RANDOM SAMPLING OF LATTICE CONFIGURATIONS USING LOCAL MARKOV CHAINS

A Thesis
Presented to
The Academic Faculty

by

Samuel G. Greenberg

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
Department of Mathematics

Georgia Institute of Technology
July, 2009

RANDOM SAMPLING OF LATTICE CONFIGURATIONS USING LOCAL MARKOV CHAINS

Approved by:

Dana Randall
College of Computing, Advisor
Department of Mathematics
Georgia Institute of Technology

Tom Trotter,
School of Mathematics
Department of Mathematics
Georgia Institute of Technology

Christine Heitsch,
School of Mathematics
Department of Mathematics
Georgia Institute of Technology

Eric Vigoda,
College of Computing
Department of Mathematics
Georgia Institute of Technology

Milena Mihail,
College of Computing
Department of Mathematics
Georgia Institute of Technology

Date Approved: 2 July 2008

To Lauren. I am constantly amazed by her dedication to this work and everything else in our life together, and I owe more to her love and strength than I can possibly write here.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor Dana Randall, who has been an amazing friend and mentor during this process. She constantly challenged me to work better, work smarter, and understand the bigger picture. I owe her so much.

The opportunity to work through problems with talented colleagues has been an integral part of the development of this thesis. I have had the pleasure of learning in the company of an incredibly gifted and friendly cadre of fellow students in several departments. Nayantara Bhatnagar and Amanda Pascoe deserve particular mention for their coauthorship of some of the work contained here. Teena Carroll and Mitch Keller were vital for their support, both professional and personal.

The faculty and staff of both the Mathematics and Computer Science departments have provided amazing guidance and support during my time at Georgia Tech. It is a rare thing for a student to have *two* home departments, and I am very grateful. It has been an honor to work with the ACO faculty; in addition to being among the top scholars in their fields, they have been exceptional in their accessibility and support for students. I was lucky enough to do research with Robin Thomas, Craig Tovey, and Prasad Tetali, but I was influenced by so many more.

I am very grateful to Eric Vigoda, who gave so much excellent feedback as the thesis reader. He went far beyond the call of duty, and greatly improved the quality of this thesis.

Finally, I thank my family, for teaching me the joy of learning, and always supporting me despite some difficult times.

SUMMARY

Algorithms based on Markov chains are ubiquitous across scientific disciplines, as they provide a method for extracting statistical information about large, complicated systems. Although these algorithms may be applied to arbitrary graphs, many physical applications are more naturally studied under the restriction to regular lattices. We study several local Markov chains on lattices, exploring how small changes to some parameters can greatly influence efficiency of the algorithms.

We begin by examining a natural Markov Chain that arises in the context of "monotonic surfaces', where some point on a surface is sightly raised or lowered each step, but with a greater rate of raising than lowering. We show that this chain is rapidly mixing (converges quickly to the equilibrium) using a coupling argument; the novelty of our proof is that it requires defining an exponentially increasing distance function on pairs of surfaces, allowing us to derive near optimal results in many settings.

Next, we present new methods for lower bounding the time local chains may take to converge to equilibrium. For many models that we study, there seems to be a phase transition as a parameter is changed, so that the chain is rapidly mixing above a critical point and slow mixing below it. Unfortunately, it is not always possible to make this intuition rigorous. We present the first proofs of slow mixing for three sampling problems motivated by statistical physics and nanotechnology: independent sets on the triangular lattice (the hard-core lattice gas model), weighted even orientations of the two-dimensional Cartesian lattice (the 8-vertex model), and non-saturated Ising (tile-based self-assembly). Previous proofs of slow mixing for other models have been based on contour arguments that allow us prove that a bottleneck in the state space constricts the mixing. The standard contour arguments do not seem to apply to these problems, so we modify this approach by introducing the notion of "fat contours" that can have nontrivial area. We use these to prove

that the local chains defined for these models are slow mixing.

Finally, we study another important issue that arises in the context of phase transitions in physical systems, namely how the boundary of a lattice can affect the efficiency of the Markov chain. We examine a local chain on the perfect and near-perfect matchings of the square-octagon lattice, and show for one boundary condition the chain will mix in polynomial time, while for another it will mix exponentially slowly. Strikingly, the two boundary conditions only differ at four vertices. These are the first rigorous proofs of such a phenomenon on lattice graphs.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Physical systems and phase transitions

Many physical systems in nature exhibit a *phase transition* as a parameter is changed. As the temperature of water is lowered, it changes abruptly from a liquid to a solid. If the temperature is held constant but the pressure decreased, water changes from a liquid to gas. Likewise, if the temperature of iron is decreased, it suddenly gains a magnetic charge, a process called "spontaneous magnetization." For each of these we see a single parameter for which a very slight (microscopic) change, affects the entire system in a dramatic (macroscopic) way.

Statistical physicists study mathematical models of such systems in an effort to discern how subtle changes in the way individual particles behave can affect the behavior of the whole system. For example, the *Ising model* is a good model for magnetism. In this model, we have a group of iron atoms on some graph, say the two dimensional Cartesian lattice, and each atom is given a "positive" or "negative" spin, corresponding to two possible directions for the spin of the electron. I.e., given a graph $G = (V, E)$, we let $\Omega_{Ising}(G)$ be the set of all assignments of spins to the vertices of G, so

$$\Omega_{Ising}(G) := \{+, -\}^V.$$

Neighboring atoms prefer to have identical spins, and the strength of this preference increases as the temperature is lowered. Specifically, the *Gibbs* (or *Boltzmann*) distribution for the Ising model assigns probability to each configuration as follows. Given $\sigma \in \Omega_{Ising}(G)$, we let

$$\pi(\sigma) = e^{\beta \sum_{(x,y)\in E} \sigma(x)\sigma(y)}/Z,$$

where β is inverse temperature, and

$$Z = \sum_{\tau \in \Omega_{Ising}(G)} e^{\beta \sum_{(x,y)\in E} \sigma(x)\sigma(y)}$$

is the normalizing constant known as the *partition function*. At high temperatures (low β), neighboring spins are loosely correlated and a typical configuration contains a fairly random assortment of positive and negative atoms, while at low temperature the influence of neighbors is stronger and there will be large clusters where spins agree. Figure 1 illustrates two such typical configurations.

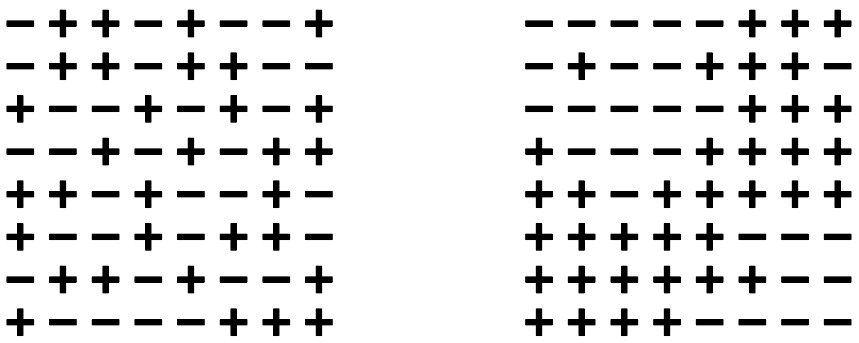

Figure 1: Two typical configurations in the Ising model, left at high temperature and right at low.

We are interested in algorithms that generate configurations from a large set according to their Gibbs probabilities, or some other meaningful probability distribution such as the uniform distribution. This approach is known as "random sampling." An alternative might be to enumerate all of the configurations in the set, but this can be prohibitively slow. For example, in the context of the Ising model on n atoms, we could try to list all 2^n possible configurations and find the likelihood of each, but for such a large set, this would take exponential time. Sampling offers an alternative to enumeration where we can learn about typical configurations while only examining a small proportion of the set of configurations.

A popular approach to random sampling is to use Markov chains, or algorithms that "walk" around the state space. For example, starting with an arbitrary Ising configuration, we change the spin of a randomly chosen individual atom, one by one. This local Markov chain is a popular approach known as *Glauber dynamics*. It is usually simple to engineer the transition probabilities so that If we iterate moves of the chain long enough, we will arrive at a random sample that will be chosen according to our goal distribution, or close to it. Another example of a Markov chain is shuffling cards to arrive at a randomly permuted deck; we start at an arbitrary ordering, say from the previous game, and randomly mix the

cards using some operation until we arrive at a "fair" deck for the next game, or close to one. We are concerned with designing efficient algorithms for which the number of steps needed is not too large.

Random sampling of such configurations arises naturally in statistical physics. Physicists are interested in thermodynamic properties of collections of particles, like the iron atoms represented in the Ising model. Recall that the partition function Z is a sum over allowable configurations of their Gibbs weight. Important thermodynamic properties of the system, such as the specific heat, heat capacity, free energy, and entropy are all functions of the partition function Z (see, e.g., [**?**]). Sampling configurations with probability proportional to their energy gives insight into the value of the partition function, and in many cases allows us to get arbitrarily good estimates of Z.

One of the most exciting discoveries arising from interactions at the interface of physics and theoretical computer science is that many natural sampling algorithms undergo a phase transition, much like the models themselves,. Just as the physical system changes from unmagnetized to magnetized as the temperature drops, these sampling algorithms are believed to undergo an abrupt change whereby they are efficient at high temperature and inefficient at low temperature. Amazingly, it is believed that the temperature at which the Glauber Dynamics for the Ising model change from fast to slow could be precisely the temperature that the physical magnet gains a charge! In this thesis we study questions about Markov chain algorithms that shed some light on the connections between physical properties of a system and algorithms for random sampling from the set of allowable configurations.

1.2 Sampling and counting

Sampling problems are fairly ubiquitous across most areas of sciences. From a combinatorial and computational perspective, one of the primary interests stems from the fundamental connection between sampling and counting.

Two classic counting problems with important connections to sampling are estimating the volume of a convex body and finding the permanent of a matrix. If we discretize space, calculating volume can be conceived as a counting problem: given a convex body, finding the

volume is essentially equivalent to counting the points in the body when the discretization is fine enough [**?**]. Finding the permanent of a $0,1$ matrix is also a counting problem, as it is equivalent to counting the number of perfect matchings of an associated bipartite graph [**?**].

Unfortunately, both of these problems are $\#P$-*Complete*, meaning they are in a complexity class of very difficult counting problems. The class $\#P$ can be thought of as a counting analogue of NP; a problem is in NP if a non-deterministic machine can determine whether a configuration exists in polynomial time, and a problem is in $\#P$ if a non-deterministic machine can determine how many configurations exist in polynomial time. (See [**?**] for more on the complexity of counting.) As computing the volume and the permanent are $\#P$-Complete, algorithms to find exact solutions are believed to require exponential time, and are therefore impractical.

Finding *approximate* solutions to these and similar problems, however, is an interesting and important challenge. Successful algorithms for approximating both the volume and the permanent were discovered in [**?**] and [**?**]. In both cases, the algorithms rely on subroutines for generating random samples. In 1986, Jerrum, Valiant, and Vazirani formalized the connection between sampling and counting for a large group of problems known as *self-reducible*, and showed that an algorithm with polynomial running time to approximate either the counting or sampling problem implies an algorithm with polynomial running time for approximating the other [**?**].

Sampling algorithms based on Markov chains simulate a random walk on the set of configurations for long enough that the sample will be chosen close to the correct distribution. We say a chain is *rapidly mixing* if the walk gets arbitrarily close to the stationary distribution in time polynomial in the size of the configurations. We say a chain is *slowly mixing* when there are starting configurations from which the random walk will take exponential time to converge to the stationary distribution. Our goal, then, is to design Markov chains which are provably rapidly mixing, as they provide efficient sampling algorithms.

We define chains by indicating which pairs of configurations are neighbors. The most straightforward of these is Glauber Dynamics, in which the chain changes the configuration

only on individual points during a step. Many problems well-studied in combinatorics can be sampled efficiently using Glauber Dynamics, at least for some settings of the parameters defining the models. Consider, for example, the set of k-colorings of a graph $G = (V, E)$, i.e., the number of possible mappings $f : V \to [k]$ where $(x, y) \in E$ implies $f(x) \neq f(y)$. If we want to sample k-colorings of graphs, we can start at any valid k-coloring. The Glauber Dynamics then proceed by choosing a vertex uniformly at random. It then changes the coloring by choosing randomly from the set of allowable colors that do not appear among the vertex's neighbors. This chain was shown to be rapidly mixing k is large enough [**?**] or when k is large and the graph has sufficiently large girth [**?**]. Another important example is sampling the independent sets of a graph. Given a graph $G = (V, E)$, an independent set is a subset $I \subseteq V$ such that $x, y \in I$ implies that $(x, y) \notin E$. The Glauber Dynamics for sampling independent sets of a graph add or remove a single vertex from the independent set, so long as it does not violate the independence property. This chain is rapidly mixing when the maximum degree is small or when the chain is weighted to favor sparse sets [**?**]. Likewise, there are also similar results for local algorithms sampling matchings on general graphs [**?**] and many other combinatorial structures.

All of the above results hold for arbitrary graphs, but we are also interested in the special case of sampling these types of combinatorial structures in lattice graphs since many applications arise in that context. The most common lattice instances are finite regions of 2-dimensional Cartesian lattice $\mathbb{Z}^2$, but other regular graphs are also of interest. In many cases there are simple algorithms that can be defined when the input graph is restricted to lie on a lattice that might not even make sense in the context of general graphs. Moreover, even for algorithms that can be defined in general, it is possible that we can prove much better bounds on the mixing time when we consider the restricted problem. For example, we cannot expect to be able to sample 3-colorings uniformly on arbitrary graphs since even producing one sample is NP-complete. However, the Glauber dynamics are known to be rapidly mixing for any finite, simply-connected region of $\mathbb{Z}^2$ [**?**]. Notice that it is easy to generate 3-colorings here because the lattice is bipartite and therefore 2-colorable. Finally, sampling on lattices for various models motivated by physics using Glauber dynamics can

give insight into the physical system itself. Identifying phase transitions is a good example of this.

An important related concept that arises in the context of sampling configurations of a physical system is how the convergence rates are affected by the environment in which they are placed. The physical phase transition often disappears when the boundary states are fixed in certain ways, although much is not known about whether this also holds for the corresponding sampling question. Martinelli, Sinclair, and Weitz [**?**] showed that for the Ising model on a complete finite tree, changing the boundary of the tree does in fact greatly affect the convergence rate dramatically. They showed that when all points on the boundary are fixed to have the same spin, then the chain converges in time that is a polynomial in the size of the tree. However, when the boundary is free, and none of the spins are fixed, the chain converges in time exponential in the size of the tree. It is conjectured that the same dichotomy holds for various lattice models motivated by statistical physics. We make progress towards verifying this conjecture by showing that measuring this effect is possible for a certain local chain on a lattice model.

1.3 Nanotechnology and self-assembly

Another field where the algorithmic underpinnings of sampling turn out to arise quite naturally is nanotechnology, and in particular self-assembly. Self-assembly is a process in which large numbers of simple objects aggregate into larger structures in random but predictable ways. One exciting approach that has received much attention is tile-based self-assembly, where very small tiles are designed with markings on each side so that two tiles are more likely to join together along an edge if their markings match. Wang studied such tiling systems and showed that they form a universal model of computation [**?**]. As the field of nanotechnology make advancements in technology, these have become an even more enticing object of study.

The challenge, then, is two fold. The first is to define tiles and markings so that tiles are likely to assemble predictably into large aggregates. The second is to control the efficiency of such an assembly, so that tiles aggregate quickly and with few errors. The approach

taken by Seeman [?, ?], Winfree [?], and others is to use DNA double-crossover molecules in order to construct marked tiles with DNA sequences on each side. These are constructed so that pairs of tiles we want next to each other have large numbers of matching base pairs along their edges, with the likelihood of their attaching being given by their hybridization energies. See, e.g., [?, ?, ?, ?, ?] for more details and definitions of hybrid energies.

Two problems we examine in this thesis are motivated by tile-based self-assembly models. In the first, we consider many tiles of the same type associating and disassociating with a large substrate, but at unequal rates. We show conditions under which the chain is rapidly mixing in any number of dimensions. Interestingly, the models that arise are equivalent to biased versions of sampling problems arising from statistical physics and combinatorics, namely lozenge tilings (modeling dimers on the triangular lattice), 3-colorings on the Cartesian lattice (modeling the anti-ferromagnetic Potts model) and biased card shuffling.

The second self-assembly model we consider can be thought of as a generalization of the Ising model. We can think of the classic Ising model as a tiling with $+$ and $-$ tiles, where tiles prefer to be adjacent to tiles of the same type. In the self-assembly context, however,it makes sense to think of a non-saturated model where these tiles associate and disassociate allowing for empty spaces where there are no tiles at all. This is equivalent to an Ising model where spins can be non-existent and we call this the "non-saturated Ising model."

1.4 Outline and summary of results

In Chapter 2, we provide an introduction to the area of sampling and finite Markov chains. We give more formal definitions for the ideas mentioned earlier, and introduce two technical methods to upper and lower bound the mixing rate.

In Chapter 3, we look at a biased tiling model motivated by self-assembly. We show conditions under which the chain is rapidly mixing in any number of dimensions. Our main innovation is to introduce a geometric distance function, which allows distances to be exponentially large. This lets us construct a coupling proof of rapid mixing. This is joint work with Dana Randall and Amanda Pascoe [?].

In Chapter 4, we examine three models where it is intuitively clear that local Markov chains will be inefficient because of a conjectured phase change in the system, but where no known tools seem sufficient for a rigorous proof. These models are the non-saturated Ising model arising in self-assembly, and independent sets and edge-orientations of the lattice, motivated by statistical physics. Establishing slow mixing for these models has been a challenge, as standard contour arguments typically used seem not to apply. We modify this approach by introducing the notion of "fat contours" that can have nontrivial area, and use these to establish slow mixing of local chains dened for these models. This is joint work with Dana Randall [**?**, **?**, **?**].

Finally, in Chapter 5, we further examine the role of boundary conditions on mixing rates. We define a local chain on perfect and near perfect matchings of the square-octagon lattice and show that with one boundary condition it will mix in polynomial time, while for another it will mix exponentially slowly. Amazingly, the two boundary conditions only differ at only four vertices. These are the first rigorous proofs of this phenomenon on lattice graphs. This is joint work with Nayantara Bhatnagar and Dana Randall [**?**].

CHAPTER II

BACKGROUND

In this chapter, we give a brief introduction to Markov chains, including definitions and some of the fundamental theorems. We then give an overview of two of the techniques used to upper and lower bound mixing rates in subsequent chapters.

2.1 General Markov Chain Mechanics

We will be concerned with sampling elements of a large set known as the *state space* Ω. To do so, we use a *Markov chain* on the elements of Ω.

Definition 2.1 *A* Markov chain *is a series of random variables* σ_0, σ_1, σ_2, ... *such that the value of* σ_i *depends on* σ_{i-1} *and no earlier variable. I.e. for all* $s_0, s_1, \cdots \in \Omega$,

$$\mathbb{P}[\sigma_i = s_i | \sigma_0 = s_0, \sigma_1 = s_1, \sigma_2 = s_2, \ldots \sigma_{i-1} = s_{i-1}] = \mathbb{P}[\sigma_i = s_i | \sigma_{i-1} = s_{i-1}].$$

The elements $\sigma_i \in \Omega$ are known as the *states* or *configurations*. It will be convenient to define the Markov chain in terms of the current state σ_i and the allowable next states σ_{i+1}, along with the likelihood of each, given that we are at σ_i. To formalize this, we define a directed graph on Ω, representing which states can immediately follow others. Our Markov chain $\mathcal{M}$ simulates a walk on that graph: σ_0 is an arbitrary configuration, and σ_1 is chosen randomly from the neighbors of σ_0. Then σ_2 is chosen randomly from the neighbors of σ_1, and so on. These choices can be uniformly distributed, but more often they are weighted. We also define a *transition matrix* P, so that for a pair of configurations σ, ρ, $P[\sigma, \rho]$ is the *transition probability*, or likelihood of $\mathcal{M}$ moving to ρ in one step if we are currently at σ. Raising this matrix to the t power, $P^t[\sigma, \rho]$ is the likelihood of $\mathcal{M}$ moving from σ to ρ in t steps.

As more and more of these steps are taken, the chance of being at any given configuration approaches the stationary distribution π, provided the Markov chain and state space satisfy a few properties:

Definition 2.2 *A Markov chain is* irreducible *if, for all $\sigma, \rho \in \Omega$ and $t > 0$, there exists a $t' > t$ such that*

$$P^{t'}[\sigma, \rho] > 0.$$

Definition 2.3 *A Markov chain is* aperiodic *if, for all $\sigma, \rho \in \Omega$,*

$$g.c.d.\{t : P^t(\sigma, \rho) > 0\} = 1.$$

We call a chain *ergodic* if it is both irreducible and aperiodic. As we can see in the following lemma, being ergodic and having a finite state space implies convergence to a unique stationary distribution π.

Lemma 2.1 *Any finite, ergodic chain on Ω with transition probability P converges to a unique stationary distribution π. I.e., for all $\sigma, \rho \in \Omega$,*

$$\lim_{t \to \infty} P^t(\sigma, \rho) = \pi(\rho).$$

If a chain is not irreducible, there could be multiple stationary distributions. If a chain is not aperiodic, there might not be a stationary distribution at all. In this sense, being ergodic is both a necessary and sufficient condition for a finite chain to have a unique stationary distribution.

For a Markov chain and probability distribution π, the *Detailed Balance* condition determines whether π is the unique stationary distribution for a given Markov chain.

Definition 2.4 *Given a Markov chain $\mathcal{M}$ on state space Ω with transition probabilities $P(\cdot, \cdot)$, we say that a probability distribution π satisfies* detailed balance *if*

$$\pi(\sigma)P(\sigma, \rho) = \pi(\rho)P(\rho, \sigma)$$

for all $\sigma, \rho \in \Omega$.

The following lemma allows us to infer what distribution a chain converges to.

Lemma 2.2 *Given a finite, ergodic chain $\mathcal{M}$ and a probability distribution π that satisfies detailed balance with respect to π, then π is the unique stationary distribution for $\mathcal{M}$.*

As example, we turn to the Ising model. The local Markov chain $\mathcal{M}_{Ising}$ takes a given configuration $\sigma_t \in \Omega_{Ising}$ and produces a new configuration σ_{t+1} by choosing a vertex of G at uniformly random and changing the spin of that vertex. It is clear that this chain is ergodic. For now, we look at the *infinite temperature* case, where vertices choose their spin independently from their neighbors. In this case, $\mathcal{M}_{Ising}$ changes the spin of the vertex uniformly among $+$ and $-$. As the probability of changing the spin of a vertex is the same as the probability of changing it back, $\mathbb{P}[\sigma, \rho] = \mathbb{P}[\rho, \sigma]$ for all $\sigma, \rho \in \Omega$. Therefore the uniform distribution satisfies detailed balance, and by Lemma 2.2, $\mathcal{M}_{Ising}$ converges to the uniform distribution at infinite temperature.

In general, we will not have such simple Markov chains that converge to the uniform distribution. For the Ising model at finite temperature, the stationary probability of a given configuration depends on the associated energy, and designing a chain that converges to such a distribution is non-trivial. However, in 1953, Metropolis et al. [**?**] showed how to define a Markov chain that does converge to an arbitrary probability distribution. This was generalized by Hastings in 1970 [**?**], and so shares their names.

Definition 2.5 *The* Metropolis-Hastings Algorithm *on a connected state space Ω with probability distribution π repeats the following steps:*

- `For current configuration` σ_t`, choose a neighbor` σ'_t `of` σ_t `uniformly at random with probability` $1/2\Delta$`, where` Δ `is the maximum number of neighbors of any configuration.`
- `With probability` $\min\left(1, \frac{\pi(\sigma'_t)}{\pi(\sigma_t)}\right)$`, let` $\sigma_{t+1} = \sigma'_t$.
- `Otherwise let` $\sigma_{t+1} = \sigma_t$*. then the*

It is straightforward to see from Lemma 2.2 that the Metropolis-Hastings Algorithm converges to the distribution π. Throughout this thesis we will use variations on the Metropolis-Hastings Algorithm to ensure our chains converge to the desired distributions.

We can now examine this notion of convergence to the stationary distribution more closely. However, to do so rigorously, we must first define a distance metric on distributions.

Definition 2.6 *The* total variation distance *between two distributions* π_1 *and* π_2 *is*

$$||\pi_1, \pi_2||_{TV} := \frac{1}{2} \sum_{\rho \in \Omega} |\pi_1(\rho) - \pi_2(\rho)|.$$

With this metric, we can now define the *mixing time* as the time after which the Markov chain gets within ε of π, for some given $\varepsilon > 0$, and all starting configurations σ.

Definition 2.7 *For a chain with transition probability* $\mathbb{P}$ *and stationary distribution* π*, we define the* mixing time $\tau(\varepsilon)$ *for* $\varepsilon > 0$ *as*

$$\tau(\varepsilon) := \max_{\sigma \in \Omega} \min\{t : ||P^{t'}(\sigma, \cdot), \pi(\cdot)||_{TV} \leq \varepsilon, \forall t' \geq t\}.$$

This allows us to define our broad definitions of fast and slow mixing.

Definition 2.8 *We call a Markov chain* rapidly mixing *if the mixing time is bounded above by a polynomial in* n *and* $\log \varepsilon^{-1}$*, where* n *is the size of each configuration in* Ω*.*

Definition 2.9 *We call a Markov chain* slowly mixing *if the mixing time is bounded below by an exponential in* n*, where* n *is the size of each configuration in* Ω*.*

Before discussing specific chains, it is important to mention that the mixing time of a Markov chain can also be described using spectral analysis of the transition matrix.

Definition 2.10 *Define the* spectral gap *of a matrix* M *to be*

$$\text{Gap}(M) := \lambda_0 - |\lambda_1|,$$

where λ_0 *and* λ_1 *are the largest (in magnitude) eigenvalues of* M*.*

The following theorem is well-known in probability theory, and relates the spectral gap and the mixing time [**?**].

Theorem 2.1 *For a Markov chain* $\mathcal{M}$ *with transition probability* $\mathbb{P}$ *and stationary distribution* π*, the following holds for all* $\varepsilon > 0$*.*

$$\tau_\sigma(\varepsilon) \leq -\frac{\ln(\pi(\sigma)\varepsilon)}{\text{Gap}(\mathbb{P})}$$

and

$$\tau(\varepsilon) \geq -\frac{|\lambda_1| \ln(2\varepsilon)}{2\,\mathrm{Gap}(\mathbb{P})}.$$

Theorem 2.1 implies that determining the spectral gap of the transition matrix characterizes the mixing time. Unfortunately, the number of rows and columns in the transition matrix is the size of the state space, and so the matrix is too large to write down, before even trying to calculate the eigenvalues directly. Instead, clever indirect methods have been introduced to bound the spectral gap. We give an overview of two of these methods, coupling and conductance, in Section 2.3. First, it will be useful to formalize the Ising model so that we can demonstrate these methods.

2.2 Spin Configurations

The majority of the models we study in this thesis were originally defined to represent simple physical systems, with an *energy function* on the space of configurations. This energy function is defined by a *Hamiltonian* $H(\sigma)$. For models where the energy is determined solely from nearest-neighbor interactions,

$$H(\sigma) = \sum_{(u,v)\in E} g(\sigma(u), \sigma(v)),$$

for some function g. Similar to a spring relaxing, systems tend to favor configurations that minimize energy, where this preference is controlled by temperature. Each configuration in Ω is given a weight

$$w(\sigma) = e^{-\beta H(\sigma)},$$

where $\beta = 1/T$ is inverse temperature. Thus, for low values of β the differences between the energy of configurations are dampened, while at large β these differences are magnified.

The likelihood of each configuration is then given by

$$\pi(\sigma) = w(\sigma)/Z,$$

where $Z = \sum_\tau w(\tau)$ is the normalizing constant known as the *partition function.* This probability measure is known as the *Gibbs (or Boltzmann) distribution.* Taking derivatives of the generating function Z (or $\ln Z$) with respect to the appropriate variables allows us

to calculate many of the interesting thermodynamic properties of the system, such as the specific heat and the free energy.

As example of such an energy function, we turn to the Ising model as mentioned in Chapter 1. For $\sigma \in \{+,-\}^n$, the Hamiltonian of the ferromagnetic Ising model is given by

$$H(\sigma) = -\sum_{(u,v)\in E} J\sigma(u)\sigma(v),$$

where J is a parameter corresponding to the energy of nearest-neighbor interactions. We define $\beta = 1/T$ as the inverse temperature. Then the stationary probability of σ is $\pi(\sigma) = \frac{e^{\beta H(\sigma)}}{Z}$, where Z is the partition function. For an informal introduction to the Ising model, see [**?**]. In this work, we use a change a variables, so that $\lambda = e^{\beta}$. Then

$$\pi(\sigma) = \lambda^{H(\sigma)}/Z.$$

For a graph G, we define the Markov chain $\mathcal{M}_{Ising}$ on $\Omega_{Ising}(G)$ by repeating the following steps:

- `Choose uniformly at random vertex` $v \in V(G)$, $s \in \{+,-\}$, `and` $p \in (0,1)$.
- `For current configuration` σ_t, `let` σ'_t `be the same configuration, but with` $\sigma'_t(v) = s$.
- `Among the neighbors of` v, `if more of` v`'s neighbors have spin` s `than spin` $-s$, `let` $\sigma_{t+1} = \sigma'_t$.
- `Otherwise let` Γ_s `be the set of neighbors with spin` s `and` Γ_{-s} `be the set of neighbors with spin` $-s$. `If` $p < \lambda^{|\Gamma_{-s}|-|\Gamma_s|}$, `then let` $\sigma_{t+1} = \sigma'_t$. `Otherwise let` $\sigma_{t+1} = \sigma_t$.

Just as with the Metropolis-Hastings Algorithm, there is a large chance of leaving σ_t alone; if $s = \sigma_t(v)$, then $\sigma_{t+1} = \sigma_t$. Also just as with the Metropolis-Hastings Algorithm, if a proposed move raises the stationary probability of the configuration, it is accepted with probability 1. If the move lowers the probability of the configuration, it is accepted with the ratio of the two probabilities.

In general, these local Markov chains that change only a small portion of the configuration at each step, are called *Glauber Dynamics*. Most of this thesis looks at algorithms of this type. We now define some of the methods we use to bound the efficiency of these algorithms.

2.3 Tools for bounding mixing time

Coupling and *conductance* are two methods of bounding the mixing rates of Markov chains. In this thesis, we use coupling in several proofs of rapid mixing, and use conductance in several proofs of slow mixing, as well as a proof of fast mixing which we elaborate on in Chapter 5.

2.3.1 Coupling

One of the most popular methods for bounding mixing times is coupling [**?**]. A *coupling* is itself a Markov chain that operates on pairs of states simultaneously so that the marginals respect the original Markov chain, with the condition that once the two states agree, they stay in agreement. The coupling time, or the expected time until agreement, can provide a good bound on the mixing time if the coupling is carefully chosen. We now formalize this notion.

Definition 2.11 *For initial states* σ, ρ, *let* $T^{\sigma,\rho} = \min\{t : \sigma_t = \rho_t \mid \sigma_0 = \sigma,\ \rho_0 = \rho\}$, *let the* coupling time *be*

$$T = \max_{\sigma,\rho} \mathbb{E}[T^{\sigma,\rho}].$$

The following theorem relates mixing and coupling times [**?**].

Theorem 2.2 *For any coupling,* $\tau(\varepsilon) \leq \lceil T \mathrm{e} \ln \varepsilon^{-1} \rceil$.

As an example of a coupling, we turn again to the Ising model. With the local chain $\mathcal{M}_{Ising}$ as defined above, one straightforward coupling is to choose the same (v, s, p) for both σ_t and ρ_t, to generate σ_{t+1} and ρ_{t+1}. Bounding the coupling time for two arbitrary configurations, however, can be quite difficult.

In general, while coupling is potentially a powerful technique, it is often difficult to measure the expected change in distance between two arbitrary configurations. The method of *path coupling*, introduced by Bubley and Dyer [**?**], greatly simplifies this approach by showing that we really need only consider pairs of configurations that are close according to some metric on the state space.

The idea behind path coupling is to consider a small set $U \subseteq \Omega \times \Omega$ of pairs of configurations that are "close" according to some distance metric φ. Suppose that we have shown that the expected change in distance is decreasing for all of the pairs in U. To now reason about arbitrary configurations $\sigma, \rho \in \Omega$, we define a shortest path $\sigma^0, \sigma^1, ..., \sigma^r$ of length r from $\sigma = \sigma^0$ to $\rho = \sigma^r$, where $(\sigma^i, \sigma^{i+1}) \in U$ for all $0 \leq i < r$. If we define U correctly, then $\varphi(\sigma, \rho) = \sum_{i=0}^{r-1} \varphi(\sigma^i, \sigma^{i+1})$. Then, by linearity of expectation, the expected change in distance between σ and ρ is the sum of the expected change between the pairs (σ^i, σ^{i+1}). Each of these has been shown to be at most zero, so the total distance is at most zero. Of course, after the update there might be a shorter path between the new configurations σ' and ρ', but this just causes the new distance to be even smaller.

Theorem 2.3 *Let ϕ be an integer-valued metric defined on $\Omega \times \Omega$ which takes values in $\{0, \ldots, B\}$. 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 $X_t = Z_0, Z_1, \ldots, Z_r = Y_t$ 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 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 there exists $\gamma \leq 1$ such that for all $(X_t, Y_t) \in U$,

$$\mathbb{E}[\phi_{t+1} - \gamma\phi_t] \leq 0.$$

(1.) If $\gamma < 1$, then the mixing time satisfies

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

(2.) If there exists $\kappa > 0$ such that $Pr[\phi_{t+1} \neq \phi_t] > \kappa$ for all t provided that $X_t \neq Y_t$, then

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

By way of example, we now use Theorem 2.3 to prove that $\mathcal{M}_{Ising}$ is rapidly mixing for the coupling just described when λ close to 1. This proof is far weaker than the current state of the art, but it provides an example of how proofs using Theorem 2.3 are constructed.

For the metric ϕ, we use the Hamming distance, or simply the number of vertices on which the two configurations differ. Let U be the pairs of configurations that have Hamming distance 1. Thus for any $(\sigma_t, \rho_t) \in U$, there is a single vertex v^* on which $\sigma_t(v^*)$ and $\rho_t(v^*)$ differ. Without loss of generality, we assume $\sigma_t(v^*) = +$ and $\rho_t(v^*) = -$.

In our coupling, where the same choice of (v, s, p) is made for both σ and ρ, any choice of v distance at least 2 from v^* will make no change in the distance $\phi(\sigma_t, \rho_t)$; the same change is made to both configurations, so the distance is not effected.

When $v = v^*$, $\mathcal{M}_{Ising}$ may decrease the distance between σ_t and ρ_t; one of the choices of s succeeds with probability 1 and the other succeeds with probability at least λ^{-4}. (For instance, if every neighbor of v^* has spin $+$, then $v = v^*$, $s = +$ succeeds with probability 1 and $v = v^*$, $s = -$ succeeds with probability λ^{-4}. Hence, given $\mathcal{M}_{Ising}$ chose $v = v^*$, the expected change in distance is at most $-1 - \lambda^{-4}$.

When v is chosen in the neighborhood of v^*, there is some chance of the distance increasing. If a move's chance of success is different on the two configurations, then there is a chance that it succeeds on one and not the other, and the distance between them increases. For this simple example, we use the fact that between σ_t and ρ_t, $\Gamma(v)$ differs on only one point. The maximum difference in success rates for each move can then differ by some $\lambda^i(1 - \lambda^2)$, depending on the neighbors of v. This is maximized when $i = 0$, so the maximum difference in success rates for each move is $1 - \lambda^{-2}$. As there are at most eight choices of v and s with this property, the chance of increasing the distance by 1 is at most $8(1 - \lambda^{-2})$. Therefore for λ close to 1,

$$\mathbb{E}[\phi(\sigma_{t+1}, \rho_{t+1}|\sigma_t, \rho_t] - \phi(\sigma_t, \rho_t) \leq 8(1 - \lambda^{-2}) - 1 - \lambda^{-4} < 0.$$

The maximum distance between any two configurations is simply the number of vertices, so B is polynomial. Therefore we may use part (2.) of Theorem 2.3, and the chain is rapidly mixing.

2.3.2 Conductance

A Markov chain must have good flow in order for it to be rapidly mixing, since a bottleneck can take a very long time to cross. This notion was formalized by Jerrum and Sinclair [**?**] and Lawler and Sokal [**?**] as the following.

Definition 2.12 *For a set $S \in \Omega$, define the* conductance of S *as*

$$\Phi_S := \frac{\sum_{\sigma \in S, \rho \notin S} \pi(\sigma) P(\sigma, \rho)}{\pi(S)}.$$

The product $\pi(\sigma)P(\sigma, \rho)$ is the "capacity" of the edge (σ, ρ), and so Φ_S is the total capacity of the edges leaving S, divided by the stationary probability of S. Note that, for chains satisfying detailed balance, $\pi(\sigma)P(\sigma, \rho) = \pi(\rho)P(\rho, \sigma)$, so the capacity into the set is the same as the capacity out.

Definition 2.13 *The conductance for the chain $\mathcal{M}$, Φ, is then defined as*

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

If a Markov chain has exponentially low conductance, then there exists a cut in the state space forming a bottleneck. This makes it intuitively clear that the chain will be slow mixing, as any chain which starts on one side of the bottleneck will take exponential time just to cross to the other side. Perhaps more surprisingly, it turns out that the converse is true as well, that every chain that is slow mixing has exponentially low conductance. The following formalizes this [**?**].

Theorem 2.4 *For any Markov chain with conductance Φ and eigenvalue gap $Gap(P)$, we have*

$$\frac{\Phi^2}{2} \leq Gap(P) \leq 2\Phi.$$

Recall that the spectral gap is inversely related to the mixing time, so this shows that a Markov chain is rapidly mixing if the conductance is at least $1/p(n)$ for some polynomial p (where n is the size of the state space), and slowly mixing if the conductance is less than e^{-cn} for some constant c. We will primarily use the second inequality in this theorem

to show slow mixing by demonstrating that the conductance is exponentially small. The following corollary will be a useful reformation of this part of the theorem.

Corollary 2.5 *For state space Ω and Markov chain $\mathcal{M}$ with transition probability P and stationary probability π, if Ω is the disjoint union of S_1, S_2 and S_3 such that*

1) $P(s_1, s_3) = 0$ for all $s_1 \in S_1$ and $s_3 \in S_3$,

2) $\pi(S_2) \leq c^n \pi(S_1)$ for some constant $c < 1$,

and

3) $\pi(S_2) \leq c^n \pi(S_3)$ for some constant $c < 1$.

Then $\mathcal{M}$ is slow mixing on Ω.

To prove Corollary 2.5, one need only notice that $\Phi < \Phi_{S_2} < c^n$. By Theorem 2.4, this implies exponentially small spectral gap, and therefore exponentially large mixing time.

CHAPTER III

SAMPLING WITH BIAS USING EXPONENTIAL METRICS

In this chapter, we look at how to use Markov chains to sample monotonic lattice surfaces where the surfaces are biased towards those with greater height. The unbiased version has been studied previously, and this variant has been introduced in several applications. We present new arguments for fast mixing based on coupling that rely on an exponentially weighted distance metric.

3.1 Introduction

Many combinatorial and physical models on lattices have a height function representation, including several 2-dimensional tilings and 3-colorings on Z^d (see, e.g., [**?**]). These height functions map a d-dimensional configuration to a surface in $d+1$ dimensions that has well-defined local properties. For example, lozenge tilings on the triangular lattice correspond to a monotonic surface on the 3-dimensional Cartesian lattice that looks like the upper envelope of a set of cubes, where each cube is supported by other cubes on its lower three sides.

It is easy to generalize this family of monotonic surfaces to d-dimensions. In 2 dimensions, the monotonic surface is a "staircase walk" that always moves to the right or down, and we will think of surface as the upper envelope of a set of unit squares in which each square is supported to its left and bottom. Likewise, in d-dimensions we consider a set of d-dimensional unit cubes within an input region such that each unit cube is supported by other cubes in the lower d directions, if these are all within the region. In other words, starting with a cube that is in our set, as we decrease in any of the d directions we must hit another cube in the set or we must hit the boundary of the region.

The Glauber dynamics on monotonic surfaces adds and removes individual cubes (squares) when possible. Informally, we choose a point on the surface uniformly. If that point lies on a cube that can be removed, then we remove it. If that point lies on a cube that can

be added, then we add it. Otherwise we do nothing. In a 2-dimensional rectangular region it is easy to see that if we can add a unit square along the staircase walk, then we cannot remove a square at the same place; this is also true in higher dimensions. We define the Markov chain much more carefully in Section 3.2.

In two dimensions, it is easy to see that Glauber dynamics are rapidly mixing using a standard coupling argument. With a three dimensional surface, the problem is considerably more interesting. Luby, Randall and Sinclair [**?**] show that a related chain is rapidly mixing and subsequently Randall and Tetali [**?**] showed that this implies that Glauber dynamics is also. However, the behavior of the chain is unknown for $d > 3$.

Recently there has been a lot of interest in *biased* version of this chain where we are more likely to add unit cubes than remove them. If $P(\sigma, \tau)$ is the probability of moving from σ to τ in one move, where τ is formed by adding one cube to σ, then $\lambda = P(\sigma, \tau)/P(\tau, \sigma)$ is the bias of the chain. Using detailed balance, it is easy to see that the stationary probability of a configuration σ will be proportional to $\lambda^{|\sigma|}$, where $|\sigma|$ is the number of unit cubes defining the surface σ. The biased versions of the chain come up in the context of nanotechnology [**?**] and biased card shuffling [**?**] and a biased version of the Glauber dynamics for 3-colorings came up in the context of asynchronous cellular automata [**?**]. We will restrict our attention to the first two examples where the monotonic surfaces correspond to sets of supported cubes.

The nanotechnology example that motivated our work comes up in a model of DNA self-assembly. "Square" tiles are constructed from strands of DNA so that each side is single-stranded. Tiles are encouraged to line up and attach by making the corresponding edges contain complementary sequences. At appropriately chosen temperatures these tiles will have a good chance of correctly assembling, and a much smaller chance of disassociating and breaking apart. The model considered by Majumder et al. [**?**] allows the left column and bottom row of a large square to form, and then allows tiles to associate with the large substrate if their left and bottom neighbors are already present. Likewise, tiles can disassociate if their upper and right neighbors are not present, and disassociation happens at a lower rate. The dynamics of this model are precisely captured by the Glauber dynamics

on 2-dimensional monotonic surfaces. The 3-dimensional analogue is also used to study self-assembly, where tiles are shaped like cubes and complementary sequences are used to encourage matching faces to self-assemble.

Benjamini et al. [**?**] gave tight bounds on the mixing rate of Glauber dynamics for the biased chain in rectangular regions of Z^2 for any constant bias. In 3 and higher dimensions substantially less is known. Majumder et al. [**?**] showed the chain mixes quickly when the bias is $O(n^2)$; apparently the case of large bias is the most interesting for the nanotechnology applications. Nothing else is known about the convergence of the biased chain, and both of these results are highly technical and do not readily generalize to other values of the bias or other dimensions.

We make progress in several aspects of the problem of sampling biased surfaces. In two dimensions we show that the baised chain is rapidly mixing for any bias, even when it is smaller than a constant. Our proof of fast mixing is significantly simpler than the arguments of Benjamini et al. and our bounds are near optimal when the bias is constant, the case they consider (losing only a factor $\log(n)$). In addition, our arguments generalize to sampling biased surfaces on any region that is simply-connected, not just the rectangular regions previously considered. Our arguments also hold in high dimensional lattice regions. In d-dimensional lattice regions we find that the Glauber dynamics are rapidly mixing when the bias $\lambda \geq d$. Again, our bounds on the mixing time are within $\log(n)$ of optimal when the regions are rectangular, but they also show fast mixing for arbitrary simply-connected regions.

The key observation underlying these results is showing that the distance between pairs of configurations is always decreasing in expectation if we define an exponential metric on the state space. It will be more convenient to think of monotonic surfaces as unions of downsets in a lattice in order to show why an exponential metric is sufficient for these problems. Finally, we show how to modify the path coupling theorem when the distances can be exponentially large. The first half of the theorem when the distance decreases by at least an inverse polynomial amount in expectation during each move. However, it is necessary to modify the second part of the theorem to deal with the case when the expected change

is less than an inverse polynomial since in the standard version the maximum distance is a factor of the running time. We present a modified version of the theorem in this case that allows us to conclude that the chain converges in polynomial time.

The remainder of the chapter is organized as follows: In Section 3.2 we define the model and the Glauber dynamics more carefully. In Section 3.3 we introduce a modified path coupling theorem that is more appropriate in the setting when distances are exponentially large. Finally, in Section 3.4 we show how we can use our new path coupling theorem to conclude that the chain is rapidly mixing.

3.2 Monotonic surfaces

Let R be any simply-connected region on the two dimensional Cartesian lattice $\mathbb{Z}^2$ that is the union of a set of unit squares. A *monotonic surface* in R is a path starting and ending on the boundary of R that only takes steps down and to the right. Such a path is illustrated in Figure 2 when R is an $n \times n$ square. By examining the coordinates of the faces of R "below" such a surface, we have a set of lattice points $\sigma \subset [0, \ldots, n]^2$ such that for some integer vector $\overline{x} \in \mathbb{Z}^2$, $\overline{x} \in \sigma$ implies $\overline{x} - \overline{u}_i \in \sigma$, provided $\overline{x} - \overline{u}_i \in R$ (where $\overline{u}_i$ is the unit vector in the i direction). In this way, each point of σ is "supported" by the point below it and to the left which are lie under the surface (or are outside of R).

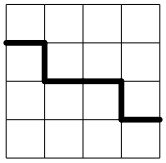

Figure 2: A monotonic surface in two dimensions.

In three dimensions, let R be a simply-connected region on $\mathbb{Z}^3$ consisting of the union of unit cubes. A monotonic surface in R is the union of two dimensional faces such that any cross-section along an axis-aligned plane is a two-dimensional monotonic surface. Such a surface is illustrated in Figure 3 when R is an $n \times n \times n$ cube. Again, examining the coordinates of the corners of the unit cubes "below" such a surface, we have a set of points $\sigma \subset [0, \ldots, n]^3$, such that for some integer vector $\overline{x} \in \mathbb{Z}^3$, $\overline{x} \in \sigma$ implies $\overline{x} - \overline{u}_i \in \sigma$, provided $\overline{x} - \overline{u}_i \in R$.

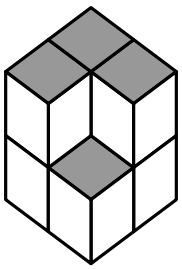

Figure 3: A monotonic surface in three dimensions.

In general, we consider simply-connected regions $R \in \mathbb{Z}^d$ composed of unit hyper-cubes, and a d-dimensional monotonic surface is a set of $d-1$ dimensional faces, such that any cross-section along an axis-aligned $(d-1)$-dimensional hyper-plane is a $d-1$-dimensional monotonic surface. The following definitions will help us define the set of regions we will be considering.

Definition 3.1 *For $v \in \mathbb{Z}^d$, we define the* ray *$r(\overline{v}^*)$ to be the infinite set*

$$r(v) = \{(k + v_1^*, k + v_2^*, \ldots, k + v_d^*) : k \in \mathbb{Z}\}.$$

Definition 3.2 *For a d-dimensional simply-connected region $R \in \mathbb{Z}^d$, we say R is* nice *if, for all $v \in R$, $R \bigcap r(v)$ is connected.*

Note that all hyper-rectangular regions are nice.

It will be convenient to think about monotonic surfaces on nice regions in terms of "downsets." Let $\overline{u}^* = (1, 1, ..., 1) \in \mathbf{Z}^d$. Then given a nice region R, we let R_L be the lower envelope of the region. That is, $R_L = \{\overline{v} \in R \text{ such that } \overline{v} - \overline{x}^* \notin R\}$.

Definition 3.3 *For a simply-connected region $R \subset \mathbf{Z}^d$ consisting of the union of unit cubes, define a* downset *to be a subset σ of R, with $R_L \subset \sigma$, such that if $v \in \sigma$ and $v - \overline{u}_i \in R$ for any i, then $v - \overline{u}_i \in \sigma$.*

For a nice region R, we define Ω_{mon} to be the set of all downsets of R. The following definition helps us formalize the Markov chain we will be using to sample from Ω_{mon}.

Definition 3.4 *Let R be any nice region and let σ be any downset of R. We say the* boundary *of σ is $\partial(\sigma) = v \in \sigma$ such that $v + \overline{u}^* \notin \sigma$.*

To sample from Ω_{mon}, we start at an arbitrary downset $\sigma_0 = R_L$. Then if we are at a given downset σ_t at time t, we pick a point $v \in \partial(\sigma)$ and a vector $b \in \pm 1$. If $b = +1$, then we let $\sigma_{t+1} = \sigma_t \cup (v + \overline{u}^*)$ if this is a valid downset; otherwise we keep σ_t unchanged. If $b = -1$, then we let $\sigma + t + 1 = \sigma_t \setminus v$ if this is a valid downset; otherwise we keep σ_t unchanged.

For nice regions, moves of this type will connect the state space Ω_{mon}. To see this, let σ be any downset, and let v^{max} be any point in σ such that $\sum_i v_i^{max}$ is maximized. Then we can always remove v^{max} and move to $\sigma' = \sigma \setminus v^{max}$ without violating the downset condition. Thus, from any valid downset σ we can always remove points and get to the "lowest" downset R_L.

Notice that for any σ and v, if $\sigma \cup v$ is a valid downset, then $|\partial(\sigma)| = |\partial(\sigma \cup v)|$. This is because we are adding v and removing $v \setminus \overline{u}^*$ from $\partial(\sigma)$ to form $\partial(\sigma \cup v)$. Therefore if follows that for any nice region R, the size of the boundary of any valid downset is fixed. We call this the *span* of R. The following two definitions will be convenient when we bound the mixing time of our Markov chain.

Definition 3.5 *The* span *of a nice region R is $\alpha = |\partial(\sigma)|$, for any downset σ of R.*

Definition 3.6 *Let R be any nice region. The* stretch *of R is*

$$\beta = \max_{v \in R} \sum_i v_i - \max_{v \in R} \sum_i v_i.$$

Thus, the stretch is the maximal distance between two points in R in the $\overline{u}^*$ direction.

Now, note that the Markov chain just described converges to the uniform distribution over downsets. We will modify the transition probabilities to define the Glauber dynamics with the appropriate Metropolis-Hastings probabilities [**?**] so that we converge to the right distribution on biased surfaces. Notice that this new chain also connects the state space by the same argument.

With all this said, for a given nice, d-dimensional nice region R, we now define the biased Glauber dynamics on Ω_{mon}.

With a given downset $\sigma_t \subset R$, each step of $\mathcal{M}_{mon}$ repeats the following steps:

- `Choose` (v, b, p) `uniformly at random from` $\partial(\sigma_t) \times \{+1, -1\} \times (0, 1)$.
- `If` $b = +1$, `let` $\sigma_{t'} = \sigma \cup (v + \overline{u}^*)$.
- `If` $b = -1$ `and` $p > \frac{1}{\lambda}$, `let` $\sigma_{t'} = \sigma \setminus v$.
- `If` σ'_t `is a valid downset, let` $\sigma_{t+1} = \sigma'_t$. `Otherwise let` $\sigma_{t+1} = \sigma_t$.

We now present the following two theorems that bound the mixing rate of M_{mon}. The first pertains to constant λ in 2 and higher dimensions. We include a corollary that gives bounds in the significant special case that the region R is a hyper-cube. The second theorem allows us to bound the mixing rate of $\mathcal{M}_{mon}$ when the bias is arbitrarily close to 1 in 2 dimensions.

Theorem 3.1 *Let R be any nice d-dimensional region with volume n. If $d = 2$, then let constant $\lambda > 1$ be the bias, or if $d \geq 3$ let $\lambda \geq d^2$ be the bias. Then the mixing time of $\mathcal{M}_{mon}$ on Ω_{mon} satisfies*

$$\tau(\varepsilon) = O(\alpha\beta \ln n \ln \varepsilon^{-1}),$$

where α is the span and β is the stretch of R.

Notice that we always have $\alpha, \beta \leq n$, so the mixing time of $\mathcal{M}_{mon}$ is at most $O(n^2 \ln n \ln \varepsilon^{-1})$ for constant λ. In addition, when R is the $h \times h \times ... \times h$ hyper-cube, with $h^d = n$, then $\alpha = O(h^{d-1})$ and $\beta = O(h)$. This gives the following corollary of Theorem 3.1.

Corollary 3.2 *Let $R = h \times h \times ... \times h$ be the d-dimensional hyper-cube with $h^d = n$. If $d = 2$ with constant $\lambda > 1$ or $d \geq 3$ with $\lambda \geq d^2$, then the mixing time of $\mathcal{M}_{mon}$ satisfies*

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

Up to a $\ln n$ factor, this matches the optimal running time. (The chain may require $O(n)$ steps just to connect the state space, so certainly no bound lower than $O(n)$ is possible.) We prove this result for all constant biases, just as in the Benjamini et al. result [**?**]. However, our result is more general, pertaining to any nice region in any dimension. We also improve their result for $d = 2$ by proving that the chain is rapidly mixing when λ is smaller than a constant. The second theorem deals with this case and is as follows.

Theorem 3.3 *Let R be any nice 2-dimensional region with volume n. Let $\lambda \geq 1$ be any small bias, which can depend on n. Then the mixing time of $\mathcal{M}_{mon}$ on Ω_{mon} satisfies*

$$\tau(\varepsilon) = O(n^5 \ln \varepsilon^{-1}),$$

where α is the span and β is the stretch of R.

This theorem is just a minor extension of Theorem 3.1, but it demonstrates how the machinery developed in this chapter can be used when we have exponentially large distance functions and we only have small (or no) decrease in the distance in expectation during a move of the chain.

To understand why it is difficult to use coupling to prove these theorems, we first examine the straightforward coupling of (σ_t, ρ_t) by simply choosing the same $(\overline{v}^*, b, p)$ to generate both σ_{t+1} and ρ_{t+1}. The natural distance metric on $\Omega_{mon} \times \Omega_{mon}$ is the Hamming distance, where $h(\sigma_t, \rho_t) = |\sigma_t \triangle \rho_t|$ (and $\triangle$ is the symmetric difference). However, with this coupling and metric, we face difficulty with even the simplest of pairs (σ_t, ρ_t).

Examine the pair of downsets in Figure 4. They differ on a single point, so $h(\sigma_t, \rho_t) = 1$. For coupling to show that $\mathcal{M}_{mon}$ is fast mixing, the expected distance $\mathbb{E}[h(\sigma_{t+1}, \rho_{t+1})]$ must be at most $h(\sigma_t, \rho_t)$. For this pair of downsets, there are two choices of $(\overline{x}^*, b)$ that decrease that distance; if $\mathcal{M}_{mon}$ succeeds on $((0), +1)$ or on $((0), -1)$ then $(\sigma_{t+1}, \rho_{t+1})$ is (ρ_t, ρ_t) or (σ_t, σ_t), respectively. In either case, the distance between σ_{t+1} and ρ_{t+1} decreases by 1.

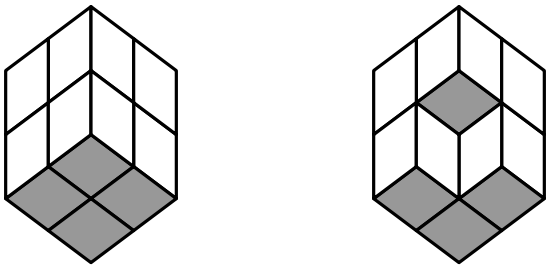

Figure 4: A pair of downsets σ_t (left) and ρ_t (right) where $\rho_t = \sigma_t \cup \{(1,1)\}$.

There are also two choices of $(\overline{x}^*, b)$ that *increase* the distance. If $\mathcal{M}_{mon}$ chooses $((1), +1)$ or $((-1), -1)$, then ρ_{t+1} gains a new point ($(2,1)$ or $(1,2)$, respectively), but σ_{t+1} remains unchanged; no addition to σ_t of a vector along that ray leaves a valid downset. With either of these choices of $(\overline{x}^*, b)$, the distance between σ_{t+1} and ρ_{t+1} increases by 1.

If $\lambda = 1$, this would be sufficient for coupling; the expected change in distance is 0. Unfortunately, with any $\lambda > 1$, the chance of $\mathcal{M}_{mon}$ choosing a p large enough to succeed on $((0), -1)$ is less than 1. Therefore the expected distance between the pair (σ_t, ρ_t) *increases* after one step.

In higher dimensions, the situation becomes even worse. For the pair of 3 dimensional downsets in Figure 5, there are *three* moves which increase the Hamming distance. (Choosing $((0,1), +1)$ adds $(1,1,0)$ to ρ_t, choosing $((-1,1), +1)$ adds $(1,0,1)$ to ρ_t, and choosing $((1,0), +1)$ adds $(0,1,1)$ to ρ_t. In all three cases the move succeeds with probability 1, and σ_t remains unchanged.) There are only two moves which decrease the distance (Choosing $((0,0), +1)$ makes $\rho_{t+1} = \sigma_t$ and choosing $((0,0), -1)$ can make $\sigma_{t+1} = \rho_t$.) Of course, the three moves that increase the distance succeed with probability 1, but one of the two moves which decreases the distance only succeeds with probability $\frac{1}{\lambda}$.

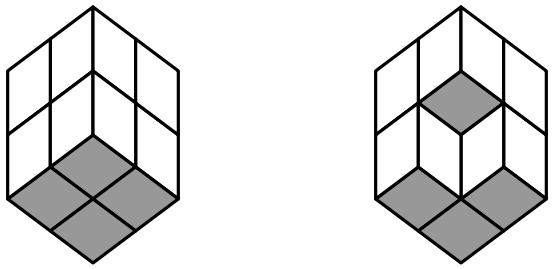

Figure 5: A pair of downsets σ_t (left) and ρ_t (right) where $\rho_t = \sigma_t \cup \{(0,0,0)\}$.

One promising solution is to alter the distance metric. When two downsets differ on some point $\overline{x}$, the two moves which decrease the distance involve removing $\overline{x}$ from $\sigma_t \triangle \rho_t$, while the moves that increase the distance involve adding $\overline{x} + \overline{u}_i$ to $\sigma_t \triangle \rho_t$ (for some i). Therefore we consider a distance metric that counts the distance between two sets that differ on $\overline{x}$ as greater than the distance between two sets that differ on $\overline{x} + \overline{u}_i$. Of course the above examples can just as easily constructed around $\overline{x} + \overline{u}_i$ as around $\overline{x}$, so our metric must measure the distance between two sets differing on $\overline{x} + \overline{u}_i$ as greater than the distance between two sets differing on $\overline{x} + \overline{u}_i + \overline{u}_j$ (for some j), and so on.

We find the following distance metric suffices. For two downsets σ, ρ, let

$$\phi(\sigma, \rho) = \sum_{\overline{x} \in \sigma \triangle \rho} (\sqrt{\lambda})^{d \cdot n - \|\overline{x}\|_1},$$

where $\|\cdot\|_1$ is the L_1 norm.

Unfortunately, this metric presents new problems as well. First, it potentially takes on non-integer values, while the Path Coupling Theorem deals solely with integer valued metrics. Second, the maximum distance between any two sets is now exponential in n, while the expected change in distance between a pair of downsets might be arbitrarily small, especially for λ very close to the bound in the theorem.

We therefore prove an extension to the Path Coupling Theorem.

3.3 An extension to path coupling

We begin by examining the two parts of Theorem 2.3. In the first, the mixing time of a chain is bounded by logarithm of the maximum distance B. The distance between two configurations may be exponentially large with respect to n and Theorem 2.3 still proves that the mixing time of the chain is polynomial in n. However, it does require that there exists a γ such that for all $(\sigma_t, \rho_t) \in U$, $\mathbb{E}[\phi(\sigma_{t+1}, \rho_{t+1})] \leq \gamma\phi(\sigma_t, \rho_t)$. This γ measures the contraction in expected distance from one step to the next,; if the contraction is within an inverse exponential of 1, e.g. $\gamma = 1 - 2^{-n}$, then the bound on mixing time is exponential and the first part of Theorem 2.3 does not prove that the chain is fast mixing.

The second part of Theorem 2.3 requires no such bound on contraction; there we need only that $\mathbb{E}[\phi(\sigma_{t+1}, \rho_{t+1})] \leq \phi(\sigma_t, \rho_t)$. However the associated bound on the mixing time is in terms of B, not $\log B$. Therefore if the distance metric can be exponentially large, e.g. $B = 2^n$, the second part of Theorem 2.3 does not prove that the chain is fast mixing either.

In this section we present the following theorem which allows for exponentially large distance metrics without requiring such a polynomially bounded contraction in expected distance.

Theorem 3.4 *Let ϕ be a real-valued metric defined on $\Omega\times\Omega$ which takes values in $[0, B]\backslash(0, 1)$. 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 $X_t = Z_0, Z_1, \ldots, Z_r = Y_t$ 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 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 there exists $\gamma \leq 1$ such that, for all $(X_t, Y_t) \in U$,

$$\mathbb{E}[\phi_{t+1} - \gamma\phi_t] \leq 0.$$

(1.) If $\gamma < 1$, then the mixing time satisfies

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

(2.) If there exists $\eta, \kappa > 0$ such that $\mathbb{P}\left[|\phi_{t+1} - \phi_t| \geq \eta\phi_t\right] \geq \kappa$ for all t provided that $X_t \neq Y_t$, then

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

There are two important distinctions between Theorem 2.3 and 3.4. The first is that Theorem 3.4 allows for non-integer metrics (provided that for all $X, Y \in \Omega$, $\phi(X, Y) < 1$ implies $\phi(X, Y) = 0$). This is a minor restructuring of the proof of Theorem 2.3 [**?**], and follows exactly from their proof.

The second is that γ may be zero (or very small) while the maximum distance B is exponentially large; this is the case for which both parts of Theorem 2.3 are insufficient for proving fast mixing. The second half of Theorem 3.4 deals with this case. We prove it with a slight modification of the original proof of Theorem 2.3, replacing the original distance $\phi(X_t, Y_t)$ with $\ln(\phi(X_t, Y_t))$. We require some technical lemmas concerning the expectation and variance of the logarithm of the distance function, which we present below, but the novelty of Theorem 3.4 is more in the statement of the result, than a new method of proof.

Note that including our modification to handle case of when $\gamma = 0$ and B is exponential requires a strong bound on the variance of ϕ_t. Without this bound on variance, Theorem 2.3 would not be not true; for example, if $\phi_0 = 2^n$ and $\phi_{t+1} = \phi_t - 1$, then clearly it will take time exponential in n for $\phi_t = 0$.

To provide the technical modifications necessary to prove Theorem 3.4, we define a new variable ψ, where $\psi_t = \ln(\phi_t)$, if $\phi_t > 0$, and $\psi_t = -1$, if $\phi_t = 0$. This means that $\psi_t \in [-1, \ln(B)]$. In Lemma 3.1 below, we will show that if the expected distance is non-decreasing, then the expected log of the distance non-decreasing. In Lemma 3.2, we show that if that the variance of ϕ is greater than a constant multiple of the current distance, then

the variance of ψ is at least a constant. Together, these lemmas give a proof of Theorem 3.4, following the exact structure of the proof of [**?**].

Lemma 3.1 *If* $\mathbb{E}[\phi_{t+1} - \phi_t] \leq 0$, *then* $\mathbb{E}[\psi_{t+1} - \psi_t] \leq 0$.

Proof:

Let $\{r_1, r_2, \dots\}$ be the possible values for ϕ_{t+1}, each occuring with probability $\{\zeta_1, \zeta_2, \dots\}$. That is, $\mathbb{P}[\phi_{t+1} = r_i | \phi_t] = \zeta_i$, with $\sum_{i=1}^{r} \zeta_i = 1$.

Then,

$$\begin{aligned}
\mathbb{E}[\psi_{t+1}] &= \sum_i \zeta_i \ln(r_i) \\
&= \ln\left(\prod_i r_i^{\zeta_i}\right) \\
&\leq \ln\left(\sum_i \zeta_i r_i\right) \\
&= \ln\left(\mathbb{E}[\phi_{t+1}]\right) \\
&\leq \ln \mathbb{E}[\phi_t] \\
&= \psi_t,
\end{aligned}$$

where the first inequality is by the Arithmetic-Geometric Mean Inequality, and the second is given in the lemma. □

Lemma 3.2 *If there exist* $\eta, \kappa > 0$ *such that* $\mathbb{P}[|\phi_{t+1} - \phi_t| \geq \eta\phi_t] \geq \kappa$, *then*

$$\mathbb{E}[(\psi_{t+1} - \psi_t)^2] \geq \eta^2\kappa.$$

Proof:

As before, let $\phi_t = \phi[\sigma_t, \sigma'_t]$. Then,

$$\begin{aligned}
\kappa &\leq \mathbb{P}\left[|\phi_{t+1} - \phi_t| > \eta\phi_t\right] \\
&= \mathbb{P}\left[\frac{\phi_{t+1}}{\phi_t} - 1 > \eta\right] + \mathbb{P}\left[\frac{\phi_{t+1}}{\phi_t} - 1 < -\eta\right] \\
&= \mathbb{P}\left[\psi_{t+1} - \psi_t > \ln(1+\eta)\right] + \mathbb{P}[\psi_{t+1} - \psi_t < \ln(1-\eta)] \\
&\leq \mathbb{P}\left[\psi_{t+1} - \psi_t > \eta\right] + \mathbb{P}[\psi_{t+1} - \psi_t < -\eta] \\
&= \mathbb{P}\left[(\psi_{t+1} - \psi_t)^2 > \eta^2\right].
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathbb{E}[(\psi_{t+1} - \psi_t)^2] &= \sum_{\ell} \ell \, \mathbb{P}[(\psi_{t+1} - \psi_t)^2 = \ell] \\
&\geq \eta^2 \kappa.
\end{aligned}$$

□

The proof of Theorem 3.4 follows the method of [**?**], replacing ϕ_t with ψ_t.

3.4 Proof that $\mathcal{M}_{mon}$ is rapidly mixing

To prove Theorems 3.1 and 3.3 , we return to the coupling of (σ_t, ρ_t) that simply supplies the same $(\overline{v}^*, b, p)$ to both σ_t and ρ_t. We let U be the set of downsets that differ on a single tile. However, instead of the Hamming distance, we define

$$\phi(\sigma, \rho) = \sum_{\overline{x} \in \sigma \triangle \rho} (\sqrt{\lambda})^{dn - \|\overline{x}\|}.$$

We will show that this distance metric is sufficient to show that ϕ_t is decreasing in expectation when λ is a constant as in the statement of Theorem 3.4. We begin by bounding the number of moves which increase the distance between surfaces.

For a pair $(\sigma_t, \rho_t) \in U$, there are two different ways the distance can increase in $(\sigma_{t+1}, \rho_{t+1})$. If $\sigma_t = \rho_t \cup \{\overline{x}\}$, we can increase the distance by attempting to add a $\overline{v}$ that succeeds in σ_t but fails in ρ_t. This occurs when $\overline{v} = \overline{x} + \overline{u}_i$ for some i, so $\overline{v}$ is "supported" in σ_t but not ρ_t. The other way to increase the distance between σ_t and ρ_t is to

remove a $\overline{v}$ that succeeds in ρ_t but not in σ_t. This occurs when $\overline{v} = \overline{x} - \overline{u}_i$ for some i, as the move creates a valid downset in ρ but not in σ. The following lemma bounds the number of such increases in distance.

Lemma 3.3 *For $\sigma_t = \rho_t \cup \{\overline{x}\}$, there are at most d choices of $(\overline{v}^*, b)$ such that ϕ_t increases.*

Proof:

We prove the Lemma by claiming that for dimensions $i \neq j$, if $\mathcal{M}_{mon}$ can increase the distance by choosing $\overline{v} = \overline{x} + \overline{u}_i$, then it cannot increase the distance by choosing $\overline{v} = \overline{x} - \overline{u}_j$.

This follows from a proof by contradiction: If $\mathcal{M}_{mon}$ can increase the distance with $\overline{x} + \overline{u}_i$, then it is because $\rho_{t+1} = \rho_t \cup \{\overline{x} + \overline{u}_i\}$ is a valid downset. That means $\overline{x} + \overline{u}_i - \overline{u}_j \in \rho_t$. On the other hand, if $\mathcal{M}_{mon}$ can increase the distance with $\overline{x} - \overline{u}_j$, it is because $\sigma_{t+1} = \sigma_t \backslash \{\overline{x} - \overline{u}_j\}$ is a valid downset, and $\overline{x} - \overline{u}_j + \overline{u}_i \notin \sigma_t$. But this contradicts the fact $\sigma_t \triangle \rho_t = \{\overline{x}\}$, justifying our claim.

This implies that, to increase the distance, $\mathcal{M}_{mon}$ may add vectors of the form $\overline{x} + \overline{u}_i$ for various dimensions i as in Figure 6, or it may remove vectors of the form $\overline{x} - \overline{u}_i$ for various dimensions i, as in Figure 7, or it may add $\overline{x} + \overline{u}_i$ and remove $\overline{x} - \overline{u}_i$ in a single dimension i, as in Figure 8. In each of these cases, there are at most d choices of $\overline{v}$ that increase the distance. □

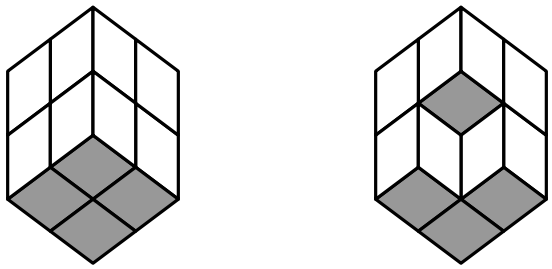

Figure 6: Downsets that differ on $\overline{x}$, where $\mathcal{M}_{mon}$ increases ϕ_t by adding $\overline{v}^* + \overline{u}_i$, for any i.

We also bound the increase in ϕ_t for each of these pairs $(\overline{v}^*, b)$.

Lemma 3.4 *If the selection of $\overline{v}^*$ and b by $\mathcal{M}_{mon}$ can cause ϕ_t to increase, then the expected increase is at most $\frac{\phi_t}{n^{d-1}2d}$*

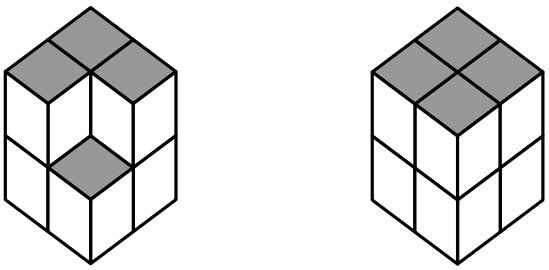

Figure 7: Downsets that differ on $\overline{x}$, where $\mathcal{M}_{mon}$ increases ϕ_t by removing $\overline{x} - \overline{u}_i$, for any i.

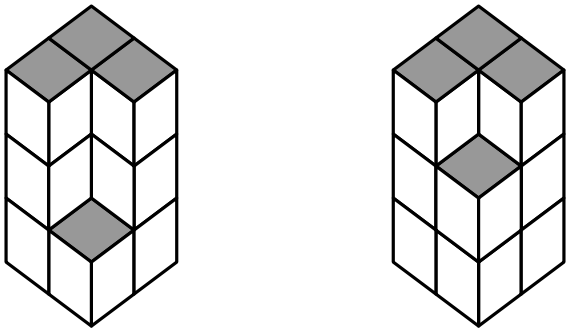

Figure 8: Downsets that differ on $\overline{x}$, where $\mathcal{M}_{mon}$ increases ϕ_t by adding the vector above $\overline{x}$ or removing the vector below $\overline{x}$.

Proof:

If the move is of the form $\overline{v} = \overline{x} + \overline{u}_i$ for some i, then the increase in distance is $\left(\frac{1}{d}\right)^{\|\overline{v}\|_1} = \frac{\phi_t}{\sqrt{\lambda}}$.

If the move is of the form $\overline{v} = \overline{x} - \overline{u}_i$ for some i, then the increase in distance is $\left(\frac{1}{\sqrt{\lambda}}\right)^{\|\overline{v}\|_1} = \phi_t\sqrt{\lambda}$, but the chance of choosing an appropriate p is $\frac{1}{\lambda}$. Therefore the expected increase is again $\frac{\phi_t}{\sqrt{\lambda}}$. □

This finally allows us to show that the expected distance is non-increasing.

Lemma 3.5 *For $\phi_t = \phi(\sigma_t, \rho_t)$,*

$$\mathbb{E}[\phi_{t+1} - \phi_t] \leq 0.$$

Proof:

As shown by Lemmas 3.3 and 3.4, there are at most d choices of $(\overline{v}^*, b)$ that can increase ϕ_t, with an expected increase of $\frac{\phi_t}{\sqrt{\lambda}}$ each. There are also two choices of $(\overline{v}^*, b)$ that *decrease* ϕ_t. These correspond to adding $\overline{x}$ and removing $\overline{x}$. These each decrease ϕ_t by ϕ_t, and

succeed with probability 1 and $\frac{1}{\lambda}$, respectively. Therefore the expected change in distance is

$$\mathbb{E}[\phi_{t+1} - \phi_t] \leq d \cdot \frac{\phi_t}{\sqrt{\lambda}} - \left(1 + \frac{1}{\lambda}\right) \phi_t$$

Substituting for $\lambda > 1$ when $d = 2$ or $\lambda > d^2$ for arbitrary d proves the lemma. □

We may now prove Theorem 3.1, showing that $\mathcal{M}_{mon}$ is rapidly mixing for appropriate choices of the bias λ. We appeal to the Theorem 3.4 to show that our coupling is sufficient.

Proof of Theorem 3.1:

We need only go through the requirements for Theorem 3.4 one by one.

For arbitrary $\sigma, \rho \in \Omega_{mon}$, if $\overline{x} \in \sigma \triangle \rho$ for some $\overline{x}$, then $\phi(\sigma, \rho) \geq \sqrt{\lambda}^{nd - \|\overline{x}\|_1} \geq 1$. Therefore if $\phi(\sigma, \rho) < 1$, $\phi(\sigma, \rho) = 0$.

We let U be the set of downsets that differ on a single vector. For arbitrary $\sigma, \rho \in \Omega_{mon}$, we can connect σ to ρ by simply adding or removing the vectors in $\sigma \triangle \rho$ one by one, and $\phi(\sigma, \rho)$ is the sum of the distances.

There are at most n^d possible vectors in $\sigma \triangle \rho$, so $\phi(\sigma, \rho) \leq n^d \sqrt{\lambda}^{n^d}$ for all σ, ρ. For any pair of σ, ρ such that $|\sigma \triangle \rho| = 1$, $\mathcal{M}_{mon}$ can always add that vector on which they differ. The appropriate $\overline{v}^*$ is chosen with probability $(2n)^{d-1}$ and the appropriate b is chosen with probability $\frac{1}{2}$ (and every p succeeds when adding). Therefore there is a $\frac{2n^{d-1}}{2}$ chance of changing $|\phi_t|$ by $|\phi_t|$.

Thus, therefore we can prove Theorem 3.1 using the first part of Theorem 3.4 □

Proof of Theorem 3.3: We prove Theorem 3.3 using the second case of The Path Coupling Theorem, Theorem 3.4. The majority of the necessary conditions were shown in the proof of Theorem 3.1. All that remains is to show the bound on the variance of the distance.

For any pair of non-equal downsets σ_t, ρ_t, examine a point $\overline{v}$ satisfying $\min_{\overline{v} \in \sigma_t \triangle \rho_t} \{\|\overline{v}\|\}$. Without loss of generality, assume $\overline{v} \in \sigma_t \backslash \rho_t$. By the minimality of $\overline{v}$, $\rho_t \cup \{\overline{v}\}$ is a valid downset. Furthermore, by the minimality of $\overline{v}$, removing $\overline{v}$ from $\sigma_t \triangle \rho_t$ reduces the distance

between the sets by at least $\frac{\phi(\sigma_t,\rho_t)}{n}$. This move occurs with probability at least $\frac{1}{n}$, so we may use the second half of Theorem 3.4, with $\ln(B) = O(n)$, $\kappa = \frac{1}{n}$ and $\eta = \frac{1}{n}$. (These extreme values of κ and η occur when the region is a long, thin rectangle, e.g. the $n \times 1$ region.) $\square$

CHAPTER IV

BOUNDING CONDUCTANCE USING "FAT FAULTS"

In this chapter we consider problems where a conjectured phase transition is causing Markov chains to require exponential convergence time at sufficiently low temperature (or low fugacity). In particular, we study three models: independent sets on the triangular lattice, weighted even orientations, and the non-saturated Ising model. (We define all of these models formally in the following section.) In each of these models, the intuition behind why the chain should be slow seems initially clear, but standard arguments do not seem sufficient for rigorous proof. We generalize these arguments to define a new tool, "fat faults," and use it to upper bound the conductance, thereby lower bounding the mixing time.

4.1 Introduction

We begin with a model for gas molecules, the so-called "hard-core lattice gas model." We let $L = (V, E)$ be the $n \times n$ Cartesian lattice, and define Ω_{Ind} to be the set of all independent sets of σ, that is, all $\sigma \subset V$ such that for all $v_1, v_2 \in \sigma$, $(v_1, v_2) \notin E$. One imagines a gas molecule at each point of σ, with the diameter of the molecule so large that there is not room for a molecule at any adjacent point.

The stationary distribution on Ω_{Ind} is parameterized by a "fugacity" $\lambda > 1$, so that larger independent sets are favored over smaller sets. The distribution is defined as

$$\pi(\sigma) = \lambda^{|\sigma|}/Z,$$

where Z is the partition function. To sample from Ω_{Ind}, we start at any initial $\sigma_0 \in \Omega_{Ind}$, say the empty set, and repeatedly add or remove single vertices, where possible, according to the correct conditional probabilities to converge to π.

More precisely, we define the following Markov chain.

Given a current independent set σ_t, a move of the chain $\mathcal{M}_{Ind}$ is defined by repeating the

following steps:

- `Choose` (v, b, p) `uniformly from` $V \times \{+1, -1\} \times (0, 1)$.
- `If` $b = +1$ `and` $\sigma_t \cup \{v\}$ `is a valid independent set, let` $\sigma_{t+1} = \sigma_t \cup \{v\}$.
- `If` $b = -1$ `and` $p \geq \frac{1}{\lambda}$, `let` $\sigma_{t+1} = \sigma_t \backslash \{v\}$.
- `Otherwise let` $\sigma_{t+1} = \sigma_t$.

The binary b determines whether $\mathcal{M}_{Ind}$ attempts to add or remove the vertex v from σ_t and the probability p gives a chance of failure when removing vertices, so that $\mathcal{M}_{Ind}$ converges to π.

An interesting phenomenon occurs as the parameter λ is varied: for sufficiently small values of λ, $\mathcal{M}_{Ind}$ converges rapidly, while for sufficiently large values of λ, the convergence will be prohibitively slow. The intuition for this transition is that when λ is sufficiently high, large independent sets dominate the stationary distribution π; most of these sets will contain predominantly vertices on the odd sublattice of L or predominantly vertices on the even sublattice, as these pack much closer than vertices of differing parity and the set may contain more vertices. However, for $\mathcal{M}_{Ind}$ to move from a set containing mostly odd vertices to a set containing mostly even vertices, it must visit a configuration that has roughly half of its vertices on each sublattice. If λ is large, these balanced configurations will have exponentially small probability. As shown in Corollary 2.5, this means that it will take exponential time to converge to equilibrium. See, e.g. [**?**, **?**, **?**].

Peierls arguments [**?**] allow us to formalize this intuition by defining "contours," or paths dividing regions of odd vertices from regions of even vertices. We then construct injections that map configurations with long contours to configurations with substantially larger stationary probability. For independent sets, the injection is constructed by shifting the interior of such a contour and adding many new vertices to the set, as in Figure 9. For formal definitions, see, e.g., [**?**, **?**, **?**, **?**].

For the Ising model at low temperature, the proof of slow mixing is similar; when temperature is sufficiently low, Ising configurations are greatly penalized for containing

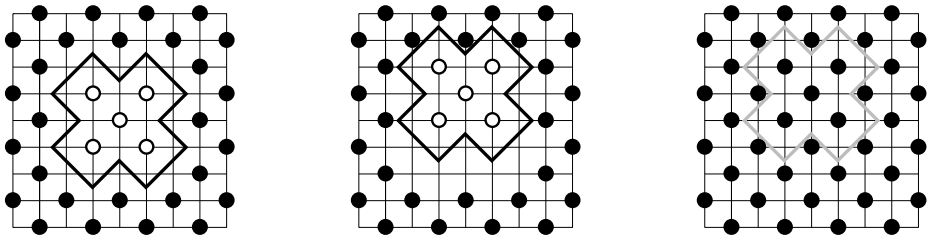

Figure 9: An independent set on L with a contour separating the odd points (left), and the injection that removes the contour to increase the size (center and right).

adjacent vertices of differing spins, so configurations will be prominently one spin or the other. However to move from a configuration with predominantly one spin to a configuration with predominantly the other, $\mathcal{M}_{Ising}$ must pass through a configuration with roughly half of each spin.

To show that this set of "balanced" configurations has exponentially low stationary probability, we again define "contours," or paths dividing regions of one spin from the other. We then again construct injections that map configurations with long contours to configurations with substantially larger stationary probability. For Ising configurations, the injection is constructed by reversing the spin of all vertices on the interior of such a contour. For formal definitions, see, e.g. [**?**, **?**, **?**].

4.1.1 Non-bipartite independent sets

The phase change observed for local Markov chains on independent sets of the Cartesian lattice is believed to persist when the underlying lattice is a finite subset of the 2-dimensional triangular lattice. (Here we use the same chain $\mathcal{M}_{Ind}$ with a different underlying graph.) For small λ, the $\mathcal{M}_{Ind}$ is known to be rapidly mixing [**?**], but for large enough λ, independent sets will tend to be denser and we again expect any local dynamics to be slow.

Notice that on the triangular lattice there are three maximum independent sets instead of two, arising from the natural tri-partition of the lattice, which we color black, white and gray as in Figure 10.

The most likely configurations should be largely monochromatic, since these will be the densest. To move from an independent set that is mostly black to one that is mostly white, for example, it is necessary to visit one that has fewer than $n/2$ vertices of each of the three

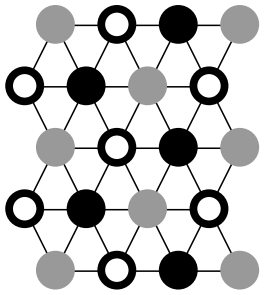

Figure 10: The three colors of the tri-partition of the triangular lattice.

colors, so there is no dominant color, and these should be unlikely when λ is large. At initial glance, it seems an easy application of Corollary 2.5.

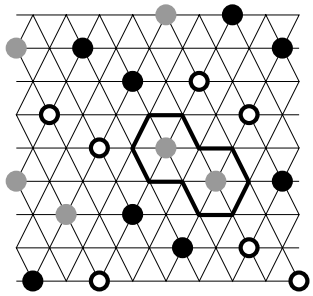

Figure 11: An independent set of the triangle lattice, with a contour separating one region of points from the others.

Unfortunately, the proof that these sets with no dominant color have small total probability is significantly more difficult than on L. Contour arguments that succeed on the Cartesian lattice do not seem to generalize readily to non-bipartite graphs such as the triangular lattice. The problem is that a contour surrounding a region whose boundary is black might be adjacent to some vertices that are white and some that are gray; there is no guarantee that any local operation such as shifting or flipping the interior of the contour will allow us to add enough new vertices to sufficiently increase the weight, as it does on bipartite lattices. An example of such a problematic balanced configuration is illustrated in Figure 11.

4.1.2 Weighted even orientations

The second model we consider should also be slow mixing, but again seems resistant to standard contour arguments. We again let $L = (V, E)$ be the $n \times n$ Cartesian lattice, and

define Ω_8 to be the set of even orientations of E, i.e., orientations of edges so that each vertex on the interior of L has even in-degree and out-degree. This is known as the *8-vertex model* in statistical physics, as there are eight possible orientations of the edges incident to any vertex.

We call vertices *sources* if they have in-degree 0 and *sinks* if they have out-degree 0; all other vertices are called *Eulerian* since their in-degree and out-degree are both 2. For $\sigma \in \Omega_8$, let $S(\sigma)$ be number of sources and sinks in σ. Given $\lambda > 0$, we define distribution π on Ω_8 by setting $\pi(\sigma) = \lambda^{S(\sigma)}/Z$ for each $\sigma \in \Omega_8$, where Z is the partition function.

To sample from Ω_8, we start at an arbitrary even orientation σ_0, say the orientation that points every edge to the right or down, and reverse the directions of the edges around a random face of L with the probabilities defined by the Metropolis-Hastings algorithm to converge to π. More precisely, we define the following Markov chain.

Given a current even orientation σ_t, a move of the chain $\mathcal{M}_8$ is defined by repeating the following steps:

- `Choose` (f, p) `uniformly from` $F(L) \times (0, 1)$`, where` $F(L)$ `is the set of faces of` L`.`
- `Let` σ'_t `be the orientation generated by reversing all four edges incident to` f`.`
- `If` $\pi(\sigma'_t) \geq \pi(\sigma_t)$ `or if` $p \geq \frac{\pi(\sigma'_t)}{\pi(\sigma_t)}$`, then let` $\sigma_{t+1} = \sigma_t$`.`
- `Otherwise, let` $\sigma_{t+1} = \sigma_t$

Notice that $\mathcal{M}_8$ reverses an even number of edges incident to any vertex of V, so σ_{t+1} is always a valid even orientation.

When $\lambda = 0$, the only allowable configurations are Eulerian orientations (known as the *6-vertex model*) where every vertex has in-degree = out-degree = 2, and the local Markov chain is known to be efficient [**?**, **?**]. When λ is close to 1, we can use simple coupling arguments to show that the chain is again rapidly mixing. However, when λ is sufficiently

large, we expect most vertices to be sources or sinks, and the chain should take exponentially long to move from a configuration that has predominantly sources which are on the odd sublattice and sinks which are on the even sublattice, to one with predominantly even sources and odd sinks.

While one would expect that configurations that are "balanced" are exponentially unlikely, this does not seem to follow from any standard contour arguments. As demonstrated by Figure 12, there is no clear map between valid configurations by flipping or shifting the interior of contour that is guaranteed to significantly increase the stationary probability, as required for the Peierls argument.

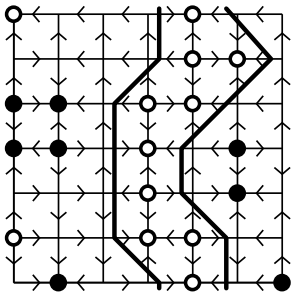

Figure 12: An even orientation of the Cartesian lattice, with contours separating one region of sources and sinks from the others.

4.1.3 Non-saturated Ising

The third model we consider is motivated by a problem in self-assembly. Recall that self-assembly is a process in which tiles are designed with markings on each side so that two are more likely to join together along an edge if they have matching markings.

The standard Ising model can be thought of as a fully-packed, or *saturated*, self-assembly model with two tiles A and B (corresponding to $+$ and $-$) occupying lattice points, where each tile prefers to be next to others of the same type. In this abstraction, we imagine that every square is occupied by one of the two tile types.

A simple modification of the standard Ising model makes it more fitting for a tile-based self-assembly model where configurations can include empty spaces. In the *non-saturated Ising model*, empty spaces are represented by a third tile type 0. As in the Ising model, we

define $\Omega_{Nonsat} = \{A, B, 0\}^V$, we define the Hamiltonian as

$$H(\sigma) = \sum_{(u,v)\in E} g(\sigma(u), \sigma(v)),$$

for some function g. We then define the Gibbs distribution as

$$\pi(\sigma) = \frac{e^{\beta H(\sigma)}}{Z},$$

where Z is the partition function and β is the inverse of the temperature T.

The range of g depends on the construction of our tiles; there is a stronger energy for pairs of tiles that contain more complementary markings. However we always assume a tile of either type most prefers to be next to another tile of the same type, but also prefers to have its neighbors occupied rather than unoccupied, even if it is with a tile of the other type. Therefore for some values $w_{match} > w_{differ} > 0$, we set the weights $g(A, A) = g(B, B) = w_{match}$, $g(A, B) = g(B, A) = w_{differ}$ and $g(0, x) = g(x, 0) = 0$ for all $x \in \{A, B, 0\}$.

For simplicity, we let $\lambda = e^{\beta w_{match}}$ and $\mu = e^{\beta w_{differ}}$. We can now describe the stationary probability of a configuration σ as

$$\pi(\sigma) = \frac{\lambda^{\#\text{match}}\mu^{\#\text{differ}}}{Z},$$

where #match is the number of nearest neighbor pairs that are both assigned A or both assigned B, and #differ is the number of pairs where one tile is A and the other is B. We will use this representation throughout the section.

For the non-saturated Ising model it might make the most sense to connect two configurations of Hamming distance one only if one of the configurations has a tile of type 0 in the position of disagreement, as in Figure 13; recall that 0 represents an empty position, so these moves correspond to tiles of type A or B attaching or detaching from the larger configuration. Switching from A to B should require two moves: first removing A and then adding B.

To sample from Ω_{Nonsat}, we start at an arbitrary tiling σ_0, say the empty tiling ($\{0\}^V$), and add or remove a tile at a random vertex, with the probabilities defined by the Metropolis-Hastings algorithm. More precisely, we define the following Markov chain.

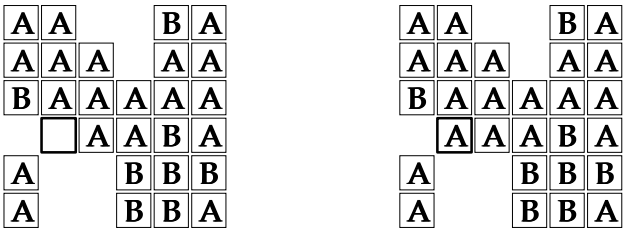

Figure 13: A move of the Markov chain $\mathcal{M}_{Nonsat}$.

Given a current tiling σ_t, a move of the chain $\mathcal{M}_{Nonsat}$ is defined by repeating the following steps:

- `Choose` (v, s, p) `uniformly from` $V \times \{A, B, 0\} \times (0, 1)$.
- `If` $s = 0$, `let` σ'_t `be the tiling identical to` σ_t `except with` $\sigma'_t(v) = 0$.
- `If` $s = A$ `or` B `and` $\sigma_t(v) = 0$, `then let` σ'_t `be the tiling identical to` σ_t `except with` $\sigma'_t(v) = s$. `If` $\sigma_t(v) \neq 0$, `let` $\sigma'_t = \sigma_t$.
- `If` $\pi(\sigma'_t) \geq \pi(\sigma_t)$ `or if` $p \geq \frac{\pi(\sigma'_t)}{\pi(\sigma_t)}$, `then let` $\sigma_{t+1} = \sigma_t$.
- `Otherwise, let` $\sigma_{t+1} = \sigma_t$

As in the models described above, the intuition for using Corollary 2.5 seems clear; let S_1 be the set dominated by A and S_3 be the set dominated by B. To pass from S_1 to S_3, one must pass through a set with no dominating tile type, in S_2. As the distribution tends towards configurations with tiles that are next to those of the same type, this set S_2 should be of exponentially small weight.

Unfortunately, as with the models above, actually proving that this set is of exponentially small weight is difficult; there is no clear way to show that containing a long contour separating A from B implies exponentially small weight.

4.1.4 Our results

We provide the first rigorous proofs that Glauber dynamics are slow for the models described above. For even orientations and the non-saturated Ising model, we show slow mixing of the local Markov chain on rectangular regions with fixed boundary conditions, while

for independent sets on the triangular lattice we consider regions with periodic (toroidal) boundary conditions; these turn out to be the simplest regions for our arguments. Our three main theorems are as follows.

Theorem 4.1 *Let Λ be an $n \times n$ rhomboidal region of the triangular lattice with periodic boundary conditions. Let Ω_{Ind} be the set of independent sets on Λ, and let $\mathcal{M}_{Ind}$ be the local Markov chain on Ω_{Ind}. There exists constants λ', k such that for all $\lambda > \lambda'$, the mixing time of $\mathcal{M}_{Ind}$ is $\Omega(e^{kn})$.*

Theorem 4.2 *There exists constants λ', k such that for all $\lambda > \lambda$, the mixing time of $\mathcal{M}_8$ is $\Omega(e^{kn})$.*

Theorem 4.3 *For all values of $w_{match} > w_{differ} > 0$, there exists constants T_1, T_2, k_1, k_2 such that for all $T > T_1$, the mixing time of $\mathcal{M}_{Nonsat}$ is $O(n^{k_1})$, but for all $T < T_2$, the mixing time of $\mathcal{M}_{Nonsat}$ is $\Omega(e^{k_2 n})$.*

Our proofs are based on several innovations. First, following [**?**], we abandon the approach of partitioning the state space so that the middle set contains "balanced" configurations in the sense described above, and instead base the partition of the state space on "topological obstructions." Roughly speaking, the middle set in our partition of the state space is defined by the presence of "fault lines," or paths across the region that pass only through "unfavorable" regions (in our cases, vacant vertices in Ω_{Ind}, Eulerian vertices in Ω_8, and edges where opposing tiles touch in Ω_{Nonsat}.). The absence of a fault line is characterized by the presence of a pair of monochromatic blocking path of "favorable" vertices, and the type of these paths determines which part of the state space a configuration lies in.

To see why this is different from the standard approach, consider an independent set that contains such a pair of perpendicular paths, composed entirely of vertices on the odd sublattice, and then also includes all possible vertices on the even sublattice in the remaining space. Such a set is illustrated in Figure 14. This independent set is considered "odd" because it contains a pair of paths on the odd sublattice across the graph. This is despite the fact that it has $O(n^2)$ even vertices and only $O(n)$ odd ones. This partition

of the state space was shown to greatly simplify the combinatorial methods underlying the Peierls argument for bipartite independent sets [**?**] and can be extended to the models we consider here as well.

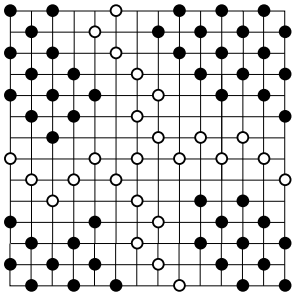

Figure 14: An independent set "dominated" by odd vertices (in white), although the majority are even (in black).

It is still the case that the 1-dimensional contours on bipartite lattices, sufficient in the context of independent sets and the Ising model, do not readily generalize to our problems for the reasons outlined above. However, a generalized notion of contours to a potentially 2-dimensional region can be made to work. Instead of defining a *minimal* connected set of unfavorable vertices, we define "fat contours" to be *maximal* connected sets of unfavorable vertices. We define an injective map from configurations in the "middle set" of the state space, i.e. those containing a fault line, by replacing an entire fat contour with favorable vertices (a maximal independent set, a maximal set of sources and sinks, or a full set of matching tiles). We then show that the gain in stationary probability is sufficient to outweigh the amount of information needed to invert the map.

Finally, in order to show slow mixing in the context of independent sets on the triangular lattice with periodic boundary conditions, it is necessary to study *multiple* non-contractible fault lines, depending on the color of the boundary vertices. On the Cartesian lattice with periodic boundary conditions, it was only necessary to find up to two non-contractible cycles and to shift (or flip) the configuration between these; on the triangular lattice it is sometimes necessary to find three non-contractible cycles, as there are three colors of regions and our fault line might not be bounded by the same color on each side.

For ease of explanation, we prove these theorems in the reverse order of how they were presented above. We prove Theorem 4.3 in Section 4.2, Theorem 4.2 in Section 4.3, and Theorem 4.1 in Section 4.4.

4.2 The non-saturated Ising model

To show fast and slow mixing for for the non-saturated Ising model at high and low temperature, we do not deal with specific temperatures T as they relate to w_{match} and w_{differ}. Instead we use the construction that $\lambda = e^{\beta w_{match}}$ and $\mu = e^{\beta w_{differ}}$, and show that $\mathcal{M}_{Nonsat}$ is rapidly mixing when λ and μ are close to 1, and $\mathcal{M}_{Nonsat}$ is slowly mixing when λ and μ are very large.

4.2.1 Rapid mixing at high temperature

We first consider the high temperature case in which tiles still prefer to be present, and nearest neighbors prefer to be of the same type, but these preferences are rather weak. In other words, we consider the case when λ and μ are close to 1, and use Theorem 2.3 to show that $\mathcal{M}_{Nonsat}$ is rapidly mixing. Specifically, we prove the following theorem.

Theorem 4.4 *For $\varepsilon > 0$ and $1 < \mu < \lambda < 1.1$, the mixing time $\tau(\varepsilon)$ of $\mathcal{M}$ is $O(n^2 \ln 1/\varepsilon)$.*

Notice that the first part of Theorem 4.3 follows as a corollary.

To prove $\mathcal{M}_{Nonsat}$ is rapidly mixing for λ and μ close to 1, we use a coupling argument; we create a new Markov chain which acts on *pairs* of configurations in $\Omega \times \Omega$, and show that the time until any two configurations become equal under this coupling is polynomial in n. To define the coupling, we choose a triple (v, s, p) exactly as before, but now, given a pair of configurations (σ_t, ρ_t) at time t, we apply the move defined by the same triple to both configurations at once, getting a new pair $(\sigma_{t+1}, \rho_{t+1})$.

We define the distance $\phi(\sigma, \rho)$ as the minimum number of moves of $\mathcal{M}_{Nonsat}$ to go from σ to ρ. (Note that this is slightly different than the Hamming distance, as changing a single vertex from A to B requires *two* moves.) Using Theorem 2.3, we may restrict our examination to the pairs that are distance 1. If we show that, for any such pair, the expected distance decreases after a single move, this will show rapid mixing.

Lemma 4.1 *Let σ_t and ρ_t be configurations such that $\phi(\sigma_t, \rho_t) = 1$. Then, for every $1 < \mu < \lambda < 1.1$,*

$$\mathbb{E}[\phi(\sigma_{t+1}, \rho_{t+1})] \leq 1 - \frac{1}{9n^2}.$$

Proof:

Let v^* be the single vertex on which σ_t and ρ_t differ. Without loss of generality, we may assume $\sigma_t(v^*) = A$ and $\rho_t(v^*) = 0$.

When performing the move prescribed by the triple (v, s, p) to the pair (σ_t, ρ_t), the distance only changes if v is v^* or a neighbor of v^*. In all other cases, the Markov chain performs the same change to σ_t and ρ_t, and $\phi(\sigma_t, \rho_t) = \phi(\sigma_{t+1}, \rho_{t+1})$.

On the other hand, if v is v^* or one of its neighbors, the distance might increase or decrease by 1 after the move. To show that the expected distance is decreasing, we first lower bound the probability of it decreasing, and then upper bound the probability of it increasing.

The two moves on which we may decrease the distance are when $v = v^*$ and s is either A or 0. In the first, ρ_t is changed to σ_t while σ_t remains the same. Therefore $\rho_{t+1} = \sigma_t = \sigma_{t+1}$. In the second, σ_t is changed to ρ_t while ρ_t stays the same, so $\sigma_{t+1} = \rho_t = \rho_{t+1}$. The probability of choosing an appropriate v and s, and an p high enough that the move succeeds, is

$$\frac{1}{n^2}\frac{1}{3}\min\left(1, \frac{\pi(\sigma_t)}{\pi(\rho_t)}\right) + \frac{1}{n^2}\frac{1}{3}\min\left(1, \frac{\pi(\rho_t)}{\pi(\sigma_t)}\right).$$

As $\pi(\sigma_t) \geq \pi(\rho_t)$, this probability is lower bounded by $\frac{1}{3n^2} + \frac{1}{3n^2\lambda^4}$.

There are many moves that potentially increase the distance. The first is when $v = v^*$ and $s = B$. Then the move automatically does nothing to σ_t, but might further change ρ_{t+1}. This occurs with at most probability $\frac{1}{n^2}\frac{1}{3}$.

Another move that may increase the distance is choosing a vertex v' adjacent to v^*. Here, the range of acceptable p might be larger for one configuration than the other, so one configuration's move succeeds while the other's fails. Such a difference occurs if $\sigma_t(v') = \rho_t(v') \neq 0$. Then, when $v = v'$ and $s = 0$, the Markov chain changes $\pi(\sigma_t)$ by a factor of (at most) λ more than it changes $\pi(\rho_t)$. Hence after choosing v and s, the probability of

succeeding on σ_t and not ρ_t is at most $\lambda^i(1-\frac{1}{\lambda})$, for some i. This is maximized when $i=0$, and the chance of choosing v and s is $\frac{1}{n^2}\frac{1}{3}$. Therefore the difference of their two chances of success is at most $\frac{1}{n^2}\frac{1}{3}(1-\frac{1}{\lambda})$.

As there are only four neighbors of v^*, we have

$$\begin{aligned}\mathbb{E}[\phi(\sigma_{t+1},\rho_{t+1})-\phi(\sigma_t,\rho_t)] &\leq \frac{1}{3n^2}+\frac{4}{3n^2}\left(1-\frac{1}{\lambda}\right)-\frac{1}{3n^2}-\frac{1}{3n^2\lambda^4}\\ &= \frac{1}{3n^2}\left(4-\frac{4}{\lambda}-\frac{1}{\lambda^4}\right)\\ &< \frac{1}{3n^2}\left(-\frac{1}{3}\right)\end{aligned}$$

□

This allows us to prove Theorem 4.4.

Proof of Theorem 4.4:

The maximum distance between any pair of tilings is $2n^2$ (from the all A tiling to the all B tiling). If we let $U \subset \Omega_{Nonsat}\times\Omega_{Nonsat}$ be the pairs of configurations that differ by a single move of $\mathcal{M}_{Nonsat}$, then the expected distance between any pair in U is non-increasing. This proves Theorem 4.4 by Theorem 2.3. □

4.2.2 Slow Mixing at low temperature

We now turn our attention to the low temperature case, when λ and μ are both large and λ is much greater than μ. When λ/μ is large, there is a large penalty to the stationary probability each time an A tile is placed next to a B tile, or when we have a 0 tile. We will show that when this ratio is large enough, typical configurations will be dense and will have a predominance of A tiles or B tiles. This allows us to show there is a bottleneck in the state space, implying that the chain is slowly mixing due to small conductance. Specifically, we show the following theorem.

Theorem 4.5 *If $\frac{\lambda}{\mu} > 28$, then there exists constant $c>1$ such that the mixing time of $\mathcal{M}$ on the non-saturated Ising model on the $n\times n$ square toroidal lattice is $\Omega(c^n)$.*

We use the conductance method described by Corollary 2.5. To move from a configuration which is dominated by A tiles to a configuration which is dominated by B tiles, $\mathcal{M}_{Nonsat}$ must pass through a configuration where there is no dominating type. Intuitively, these configurations without a dominating type are heavily penalized when A and B interact often. We therefore partition the state space into three sets, and show that the set of configurations with no dominating type forms a bottleneck.

We also take the approach of [**?**] and focus on finding the simplest possible description of the cut to simplify the subsequent analysis. We define middle set $\mathcal{F}$ as configurations that have a "fault line": a long path consisting entirely of "unfavorable" edges, where either a tile is unoccupied or where A meets B. We formalize these definitions momentarily, but any configuration with such a path will be penalized for each edge, and therefore have very low stationary probability.

The key idea is that the state space Ω now can be partitioned into three sets $\mathcal{A}$, $\mathcal{F}$ and $\mathcal{B}$. Any configuration that does not have a fault line can be shown to have a large connected component of a single type of tile crossing the lattice region; these configurations (which actually comprise the majority of Ω) lie in $\mathcal{A}$ or $\mathcal{B}$ according to the tile type.

In the proof, we show that it is necessary for the Markov chain to pass through configurations in $\mathcal{F}$ to move from $\mathcal{A}$ to $\mathcal{B}$, and yet $\mathcal{F}$ has exponentially smaller Gibbs measure than each of the other two sets; this guarantees that the chain is slowly mixing since it is enough to show that the chain has exponentially small conductance. This idea of using "topological obstructions" based on fault lines to partition the state space for a proof of slow mixing has been applied previously in the context of independent sets [**?**] and the standard Ising model [**?**].

We need to define several terms to formalize the argument.

Definition 4.1 *We call an edge* good *if it lies between A and A or B and B, and call it* bad *otherwise.*

Definition 4.2 *Here we define a* vertical fault line *to be a path of bad edges from the top of L to the bottom. Define a* horizontal fault line *similarly.*

We say two faces of the lattice are *adjacent* if they share an edge. Define an *A-bridge* to be a path of adjacent A faces from one side of L to the other.

Definition 4.3 *Define an* A-cross *to be a pair of A-bridges, one vertical and one horizontal. Define a* B-cross *similarly.*

Let $\mathcal{F}$ be the set of configurations that contain a fault line. Let $\mathcal{A}$ and $\mathcal{B}$ be the configurations which have an A-cross and a B-cross, respectively. We now show that these three sets partition Ω.

Lemma 4.2 *The sets $\mathcal{A}$, $\mathcal{F}$, and $\mathcal{B}$ are disjoint, with $\mathcal{A} \cup \mathcal{F} \cup \mathcal{B} = \Omega_{Nonsat}$. Furthermore, for any $\sigma \in \mathcal{A}$ and $\rho \in \mathcal{B}$, $P(\sigma, \rho) = 0$.*

Proof:

Since we are in two dimensions, no configuration can have both an A-cross and a B-cross, as their intersection would have both types of tiles. Similarly, a configuration cannot have both a cross and a fault line, as the fault line would put a bad edge across a bridge, contradicting the notion of a bad edge. Therefore $\mathcal{A}$, $\mathcal{F}$, and $\mathcal{B}$ are disjoint.

To show that the three sets make up all of Ω_{Nonsat}, examine a configuration σ with no horizontal bridge. Let T be the set of bichromatic vertices of σ that have a face-connected path to the top of L. As the top of L is not a bridge, T is non-empty (although not necessarily connected). By the definition of T, each edge incident to T is good, or it would be included in T. If T never reaches the bottom of L, then the vertices incident to T contain a horizontal bridge (and therefore a contradiction). On the other hand, if T reaches the bottom of L, then there is a vertical fault line.

In a similar argument, if σ has no vertical bridge then σ has a horizontal fault line. Therefore $\Omega_{Nonsat} = \mathcal{A} \cup \mathcal{F} \cup \mathcal{B}$.

To show that $\mathcal{M}_{Nonsat}$ may not move from $\mathcal{A}$ to $\mathcal{B}$, note that every move of $\mathcal{M}_{Nonsat}$ either adds or removes a tile. Moving from $\sigma \in \mathcal{A}$ to $\rho \in \mathcal{B}$ would entail first breaking an A-cross and then creating a B-cross; this requires at least two moves. Hence $\mathcal{F}$ divides $\mathcal{A}$ from $\mathcal{B}$. □

Clearly $\pi(\mathcal{A}) = \pi(\mathcal{B})$ by symmetry. To use Corollary 2.5, we need only show that for large n, $\pi(\mathcal{F}) < C^n$, for some fixed constant $C < 1$. To this end, we define a function ψ that maps configurations in $\mathcal{F}$ to new configurations with exponentially larger stationary probability. Although this function will not be one-to-one, we will show that, for each configuration in Ω_{Nonsat}, the total stationary probability of the pre-image is exponentially less than the weight of the configuration itself. This will allow us to show that $\mathcal{F}$ is exponentially smaller than Ω_{Nonsat}, and therefore significantly smaller than each of $\mathcal{A}$ and $\mathcal{B}$.

We know that any configuration $\sigma \in \mathcal{F}$ has at least one fault line, and possibly many. The function ψ will select one and then will remove it by adding A tiles everywhere that is currently unoccupied, thereby forcing the new configuration to have much greater Gibbs weight. This would be enough to show that $\pi(\mathcal{F})$ were small if ψ were injective, but of course it typically is not. Therefore we need to argue that for each $\sigma \in \mathcal{F}$, the increase in weight of $\psi(\sigma)$ is large enough to account for all the pre-images in $\mathcal{F}$. In addition, we need to generalize the notion of a fault line to a maximal connected component, in order to define a map that has the described properties.

Definition 4.4 *Define a* fat fault *to be a maximal connected set of adjacent bad edges containing a horizontal or vertical fault line.*

An example of a fault line and a fat fault is illustrated in Figure 15.

Figure 15: A fault line (black) and the fat fault containing it (gray).

For ψ, we choose an arbitrary fat fault T of σ, say the lexicographically first one, and replace all of the bad edges of T with good edges.

To define ψ formally, first notice that the regions separated by the edges of T are either individual faces of 0 or larger collections of faces that are bordered with entirely A or entirely B. We call a region bounded by A tiles to be *A-bordered* and define *B-bordered* regions similarly. To find $\psi(\sigma)$, fill in every 0 face with a A, and flip the sign of every A or B face in the B-bordered regions. Therefore every edge that was bad in σ is between two A tiles

in $\psi(\sigma)$. Such a change is illustrated in Figure 16.

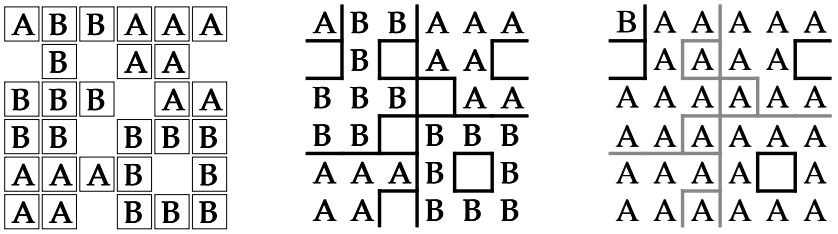

Figure 16: An example of σ, σ with bad edges highlighted, and $\psi(\sigma)$.

Now suppose that we are given $\sigma' \in \mathrm{Img}(\psi)$. Notice that in order to invert ψ to recover σ, it suffices to have an encoding of T and a labeling of the distinct regions of T dictating for each whether they were 0s, A-bordered, or B-bordered before the map made them all A-bordered. In what follows we will continue to call the faces of T *regions* to distinguish them from the faces of the underlying lattice; a region of T may contain far more than one face of L.

We now proceed to prove several lemmas that provide the necessary bounds to put these pieces together.

Lemma 4.3 *Let $\mathcal{T}_m$ be the set of fat faults with m bad edges. The number of fat faults in $\mathcal{T}_m$ is at most $n^2 16^m$.*

Proof:

For some $T \in \mathcal{T}_m$, a depth-first search of the edges of T hits each edge at twice, so takes at most $2m$ steps. For each step of the search, we follow an edge north, south, east, or west. If we are given the starting point and these $2m$ steps, we can recreate T. There are at most n^2 possible starting points, and 4^{2m} possible depth-first search traversals, so we have the lemma. □

Lemma 4.4 *For a given $T \in \mathcal{T}_m$, the number of labellings of the regions of T as A-bordered, B-bordered, or 0 is at most $\sqrt{3}^m$.*

Proof:

Every region of T is bounded by at least four edges. Each edge is on the boundary of two regions. Therefore the number of regions is at most $m/2$. Hence there are at most $3^{\frac{m}{2}}$ choices of labellings for each region. □

We may now use Lemmas 4.3 and 4.4 to bound the weight of the pre-image of ψ.

Lemma 4.5 *For each $\sigma' \in \mathrm{Img}(\psi)$,*

$$\pi(\psi^{-1}(\sigma')) < n^4 \frac{28\mu}{\lambda}^n \pi(\sigma').$$

Proof:

We first partition pre-images based on the size of the fat fault defining ψ. Let $\psi_m^{-1}(\sigma')$ be the set of all σ such that the fat fault chosen by ψ has m edges and $\psi(\sigma) = \sigma'$. The function ψ turns every bad edge of T into an edge separating A and A, thereby increasing the weight of each by at least a factor of $\frac{\lambda}{\mu}$. So, for each $\sigma \in \psi_m^{-1}(\sigma')$,

$$\pi(\sigma) \leq \left(\frac{\mu}{\lambda}\right)^m \pi(\sigma').$$

On the other hand, $|\psi_{m,\ell,s}^{-1}(\sigma')|$ is bounded by the number of possible fat faults T and labellings of the regions of T. By Lemmas 4.3 and 4.4,

$$|\psi_m^{-1}(\sigma')| < n^2 16\sqrt{3}^m < n^2 28^m.$$

Putting these pieces together, we have

$$\begin{aligned}
\pi(\psi^{-1}(\sigma')) &= \sum_{m=n}^{n^2} \pi(\psi_m^{-1}(\sigma')) \\
&= \sum_{m=n}^{n^2} \sum_{\sigma \in \psi_m^{-1}(\sigma')} \pi(\sigma) \\
&= \sum_{m=n}^{n^2} |\psi_{m,\ell,s}^{-1}(\sigma')| (\frac{\mu}{\lambda})^m \pi(\sigma') \\
&< \sum_{m=n}^{n^2} n^2 28^m (\frac{\mu}{\lambda})^m \pi(\sigma') \\
&< n^2 n^2 (\frac{28\mu}{\lambda})^n \pi(\sigma').
\end{aligned}$$

□

Lemma 4.5 allows us to now prove our main Theorem, that at low temperatures the Markov chain mixes slowly.

Proof of Theorem 4.5:

We will show that when $\lambda > 28\mu$, $\mathcal{F}$ has exponentially low stationary probability, and thus the chain has exponentially small conductance.

$$\begin{aligned}
\pi(\mathcal{F}) &= \sum_{\sigma' \in \mathrm{Img}(\psi)} \pi(\psi^{-1}(\sigma')) \\
&< \sum_{\sigma' \in \mathrm{Img}(\psi)} n^4 \left(\frac{28\mu}{\lambda}\right)^n \pi(\sigma') \\
&\leq n^4 \left(\frac{28\mu}{\lambda}\right)^n \sum_{\sigma' \in \Omega_{Nonsat}} \pi(\sigma') \\
&= n^4 \left(\frac{28\mu}{\lambda}\right)^n .
\end{aligned}$$

where the first inequality is by Lemma 4.5. By Corollary 2.5, this implies that the mixing time of $\mathcal{M}_{Nonsat}$ is exponential in n. □

4.3 Weighted Even Orientations

We now return to even orientations of the lattice. Recall, given a fixed constant $\lambda > 0$, for each $\sigma \in \Omega_8$, $\pi(\sigma) = \lambda^{S(\sigma)}/Z$, where $S(\sigma)$ is the number of sources and sinks in σ and Z is the partition function.

When $\lambda = 0$, it is known that $\mathcal{M}_8$ is rapidly mixing, as this corresponds to Eulerian orientations on Cartesian lattice regions (since there cannot be any sources or sinks) [**?**, **?**].

When $\lambda = 1$, all even orientations are equally likely and the probability of flipping the orientation of edges around any face is the same. If we define the distance between two configurations to be the number of edges in which their orientations differ, then it is easy to construct a coupling argument to show that the chain is rapidly mixing. In fact, when $\lambda = 1$, all moves occur with probability $1/2$. We can define a coupling so that the distance function never increases during moves of the coupled chain.

When λ is close to 1 but not equal, the distance function can increase as well as decrease, but the coupling argument still can be made to work when λ is sufficiently close to 1. We present this straightforward argument in Section 4.3.2.

However, when λ is large, the Markov chain behaves quite differently and we verify that the convergence to equilibrium requires exponential time. This proof is based on the fat-faults presented in the earlier, but before proceeding with our analysis of the mixing time, we present a reinterpretation of Ω_8 as an edge coloring. This edge coloring model will aid in the clarity of both the proof of rapid mixing for low λ and of slow mixing for high λ.

4.3.1 Edge orientations as edge colorings.

For a given even orientation, color an edge white if it points from a vertex on the even sublattice to a vertex on the odd one, and color it black if it points from an odd vertex to an even one. An example of this transformation is shown in Figure 17.

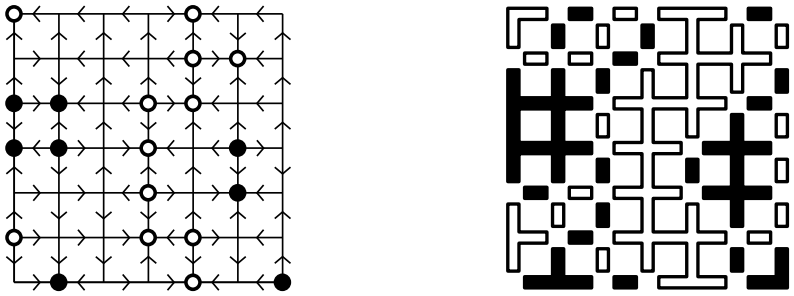

Figure 17: A configuration $\sigma \in \Omega_8$ (with sources and sinks marked) and the corresponding edge-coloring.

Now Ω_8 can be seen as the set of edge-colorings where every internal vertex has an even number of edges of each color. The sources and sinks are now *monochromatic* vertices, i.e., all incident edges are the same color. We call vertices that are incident to both black and white edges *bichromatic*.

4.3.2 Fast Mixing at low λ

We first deal with λ close to 1, and show that the chain is rapidly mixing via Path Coupling.

Theorem 4.6 *For $\lambda < 1.1$, $\mathcal{M}_8$ is rapidly mixing.*

Proof:

This follows from a standard Path Coupling argument. We use the most straightforward coupling, distance metric, and set U. For the coupling, we simply supply the same face F and real value p for both σ_t and ρ_t, when generating σ_{t+1} and ρ_{t+1}. For our distance metric $\phi(\sigma, \rho)$, we take simply the number of moves of $\mathcal{M}_8$ to convert σ to ρ. For our set $U \subset \Omega_8 \times \Omega_8$, we take the configurations that differ on a single move.

For such a pair σ_t, ρ_t, let f^* be the face on which they differ. There are two moves for which $\mathcal{M}_8$ decreases the distance in σ_{t+1}, ρ_{t+1}, flipping f^* one way or flipping it back. The probability of these two is at least $1 + \frac{1}{\lambda^4}$.

The moves for which $\mathcal{M}_8$ could increase the distance are when it chooses a face adjacent to f^*. If these moves are more likely to succeed on, say, σ than ρ, then the distance could increase if p is chosen between the two acceptance probabilities. For each of these eight faces, the chance of p being chosen in such a range is at most $1 - \frac{1}{\lambda^2}$.

Summing these two probabilities, the expected change in distance between σ_t and ρ_t is at most

$$8\left(1 - \frac{1}{\lambda^2}\right) - 1 - \frac{1}{\lambda^4}.$$

Substituting for the bounds on λ, we prove the theorem. □

This proof can be significantly improved, but this straightforward argument shows that $\mathcal{M}_8$ is rapidly mixing for λ in some range around 1.

4.3.3 Slow Mixing at high λ

We may now proceed to the main proof, which is that the chain is slowly mixing. The intuition is that it takes a long time to move from a configuration that is predominantly white to one that is predominantly black. This is because it is necessary to pass through configurations that have a large number of bichromatic vertices, and these have much smaller stationary probability. The goal is to show that there is a partition of the state space that defines a bad cut so that we can use the conductance theorem to show that the chain mixes slowly. Rather than the natural choice of partitioning the state space according to

the relative numbers of black and white vertices in the configurations, we instead use the approach of [**?**] and partition according to "fault lines."

We need to define a looser definition of "adjacency" for vertices of L. Therefore we call two vertices *edge-adjacent* if they are adjacent in the traditional sense (they share an edge of L) and call two vertices of L *face-adjacent* if they lie on a common face of L. ("Face-adjacent" is a broader term, as the vertices can be edge-adjacent or diagonally opposite across a face.)

Definition 4.5 *Let a* vertical fault line *to be a connected path of bichromatic face-adjacent vertices from the top of L to the bottom. A* horizontal fault line *is defined similarly.*

Let $\mathcal{F} \subset \Omega_8$ be the set of all configurations containing a fault line. We may now define the construction which blocks the existence of a fault line.

Definition 4.6 *Define a* vertical bridge *to be a edge-connected path of monochromatic vertices which touches both the top and bottom of L. A* horizontal bridge *is defined similarly. We say that a configuration has a* cross *if it contains both a horizontal and a vertical bridge.*

Note that the pair of crossing bridges must be of the same color. We let $\mathcal{W}$ be the set of configurations containing a white cross and $\mathcal{B}$ be the set of configurations containing a black cross. We now show that these three sets $\mathcal{F}$, $\mathcal{W}$, and $\mathcal{B}$ are disjoint and characterize all of Ω_8.

Lemma 4.6 *We may partition Ω_8 into $\mathcal{F}$, $\mathcal{W}$, and $\mathcal{B}$. That is, every configuration of Ω_8 has either a fault line, a white cross, or a black cross (but no two of these).*

Proof:

The sets $\mathcal{W}$ and $\mathcal{B}$ are clearly disjoint, as every configuration in $\mathcal{W}$ has a vertical white bridge and every configuration in $\mathcal{B}$ has a horizontal black bridge; no configuration can have both. Similarly $\mathcal{F}$ is disjoint from $\mathcal{W}$ and $\mathcal{B}$ because fault lines obstruct crosses; if a configuration has a horizontal fault line it cannot have a vertical bridge. What remains to be shown is that these three sets cover all of Ω_8, that any configuration without a cross must have a fault line.

Just as in Section 4.2, let σ be a configuration with no horizontal bridge and T be the set of bichromatic vertices of σ that have a face-connected path to the top of L. As the top of L is not a bridge, T is non-empty (although not necessarily connected). By the definition of T, each vertex adjacent to T is monochromatic, or it would be included in T. If T never reaches the bottom of L, then the vertices adjacent to T contain a horizontal bridge (and therefore a contradiction). On the other hand, if T reaches the bottom of L, then there is a vertical fault line. In a similar argument, if σ has no vertical bridge then σ has a horizontal fault line. □

We now show that for $\mathcal{M}_8$ to pass from $\mathcal{W}$ to $\mathcal{B}$, it must pass through $\mathcal{F}$, so the edges incident to $\mathcal{F}$ define a cut in the state space.

Lemma 4.7 *For transition probability $P(\cdot,\cdot)$ of $\mathcal{M}_8$, we have $P(\sigma_W, \sigma_B) = 0$ for all $\sigma_W \in \mathcal{W}$ and $\sigma_B \in \mathcal{B}$.*

Proof:

Assume we do have configurations $\sigma_W \in \mathcal{W}$ and $\sigma_B \in \mathcal{B}$ that differ by a single move of $\mathcal{M}_8$, say on face f. Configuration σ_W has both vertical and horizontal white bridges and σ_B has vertical and horizontal black bridges. Outside of f, $\sigma_W = \sigma_B$, so each contain both white and black paths from f to the top, bottom, left, and right of L. Moreover, for f to be critical to there being a black or a white cross, one edge e_B of f must be incident to only black edges outside of f, and the opposite edge e_W of f must be incident to only white edges outside of f. Coloring e_B black and e_W white, we must then have both a horizontal white bridge and a vertical black bridge, which is a contradiction. We therefore can conclude that $\mathcal{W}$ and $\mathcal{B}$ are not connected by a single move. □

Next we show that the stationary probability of $\mathcal{F}$ is exponentially small.

Definition 4.7 *Define a* fat contour *to be a maximally face-connected set of bichromatic vertices containing a fault line.*

As a fault line is a face-connected set of bichromatic vertices, any configuration with a fault line has a fat contour. To bound $\pi(\mathcal{F})$, we define a mapping $\psi : \mathcal{F} \to \Omega_8$ which takes $\sigma \in \mathcal{F}$ and re-colors edges incident to a fat contour so that the contour contains only white (or black) vertices. Although ψ is not one-to-one, we will show that for every $\sigma' \in \text{Img}(\psi)$, the stationary probability of the pre-image $\psi^{-1}(\sigma') = \{\sigma \in \mathcal{O} : \psi(\sigma) = \sigma'\}$ is exponentially less than the stationary probability of σ'.

Our definition of $\psi(\sigma)$ proceeds as follows. First, choose an arbitrary fat contour of σ, F (e.g. farthest to the left or top). In the complement of F, each of the connected components (or "islands" of F) has a boundary which is either entirely white or entirely black. We reverse the color of every edge within the white islands (leaving black islands as they are). Now the edges incident to F are entirely black, and F can be re-colored completely with black monochromatic vertices. An example of this modification is in Figure 18. The resulting configuration has $|F|$ more monochromatic vertices than the pre-image, corresponding to additional sources and sinks in the original even orientation.

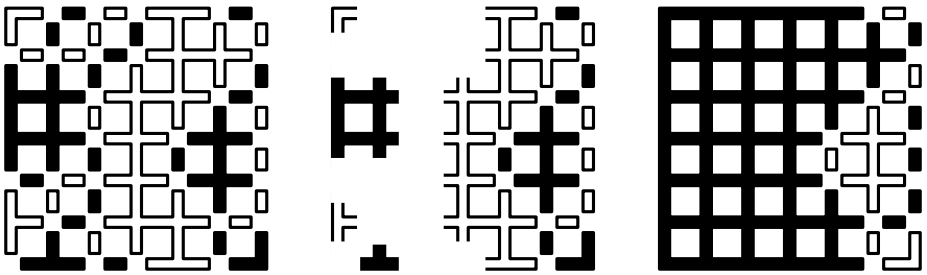

Figure 18: A coloring σ, σ with the fat contour removed, and $\psi(\sigma)$.

Notice that to find the inverse of ψ, we need only the location of F and the colorings of the edges incident to vertices of F. The edges on the boundary of the fat contour will then define whether a island was originally white or black, and we can recover the colors of edges in those islands accordingly. In Lemma 4.8 below, we bound the number of possible fat contours and their edge colorings. This allows us to bound the size of $\mathcal{F}$ in Lemma 4.9.

Lemma 4.8 *If F is a fat contour such that $|F| = \ell$, then there are fewer than $n^2 \cdot 64^\ell$ possible choices for the vertices of F and fewer than 6^ℓ colorings of the edges incident to those vertices.*

Proof:

The location of a vertex of F, together with a description of a DFS traversal of F starting at this vertex, is sufficient to reconstruct the vertices in F. As there are at most eight choices at each step in the DFS (four adjacencies along edges and four diagonals across faces) there are at most $8^{2\ell}$ such traversals. With n^2 choices for the starting vertex, we have the bound on F.

Finally, for each vertex in F, there are at most 6 possible Eulerian orientations of the incident edges, so 6^ℓ is an immediate (albeit weak) upper bound on the number of colorings of the edges incident to F. □

Lemma 4.9 *There exist constants $\lambda_0, n_0, c > 1$ such that, if $\lambda > \lambda_0$ and $n > n_0$, then $\pi(\mathcal{F}) < c^{-n}$.*

Proof:

For each $\ell \in [n, n^2]$, let $\mathcal{F}_\ell$ be the edge-colorings in $\mathcal{F}$ where the fat contour chosen by ψ is of size ℓ. Then, for each $\sigma' \in \mathrm{Img}(\psi)$,

$$\begin{aligned}
\pi(\psi^{-1}(\sigma')) &= \sum_{\ell=n}^{n^2} \sum_{\sigma \in \mathcal{F}_\ell : \psi(\sigma) = \sigma'} \pi(\sigma) \\
&= \sum_{\ell=n}^{n^2} \sum_{\sigma \in \mathcal{F}_\ell : \psi(\sigma) = \sigma'} \pi(\sigma') \cdot \lambda^{-\ell} \\
&\leq \sum_{\ell=n}^{n^2} n^2 64^\ell 6^\ell \cdot \pi(\sigma') \cdot \lambda^{-\ell} \\
&< n^5 \left(\frac{384}{\lambda}\right)^n \pi(\sigma').
\end{aligned}$$

where the first inequality is by Lemma 4.8.

This bound on the pre-image allows us to bound $\pi(\mathcal{F})$ as follows:

$$\begin{aligned}\pi(\mathcal{F}) &= \sum_{\sigma' \in \mathrm{Img}(\psi)} \pi(\psi^{-1}(\sigma')) \\ &< \sum_{\sigma' \in \mathrm{Img}(\psi)} n^5 \left(\frac{384}{\lambda}\right)^n \pi(\sigma') \\ &= n^5 \left(\frac{384}{\lambda}\right)^n \pi(\mathrm{Img}(\psi)) \\ &< n^5 \left(\frac{384}{\lambda}\right)^n .\end{aligned}$$

Taking $\lambda > 384$ yields the lemma. □

By Corollary 2.5, this proves Theorem 4.2.

4.4 Independent Sets on the Triangular Lattice

A similar approach to that of the previous sections can be applied to the context of independent sets on the triangular lattice. The primary challenge in extending the argument to this context is that the map now needs to be based on *shifting* the interior of a contour to change the contents, and here the lattice regions have periodic boundary conditions.

We let Λ be a $3n \times 3n$ rhomboidal region of the triangular lattice with periodic boundary conditions and let Ω_{Ind} be the set of all independent sets of Λ. The lattice has a natural tri-partition, which we color black, white, and gray. Call a face of Λ *empty* if it is not incident to any vertex of I, as illustrated in Figure 19. Call two faces of Λ adjacent if they share at least one vertex. We then define a *fault line* to be a non-contractible cycle of empty faces. We let $\mathcal{F} \subset \Omega_{Ind}$ be the set of all independent sets with at least one fault line.

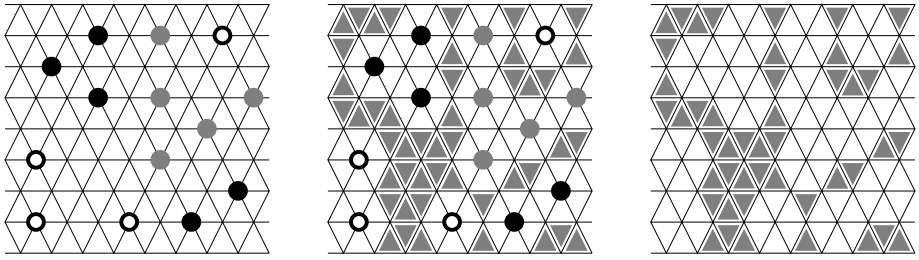

Figure 19: An independent set on Λ and the corresponding empty faces.

The obstruction preventing a fault line must be a set of tightly packed vertices of I.

Call two vertices of an independent set *touching* if they are incident to faces which share an edge. Note that touching vertices must have the same color. We define a monochromatic *bridge* to be a non-contractible cycle of touching vertices of I. For any non-contractible cycle, the *winding vector* is an ordered pair of integers (w_x, w_y), where w_i represents the net number of times the cycle intersects an elementary loop in the ith lattice direction. For instance, the elementary loops have winding vector $(0, 1)$ and $(1, 0)$. We say a configuration has a monochromatic *cross* if it contains two bridges with different winding vectors. (As cycles with different winding vectors must intersect, these two bridges are automatically of the same color.) We let $\mathcal{B}$, $\mathcal{W}$, and $\mathcal{G}$ be the set of independent sets containing black, white, and gray crosses, respectively.

We first show that these sets define a partition of Ω_{Ind}.

Lemma 4.10 *Every independent set in Ω_{Ind} has a fault line or a white, black, or gray cross, but no two of these.*

Proof:

This proof is similar to the argument in [**?**] that shows that independent sets on the Cartesian lattice without fixed boundary conditions must have horizontal and vertical bridges of one color or there must be a fault line. Again, a key point we use here is that any two non-contractible cycles of different winding vectors must intersect.

The sets $\mathcal{B}$, $\mathcal{W}$, and $\mathcal{G}$ are disjoint, as an independent set cannot have two crosses of different colors; that would involve two bridges of different winding vectors and different colors, whose intersection would lead to a contradiction. Similarly, $\mathcal{F}$ is disjoint from $\mathcal{B}$, $\mathcal{W}$, and $\mathcal{G}$, as no set can have both a cross and a fault line; the fault line must intersect at least one of the bridges, which is impossible.

To see that there must be either a cross or a fault line, examine the torus after the removal of a bridge. The remaining space is a non-contractible strip of the torus of the same winding vector as the bridge. If there exists a path of touching vertices across this strip, we find a second bridge of a different winding vector and therefore a cross. This is illustrated in Figure 20. If no such path exists, then there must exist a fault line along the

strip.

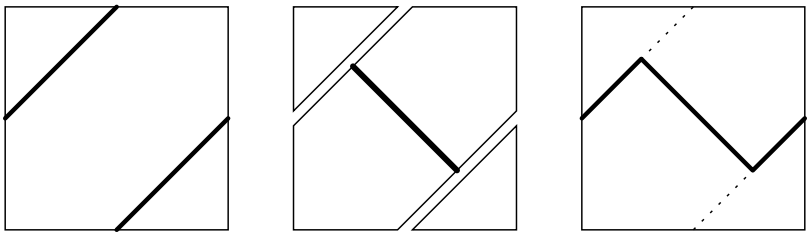

Figure 20: A cycle of winding vector $(1, 1)$, the strip remaining after its removal, and the cycle of winding vector $(0, 1)$ created using the path across the strip.

This proves the lemma. □

We now show that removing $\mathcal{F}$ disconnects $\mathcal{W}$, $\mathcal{B}$, and $\mathcal{G}$. We let $P(\cdot, \cdot)$ be the transition probabilities of $\mathcal{M}_{Ind}$.

Lemma 4.11 *Let $I_1 \in \mathcal{W}$, $I_2 \in \mathcal{B}$ and $I_3 \in \mathcal{G}$ be three independent sets. Then $P(I_i, I_j) = 0$ for all $i \neq j$.*

Proof:

Individual moves of $\mathcal{M}_{Ind}$ either add or remove a single vertex. Clearly it requires multiple moves to eliminate one cross and complete another. □

We now show that the stationary probability of $\mathcal{F}$ is exponentially small. In doing so, we again extend fault lines to 2-dimensional regions. Define a *fat contour* to be a maximally connected set of empty faces containing a fault line. We define a mapping $\psi : \mathcal{F} \to \Omega_{Ind}$ which eliminates at least one fat contour. Although the mapping is not one-to-one, we will show that each $I' \in \text{Img}(\psi)$ has a pre-image whose total weight is exponentially smaller.

To bound the number of sets in this pre-image, we first bound the number of fat contours.

Lemma 4.12 *If F is a fat contour with ℓ faces, then there are at most $2n^2 36^\ell$ choices for the locations of those faces.*

Proof:

We begin by limiting the notion of adjacencies in F. Define two adjacent faces to be *edge-adjacent* if they share an edge. Call them *point-adjacent* if they share a single vertex

and yet are not both adjacent to a common face. As illustrated in Figure 21, not all pairs of adjacent faces are edge- or point-adjacent. However, note that edge- and point-adjacencies suffice to connect F, as a vertex of I removes a complete hexagon from F. We can therefore find a traversal of F that uses only edge- and point-adjacencies.

There are $2n^2$ choices for a face to start the DFS of F. Then each step of the DFS has six possible directions (three edge-adjacencies and three point-adjacencies), so there are at most 6^{2l} possible traversals starting at f. □

Figure 21: Adjacent faces which are point-adjacent (left), edge-adjacent (center), and neither (right).

Our definition of ψ is slightly more complicated than in Section 4.3 because we are considering toroidal regions. Suppose first that F contains two fault lines with different winding vectors. Then the complement of F contains only regions whose boundaries are contractible. By the maximality of F, each of these regions has a monochromatic boundary, so we may refer to these connected components (or "islands") by the colors of their boundaries. Note that if the lattice partition is colored as in Figure 22, if we shift all white islands one space East so that their boundaries become gray, shift all black islands one space to the North-East so that their boundaries also become gray, and leave gray islands as they are, we form a new independent set of the same size. After this shift, all vertices incident to F are gray, so ψ may fill the entire fat contour with gray vertices. If F has ℓ faces, ψ adds exactly $\ell/6$ vertices to I, each the center of a vacant hexagon. Such a transformation is illustrated in Figure 22.

Unfortunately, fat contours need not contain multiple fault lines with differing winding vectors and, indeed the complement of F can contain regions whose boundaries are non-contractible. If this is the case, then F has a bridge on each side. Define a *bulge* to be a maximal set of touching vertices of I which contain a bridge. In this case F must be incident to a bulge on each side. If these are of the same color, ψ can shift all islands within

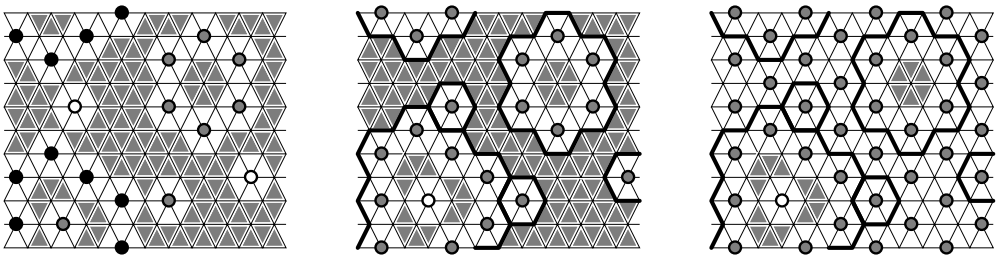

Figure 22: A set σ containing a pair of fault lines with differing winding vectors, and $\psi(\sigma)$.

F to that color and fill F, as in Figure 23.

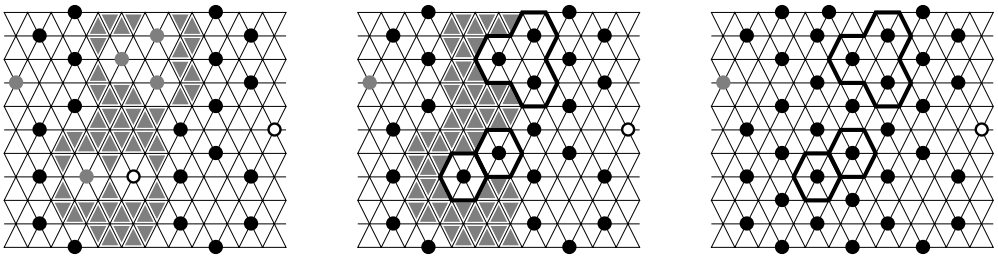

Figure 23: A set σ containing a fault line with a black bulge on each side, and $\psi(\sigma)$.

Complicating matters further, there may be no fat contour incident to two bulges of the same color. For instance, let F_1 be a fat contour incident to a white bulge on the left and a black one on the right. To the right of the black bulge there must be another fat contour, F_2. If F_2 is incident to a white bulge, ψ can shift the islands of F_1 and F_2 and the black bulge all to white. Then ψ can fill both fat contours with white vertices. If F_1 and F_2 contain ℓ_1 and ℓ_2 faces (respectively), then ψ adds $(\ell_1 + \ell_2)/6$ vertices to I.

In one final case, suppose we have no pair of neighboring fat contours bordered by bulges of the same color. Then there must then be a third fat contour F_3 which is incident to still another bulge. Luckily we only have three colors of bulges; at some point these colors must repeat. For example, if we have, in order, a black bulge, F_1, a white bulge, F_2, a gray bulge, F_3, and then a black bulge, then ψ can shift the white bulge, the gray bulge, and all islands of the fat contours to black. We may then fill all three fat contours with black vertices.

To find the inverse of ψ, note that we need only the faces of the fat contour(s); the colors of the neighboring vertices can be inferred from the shape of F, and these colors define the direction of the shift.

We may now bound the stationary probability of $\mathcal{F}$.

Lemma 4.13 *There exist constants $\lambda_0, n_0, c > 1$ such that, if $\lambda > \lambda_0$ and $n > n_0$, then $\pi(\mathcal{F}) < c^{-n}$.*

Proof:

For each $\ell \in [n, 2n^2]$, let $\mathcal{F}_\ell \subset \mathcal{F}$ be the independent sets where the fat contours chosen by ψ contain a total of ℓ faces. Given ℓ_1,ℓ_2,ℓ_3 such that $\ell_1 + \ell_2 + \ell_3 = \ell$, Lemma 4.12 shows that there are at most $\prod_{i=1}^{3} 2n^2 36^{\ell_i} = 8n^6 36^\ell$ choices of faults such that $|F_i| = \ell_i$. Then, for each $I' \in \mathrm{Img}(\psi)$,

$$\begin{aligned}
\pi(\psi^{-1}(I')) &= \sum_{\ell=n}^{2n^2} \sum_{I \in \mathcal{F}_\ell : \psi(I) = I'} \pi(I) \\
&= \sum_{\ell=n}^{2n^2} \sum_{I \in \mathcal{F}_\ell : \psi(I) = I'} \pi(I') \cdot \lambda^{-\frac{\ell}{6}} \\
&< \sum_{\ell=n}^{2n^2} l^3 8n^6 36^\ell \cdot \pi(I') \cdot \lambda^{-\frac{\ell}{6}} \\
&< 64n^{12} \left(\frac{36^6}{\lambda}\right)^{\frac{2n^2}{6}} \pi(I').
\end{aligned}$$

The bound on the pre-image then allows us to bound $\pi(\mathcal{F})$ as follows:

$$\begin{aligned}
\pi(\mathcal{F}) &= \sum_{I' \in \mathrm{Img}(\psi)} \pi(\psi^{-1}(I')) \\
&< \sum_{I' \in \mathrm{Img}(\psi)} 64n^{12} \left(\frac{36^6}{\lambda}\right)^{2n^2} \pi(I') \\
&= 64n^{12} \left(\frac{36^6}{\lambda}\right)^{\frac{2n^2}{6}} \pi(\mathrm{Img}(\psi)) \\
&< 64n^{12} \left(\frac{36^6}{\lambda}\right)^{\frac{2n^2}{6}}.
\end{aligned}$$

Taking $\lambda > 36^6$ completes the proof. □

Note that by symmetry the sets $\mathcal{W}, \mathcal{B}$ and $\mathcal{G}$ have equal stationary probability. Observing now that their total weight is at least $1 - c^{-n}$, Corollary 2.5 allow us to complete the proof

of Theorem 4.1. Thus, we have shown that the Glauber dynamics on independent sets of the triangular lattice converges slowly to equilibrium for large λ.

CHAPTER V

HOW BOUNDARY CONDITIONS CAN AFFECT MIXING RATE

We now examine how the boundary of a graph can influence the mixing rate of a finite Markov chain. The motivation for this question comes from physics, where the boundary represents an external environment, and differences in the boundary are known to strongly influence whether or not there is a phase transition in the underlying infinite model. It is not yet understood whether we see a similar change for the corresponding finite computational models. I.e., can a Markov chain be fast for a region with one boundary and slow with another? We make progress towards answering this question affirmatively in this chapter.

5.1 Introduction

We return again to the Ising model on the $n \times n$ Cartesian lattice. Recall that each configuration σ in the state space $\Omega_{Ising} = \{+,-\}^{n^2}$ consists of an assignment of a $+$ or $-$ spin to each of the vertices, and the Gibbs distribution assigns weight

$$\pi(\sigma) = \lambda^{-D(\sigma)}/Z,$$

where $D(\sigma) = |\{(i,j) \in E \mid \sigma(i) \neq \sigma(j)\}|$ and Z is the partition function.

Physicists study whether there is a unique limiting distribution as $n \to \infty$ to characterize when there is a phase transition in the underlying model. For each value of n, the vertices on the boundary of an $n \times n$ grid are hard-wired to be $+$ in one case, and $-$ in another case, and the Gibbs measure on the interior is defined as the limiting conditional probabilities of these two distributions. It is well known that there is a critical value λ_c, such that for $\lambda < \lambda_c$, the limiting distribution is unique, yet for $\lambda > \lambda_c$, correlations between the spins of vertices inside a finite region and the spins on the boundary persist over long distances, and there are multiple limiting distributions (see, e.g., [**?**]). This appears related to the phase change observed in the context of mixing rates of local chains on finite regions. When λ is sufficiently small, local dynamics are efficient, while when λ is sufficiently large, local chains

require exponential time to converge to equilibrium [**?**]. It is believed that these physical and computational models both have abrupt changes at a critical point, that is the same in the finite and infinite case.

The problems addressed in this chapter examine this conjectured connection. *How much influence can the choice of the boundary have on the mixing rate of a finite Markov chain?* Martinelli, Sinclair and Weitz [**?**, **?**] took a significant step towards exhibiting the influence of the boundary on mixing rates by demonstrating this phenomenon when the underlying graph is a tree, known in physics as the *Bethe-approximation* to the lattice. They left as an important open question whether this also occurs when the underlying graph is a lattice. This is a much more challenging question because lattice regions are *non-amenable* graphs [**?**], i.e. the boundary forms a much smaller proportion of the vertices than on trees. More importantly, the presence of short cycles dramatically increases the difficulty of the analysis.

5.1.1 Our Results

We present the first proof that the choice of boundary can cause an exponential difference in the mixing rate of a local chain defined on the 2-dimensional lattice. We analyze a chain defined by Broder [**?**], used to sample matchings, or sets of edges with no common vertices. To sample perfect and near-perfect matchings (where the edge set covers all or all but two of the vertices), we start at an arbitrary perfect or near-perfect matching σ_0 and add, remove, or trade edges. More precisely, we define the Broder Markov chain as follows.

Given a current perfect or near-perfect matching σ_t, a move of the chain $\mathcal{M}_{Broder}$ is defined by repeating the following steps:

- `Choose` e `uniformly from the edges of the graph.`
- `If both endpoints of` e `are unmatched in` σ_t`, let` $\sigma_{t+1} = \sigma_t \cup \{e\}$.
- `If` $e \in \sigma_t$ `and` σ_t `is perfect, let` $\sigma_{t+1} = \sigma_t \backslash \{e\}$.
- `If exactly one endpoint of` e `is matched in` σ_t`, say in some edge` e'`, let` $\sigma_{t+1} = \sigma_t \cup \{e\} \backslash \{e'\}$.

- `Otherwise let` $\sigma_{t+1} = \sigma_t$.

Jerrum and Sinclair [**?**] showed that proving a polynomial bound on the number of near-perfect matchings for a graph, in terms of the number of perfect matchings on that graph, is sufficient to prove that $\mathcal{M}_{Broder}$ is fast mixing on the perfect and near-perfect matchings of that graph.

Theorem 5.1 *Let $\widehat{\Omega}_{\mathcal{P}}$ be the set of perfect matchings and $\widehat{\Omega}_{\mathcal{N}}$ be the set of near-perfect matchings. If $|\widehat{\Omega}_{\mathcal{N}}| \leq p(n)|\widehat{\Omega}_{\mathcal{P}}|$ for some polynomial p, then Φ is at least inverse-polynomial.*

We study $\mathcal{M}_{Broder}$ on a certain lattice called the "square-octagon lattice," generated by tiling the plane with regular octagons filling the gaps with squares, as in Figure 24. For a given a finite region of this lattice, we specify which boundary vertices are matched internally, thereby remaining part of the finite region, and which are matched externally, in which case they are effectively deleted. We then analyze $\mathcal{M}_{Broder}$ on the remaining vertices. We find that there are two boundary conditions such that the chain is fast mixing with one and slow mixing with the other. More remarkably, these two boundary conditions differ only by the inclusion or exclusion of four vertices! This suggests that the sensitivity to boundary conditions does not necessarily scale with the size of the region and may depend on only a constant number of vertices.

5.1.2 Techniques

The matchings of the lattice can be reformulated as a "contour model" on finite regions of the Cartesian lattice. We show that we can represent matchings as even subgraphs of the Cartesian lattice, with a penalty in Gibbs measure for families with many edges, favoring configurations with a relatively small subgraph. The important feature is that the union of these contours form a subgraph of even degree vertices, except for at most two on the interior. The Broder chain allows us to remove or add an edge in each step, provided we only introduce at most two new odd degree vertices.

In the first (fast) case, the boundary is defined so that initially all vertices, including the boundary, have even degree in the contour representation, and during the simulation

configurations correspond to contours with at most one path. (All others are closed.) We show that the weight of perfect and almost-perfect configurations are polynomially related. Following the canonical path technique introduced by Jerrum and Sinclair [**?**], this suffices to show polynomial mixing for both models.

In the second (slow) case, we introduce four designated vertices on the boundary that have odd degree in the contour representation, forming the endpoints of two paths. As we simulate the Broder-chain, at most two vertices change parity: if there are 6 odd-degree vertices during the simulation, these include all four of the designated boundary vertices. Once one of the initial paths is disconnected, the two pieces tend to shrink. We prove, via sensitive injections and combinatorial inequalities, that once two designated boundary vertices become disconnected, it takes exponential time for them to reconnect in the other pairing. This is sufficient to show slow mixing.

The proofs of slow mixing rely on identifying a small cut set in the state space and appealing to Corollary 2.5. We define an injection that map configurations in the cut to a set of configurations that each have much higher stationary probability. This allows us to conclude that the cut set has exponentially small weight.

5.2 Perfect matchings in the square-octagon lattice

We now demonstrate the effect of this boundary on mixing rates for near-perfect and perfect matchings on the square-octagon lattice. Let $\widehat{L}$ to be the diamond shaped region of the square-octagon lattice that is n squares across, for odd, arbitrarily large n, as illustrated in Figure 24. We will designate the four corners of this diamond shaped region as $\{v_N, v_S, v_E, v_W\}$ for the cardinal directions. Define $\widehat{L}'$ to be the same region of the square-octagon lattice as $\widehat{L}$, but with the exclusion of one vertex from each of corner. This graph is also illustrated in Figure 24. Regions $\widehat{L}$ and $\widehat{L}'$ capture the two boundary conditions we consider.

We let $\widehat{\Omega}$ (resp. $\widehat{\Omega}'$) be the set of perfect and almost-perfect matchings on $\widehat{L}$ (resp. $\widehat{L}'$) and let $\widehat{\mathcal{M}}_{Broder}$ be the Broder-chain on these state spaces. Our main result in this chapter is the following:

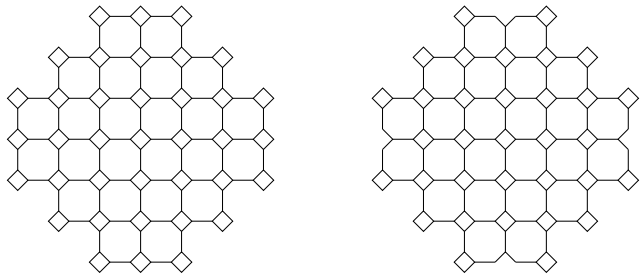

Figure 24: Examples of $\widehat{L}$ (left) and $\widehat{L}'$ (right) for $n = 5$.

Theorem 5.2 *There exist constants $c_1, c_2 > 1$ such that the mixing time of $\widehat{\mathcal{M}}_{Broder}$ on $\widehat{\Omega}$ is $O(n^{c_1})$ while the mixing time of $\widehat{\mathcal{M}}_{Broder}$ on $\widehat{\Omega}'$ is $\Omega(e^{c_2 n})$.*

This theorem provides the first rigorous proof that the choice of boundary can greatly influence the mixing rate of a lattice model.

5.2.1 Contraction

Before presenting the proof of Theorem 5.2, it will be convenient to define a bijection between perfect matchings and a related contour representation. We contract the lattice regions $\widehat{L}$ and $\widehat{L}'$ by replacing each of the squares with vertices, so only the edges bounded on each side by octagons remain. The resulting graph is a region of the Cartesian lattice, as illustrated in Figure 25.

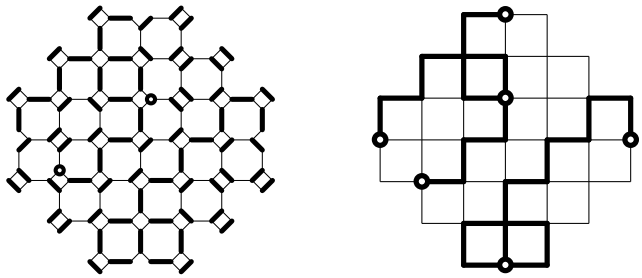

Figure 25: A almost-perfect matching of $\widehat{L}'$ and that matching's contraction.

We use hat notation ($\widehat{L}$, $\widehat{\Omega}$, $\widehat{\mathcal{M}}_{Broder}$, etc.) when referring to the square-octagon graphs, and regular notation (L, Ω, $\mathcal{M}_{Broder}$, etc.) for the contracted case. To that end, define L and L' to be the diamond shaped regions of the cartesian lattice with n vertices across.

Consider the effect of this contraction on a perfect matching of $\widehat{L}$. Figure 26 illustrates all possible matchings of the edges incident to a diamond in $\widehat{L}$, and their contracted form

in L. In all cases, the contraction is an even degree vertex. Furthermore, if that vertex is degree 2, than the endpoints of the incident edges are not colinear in the contracted lattice. In this way the subgraph of L *turns* at each point.

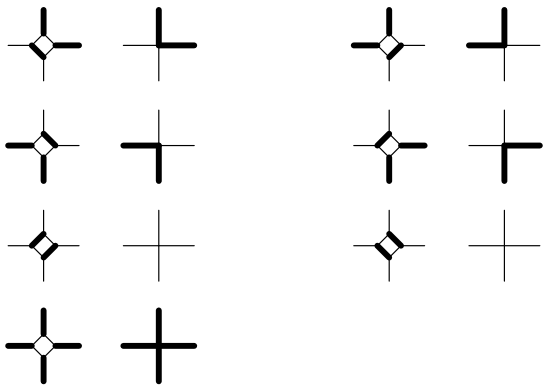

Figure 26: All seven possible matchings of the edges incident to a diamond in $\widehat{L}$, and the contracted form of each in L.

Finally, note that almost-perfect matchings of $\widehat{L}$ and $\widehat{L}'$ contract as perfect matchings do, except for the parity at some 2 vertices, corresponding to the squares containing the original unmatched vertices. These vertices have odd degree in the contracted graph, and are marked with circles in Figure 25.

We define the state spaces Ω and Ω' with this contraction in mind.

Definition 5.1 *We call a subgraph of L a* turning graph *if all vertices have even degree and vertices of degree* 2 *are incident to orthogonal edges. We call a subgraph of L' a* turning graph *if this applies to all vertices but the ones on the corners of the diamond, $\{v_N, v_S, v_E, v_W\}$.*

We call these "turning graphs" since, applying a variation of Euler's Theorem, each component can be decomposed into a set of Eulerian cycles which turn at every step. We call a subgraph of L or L' an *almost-turning graph* if all vertices of degree 2 turn, and exactly two vertices have parity different from what was prescribed for turning graphs. Let $\Omega_{\mathcal{T}}$ and $\Omega_{\mathcal{A}}$ (resp. $\Omega'_{\mathcal{T}}$ and $\Omega'_{\mathcal{A}}$) be the turning and almost-turning graphs of L (resp. L'). Let Ω and Ω' be the union of the turning and almost-turning graphs in each case.

Notice that the uniform distribution on $\widehat{\Omega}$ before the contraction does not correspond

to the uniform distribution on Ω afterwards. Let $G \in \Omega$ and let $v \in L$. As Figure 26 illustrates, if the degree of v in G is 2, 3, or 4, there is a unique way to expand v to a square face of $\widehat{G} \in \widehat{\Omega}$. However, if $d(v) = 0$ or 1, then there are two ways to recover the matched edge(s) that were deleted from the corresponding square. For $G \in \Omega \cap \Omega'$, the number of matchings which contract to G is within a polynomial factor of $2^{|V(L)|-|V(G)|}$. (This polynomial factor is due to the effect of the possible pair of internal vertices with odd degree if $G \in \Omega'$. For clarity we use the $2^{|V(L)|-|V(G)|}$ figure, as our methods need only be tight up to a polynomial factor.)

Our first goal is to show that the Broder-chain is fast on $\widehat{\Omega}$, and for this we need to show a polynomial relationship between the numbers of almost-perfect and perfect matchings. For simplicity, we instead consider their contracted versions in L and introduce a probability distribution π on turning and almost-turning contours as follows: for $G \in \Omega \cup \Omega'$ let $\pi(G) := 2^{-|V(G)|}/Z$, where Z is the partition function. Then $\pi(G)$ is within a polynomial factor of the number of matchings which contract to G. It follows that showing a polynomial relationship between $\sum_{G \in \Omega_{\mathcal{T}}} \pi(G)$ and $\sum_{G \in \Omega_{\mathcal{A}}} \pi(G)$ is sufficient to establish fast mixing of the Broder-chain.

A key lemma used in our proofs of fast and slow mixing bounds how much information is necessary to reconstruct a single contour (or path) if we removed it. Our hope is that the total amount of information is less that the ratio of the weight of the configuration after the contour is removed to the weight of the original contour. The tricky part here is that the measure is in terms of the number of vertices on the contour, while the information to encode the contour is naively related to its length; these two might be incomparable.

Lemma 5.1 *Given $G \in \Omega \cup \Omega'$, let $\mathcal{A}_a(G)$ be the set of almost-turning components A, edge-disjoint from G, such that $|V(G \cup A)| = |V(G)| + a$. Then $|\mathcal{A}_a(G)| \leq 4n^4 2^a$.*

Proof:

Using an Euler-decomposition, the almost-turning component A can be described as a single turning path between the odd-degree vertices, where the direction of each edge is orthogonal to the edge preceding it in the path. If we are given the coordinates of the first

odd vertex and the initial direction of the path, we can encode the rest in a binary string, with 0 representing a left-turn and 1 representing a right-turn. This is the most natural encoding of the turning-path, but requires $|E(A)|$ bits, possibly more than a. It is therefore insufficient to bound the number of almost-turning components in terms of their number of *vertices*, so we must find a different encoding of A which focuses on vertices instead of edges.

The key idea is that not all turns must be encoded; whenever the path touches the graph, either because it turns back on itself or because it touches $G \backslash A$, the next turn is forced, since walks cannot repeat edges. We therefore only encode those turns when our path hits previously-empty vertices, creating a bitstream of length $a - 1$.

Unfortunately, this encoding is not completely unique; after the last recorded turn, the path might proceed for any number of forced-turns; our encoding would fail to represent how many. Luckily, knowing the length of the turning-path would determine this uniquely, and we may surely upper bound the length by $2n^2$ edges. Hence each binary string corresponds to at most a polynomial number of turning-paths.

The total size of $\mathcal{A}_a(G)$ is therefore upper bounded by the number of possible lengths times the number of starting vertices and directions, times the number of binary strings of length $a - 1$. This gives us $|\mathcal{A}_a(G)| \leq 2n^2 \cdot n^2 \cdot 4 \cdot 2^{a-1}$, and the lemma follows. □

5.2.2 Fast Mixing of $\widehat{\mathcal{M}}_{Broder}$ on $\widehat{\Omega}$

For the rapid mixing result of Theorem 5.2, it is sufficient to show a polynomial relationship between the size of $\widehat{\Omega}_{\mathcal{P}}$ and the size of $\widehat{\Omega}_{\mathcal{A}}$. Looking at contractions, this is equivalent to the following:

Lemma 5.2 *For some polynomial p, $\pi(\Omega_{\mathcal{A}}) \leq p(n) \cdot \pi(\Omega_{\mathcal{T}})$.*

Proof:

We define a function $\psi : \Omega_{\mathcal{A}} \to \Omega_{\mathcal{T}}$. For $G' \in \Omega_{\mathcal{T}}$, define the pre-image of G', namely $\psi^{-1}(G') = \{G \in \Omega_{\mathcal{A}} : f(G) = G'\}$. We define ψ in such a way that, although $\psi^{-1}(G')$ contains many graphs, their total weight is within a polynomial factor of the weight of G'.

Let $A(G)$ be the component of G containing the two odd vertices and let $\psi(G) = G \backslash A(G)$. Then, partitioning on the size of $A(G)$, and we see that, for $G' \in \mathrm{Img}(f)$,

$$\begin{aligned}\pi(f^{-1}(G')) &= \sum_{a=1}^{n^2} \sum_{A \in \mathcal{A}_a(G')} \pi(G' \cup A) \\ &= \sum_{a=1}^{n^2} |\mathcal{A}_a(G')| \cdot \pi(G') \cdot 2^{-a} \\ &\leq 16n^4 \pi(G').\end{aligned}$$

where the last inequality is due to Lemma 5.1. Then

$$\begin{aligned}\pi(\Omega_{\mathcal{A}}) &= \sum_{G' \in \mathrm{Img}(f)} \pi(f^{-1}(G')) \\ &\leq 16n^4 \sum_{G' \in \mathrm{Img}(f)} \pi(G') \\ &\leq 16n^4 \pi(\Omega_{\mathcal{T}}).\end{aligned}$$

□

By Theorem 5.1, we can conclude that $\mathcal{M}_{Broder}$ is fast on $\widehat{L}$, verifying the first half of Theorem 5.2.

5.2.3 Slow Mixing of $\widehat{\mathcal{M}}_{Broder}$ on $\widehat{\Omega}'$

We turn now to the behavior of the Broder-chain on $\widehat{L}'$. Recall that a perfect matching of $\widehat{L}'$ contracts to a turning graph of L' that has four odd-degree vertices in each of the corners. These must pair up to form two paths we call *bridges*.

Definition 5.2 *For two corner vertices a and b, define an (a, b) bridge to be a turning path connecting a and b.*

Clearly a perfect matching of $\widehat{L}'$ is contracted to a turning subgraph of L' with an (a, b) and (c, d) bridge, for distinct corners (a, b, c, d). However, almost-perfect matchings of $\widehat{L}'$

may be more complex; if two of the corners each connect to an extra odd vertex, then there may be a bridge only between one pair of corners. If three corners connect to a single odd vertex, then there may be an (a, b) bridge and a (b, c) bridge, but no bridge between any other pair of corners. Examples of these are in Figure 27.

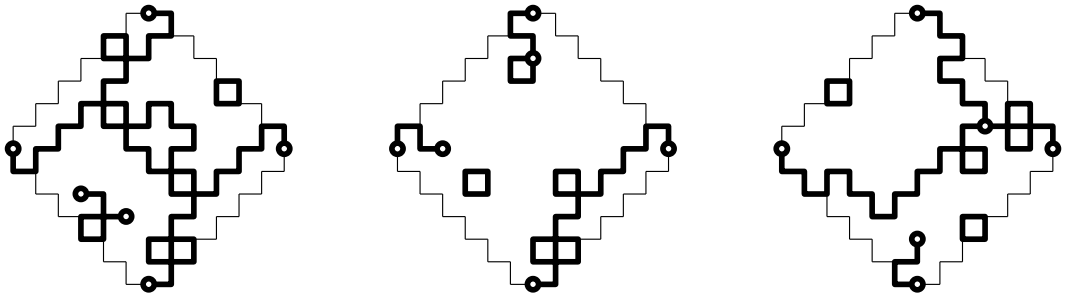

Figure 27: Examples of $\mathcal{H}$ (left), $\mathcal{I}$ (center), and $\mathcal{J}$ (right).

To classify these further, let $\mathcal{H}$ be the set of graphs in Ω' where every corner is in some bridge. For vertices $i, j \in \{v_N, v_S, v_E, v_W\}$, let $\mathcal{I}(i, j)$ be the graphs in Ω' where the only bridges are between i and j. For $i, j, k \in \{v_N, v_S, v_E, v_W\}$, let $\mathcal{J}(i, j, k)$ be the graphs where only corners i, j, and k are in a bridge. Examples of all three of these sets are in Figure 27.

For ease of notation, we let $\mathcal{I} = \cup_{a,b}\mathcal{I}(a, b)$ and $\mathcal{J} = \cup_{a,b,c}\mathcal{J}(a, b, c)$. These will be the three sets for Corollary 2.5. We first show that these sets partition all of Ω', with each $\mathcal{I}(a, b)$ separated from the others by $\mathcal{H}$ and $\mathcal{J}$.

Lemma 5.3 *The state space Ω' is the disjoint union of $\mathcal{H}$, $\mathcal{I}$, and $\mathcal{J}$. Furthermore, for $I_1 \in \mathcal{I}(a, b)$ and $I_2 \in \mathcal{I}(c, d)$, $P(I_1, I_2) = 0$ unless $(a, b) = (c, d)$.*

Proof:

By the definition of $\mathcal{H}$, $\mathcal{I}$, and $\mathcal{J}$, the sets are obviously disjoint. To show that every graph G in Ω' is in one of these sets, we need only show that G has at least one bridge. However, as G has vertices of odd degree at v_N, v_S, v_E, v_W, and at most two other points, there must be a bridge between two corners.

To show $P(I_1, I_2) = 0$, note that to move from a graph in $\mathcal{I}(a, b)$ to a graph in $\mathcal{I}(c, d)$, one bridge must be created and one destroyed; as every move of $\mathcal{M}_{Broder}$ either adds *or* removes an edge, this cannot be accomplished in a single step. □

To show that these sets $\mathcal{J}$ and $\mathcal{H}$ are exponentially small, we will define a function ψ that maps graphs in $\mathcal{J}$ or $\mathcal{H}$ into $\mathcal{I}$. This function removes a long path and then inserts 4-cycles in the resulting empty faces. An example of this transformation is in Figure 28. However, before fully defining ψ, we define some long paths in these graphs and give a technical lemma on the number of empty faces that remain after such paths are removed.

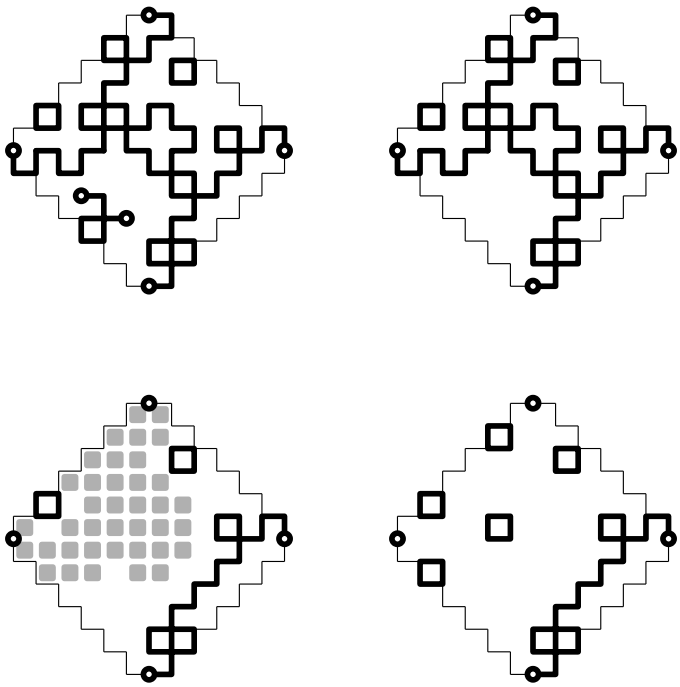

Figure 28: A graph $H \in \mathcal{H}$, $H \backslash A(H)$, $H \backslash A(H) \backslash P(H)$ and $\psi(H)$.

For $H \in \mathcal{H}$, define $P_{NE}(H)$ to be the maximal bridge from v_N to v_E. Define P_{NW}, P_{SE}, and P_{SW} similarly.

For $J \in \mathcal{J}(v_W, v_N, v_E)$, let v_M be the middle vertex of odd degree connected to both v_W and v_E, and let $P_W(J)$ be the maximal turning path from v_M to v_W. Define $P_E(J)$ similarly, from v_M to v_E. Examples of P_W and P_E are in grey on the right and center of Figure 29, respectively.

Definition 5.3 *For a given graph $G \in \Omega'$, path $P \in G$, and face f of L', we call f $\underline{P\text{-available}}$ if f shares a vertex with P but all four edges of f are empty in $G \backslash P$.*

Such a face is adjacent to P, but after ψ removes P, we may insert a 4-cycle on f without violating the turning property. Such faces are represented by gray squares in the

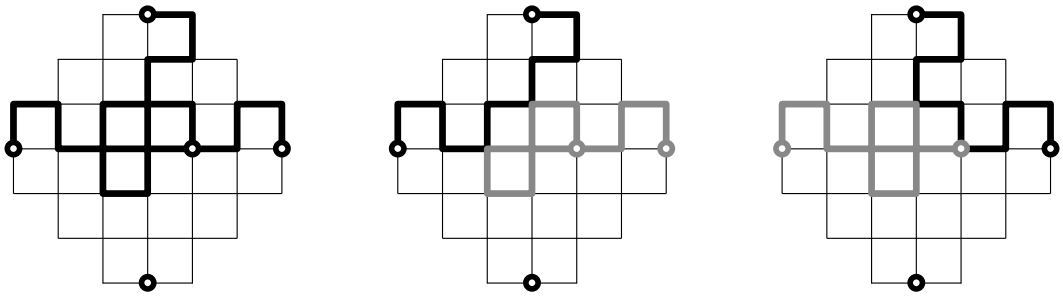

Figure 29: Examples of $H \in \mathcal{H}$ and $P_E(H)$ and $P_W(H)$ (in gray), with $P^-(H)$ (in black).

third part of Figure 28. We will show that for all graphs in $\mathcal{H}$ or $\mathcal{J}$, there is some path with many available faces.

For $H \in \mathcal{H}$, let $P^*(H)$ be the path in $\{P_{NE}, P_{NW}, P_{SE}, P_{SW}\}$ with the most available faces. For $J \in \mathcal{J}(v_W, v_N, v_E)$, let $P^*(J)$ be the path in $\{P_E, P_W\}$ with the most available faces. This path P^* will be the path removed by ψ. Note that whether $G \in \mathcal{H}$ or $G \in \mathcal{J}$, $G \backslash P^*$ contains a single bridge. Call that bridge P^-, illustrated in black in Figure 29.

By the maximal nature of P^*, any component of even-degree vertices touching P^* is included in P^*. The only edges which may touch P^* and not be included are those of P^-. Unfortunately, combined with the boundary of L', P^* and P^- can be constructed to give fewer P^*-available faces than one might expect. Luckily, there is a limit to the number of edges whose adjacent faces can be blocked by both P^- and the boundary.

Definition 5.4 *Call an edge of P^** obstructed *if it is both along the boundary of L' and is within distance* 1 *of P^-. Call it* unobstructed *otherwise.*

The following lemma shows the relationship between unobstructed edges and P^* available faces.

Lemma 5.4 *Given a vertex v in an unobstructed edge of P^*, at least one of the faces containing v is P^*-available.*

Proof:

We divide our analysis between edges whose neighbors in P^* turn in the same direction, and those whose neighbors in P^* turn in opposite directions.

If both neighbors turn the same direction, then there is a face bounded on three sides by P^*. The only way for this face to not be P^*-available is if P^- includes that fourth side, as illustrated on the left of Figure 30. However, if P^- touches P^* on that side, then for parity issues, it cannot touch P^* on the other side. As long as the other side is not outside the boundary of L', there are two P^*-available faces. (The dotted edges in the figure can only contain edges of P^*, by the maximality of P^*.)

Figure 30: Examples of P^* (gray) touching P^- (black). In both cases, if P^* is not along the boundary, there is a P^*-available face.

Therefore if both edges adjacent to an unobstructed edge in P^* turn the same direction, both vertices of that unobstructed edge are incident to a P^*-available face.

If the neighbors turn opposite directions, then there might be one side of P^* outside the boundary of L' or hit by P^-, as on the right of Figure 30. However, on the other side of P^*, there must be a P^*-available face. (Again, the dotted edges can only contain edges of P^*, by the maximality of P^*.) This proves the lemma. □

This allows us to find a large set of P^*-available faces in which we may insert 4-cycles. The next lemma provides a lower-bound on the size of this space.

Lemma 5.5 *There are at least $\frac{n}{8}$ independent P^*-available faces.*

Proof:

By the construction of P^*, it must contain at least n vertices in unobstructed edges. (It takes a turning path $2n$ vertices just to go from v_E to v_W.) By Lemma 5.4, each of these is contained in a P^*-available face. As each face contains at most four vertices of P^*, this implies that P^* has at least $\frac{n}{4}$ available faces, and at least half of these form an independent set. □

Similar to the proof of Lemma 5.2, in proving that $|\mathcal{H}| + |\mathcal{J}|$ is exponentially small, we define a function $\psi : \mathcal{J} \cup \mathcal{H} \to \mathcal{I}$ in such a way that the weight of all of $\psi^{-1}(G)$ is exponentially smaller than the weight of G. The outline of ψ is as follows: First, if $G \in \mathcal{H}$, remove the odd component $A(G)$ if one exists. Then, whether G is in $\mathcal{H}$ or $\mathcal{J}$, remove the long path P^*, leaving only one bridge. Finally, to encode some small segment of P^*, add a subgraph E along the gap P^* left behind. Again, this entire process is illustrated in Figure 28. Note that this final encoding is crucial to the difference between showing that the inverse is *polynomially* smaller, and *exponentially* smaller.

When encoding part of the path, we will be reducing the size of $\psi^{-1}(G')$, by reducing the number possible paths P^* that had been removed from G. However, in adding edges to the graph $\psi(G)$, we unfortunately reduce the weight of $\psi(G)$. Therefore it's crucial that the information gained by encoding the path is larger than the weight lost by adding edges.

Because we are encoding part of P^* in the location of P^* itself, we use only the first half of the available faces. These first $n/16$ empty faces along P^* from Lemma 5.5 encode the final segment of P^* as follows: Break the faces into groups of size 2^5. We will add exactly one facial 4-cycle to each group, encoding the last five bits of the bit-string describing the location of P^*. (For instance, if the last five bits of P^* are 01010, we add a cycle on the $0 \cdot 2^0 + 1 \cdot 2^1 + 0 \cdot 2^2 + 1 \cdot 2^3 + 0 \cdot 2^4 = 10$th available face.) We do this $\frac{n}{16 \cdot 32}$ times. In this way, we encode the last $5n/512$ bits of P^* at a cost of only $4n/512$ additional edges (4 per face). To find P^* given $\psi(G)$, we need only the first $|P^*| - \frac{5n}{512}$ bits encoding P^*.

With ψ fully defined, we are finally able to show that the inverse of $\psi(G)$ is exponentially smaller than $\psi(G)$.

Lemma 5.6 *There exists constant $c > 1$ such that, for all $G' \in \mathrm{Img}(\psi)$,*

$$\pi(\psi^{-1}(G')) \leq 4n^6 \pi(G') c^n.$$

Proof:

As in Section 5.2.2, we partition $\psi^{-1}(G)$ based on the size of the path removed. Let $\psi_\ell^{-1}(G') = \{G : \psi(G) = G', |V(G')| = |G| + \ell\}$.

Given $\psi(G)$, to retrieve G we need only the path P^*. By Lemma 5.1, we need only the starting location and ℓ bits of information to reconstruct P^*, where $|V(G \cup P^*)| = |V(G)| + \ell$. However, by the construction of ψ, given the first $\ell - \frac{5n}{512}$ bits, we may find the final $\frac{5n}{512}$. Therefore

$$|\psi_\ell^{-1}(G')| \leq 4n^4 2^{\ell - \frac{5n}{512}}.$$

On the other hand, we can bound the weight of $G \in \psi_\ell^{-1}(G')$. G is missing at most $\frac{4n}{512}$ vertices by removing the encoding, and then adds ℓ vertices by adding P^*. Therefore $\forall G \in \psi_\ell^{-1}(G')$,

$$\pi(G) \leq \pi(G') \cdot 2^{-(\ell - \frac{4n}{512})}$$

Together, these two inequalities show that

$$\begin{aligned} \pi(\psi_\ell^{-1}(G')) &\leq 4n^4 \pi(G') 2^{-\frac{n}{512}} \\ &< 4n^4 \pi(G') 0.9987^n. \end{aligned}$$

Summing over all ℓ, we prove the lemma. □

We may now prove Theorem 5.2.

Proof:

$$\begin{aligned} \pi(\mathcal{H}) + \pi(\mathcal{T}) &= \sum_{I \in \mathrm{Img}(f) \cap \mathcal{H}_\mathcal{T}} \pi(f^{-1}(I)) + \sum_{I \in \mathrm{Img}(f) \cap \mathcal{H}_\mathcal{A}} \pi(f^{-1}(I)) + \sum_{I \in \mathrm{Img}(f) \cap \mathcal{J}} \pi(f^{-1}(I)) \\ &< c^n \sum_{I \in \mathrm{Img}(f)} \pi(I) \\ &\leq c^{-n} \pi(\mathcal{I}). \end{aligned}$$

This establishes an exponentially small cut in Ω'. By symmetry, $\pi(I(a,b)) = \pi(I(c,d))$ for all a, b, c, d. However, for $\mathcal{M}_{Broder}$ to move from one set to the other, it must pass through the exponentially small set $\mathcal{H} \cup \mathcal{J}$. This bounds conductance, and verifies the second result in Theorem 5.2. □