–271, 1963.

- [85] VERA, J., VIGODA, E., and YANG, L., “Improved bounds on the phase transition for the hard-core model in 2-dimensions,” in *Proc. 17th Workshop on Randomization and Computation*, pp. 699–713, 2013.
- [86] 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.
- [87] WANG, J. and ZHOU, H., “Linear constraint graph for floorplan optimization with soft blocks,” in *International Conference on Computer-Aided Design (ICCAD’08)*, 2008.
- [88] WELZL, E., “The number of (and random) triangulations of planar point sets,” *Presentation at Oberwolfach Workshop on Combinatorics, Probability and Computing*, 2006.
- [89] WILSON, D., “Mixing times of lozenge tiling and card shuffling Markov chains,” *The Annals of Applied Probability*, vol. 1, pp. 274–325, 2004.
- [90] WOLFF, U., “Collective Monte Carlo updating for spin systems,” *Phys. Rev. Lett.*, vol. 62, pp. 361–364, 1989.
- [91] YOUNG, H., *Individual strategy and social structure: an evolutionary theory of institutions*. Princeton paperbacks, Princeton University Press, 2001.
- [92] ZHANG, J., “A dynamic model of residential segregation,” *Journal of Mathematical Sociology*, vol. 28, pp. 147–170, 2004.
- [93] ZOBRIK, G., *Routing, Placement, and Partitioning*. Intellect Ltd., 1994.