Combinatorics, Probability and Computing (2004) 00, 000–000.
DOI: 10.1017/S0000000000000000 Printed in the United Kingdom

Disjoint decomposition of Markov chains and sampling circuits in Cayley graphs

Russell Martin[1†] and Dana Randall[2†]

[1] Department of Computer Science
University of Warwick
Coventry CV4 7AL United Kingdom
and
[2] College of Computing
Georgia Institute of Technology
Atlanta, GA 30332-0280 USA

Markov chain decomposition is a tool for analyzing the convergence rate of a complicated Markov chain by studying its behavior on smaller, more manageable pieces of the state space. Roughly speaking, if a Markov chain converges quickly to equilibrium when restricted to subsets of the state space, and if there is sufficient ergodic flow between the pieces, then the original Markov chain also must converge rapidly to equilibrium. We present a new version of the decomposition theorem where the pieces *partition* the state space, rather than forming a cover where pieces overlap, as was previous required. This new formulation is more natural and better suited to many applications. We apply this disjoint decomposition method to demonstrate the efficiency of simple Markov chains designed to uniformly sample circuits of a given length on certain Cayley graphs. The proofs further indicate that a Markov chain for sampling adsorbing staircase walks, a problem arising in statistical physics, is also rapidly mixing.

1. Introduction

Markov chain Monte Carlo methods have become ubiquitous for sampling combinatorial structures. Given a large (finite) set Ω, we define a Markov chain whose transitions consist of small local changes in the structures that allows us to move from one element to any other in Ω. The (possibly directed) graph G with $V(G) = \Omega$ and edges representing the transitions of the chain is called the *Markov kernel*, and the Markov chain performs a random walk on this graph. One popular way to sample from Ω according to a probability distribution π is to use the Metropolis algorithm. Letting Δ denote the largest degree of G, the Metropolis transition probabilities are $P(x,y) = \frac{1}{2\Delta}\min\{1, \frac{\pi(y)}{\pi(x)}\}$ for $x \neq y$ and $(x,y) \in E(G)$. With these transition probabilities the Markov chain is time-reversible and

† Supported in part by NSF Grant No. CCR-0105639 and an Alfred P. Sloan Research Fellowship.

we may view G as an undirected graph where the edge (x, y) has weight $\pi(x)P(x, y) = \pi(y)P(y, x)$. If G is also connected then we are ensured that the chain has π as its stationary distribution, so starting at any state in Ω and performing the random walk on G according to these probabilities, we converge to the distribution π.

For this Markov chain to be useful for sampling it must be that after a small number of transitions the resulting distribution on Ω is close to the desired distribution π. Much recent work has focused on the task of deriving bounds on the *mixing time* of Markov chains, i.e. the number of iterations until the distance between the current distribution and π is smaller than ε, for some $\varepsilon > 0$. A Markov chain is *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 element in Ω.

1.1. Decomposition of Markov chains

Popular methods for bounding the mixing time include coupling and path coupling [2, 4], comparison of Markov chains [3, 13], and bounding the conductance or other isoperimetric constants related to a chain [5, 16]. Each of these methods, while successful in many situations, are specialized and demand strong characteristics of the chain to provide useful bounds. Decomposition is a recent tool that has proven useful in the analysis of Markov chains and can be used in conjunction with any of these other methods [9, 10]. The decomposition method relates the mixing rate of a Markov chain to the mixing rate of simpler chains restricted to each of the pieces and to a measure of the ergodic flow between the pieces. Typically the restrictions of a Markov chain on these pieces are easier to analyze with the above methods than the original chain, so decomposition provides a systematic top-down approach towards bounding the convergence rate of a chain.

The main contribution of this paper is a new version of the decomposition theorem based on a disjoint partition of the state space. Our new technique is similar to the decomposition theorem of Madras and Randall [9, 10], but is more natural and can be simpler to apply. Suppose the state space can be naturally partitioned into sets $\{\Omega_i\}$. Further, suppose that the Markov chain is rapidly mixing when restricted to any of the Ω_i. Finally, suppose that a *projection* (defined in Section 3.2) of these sets is rapidly mixing, suggesting that it is easy to travel from any of the Ω_i to any other Ω_j. Then we can conclude that the original Markov chain is rapidly mixing as well. This is quite similar in spirit to the Madras-Randall result, however their decomposition theorem requires that the $\{\Omega_i\}$ form a *cover* and must in fact have considerable overlaps. We have found that the new theorem is far more natural for several applications. The key idea is that the restrictions and projection chain are often much simpler than the original Markov chain, so that the other more standard methods (like coupling) are easier to apply to these simpler chains. The examples we consider in this paper show that using decomposition allows us to prove rapid mixing for some Markov chains where other established methods fail to show this directly.

1.2. The models

We explore this new decomposition technique by studying Markov chains on the set of closed, fixed length circuits on various Cayley graphs. *Circuits* are walks in a graph

that begin and end at a specified (root) vertex. In our first application, we consider circuits in $\mathbb{Z}^d$ with $2n$ edges that start and end at the origin. We can think of the d-dimensional lattice as the Cayley graph of an infinite abelian group with d generators. Every path in the lattice is a product of generators and their inverses, one per edge, and the words corresponding to different trajectories from the origin to a particular point are equivalent words in the group. Circuits then correspond precisely to the trivial words that are equivalent to the identity in the group.

One method for uniformly sampling circuits is to define a Markov chain on this state space. We study a Markov chain that, in a single step, either transposes two adjacent edges of the walk, replaces an adjacent pair of edges in the a and then a^{-1} direction with a pair in the b, b^{-1} direction, for two generators (or basis vectors) a and b, or does nothing. Both nontrivial transitions can be regarded as "bending" a part of the circuit so as to change the orientation of two neighboring edges. This local Markov chain connects the state space. We examine whether it converges quickly to stationarity by appealing to the disjoint decomposition method.

A second example we study is sampling circuits with $2n$ edges in infinite d-regular trees. The d-ary tree can be interpreted as the Cayley graph of the nonabelian free group with d generators, where every generator is its own inverse. Again circuits are the trivial words in the group.

The approach we use to generate circuits in trees is indirect and based on a relationship between circuits and another combinatorial structure called staircase walks, defined below. We define a Markov chain on the set of staircase walks, and show that it converges quickly to the stationary distribution. In fact, it is possible to sample staircase walks according to specified weights and then assign labels to the edges of the walk corresponding to generators of the nonabelian group; these labeled walks then represent circuits in the d-ary tree and can be sampled uniformly by suitably weighting the walks.

Staircase walks (also called *Dyck paths*) are walks in $\mathbb{Z}^2$ from $(0,0)$ to (n,n) which do not fall below the diagonal $x = y$. Rotating by 45°, they correspond to walks from $(0,0)$ to $(2n,0)$ which take diagonal steps by adding $(1,1)$ or $(1,-1)$ at each step and which never fall below the x-axis (see Figure 1). The number of staircase walks is $C(n)$, the nth Catalan number, which can be calculated exactly, so sampling can be performed recursively without a Markov chain. However, we are also interested in a simple Markov chain on the set of staircase walks as it has proven useful for sampling other combinatorial objects including triangulations [12] and planar matchings [8, 18]. The Markov chain consists of "mountain/valley" flips, defined by choosing $i \in [2n]$ and, if the ith step of the walk is a local optimum (a mountain), inverting it so that it is a local minimum (a valley), or vice-versa.

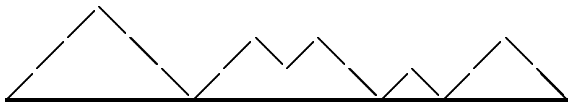

Figure 1. A staircase walk ($n = 9$)

A weighted version, studied in the statistical physics community, assigns weight λ^k to a

staircase walk which hits the x-axis k times[14, 15]. The Gibbs measure normalizes these weights so that the probability of a walk w is $\pi(w) = \frac{\lambda^{k_w}}{\sum_v \lambda^{k_v}}$, where the sum is taken over all walks v and k_v is the number of times v hits the x-axis. When $\lambda = 1$ this is just the uniform probability where $\pi(w) = 1/C(n)$. Taking $\lambda < 1$ favors walks which tend to avoid the x-axis, and taking $\lambda > 1$, known as *adsorbing walks*, favors walks which hit the x-axis many times. It was shown by van Rensburg [14] that there is a phase transition at $\lambda = 2$: when $\lambda < 2$, the walks wander $\mathrm{O}(\sqrt{n})$ away from the x-axis, whereas when $\lambda > 2$ the walks never wander more than $\mathrm{o}(\sqrt{n})$ away.

The mountain/valley Markov chain can be modified to incorporate the weights so that the stationary distribution is the Gibbs measure. The Markov chain for circuits on the d-ary tree turns out to be very closely related to this mountain/valley chain on staircase walks, where we take $\lambda = d/(d-1)$. The circuits arise from labeling the northeast edges of staircase walks with generators of the Cayley graph, representing the possible directions of steps away from the root. The weighting arises from the observation that northeast edges incident to the x-axis must have d choices, corresponding to the d generators representing possible directions away from the root of the d-ary tree, while all other edges have only $d-1$ choices, as one generator will take a step back towards the root and only $d-1$ will take steps away. We define this correspondence more carefully in Section 6.

It is straightforward to show that the mountain/valley chain is rapidly mixing when $\lambda \leq 1$. Wilson [18] gives a tight bound of $\mathrm{O}(n^3(\log n + \log \frac{1}{\varepsilon}))$ when $\lambda = 1$, which also provides an upper bound in the case $\lambda < 1$.

When $\lambda > 1$ a simple coupling argument is insufficient. Informally, for coupling to succeed we need to construct a coupled Markov chain so that close configurations tend to come closer together. However, in the adsorbing case, pairs of walks that differ near the x-axis will tend to diverge initially.

1.3. Outline of the decompositions

In this paper, we show the Markov chains on the set of circuits in $\mathbb{Z}^d$ and the mountain/valley chain for staircase walks (in particular the case $\lambda > 1$) are both rapidly mixing. The disjoint decomposition method is integral in our analysis of these chains.

For circuits in $\mathbb{Z}^d$ we partition the state space $\mathcal{C}$ of all circuits into sets $\mathcal{C}_X$, where $X = (x_1, x_2, \cdots, x_d)$. Each walk in $\mathcal{C}_X$ has exactly x_i positive steps in the direction x_i, for all i. Showing the restriction to each set is rapidly mixing follows using path coupling and is very similar to the analysis of Bubley and Dyer [1] for linear extensions of partial orders.

The projection chain arising from this partition is a random walk on an integer simplex, the collection of non-negative solutions to the integer equation $x_1+x_2+\cdots+x_d = n$, where site $X = (x_1, \cdots, x_d)$ has stationary probability proportional to $\binom{2n}{x_1,x_1,x_2,x_2,\ldots,x_d,x_d}$. Rapid mixing of this projection follows from another application of path coupling. Taken together, these results imply rapid mixing of the original chain on all of $\mathcal{C}$.

The analysis of the mountain/valley chain on staircase walks is more complicated. We first decompose the state space $\mathcal{S}$ into $\bigcup \mathcal{S}_k$, where $\mathcal{S}_k$ is the set of staircase walks which hit the x-axis exactly k times. First we show that $|\mathcal{S}_k|$ is log–concave in k. This immediately implies that the projection (according to the decomposition theorem) is

mixing in polynomial time, so it suffices to show that the Markov chain restricted to $\mathcal{S}_k$, $\mathcal{M}_{\mathcal{S}_k}$, is rapidly mixing for each k.

To show that $\mathcal{M}_{\mathcal{S}_k}$ is rapidly mixing, we apply the decomposition theorem a second time. This time we partition the state space (i.e., the set of staircase walks which hit the x-axis exactly k times) into $\binom{n-1}{k}$ sets according to which k points on the x-axis the paths hit. The conditional distribution on elements of each of these $\binom{n-1}{k}$ sets is uniform. Hence, showing that $\mathcal{M}$ is rapidly mixing when restricted to any of these sets is straightforward and follows from the unbiased case when $\lambda = 1$.

The projection arising from the second decomposition can be viewed as an interesting particle process: we want to sample from the $\binom{n-1}{k}$ ways to place k particles on the x-axis between 0 and n so that each configuration occurs with probability proportional to $\Pi_i C(x_i)$, where x_i is the length of the gap between the ith and $i+1$st particles, and $C(n)$ is the nth Catalan number. The Markov chain arising from this projection is quite natural: choose $(i, d) \in [k] \times \{\ell, r\}$ and move the ith particle in the left or right direction according to d with the appropriate Metropolis probability.

This particle process isolates the difficulty in the original mountain/valley Markov chain on weighted staircase walks; indeed a simple path coupling argument fails on this simpler chain as well for an analogous reason. The final step of our analysis is noticing that a heat bath algorithm is easy to define for this particle system which does have the desirable properties. Namely, it converges to the correct stationary distribution and, moreover, we can show it mixes in polynomial time. By a standard comparison argument we can show that the mixing rate of the heat bath algorithm is close to the mixing rate of the original particle process. Hence, we can apply the decomposition theorem for a second time, thereby establishing the polynomial-time mixing rate of the original mountain/valley chain on the entire state space of staircase walks, as desired.

Our overall goal for this paper is to describe this new decomposition theorem and show how it may be applied. It should be noted that there are alternative, recursive methods for sampling circuits in Cayley trees (and, more generally, adsorbing staircase walks) based on generating functions. The objective of this paper is to provide insight into these very natural local Markov chains and ways to analyze their mixing rates. Our interest stems primarily from their similarity to many other local Markov chains used when analytical methods are not available. We also note that we have not tried to optimize the bounds on the mixing rates of these Markov chains.

1.4. Related work

In the time period between the appearance of the preliminary version [11] of this work and the preparation of this newer, expanded version, other decomposition results have appeared. In particular, using an inductive type of argument, Jerrum and Son [6] have given a better upper bound on the mixing time for the random walk on the basis exchange graph of a balanced matroid. That chain is an instance of what the authors call a π-recursive chain, and their argument establishes a bound on both the spectral gap and log-Sobolev constant for this type of Markov chain.

Building on that work, Jerrum, Son, Tetali, and Vigoda [7] give a general decomposition result for bounds on the spectral gap and log-Sobolev constants using similar restriction

and projection chains as those defined in this paper. Their result, however, relies on considering the Dirichlet form that defines the spectral gap, and the entropy-like quantity that appears in log-Sobolev inequalities. Interested readers are invited to consult those references for details on their methods and the applications they consider.

2. Mixing machinery

In what follows, we assume that $\mathcal{M}$ is an ergodic (i.e. irreducible and aperiodic), reversible Markov chain with finite state space Ω, transition probability matrix P, and stationary distribution π.

The time a Markov chain takes to converge to its stationary distribution, the mixing time of the chain, is measured in terms of the distance between the distribution at time t and the stationary distribution. Letting $P^t(x,y)$ denote the t-step probability of going from x to y, the *total variation distance* at time t is

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

For $\varepsilon > 0$, the *mixing time* $\tau(\varepsilon)$ is

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

We say a Markov chain is *rapidly mixing* if the mixing time is bounded above by a polynomial in n and $\log \frac{1}{\varepsilon}$, where n is the size of each configuration in the state space.

It is well known that the mixing rate is related to the *spectral gap* of the transition matrix P. We let $Gap(P) = \lambda_0 - |\lambda_1|$ denote the spectral gap, where $\lambda_0, \lambda_1, \ldots, \lambda_{|\Omega|-1}$ are the eigenvalues of P and $1 = \lambda_0 > |\lambda_1| \geq |\lambda_i|$ for all $i \geq 2$. The following result relates the spectral gap with the mixing time of the chain (see, e.g., [16]):

Theorem 2.1. *Let $\pi_* = \min_{x \in \Omega} \pi(x)$. For all $\varepsilon > 0$ we have*

(a) $\tau(\varepsilon) \leq \frac{1}{Gap(P)} \log(\frac{1}{\pi_* \varepsilon})$.

(b) $\tau(\varepsilon) \geq \frac{|\lambda_1|}{2Gap(P)} \log(\frac{1}{2\varepsilon})$.

Remark. For simplicity, we typically add self-loops with probability $\frac{1}{2}$ to each point in the state space, ensuring that $\lambda_1 \geq 0$ as well as the aperiodicity of the Markov chain.

We give a brief review of some of the techniques that are used to bound the mixing time (or spectral gap) of a Markov chain, before introducing our new method in Section 3.2.

2.1. Path coupling

A *coupling* is a new Markov chain on $\Omega \times \Omega$ with the following properties: Instead of updating the pair of configurations independently, the coupling updates them so that the two processes will tend to coalesce, or "move together" under some measure of distance, but each process, viewed in isolation, is just performing transitions of the original Markov chain $\mathcal{M}$. Also, once the pair of configurations agree at some time, the coupling guarantees

they agree from that time forward. The mixing time can be bounded by the expected time for configurations to coalesce under any valid coupling.

More simply, *path coupling* lets us bound the mixing time by analyzing a subset of $\Omega \times \Omega$. The method of path coupling is described in the next theorem, adapted from [4]. We use the notation (x', y') for the configuration obtained from the pair (x, y) after one step of the coupling.

Theorem 2.2. [4] *Let Φ be an integer valued metric defined on $\Omega \times \Omega$ taking values in $\{0, \ldots, D\}$. Let U be a subset of $\Omega \times \Omega$ such that for all $(x, y) \in \Omega \times \Omega$ there exists a path $x = \omega_0, \omega_1, \ldots, \omega_k = y$ between x and y such that $(\omega_i, \omega_{i+1}) \in U$ for $0 \leq i < k$ and*

$$\sum_{i=0}^{k-1} \Phi(\omega_i, \omega_{i+1}) = \Phi(x, y).$$

Define a coupling $(x, y) \to (x', y')$ of $\mathcal{M}$ on all pairs $(x, y) \in U$. Suppose that there exists $\alpha \leq 1$ such that $E(\Phi(x', y')) \leq \alpha \Phi(x, y)$ for all $(x, y) \in U$.

(a) If $\alpha < 1$ then the mixing time of $\mathcal{M}$ satisfies

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

(b) If $\alpha = 1$ (alternatively, $E(\Phi(x', y') - \Phi(x, y)) \leq 0$) and there exists $\eta > 0$ such that $Pr[\Phi(x', y') \neq \Phi(x, y)] \geq \eta$ for all $(x, y) \neq (x', y')$, then the mixing time satisfies

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

Remark. In the above theorem, it is assumed that the pairwise coupling on U may be extended to a *complete coupling* (or a *grand coupling*), where at each time step there is a random function F_t, defined on the whole state space, such that $x' = F_t(x)$ and $y' = F_t(y)$. For example, in a simple random walk on $[n] = \{1, 2, \ldots, n\}$, the rule may be to flip a coin showing 'heads' and 'tails' with equal probability and try to move left or right depending on if heads or tails appears on the coin. In this case, it may not necessarily be true that there is an η in part (b) of the theorem that bounds the change in distance away from zero. In order to circumvent this problem, as will be necessary to apply the path coupling lemma later on, we may utilize the following trick at the cost to the running time of only a small constant factor. Before each iteration of the chain, for each state in Ω, we flip another coin that shows heads and tails with equal probability, and the coin flip at each site is independent of all the others. If the coin shows heads, we perform the transition as usual, otherwise we stay at that state. Using this rule (or a variation of it), we can ensure the existence of $\eta > 0$, so that (b) applies to obtain a bound on the mixing time.

2.2. The comparison method

The comparison method [3, 13] is useful for relating the mixing rates of two similar Markov chains. It is powerful in cases where it is easier to analyze a Markov chain if

some auxiliary moves are added or some transition probabilities are altered or amplified, but then it is necessary to derive bounds on the original Markov chain.

Let $\widetilde{P}$ and P two reversible Markov chains on the same state space Ω with the same stationary distribution π. It is assumed that the mixing time, $\tau_{\widetilde{P}}(\varepsilon)$, of $\widetilde{P}$ is known (or a suitable bound) and we desire to obtain a bound for the mixing time, $\tau_P(\varepsilon)$, of P.

Let $E(P) = \{(x,y) : P(x,y) > 0\}$ and $E(\widetilde{P}) = \{(x,y) : \tilde{P}(x,y) > 0\}$ denote the sets of edges of the two chains, viewed as directed graphs. For each x, y with $\widetilde{P}(x,y) > 0$, define a *path* γ_{xy} using a sequence of states $x = \omega_0, \omega_1, \ldots, \omega_k = y$ with $P(\omega_i, \omega_{i+1}) > 0$, and let $|\gamma_{xy}|$ denote the length of the path. Let $\Gamma(z,w) = \{(x,y) \in E(\widetilde{P}) : (z,w) \in \gamma_{xy}\}$ be the set of paths that use the transition (z,w) of P. Finally, define

$$A = \max_{(z,w)\in E(P)} \left\{ \frac{1}{\pi(z)P(z,w)} \sum_{\Gamma(z,w)} |\gamma_{xy}| \pi(x) \widetilde{P}(x,y) \right\}.$$

The following is from [13, Thms. 3 & 4] and closely follows the work in [3]:

Theorem 2.3. *With the above notation, we have*

(a) $Gap(P) \geq \frac{1}{A} Gap(\widetilde{P})$.

(b) Assuming that $\lambda_1(\widetilde{P}) \geq \frac{1}{2}$, *for* $0 < \varepsilon < 1$ *we have that*

$$\tau_P(\varepsilon) \leq \frac{4\log(1/(\varepsilon\pi_*))}{\log(1/(2\varepsilon))} A \tau_{\widetilde{P}}(\varepsilon),$$

where $\pi_* = \min_{x\in\Omega} \pi(x)$.

Another useful lemma relates the gap of two chains more directly. This result follows immediately from the "functional definition" of the spectral gap (e.g. see [10, Eq. (7)]):

Lemma 2.4. *Let* P *and* $\widetilde{P}$ *be Markov chains on the same state space, each reversible with respect to the distribution* π. *Suppose there are constants* c_1 *and* c_2 *such that* $c_1 P(x,y) \leq \widetilde{P}(x,y) \leq c_2 P(x,y)$ *for all* $x \neq y$. *Then* $c_1 Gap(P) \leq Gap(\widetilde{P}) \leq c_2 Gap(P)$.

2.3. The decomposition method

The Madras-Randall *decomposition method* [10] offers a different approach for bounding the mixing time of a Markov chain and will be the main motivation behind our analysis in this paper. The intuition behind this method is that we look at subsets of the state space and show that the Markov chain restricted to each subset is mixing. Then, if the sets overlap enough (and cover all of Ω), we can deduce a bound on the mixing rate of the original chain on the entire state space.

Following [10], let $\Omega_1, \ldots, \Omega_m$ be subsets of Ω such that $\cup_i \Omega_i = \Omega$. We are interested in two classes of induced Markov chains. The first is a set of *restricted* Markov chains, obtained by restricting $\mathcal{M}$ to each subset Ω_i, i.e., any move of $\mathcal{M}$ that would take us from an element $x \in \Omega_i$ to some $y \notin \Omega_i$ is rejected. In particular, the restriction to Ω_i is a Markov chain, $\mathcal{M}_i$, where the transition matrix P_{Ω_i} is defined as follows: If $x \neq y$ and $x, y \in \Omega_i$ then $P_{\Omega_i}(x,y) = P(x,y)$; if $x \in \Omega_i$ then $P_{\Omega_i}(x,x) = 1 - \sum_{y\in\Omega_i, y\neq x} P_{\Omega_i}(x,y)$.

The second Markov chain is the *projection* $\mathcal{M}_H$ of the cover $\{\Omega_1, \ldots, \Omega_m\}$, defined on the set $[m] = \{1, \ldots, m\}$, where each point i is associated with the set Ω_i. Let $\Theta = \max_{x \in \Omega} |\{i : x \in \Omega_i\}|$. The transition matrix P_H for Markov chain $\mathcal{M}_H$ is defined by letting $P_H(i,j) = \frac{\pi(\Omega_i \cap \Omega_j)}{\Theta \pi(\Omega_i)}$ for $i \neq j$, and $P_H(i,i) = 1 - \sum_{j \neq i} P_H(i,j)$. The limiting distribution ρ of this chain is given by $\rho(i) = \pi(\Omega_i)/\hat{Z}$, where $\hat{Z} = \sum_i \sum_{x \in \Omega_i} \pi(x) \leq \Theta$. From [10] we have

Theorem 2.5. [10] *In the preceding framework,*

$$Gap(P) \geq \frac{1}{\Theta^2} Gap(P_H) \min_{i=1,\ldots,m} Gap(P_{\Omega_i}).$$

3. A new decomposition result

Our goal is to give a method, analogous to that in Section 2.3, but using a partition of Ω into *disjoint* pieces. We relate the spectral gap of the original chain to the spectral gap of the restriction to each set in the partition, and that of a new projection of this partition. We first briefly introduce the framework that Caracciolo, Pelisetto, and Sokal (CPS) use in the context of simulated tempering (see [10] for further details).

3.1. The CPS tempering method

Let P denote a transition matrix of a Markov chain on the finite state space Ω that is reversible with respect to the probability distribution π. Suppose that the state space is partitioned into m disjoint pieces $\Omega_1, \ldots, \Omega_m$. For each $i = 1, \ldots, m$, define P_{Ω_i}, the restriction of P to Ω_i, by rejecting jumps that leave Ω_i (as in Section 2.3). Let π_i be the normalized restriction of π to Ω_i, i.e., $\pi_i(A) = \frac{\pi(A \cap \Omega_i)}{b_i}$ where $b_i = \pi(\Omega_i)$. Let Q be another transition matrix that is also reversible with respect to π. Define $\overline{Q}$ to be the following aggregated transition matrix on the state space $[m]$:

$$\overline{Q}(i,j) = \frac{1}{b_i} \sum_{\substack{x \in \Omega_i, \\ y \in \Omega_j}} \pi(x) Q(x,y).$$

We note that

$$b_i \overline{Q}(i,j) = b_j \overline{Q}(j,i)$$

so $\overline{Q}$ is reversible with respect to the probability measure $b = (b_1, \ldots, b_m)$ on $[m]$.

Theorem 3.1 ([10], Thm A.1). *Assume Q is positive semi-definite. Let $Q^{1/2}$ denote the nonnegative square root of Q. Then*

$$Gap(Q^{1/2} P Q^{1/2}) \geq Gap(\overline{Q}) \min_{i=1,\ldots,m} Gap(P_{\Omega_i}).$$

3.2. Disjoint decomposition

We use Theorem 3.1 to derive a bound on the spectral gap of P. Note that we assume the eigenvalues of P are all non–negative. Recalling our previous remark following Theorem 2.1, we assume that $P(x,x) \geq \frac{1}{2}$ for all $x \in \Omega$. This ensures that the eigenvalues of P are non-negative.

Theorem 3.2. *Let P_{Ω_i} be as above, and let $\overline{P}$ be defined as above with P in place of Q. Then*

$$Gap(P) \geq \frac{1}{2} Gap(\overline{P}) \min_{i=1,\ldots,m} Gap(P_{\Omega_i}).$$

Proof. Take $Q = P$ in Theorem 3.1 above. This gives the inequality

$$Gap(P^2) \geq Gap(\overline{P}) \min_{i=1,\ldots,m} Gap(P_{\Omega_i}). \tag{3.1}$$

Note that the eigenvalues of P^2 are the squares of the eigenvalues of P. Since $Gap(P) = 1 - \lambda_1$ and $Gap(P^2) = 1 - \lambda_1^2$, we find that $Gap(P) = 1 - \sqrt{1 - Gap(P^2)}$.

To complete the proof, consider the function $1 - \sqrt{1-x}$. The Taylor series of this function is

$$1 - \sqrt{1-x} \;=\; \frac{1}{2}x + \frac{1}{8}x^2 + \frac{1}{16}x^3 + \frac{5}{128}x^4 + \cdots. \tag{3.2}$$

From (3.2) we see that $1 - \sqrt{1-x} \geq \frac{1}{2}x$ for $x \geq 0$. Taking $x = Gap(P^2)$ gives us the inequality $Gap(P) = 1 - \sqrt{1 - Gap(P^2)} \geq \frac{1}{2} Gap(P^2)$. Multiplying (3.1) by $\frac{1}{2}$, we obtain the result. □

Remark. We may justify taking $Q = P$ in Theorem 3.1 by working in the inner product space $\ell^2(\pi)$, the set of all square summable functions on Ω. In this setting, the inner product for two functions f and g is defined as

$$\langle f, g \rangle = \sum_{x \in \Omega} f(x) g(x) \pi(x).$$

Since P is the transition matrix of a reversible Markov chain, it defines a self-adjoint operator on $\ell^2(\pi)$, i.e.,

$$\langle Pf, g \rangle = \langle f, Pg \rangle \quad \forall f, g \in \ell^2(\pi)$$

where $Pf(x) = \sum_{y \in \Omega} P(x,y) f(y)$. Thus, in this space, P is a positive semi–definite matrix (operator), $P^{1/2}$ is well–defined, and we can apply Theorem 3.1.

We also derive a useful corollary. One difficulty of directly applying Theorem 3.2 is that we must find (or bound) the spectral gap of the aggregate chain $\overline{P}$. Suppose instead, keeping the same Markov kernel, we replace the matrix $\overline{P}$ with a transition matrix P_M on the set $\{1, \ldots, m\}$, with Metropolis transitions, i.e.,

$$P_M(i,j) = \frac{1}{2\Delta} \min\{1, \frac{\pi(\Omega_j)}{\pi(\Omega_i)}\}.$$

In the expression for the transition probabilities, Δ is the maximum degree of the vertices in the Markov kernel of $\overline{P}$, i.e. the maximum degree in the graph having vertex set $[m]$ and edges (i,j) where $\overline{P}(i,j) > 0$. In our corollary, the essential idea is that the sets $\{\Omega_i\}$ have large boundaries, so that we can cross from one set to another efficiently. To formalize this, let

$$\partial_i(\Omega_j) = \{y \in \Omega_j : \exists x \in \Omega_i \text{ with } P(x,y) > 0\}.$$

Note that $\partial_i(\Omega_j)$ denotes the set of all elements in Ω_j that can be obtained from some element of Ω_i by one step of the Markov chain $\mathcal{M}$.

Corollary 3.3. *With P_M defined as above where Δ is the maximum degree of the Markov kernel (of this projection chain), suppose there exists $\beta > 0$ and $\gamma > 0$ such that*

(a) $P(x,y) \geq \beta$ for all $x \sim y$ in P.

(b) $\pi(\partial_i(\Omega_j)) \geq \gamma\pi(\Omega_j)$ for all pairs $i \sim j$ in the Markov chain defined by $\overline{P}$.

Then

$$Gap(P) \geq \beta\gamma\Delta\, Gap(P_M) \min_{i=1,\ldots,m} Gap(P_{\Omega_i}).$$

Proof. Note that

$$\begin{aligned}
\sum_{\substack{x\in\Omega_i,\\ y\in\Omega_j}} \pi(x)P(x,y) &= \sum_{\substack{x\in\partial_j(\Omega_i),\\ y\in\partial_i(\Omega_j)}} \pi(x)P(x,y) \\
&= \sum_{\substack{x\in\partial_j(\Omega_i),\\ y\in\partial_i(\Omega_j)}} \pi(y)P(y,x) \qquad (3.3)\\
&\geq \sum_{y\in\partial_i(\Omega_j)} \pi(y)\beta \qquad (3.4)\\
&\geq \beta\gamma\pi(\Omega_j) \qquad (3.5)
\end{aligned}$$

where (3.3) follows from reversibility and inequalities (3.4) and (3.5) follow from conditions (a) and (b). Multiplying by $\frac{1}{\pi(\Omega_i)}$, we see

$$\overline{P}(i,j) = \frac{1}{\pi(\Omega_i)} \sum_{\substack{x\in\Omega_i,\\ y\in\Omega_j}} \pi(x)P(x,y) \geq \beta\gamma\frac{\pi(\Omega_j)}{\pi(\Omega_i)} \geq 2\beta\gamma\Delta P_M(i,j).$$

Therefore, by Lemma 2.4, we have $Gap(\overline{P}) \geq 2\beta\gamma\Delta\, Gap(P_M)$. Using this bound on $Gap(\overline{P})$ in Theorem 3.2 gives us the result. □

Remark. This corollary may indeed be thought of as one of many possible corollaries of Theorem 3.2. For example, condition (a) of the corollary may appear exceedingly restrictive. It might happen that there are transitions of the chain for which the transition probabilities are exponentially small. This would mean that β is exponentially small, and the corollary would not give a useful lower bound on the spectral gap of P. In these cases, we might choose to disregard these small transition probabilities (with a possible slight alteration to the value of γ) to obtain better results than this general corollary may initially supply us.

Similarly, it may be possible that a more careful analysis, or insight into the Markov kernel of the chain, might lead to a better result in passing from (3.3) to inequality (3.4) in the proof of the corollary, thereby obtaining a tighter lower bound on the sum in (3.3). For example, suppose we replace condition (a) with a stronger condition of the form "For all $i,j \in [m], i \neq j$ and all $y \in \partial_i(\Omega_j)$, we have $\sum_{x\in\partial_j(\Omega_i)} P(y,x) \geq \beta$." This condition

might be helpful in passing from (3.3) to (3.4) to get a better lower bound on the spectral gap of P.

4. Sampling circuits in $\mathbb{Z}^d$

Our first application of the disjoint decomposition method of Section 3.2 is that of sampling circuits in $\mathbb{Z}^d$. We fix a vertex in $\mathbb{Z}^d$, which we call the *origin*, and wish to sample, uniformly at random, from the set of *circuits* in $\mathbb{Z}^d$ with $2n$ edges, i.e., closed walks of length $2n$ that begin and end at the origin.

We may represent each circuit in $\mathbb{Z}^d$ as a string of letters using the symbols $a_1, \ldots, a_d$ and $a_1^{-1}, \ldots, a_d^{-1}$ where a_i denotes a unit step in the positive direction of the ith coordinate and a_i^{-1} a unit step in the corresponding negative direction. Note that since we consider circuits in $\mathbb{Z}^d$, the number of occurrences of a_i in the string equals the number of occurrences of a_i^{-1} for all i.

Fix $n \geq 1$ and let $\mathcal{C}$ denote the set of all circuits with $2n$ edges. We note that $|\mathcal{C}| = \sum \binom{2n}{x_1, x_1, x_2, x_2, \ldots, x_d, x_d}$ where the sum is over all non–negative solutions to $x_1 + x_2 + \cdots + x_d = n$.

4.1. The Markov chain on circuits in $\mathbb{Z}^d$

The Markov chain $\mathcal{M}$ on $\mathcal{C}$ consists of *transpositions* which swap neighboring elements of the string, and *rotations* that change a neighboring pair of elements a_i, a_i^{-1} to a pair a_k, a_k^{-1}. Let $\sigma \in \mathcal{C}$ denote a circuit, and σ_i, $i \in [2n]$, the elements of σ (in the string representing σ). In what follows, we use notation like "$t \in_u T$" to mean that t is selected *uniformly at random* from the set T. The Markov chain on $\mathcal{C}$ is:

One step of Markov chain $\mathcal{M}$:

1. Pick $t \in_u \{1, \ldots, 2n-1\}, b \in_u \{0, 1\}$.
2. - If $b = 0$, then if $\sigma_t = a_i$ and $\sigma_{t+1} = a_i^{-1}$ for some $i \in \{1, \ldots, d\}$, then choose $k \in_u \{1, \ldots, d\}$ and with probability $1/2$ set $\sigma_t = a_k$ and $\sigma_{t+1} = a_k^{-1}$.
   - If $b = 1$, then with probability $1/2$ transpose σ_t and σ_{t+1}.
   - In all other cases, do nothing.

The procedure outlined above shows that the transition probabilities for $\mathcal{M}$ are:

$$P(\sigma_1, \sigma_2) = \begin{cases} \frac{1}{4d(2n-1)} & \text{if } \sigma_1, \sigma_2 \text{ differ by a rotation,} \\ \frac{1}{4(2n-1)} & \text{if } \sigma_1, \sigma_2 \text{ differ by a transposition,} \\ 1 - \sum_{\sigma' \neq \sigma_1} P(\sigma_1, \sigma') & \text{if } \sigma_1 = \sigma_2. \end{cases} \tag{4.1}$$

This Markov chain is ergodic and aperiodic (there are self-loops at each state with probability at least $1/2$). Also, since the transition probabilities are symmetric, we easily see that the limiting distribution π is uniform on $\mathcal{C}$.

4.2. Bounding the mixing time of $\mathcal{M}$

We apply the disjoint decomposition method of Section 3.2 to obtain a bound on the mixing time of $\mathcal{M}$. The partition $\mathcal{C}$ is based on the number of steps in each coordinate direction. More precisely, let $\sigma \in \mathcal{C}$ be a circuit and let x_i denote the number of times a_i (and hence a_i^{-1}) appears in σ. The *trace*, $tr(\sigma)$, is defined as $tr(\sigma) = (x_1, x_2, \ldots, x_d)$. Let $\widetilde{\mathcal{C}}$ denote the set of all possible traces, i.e., the set of possible solutions in non–negative integers to $x_1 + x_2 + \cdots + x_d = n$. We see that $|\widetilde{\mathcal{C}}| = \binom{n+d-1}{n}$ since this is the number of distinct traces. With $X = (x_1, \ldots, x_d)$ denoting a trace, we use $\mathcal{C}_X$ to denote the subset of $\mathcal{C}$ where each element has trace X. The collection of restricted Markov chains $\mathcal{M}_X$, with transition matrices P_X, one for each set $\mathcal{C}_X$, only allow transpositions of letters in the string σ, keeping the number of each a_i (and a_i^{-1}) fixed. The projection chain can be represented by an integer simplex where each vertex represents a trace in $\widetilde{\mathcal{C}}$ and neighbors in the chain are vertices whose traces have an ℓ^1 distance of 2.

4.2.1. The restrictions We first examine the mixing time of the restricted chains $\mathcal{M}_X$. We shall see that the mixing time of the restrictions follows from previous work by Bubley and Dyer [1] on sampling linear extensions of partial orders.

The set of circuits with fixed trace X can be seen to be the number of ways of interleaving x_i copies of a_i and a_i^{-1}, for each i. Up to multiplicities, the set of circuits is just the set of permutations on this multiset. To account for multiplicities, however, we can think of the set of x_i copies of a_i as x_i distinct elements $a_{i,1} < a_{i,2} < \ldots < a_{i,x_i}$ and the copies of a_i^{-1} as $A_{i,1} < \ldots < A_{i,x_i}$. Let T_X be the set of total linear orders consistent with this set of partial orders. The following lemma is easy to verify:

Lemma 4.1. *There is a bijection between $\mathcal{C}_X$, the set of circuits with trace X, and T_X, the set of total orders.*

Moreover, the restricted Markov chain $\mathcal{M}_X$ performs a walk on the state space T_X of total orders by performing random transpositions (with at least a $3/4$ self-loop probability at each state because rotations are disallowed in the restricted chain). Such a chain was studied previously by Bubley and Dyer [1]. Together with part (b) of Theorem 2.1, we get this result:

Theorem 4.2. *The transposition Markov chain P_X has mixing time* $\mathrm{O}(n^3(\log n + \log \varepsilon^{-1}))$. *Consequently we have $Gap(P_X) \geq \frac{c}{n^3}$ for some constant c (for all possible traces X).*

Proof. For the proof of the mixing time, the reader is referred to [1]. We only describe how to obtain the lower bound on the eigenvalue gap, given the upper bound on the mixing time. For $0 < \varepsilon < 1$, Theorem 2.1(b) tells us that

$$Gap(P_X) \geq \frac{\lambda_1(P_X)}{2\,\tau_{P_X}(\varepsilon)} \log(2\varepsilon)^{-1}.$$

Since P_X has self-loop probabilities of at least $3/4$, this implies that $\lambda_1(P_X) \geq 1/2$.

Hence, there is some constant $c_0 > 0$ such that

$$Gap(P_x) \geq \frac{c_0}{4n^3 \log(n/\varepsilon)} \log(2\varepsilon)^{-1}.$$

This inequality must hold for all $0 < \varepsilon < 1$, so in particular we can take $\varepsilon = \frac{1}{2n}$, which gives the stated bound on the eigenvalue gap of P_X. □

4.2.2. The projection As described above, the projection of the partition is a random walk on an integer simplex in $\mathbb{Z}^d$. With the previous notation, we let $X = (x_1, \ldots, x_d)$ denote an element of $\widetilde{\mathcal{C}}$ (a trace). The stationary probability of X is $\widetilde{\pi}(X) = \binom{2n}{x_1, x_1, \ldots, x_d, x_d} / \widetilde{Z}$ where the normalizing constant is $\widetilde{Z} = \sum \binom{2n}{q_1, q_1, \ldots, q_d, q_d}$, and the sum is over all non–negative solutions to $q_1 + q_2 + \cdots + q_d = n$. The transitions in the projection are moves where we choose $i, j \in_u \{1, \ldots, d\}$ and $b \in_u \{-1, +1\}$, then add b to x_i and $-b$ to x_j, provided this resulting vector still lies in the simplex. In keeping with the decomposition strategy of Corollary 3.3, we want to use Metropolis transitions for this chain and add self-loop probabilities to ensure aperiodicity. For simplicity, we analyze this Markov chain indirectly. In order to apply path coupling to show rapid mixing, we first consider a modified chain with slightly different transition probabilities; rapid mixing of the Metropolis projection chain follows by a simple comparison of transition probabilities using Lemma 2.4.

Let $X = (x_1, \ldots, x_d)$ and $Y = (y_1, \ldots, y_d)$ denote two elements of $\widetilde{\mathcal{C}}$. The distance between X and Y, denoted $\delta(X, Y)$, is the smallest number of transitions that must be performed to change one into the other. This can also be written as

$$\delta(X, Y) = \frac{1}{2} \|X - Y\|_1 = \frac{1}{2} \sum_{i=1}^{d} |x_i - y_i|$$

where $\| \cdot \|_1$ denotes the ℓ^1 metric.

If $X = (x_1, \ldots, x_i, \ldots, x_j, \ldots, x_d)$ and $Y = (x_1, \ldots, x_i + 1, \ldots, x_j - 1, \ldots, x_d)$ (so $\delta(X, Y) = 1$) the transition probability in the Metropolis projection chain, P_M, is

$$P_M(X, Y) = \frac{1}{4d^2} \min\{1, \frac{\widetilde{\pi}(Y)}{\widetilde{\pi}(X)}\} = \frac{1}{4d^2} \min\{1, (\frac{x_j}{x_i + 1})^2\} \tag{4.2}$$

since the maximum degree in the projection chain is $2d^2$.

As mentioned, we first show a Markov chain with different transition probabilities is rapidly mixing, then use the comparison method to show that P_M is rapidly mixing. With X and Y as above, differing by a single transposition, the new chain P'_M has transition probability

$$P'_M(X, Y) = \frac{1}{8d^2} \frac{1}{(x_i + 1)^2}. \tag{4.3}$$

In other words, the denominator is determined by the coordinate that increases in the transition. (As usual, we define $P'_M(X, X) = 1 - \sum_{Y \neq X} P'_M(X, Y)$.) The change from 4 to 8 in the denominator of (4.3) is for a technical reason that arises in the proof of Lemma 4.4 below. Doing this ensures that $P'_M(X, X) \geq 3/4$ for all X, and hence

$\lambda_1(P'_M) \geq 1/2$. We note that detailed balance is satisfied, i.e.,

$$\frac{\widetilde{\pi}(Y)}{\widetilde{\pi}(X)} = (\frac{x_j}{x_i+1})^2 = \frac{P'_M(X,Y)}{P'_M(Y,X)}.$$

This ensures that P'_M has the same stationary distribution $\widetilde{\pi}$ as P_M.

Now we use path coupling to bound the mixing time of this modified chain. In the notation of Theorem 2.2, we take $U \subset \widetilde{\mathcal{C}} \times \widetilde{\mathcal{C}}$ to be the pairs (X,Y) with $\delta(X,Y) = 1$. We couple by choosing the pair of indices i and j and $b \in \{-1,+1\}$, and try to update X and Y simultaneously by adding b to x_i and y_i and $-b$ to x_j and y_j, according to the transitions of P'_M (assuming they give new elements of $\widetilde{\mathcal{C}}$).

Lemma 4.3. *Let $(X,Y) \in U$. Under one step of the coupled Markov chain we have $E(\delta(X',Y')) \leq (1 - \frac{1}{4d^2(n+2)^3})\delta(X,Y)$.*

Proof. Since $\delta(X,Y) = 1$ there exists some k and k' such that $y_k = x_k + 1$ and $y_{k'} = x_{k'} - 1$. Without loss of generality, we may assume that $k = 1$ and $k' = 2$. We consider three cases, depending on the choice of the indices i and j under the coupling described above.

Case 1: If $|\{i,j\} \cap \{1,2\}| = 0$, then both processes accept the move with the same probability, so $\delta(X',Y') = 1$.

Case 2: If $|\{i,j\} \cap \{1,2\}| = 1$, we shall see that the expected change is also zero. Assume without loss of generality that $i = 1$ and $j = 3$, and first consider the case where $b = +1$. It may happen that $x_3 = 0$. In this situation, we can change neither X nor Y (i.e. $X' = X$ and $Y' = Y$), so the distance remains unchanged. So now assume that $x_3 \geq 1$.

In this case we move from X to $X' = (x_1 + 1, x_2, x_3 - 1, \ldots, x_d)$ with probability $\frac{1}{8d^2(x_1+1)^2}$ and from Y to $Y' = (x_1 + 2, x_2 - 1, x_3 - 1, \ldots, x_d)$ with probability $\frac{1}{8d^2(x_1+2)^2}$. Since $P'_M(X,X') > P'_M(Y,Y')$, with probability $P'_M(Y,Y')$ we update both X and Y; with probability $P'_M(X,X') - P'_M(Y,Y')$ we update just X; and with all remaining probability, we update neither. In the first case we end up with X' and Y', in the second we end up with X' and Y, and in the final case we stay at X and Y. In all cases, these pairs are unit distance apart, so the expected change in distance is zero. If $b = -1$, then $P'_M(X,X'') = P'_M(Y,Y'') = \frac{1}{8d^2(x_3+1)^2}$, where $X'' = (x_1 - 1, x_2, x_3 + 1, \ldots, x_d)$ and $Y'' = (x_1, x_2 - 1, x_3 + 1, \ldots, x_d)$. If both moves are allowable (i.e., $x_1 \geq 1$), the coupling keeps the updated configurations unit distance apart. If $x_1 = 0$, then X will not change (since the update will give a negative first coordinate), and in this case we have $Y'' = (0, x_2 - 1, x_3 + 1, \ldots, x_d)$. Noting that $\delta(X,Y'') = 1$, the updated configurations still have unit distance.

Case 3: If $|\{i,j\} \cap \{1,2\}| = 2$, then we show that the expected change in distance is at most zero, and is, in fact, strictly negative. Assume without loss of generality that $i = 1, j = 2$, and $b = +1$. The probability of moving from X to $X''' = (x_1 + 1, x_2 - 1, \ldots, x_d) = Y$ is $P'_M(X,X''') = \frac{1}{8d^2(x_1+1)^2}$. The probability of moving from Y to $Y''' = (x_1 + 2, x_2 - 2, \ldots, x_d)$ is $P'_M(Y,Y''') = \frac{1}{8d^2(x_1+2)^2}$. Therefore, with probability

$P'_M(Y, Y''')$ we update both configurations, keeping them at unit distance, and with probability $P'_M(X, X''') - P'_M(Y, Y''') \geq \frac{1}{4d^2(n+2)^3}$ we update only X, decreasing the distance to zero. (Note that if $x_2 = 1$, then Y cannot move to Y''', but X can still be updated to $X''' = Y$, and the bound given is still valid.) When $b = -1$, a similar argument shows that there is again a small chance of decreasing the distance.

Summing over all three cases yields the lemma. □

Using Theorem 2.2 with $\alpha = 1 - \frac{1}{4d^2(n+2)^3}$, noting that $\delta(X, Y) \leq n$ for all X and Y, together with Theorem 2.1, we have the following result.

Lemma 4.4. *The Markov chain P'_M is rapidly mixing and*

(a) $\tau_{P'_M}(\varepsilon) = \mathrm{O}(d^2 n^3(\log n + \log \varepsilon^{-1}))$.

(b) $Gap(P'_M) \geq \frac{c'}{d^2 n^3}$ *for some constant* c'.

Proof. Part (a) follows from Theorem 2.2 immediately, using the bound on $\delta(X, Y)$ above and α as described.

For part (b), we can proceed as we did in Theorem 4.2. ¿From Theorem 2.1(b), we know that

$$Gap(P'_M) \geq \frac{\lambda_1(P'_M)}{2\,\tau_{P'_M}(\varepsilon)} \log(2\varepsilon)^{-1}.$$

Here we use the fact that $\lambda_1(P'_M) \geq 1/2$ we mentioned earlier. Combining this with the bound on the mixing time just established, we have

$$Gap(P'_M) \geq \frac{k}{2 \cdot 2 \cdot d^2 n^3 \log(n/\varepsilon)} \cdot \log(2\varepsilon)^{-1} \tag{4.4}$$

for some positive constant k. Since (4.4) must hold for all $0 < \varepsilon < 1$, as in Theorem 4.2 we may take $\varepsilon = \frac{1}{2n}$ to establish the lower bound on $Gap(P'_M)$. □

Finally, using the comparison method, the original Metropolis chain P_M for the projection is seen to be rapidly mixing. In fact, since all transitions of P_M are as large as those of P'_M (recall expressions (4.2) and (4.3)), by Lemma 2.4 we have the same bounds for P_M.

Theorem 4.5. *The projection chain P_M on $\widetilde{\mathcal{C}}$ is rapidly mixing and*

(a) $\tau_{P_M}(\varepsilon) = \mathrm{O}(d^2 n^3(\log n + \log \varepsilon^{-1}))$.

(b) $Gap(P_M) \geq \frac{c'}{d^2 n^3}$, *for some constant* c'.

Putting together this result with the previous observation about the restricted chains gives us this theorem:

Theorem 4.6. *The Markov chain $\mathcal{M}$ on the set $\mathcal{C}$ of all circuits of length $2n$ in $\mathbb{Z}^d$ is rapidly mixing. Furthermore, assuming $n \geq d$, we have*

(a) $Gap(\mathcal{M}) \geq \frac{c''}{dn^8}$, *for some constant* c''.

(b) $\tau_{\mathcal{M}}(\varepsilon) = \mathrm{O}\left(dn^8(2n\log(2d) + \log\varepsilon^{-1})\right)$

Proof. In the notation of Corollary 3.3, we need to determine (bounds for) β and γ. Looking back at the transition probabilities of $\mathcal{M}$ (equation (4.1) in Section 4.1), we note that $\beta \geq \frac{1}{8nd}$. If $\delta(X,Y) = 1$, where $X = (x_1, \ldots, x_d)$ and $Y = (x_1, \ldots, x_i + 1, \ldots, x_j - 1, \ldots, x_d)$, we need to determine what fraction of words in X are neighbors of words in Y to find a bound for γ. This fraction is exactly the likelihood that a word in C_X has an a_j followed by an a_j^{-1}. This is at least $1/(2n)$, hence $\gamma \geq 1/(2n)$.

Finally, provided $n \geq d$, the Markov kernel of the projection chain has maximum degree $\Delta = 2d^2$ (otherwise the maximum degree is $\Delta = 2n^2$).

Combining these bounds with Theorems 4.2 and 4.5 for the restrictions and the projection chain, we obtain the bound that

$$Gap(\mathcal{M}) \geq \frac{1}{8nd} \cdot \frac{1}{2n} \cdot 2d^2 \cdot \frac{c'}{d^2n^3} \cdot \frac{c}{n^3} \geq \frac{c''}{dn^8}.$$

This establishes (a), and (b) follows using the first part of Theorem 2.1 and noting that $|\mathcal{C}| \leq (2d)^{2n}$, hence (in the notation of that theorem) $\log(1/\pi_*) \leq 2n\log(2d)$. □

5. Weighted exclusion processes

Before we present our next application of the decomposition theorem, we take a combinatorial excursion to study a certain exclusion process. These results will provide tools necessary for Section 6 where we discuss sampling circuits in d-regular trees. Moreover, since exclusion processes are such fundamental models, we believe that the mixing results obtained here are independently interesting.

An *exclusion process* is a stochastic process involving a graph, typically a lattice, where some of the vertices are occupied by single particles. Over time, particles can move about, but two particles can never occupy the same location simultaneously. We will consider a weighted model in which each configuration is assigned a weight. Our goal is to sample configurations with probabilities proportional to their weights. In other words, we would like to define a Markov chain that has as its stationary distribution probabilities proportional to these weights. Naturally, we want to bound the mixing rate of the chain.

The weighted exclusion process we study is motivated by the applications in Section 6. In particular, consider a set of n sites arranged in linear order, k of which contain particles, and let Ω_k denote the set of all $\binom{n}{k}$ such configurations. In what follows it will be convenient to augment each configuration with sites at locations 0 and $n+1$, each containing a fixed particle. Let $f = \{f_i\}$ denote a sequence of positive numbers for $i \geq 0$. Using this sequence, we assign a weight to each $\sigma \in \Omega_k$, that is $wt(\sigma) = f_{x_1} f_{x_2} \cdots f_{x_{k+1}}$ where $x_i \in \{0, \ldots, n-k\}$ is the number of unoccupied sites between particles i and $i+1$ for $i = 1, \ldots, k+1$. Normalizing, we have a probability distribution ρ on Ω_k, with $\rho(\sigma) = \frac{wt(\sigma)}{Z_k}$ where $Z_k = \sum_{\sigma' \in \Omega_k} wt(\sigma')$. For example, if $f_i = 1, \forall i$, then ρ is the uniform distribution on Ω_k, but, for our application in this paper we are more interested in the non–uniform case. To sample from Ω_k according to ρ, we define a Markov chain having ρ as its stationary distribution and aim to show this chain is rapidly mixing.

5.1. A particle process with nearest-neighbor dynamics

A natural Markov chain $\mathcal{M}_{Gl}$ on Ω_k, called *nearest-neighbor (or Glauber) dynamics*, consists of picking a particle at random and moving it one space to the left or right (provided the destination is unoccupied). We note that the two particles we added at sites 0 and $n+1$ are fixed and cannot move. For $\sigma \in \Omega_k$, we use $\sigma(i)$ to denote the position of particle i, and we define the distance between pairs σ_1, σ_2 as $\delta(\sigma_1, \sigma_2) = \sum_i |\sigma_1(i) - \sigma_2(i)|$, the sum of distances between corresponding particles. We use Metropolis transitions for the nearest-neighbor dynamics. We let σ_1 denote the current particle configuration, and σ_2 denote the configuration after one step of the chain $\mathcal{M}_{Gl}$.

One step of the nearest-neighbor Markov chain $\mathcal{M}_{Gl}$:

1. Select a particle uniformly at random (excluding the two fixed particles), and a direction $d \in_u \{\ell, r\}$, where ℓ means "left" and r "right". Let σ' denote the configuration obtained from σ_1 by moving the chosen particle one step in the direction d *if* the destination is unoccupied. If the destination is occupied, set $\sigma' = \sigma_1$.
2. With probability $\frac{1}{2}\min\{1, \frac{\rho(\sigma')}{\rho(\sigma_1)}\}$ set $\sigma_2 = \sigma'$.
3. In all other cases, set $\sigma_2 = \sigma_1$.

With the chain defined above, the transition matrix P_{Gl} for the Glauber chain is given by

$$P_{Gl}(\sigma_1, \sigma_2) = \begin{cases} \frac{1}{4k}\min\{1, \frac{\rho(\sigma_2)}{\rho(\sigma_1)}\} & \text{if } \delta(\sigma_1, \sigma_2) = 1, \\ 0 & \text{if } \delta(\sigma_1, \sigma_2) > 1, \\ 1 - \sum_{\sigma' \neq \sigma_1} P_{Gl}(\sigma_1, \sigma') & \text{if } \sigma_1 = \sigma_2. \end{cases}$$

The transitions ensure that $\mathcal{M}_{Gl}$ is aperiodic since $P_{Gl}(\sigma, \sigma) \geq \frac{1}{2}$ for all $\sigma \in \Omega_k$.

Our approach to establishing a bound for the mixing time of $\mathcal{M}_{Gl}$ is indirect. We first define a new Markov chain that includes some non-local moves and show this chain is rapidly mixing. Using the comparison method of Section 2.2 we can then hope to establish rapid mixing for the simpler Glauber chain. In subsequent sections, we demonstrate the heat bath chain is rapidly mixing when the sequence f is log–concave, or is log–convex and satisfies an additional condition.

5.2. A particle process with heat bath dynamics

A second natural Markov chain on Ω_k is *heat bath dynamics*. The transitions differ from nearest-neighbor dynamics in that particles may move more than one space at a time, but can never pass over another particle in doing so. We again use σ_1 to denote the current particle configuration, and σ_2 for the configuration after one step of the heat bath chain. Recall that $\sigma_1(j)$ denotes the location of particle j in the configuration σ_1.

One step of the heat bath Markov chain $\mathcal{M}_{HB}$:

1. With probability $\frac{3}{4}$, set $\sigma_2 = \sigma_1$.
2. Otherwise,
   - Select a particle uniformly at random (excluding the two fixed particles). Suppose

we select particle j. Let $m = \sigma_1(j+1) - \sigma_1(j-1) - 1$. Then m is the number of sites between particles $j-1$ and $j+1$.

- Remove particle j.
- Replace particle j in location $\sigma_1(j-1)+t$ to obtain σ_2, where t is selected according to the probability distribution

$$\mathcal{P}r(t=i) = \frac{f_{i-1}f_{m-i}}{Z} \qquad \forall i \in \{1,\ldots,m\}$$

and $Z = \sum_{i=1}^{m} f_{i-1}f_{m-i}$ is the normalizing constant.

Step 1 is introduced for technical reasons (namely, to ensure that $\lambda_1(P_{HB}) \geq \frac{1}{2}$ which we use later on), and it only slows down the chain by a factor of four. We see that in one step of $\mathcal{M}_{HB}$ we remove a particle and reinsert it between its neighbors according to the desired conditional distribution. The transition probabilities of $\mathcal{M}_{HB}$ are

$$P_{HB}(\sigma_1,\sigma_2) = \begin{cases} \frac{1}{4k}\frac{f_{x_{j-1}}f_{x_j}}{Z} & \text{if } \sigma_1,\sigma_2 \text{ differ solely at particle } j, \\ 0 & \text{if } \sigma_1,\sigma_2 \text{ differ by two or more particles,} \\ 1 - \sum_{\sigma' \neq \sigma_1} P_{HB}(\sigma_1,\sigma') & \text{if } \sigma_1 = \sigma_2, \end{cases} \tag{5.1}$$

where x_j is the number of empty sites between particles j and $j+1$ in σ_2 and Z is the normalizing constant defined above in step 2 of this chain. It is important to reiterate that the ordering of the particles in the interval does not change during the transitions. Any particle that moves must remain between its neighbors and cannot jump over them.

Finally, note this Markov chain is reversible and the stationary probabilities are those defined earlier, namely $\rho(\sigma) = f_{x_1}f_{x_2}\cdots f_{x_{k+1}}/Z_k$, where $Z_k = \sum f_{y_1}f_{y_2}\cdots f_{y_{k+1}}$ is the normalizing constant and the sum is over all non–negative solutions to $y_1 + y_2 + \ldots + y_{k+1} = n-k$. Having defined the heat bath Markov chain, we prove it is rapidly mixing for some natural classes of sequences. The general approach we use is to apply path coupling to show this.

5.3. Log–concave sequences

A log–concave sequence is one satisfying $f_{i-1}f_{i+1} \leq f_i^2$ for all i. For these sequences, the configurations having the largest weight (stationary probability) are those where the particles are distributed as evenly as possible in the range of sites. Intuitively, this suggests that regardless of the initial arrangement of the particles, they will tend to spread out over time.

We have already defined a distance metric on Ω_k, and note that $0 \leq \delta(\sigma_1,\sigma_2) \leq k(n-k)$ for any pair $\sigma_1,\sigma_2 \in \Omega_k$. We examine the heat bath dynamics using path coupling. In the notation of Theorem 2.2 we take $U \subseteq \Omega_k \times \Omega_k$ to be pairs (σ_1,σ_2) with $\delta(\sigma_1,\sigma_2) = 1$, so the pair differs by a single (nearest-neighbor) transition.

Figure 2 shows parts of two configurations with k particles, differing only at particle j. In our coupling, if particle j is chosen in the first step of the move, we can reinsert it at the same position in each configuration, decreasing the distance by one. Also, choosing any other particle except $j-1$ or $j+1$ allows us to reinsert it at the same position

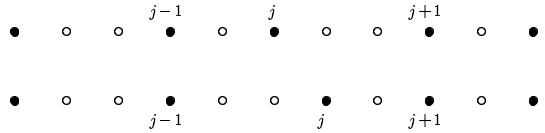

Figure 2. Typical situation for path coupling

in each configuration with identical probabilities, leaving the distance unchanged. So we need to consider how to couple the moves if we choose particle $j-1$ or $j+1$. For example, consider Figure 3 which shows the configuration of Figure 2 with particle $j-1$ removed. When we couple the chains together, it would be very bad to replace particle $j-1$ at the first position in σ_1 and at the fifth position in σ_2 since the new distance between the configurations is now four where it was previously one.

The dashed lines of Figure 3 indicate how we would like to couple the moves. Replacing particle $j-1$ between its neighboring particles in the manner indicated either keeps

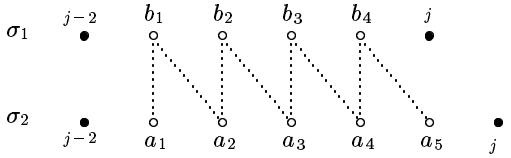

Figure 3. The coupling for the particle system

the distance between the pair of configurations unchanged or increases it by only one. We show such a coupling is possible by demonstrating majorization inequalities for the transition probabilities a_i and b_i for the pair of configurations.

Referring to Figure 3 as a representative case, let m denote the number of sites between particles $j-2$ and j in σ_1, so $m = \sigma_1(j) - \sigma_1(j-2) - 1$. For $i \in [m]$ let $b_i = f_{i-1} f_{m-i}/Z_b$ be the probability to insert particle $j-1$ at position i between $j-2$ and j in the upper configuration σ_1, and for $i \in [m+1]$, let $a_i = f_{i-1} f_{m+1-i}/Z_a$ be the probability to insert particle $j-1$ in the lower configuration σ_2, where $Z_b = \sum_{i=1}^{m} f_{i-1} f_{m-i}$ is the normalizing constant for σ_1 and $Z_a = \sum_{i=1}^{m+1} f_{i-1} f_{m+1-i}$ is the normalizing constant for σ_2. We have the following surprising combinatorial lemma:

Lemma 5.1. *Suppose $f = \{f_i\}$ is a log-concave sequence. With a_i and b_i as above, for all $t \in [m]$, we have*

$$\sum_{i=1}^{t} a_i \le \sum_{i=1}^{t} b_i \le \sum_{i=1}^{t+1} a_i.$$

Proof. We need only verify the first set of inequalities,

$$\sum_{i=1}^{m-t} a_i \le \sum_{i=1}^{m-t} b_i, \tag{5.2}$$

since they imply the remaining ones. To see this, recall that both sequences are symmetric,

i.e., $a_i = a_{m+2-i}$ and $b_i = b_{m+1-i}$, and that $\sum_{i=1}^{m+1} a_i = \sum_{i=1}^{m} b_i = 1$. This tells us that

$$\sum_{i=1}^{t} b_i = 1 - \sum_{i=t+1}^{m} b_i = 1 - \sum_{i=1}^{m-t} b_i \leq 1 - \sum_{i=1}^{m-t} a_i = 1 - \sum_{i=t+2}^{m+1} a_i = \sum_{i=1}^{t+1} a_i.$$

We spend the remainder of the proof verifying the inequalities in (5.2). In other words, recalling the definitions of a_i and b_i, we want to show

$$\frac{\sum_{i=0}^{t-1} f_i f_{m-i}}{\sum_{i=0}^{m} f_i f_{m-i}} \leq \frac{\sum_{j=0}^{t-1} f_j f_{m-j-1}}{\sum_{j=0}^{m-1} f_j f_{m-j-1}}$$

or, equivalently,

$$(\sum_{i=0}^{t-1} f_i f_{m-i})(\sum_{j=0}^{m-1} f_j f_{m-j-1}) \leq (\sum_{i=0}^{m} f_i f_{m-i})(\sum_{j=0}^{t-1} f_j f_{m-j-1}). \tag{5.3}$$

Case 1: $t \leq m/2$. Referring to equation (5.3), we have

$$\begin{aligned} LHS &= (\sum_{i=0}^{t-1} f_i f_{m-i})(\sum_{j=0}^{t-1} f_j f_{m-j-1}) + (\sum_{i=0}^{t-1} f_i f_{m-i})(\sum_{j=t}^{m-t-1} f_j f_{m-j-1}) \\ &\quad + (\sum_{i=0}^{t-1} f_i f_{m-i})(\sum_{j=m-t}^{m-1} f_j f_{m-j-1}) \\ &= (\sum_{i=0}^{t-1} f_i f_{m-i})(\sum_{j=0}^{t-1} f_j f_{m-j-1}) + (\sum_{i=0}^{t-1} f_i f_{m-i})(\sum_{j=t}^{m-t-1} f_j f_{m-j-1}) \\ &\quad + (\sum_{i=m-t+1}^{m} f_i f_{m-i})(\sum_{j=0}^{t-1} f_j f_{m-j-1}). \end{aligned} \tag{5.4}$$

We obtained this last expression by rewriting the third product in an equivalent manner.

Now, using the log–concavity of f, we find that for $i \leq j$,

$$f_{m-j-1} f_{m-i} \leq f_{m-j} f_{m-i-1},$$

so, adding additional factors,

$$f_j f_{m-j-1} f_i f_{m-i} \leq f_j f_{m-j} f_i f_{m-i-1}. \tag{5.5}$$

This implies

$$(\sum_{j=t}^{m-t-1} f_j f_{m-j-1})(\sum_{i=0}^{t-1} f_i f_{m-i}) \leq (\sum_{j=t}^{m-t-1} f_j f_{m-j})(\sum_{i=0}^{t-1} f_i f_{m-i-1}). \tag{5.6}$$

Substituting inequality (5.6) into the second of the three products in (5.4) (and reversing the roles of i and j), we have

$$\begin{aligned} LHS &\leq (\sum_{i=0}^{t-1} f_i f_{m-i})(\sum_{j=0}^{t-1} f_j f_{m-j-1}) + (\sum_{i=t}^{m-t-1} f_i f_{m-i})(\sum_{j=0}^{t-1} f_j f_{m-j-1}) \end{aligned}$$

$$+\Big(\sum_{i=m-t+1}^{m} f_i f_{m-i}\Big)\Big(\sum_{j=0}^{t-1} f_j f_{m-j-1}\Big) \;\le\; RHS.$$

Note that the final inequality comes from adding in the missing term(s) when $i = m - t$ in the second product.

Case 2: $t > m/2$. Breaking up (5.3) differently this time, we have

$$\begin{aligned} LHS &= \Big(\sum_{i=0}^{t-1} f_i f_{m-i}\Big)\Big(\sum_{j=0}^{t-1} f_j f_{m-j-1}\Big) + \Big(\sum_{i=0}^{m-t-1} f_i f_{m-i}\Big)\Big(\sum_{j=t}^{m-1} f_j f_{m-j-1}\Big) \\ &\quad + \Big(\sum_{i=m-t}^{t-1} f_i f_{m-i}\Big)\Big(\sum_{j=t}^{m-1} f_j f_{m-j-1}\Big) \\ &= \Big(\sum_{i=0}^{t-1} f_i f_{m-i}\Big)\Big(\sum_{j=0}^{t-1} f_j f_{m-j-1}\Big) + \Big(\sum_{i=t+1}^{m} f_i f_{m-i}\Big)\Big(\sum_{j=0}^{m-t-1} f_j f_{m-j-1}\Big) \\ &\quad + \Big(\sum_{i=m-t}^{t-1} f_i f_{m-i}\Big)\Big(\sum_{j=t}^{m-1} f_j f_{m-j-1}\Big) \end{aligned} \tag{5.7}$$

where we have rewritten the second term with an equivalent expression.

Using the log–concavity of f with $i \le j$, we again have (5.5) and find

$$\Big(\sum_{i=m-t}^{t-1} f_i f_{m-i}\Big)\Big(\sum_{j=t}^{m-1} f_j f_{m-j-1}\Big) \le \Big(\sum_{j=t}^{m-1} f_j f_{m-j}\Big)\Big(\sum_{i=m-t}^{t-1} f_i f_{m-i-1}\Big) \tag{5.8}$$

Similar to before, we swap the roles of i and j in the final expression of (5.8), and then use it to bound the third product in (5.7). This gives the new inequality

$$\begin{aligned} LHS &\le \Big(\sum_{i=0}^{t-1} f_i f_{m-i}\Big)\Big(\sum_{j=0}^{t-1} f_j f_{m-j-1}\Big) + \Big(\sum_{i=t+1}^{m} f_i f_{m-i}\Big)\Big(\sum_{j=0}^{m-t-1} f_j f_{m-j-1}\Big) \\ &\quad + \Big(\sum_{i=t}^{m-1} f_i f_{m-i}\Big)\Big(\sum_{j=m-t}^{t-1} f_j f_{m-j-1}\Big). \end{aligned}$$

Adding in the extra terms when $i = t$ in the middle product and the terms when $i = m$ in the third one, we can combine the second and third terms to get

$$LHS \;\le\; \Big(\sum_{i=0}^{t-1} f_i f_{m-i}\Big)\Big(\sum_{j=0}^{t-1} f_j f_{m-j-1}\Big) + \Big(\sum_{i=t}^{m} f_i f_{m-i}\Big)\Big(\sum_{j=0}^{t-1} f_j f_{m-j-1}\Big) \;=\; RHS.$$

□

Lemma 5.1 allows us to couple moves in the "zig–zag" manner shown in Figure 3. We describe the coupling for pairs σ_1, σ_2 with $\delta(\sigma_1, \sigma_2) = 1$. As before, we suppose that they differ at particle j. We use the notation (σ_1', σ_2') for the configuration obtained from the

pair (σ_1, σ_2) after one step of the coupling. Let $\Delta\delta(\sigma_1, \sigma_2) = \delta(\sigma_1', \sigma_2') - \delta(\sigma_1, \sigma_2)$ denote the change in distance after one step of the coupled chain.

Coupling for the heat bath chain $\mathcal{M}_{HB}$:

1. Pick a particle, say particle i, uniformly at random.
2. If $i \notin \{j-1, j, j+1\}$ we remove particle i from σ_1 and σ_2 and reinsert it in the same location in both, with the exact same transition probabilities (according to equation (5.1)). In this case we have $\Delta\delta(\sigma_1, \sigma_2) = 0$.
3. If $i = j$ we can replace particle j in the same position in each configuration so that they now agree. Conditioned on $i = j$ we have $\Delta\delta(\sigma_1, \sigma_2) = -1$.
4. If $i \in \{j-1, j+1\}$ pick an ordered pair (u, v) according to the following probability distribution:

$$\mathcal{P}r\{(u,v) = (t,t)\} = \textstyle\sum_{r=1}^{t} a_r - \sum_{r=1}^{t-1} b_r \quad \text{for } t = 1, \ldots, m$$
$$\mathcal{P}r\{(u,v) = (t,t+1)\} = \textstyle\sum_{r=1}^{t} b_r - \sum_{r=1}^{t} a_r \quad \text{for } t = 1, \ldots, m$$

   Here we assume that σ_1 has the shorter interval between particles $i-1$ and $i+1$ and $m = \sigma_1(i+1) - \sigma_1(i-1) - 1$ denotes the number of sites between the neighbors of particle i, and a_r, b_r are defined as before. Having picked the ordered pair (u, v), we place particle i at position u in σ_1 and at position v in σ_2 (between its neighboring particles $i-1$ and $i+1$). In this case we have $\Delta\delta(\sigma_1, \sigma_2) \in \{0, +1\}$.

Lemma 5.2. *Suppose $f = \{f_i\}$ is a log–concave sequence. Let $\sigma_1, \sigma_2 \in \Omega_k$ with $\delta(\sigma_1, \sigma_2) = 1$. After one step of the coupled Markov chain we have $E(\Delta\delta(\sigma_1, \sigma_2)) \leq 0$.*

Proof. We have noted that selecting a particle other than $j-1$, j, or $j+1$ does not change the distance, and that by choosing particle j the distance decreases by one. The bad moves that increase the distance are those in which we insert $j-1$ (or $j+1$) at different positions in σ_1 and σ_2. If we select either of these two particles, say $j-1$, the expected change in distance is

$$\begin{aligned} \beta &= \sum_{i=1}^{m} \left(\sum_{r=1}^{i} b_r - \sum_{r=1}^{i} a_r \right) \\ &= m\, b_1 + (m-1) b_2 + \cdots + b_m - m\, a_1 - (m-1) a_2 - \cdots - a_m. \end{aligned}$$

Recalling that $b_i = b_{m+1-i}$ and $a_i = a_{m+2-i}$, we can also write

$$\beta = m\, b_m + (m-1) b_{m-1} + \cdots + b_1 - m\, a_{m+1} - (m-1) a_m - \cdots - a_2.$$

Summing these equations and simplifying yields $\beta = \frac{1}{2}$. By symmetry this also represents the expected change in distance for particle $j+1$.

Putting these pieces together to determine the overall expected change, we find that $E(\Delta\delta(\sigma_1, \sigma_2)) \leq \frac{1}{4k}(-1 + \frac{1}{2} + \frac{1}{2}) = 0$. (An inequality, as it is possible, say, that $j = 2$ so there is no bad move for the fixed particle 1.) □

The preceding lemmas result in the following theorem:

Theorem 5.3. *The heat bath Markov chain $\mathcal{M}_{HB}$ is rapidly mixing for any log–concave sequence $f = \{f_i\}$ and:*

(a) $\tau_{\mathcal{M}_{HB}}(\varepsilon) = \mathrm{O}(n^5 \log(\frac{1}{\varepsilon}))$.

(b) $Gap(\mathcal{M}_{HB}) \geq \frac{c}{n^5}$ *for some constant* $c > 0$.

Proof. Whenever $\delta(\sigma_1, \sigma_2) > 0$, the probability that the distance changes in one step is at least $1/4k$. The path coupling theorem gives a bound on the mixing time for the heat bath Markov chain $\mathcal{M}_{HB}$ on Ω_k:

$$\tau_{\tilde{P}_k}(\varepsilon) \quad \leq \quad \lceil \frac{e(k(n-k-1))^2}{\frac{1}{4k}} \log(\frac{1}{\varepsilon}) \rceil = \mathrm{O}(n^5 \log(\frac{1}{\varepsilon})).$$

For part (b), using Theorem 2.1(b) we can lower bound $Gap(\mathcal{M}_{HB})$ (using the trivial bound $\lambda_1(P_{HB}) \geq \frac{1}{2}$):

$$Gap(P_{HB}) \geq \frac{1}{4\tau_{\tilde{P}_k}(\varepsilon)} \log(\frac{1}{2\varepsilon}) = c \frac{\log(\frac{1}{2\varepsilon})}{n^5 \log(\frac{1}{\varepsilon})},$$

for some constant $c > 0$. As this inequality must hold for every $0 < \varepsilon < 1$, letting $\varepsilon \to 0$ (so $\frac{\log(\frac{1}{2\varepsilon})}{\log(\frac{1}{\varepsilon})} \to 1$) we have $Gap(\mathcal{M}_{HB}) \geq \frac{c}{n^5}$. □

Recall the original Markov chain $\mathcal{M}_{Gl}$ on Ω_k that has nearest-neighbor dynamics. We can use the comparison theorem to give a bound on the mixing (or, equivalently, the spectral gap) of $\mathcal{M}_{Gl}$. Instead of trying to give a general upper bound for an arbitrary log–concave sequence f, we will defer this task until those times that we have specified a particular sequence.

5.4. Log–convex sequences

Another natural class of sequence to consider are log–convex sequences, those satisfying $f_i^2 \leq f_{i-1} f_{i+1}$ for all i. Unfortunately, the analysis we used in the previous section does not hold for an arbitrary log–convex sequence.

The difficulty arises in that log–convex sequences can grow "too fast" so the majorization Lemma 5.1 does not generalize to any log–convex sequence. To see this, consider a sequence of numbers beginning $1, 1, 1, 1, \alpha$, where $\alpha \gg 1$ which is a log–convex sequence. Suppose that when we remove particle j the interval between its neighbors in σ_2 has five sites in it as in Figure 2. To use this same coupling argument, we would require, for example, that $\frac{1 \cdot \alpha}{1 \cdot \alpha + 1 \cdot 1 + 1 \cdot 1 + 1 \cdot 1 + \alpha \cdot 1} \leq \frac{1 \cdot 1}{1 \cdot 1 + 1 \cdot 1 + 1 \cdot 1 + 1 \cdot 1}$ $= \frac{1}{4}$ (showing $a_1 \leq b_1$). Yet letting $\alpha \to \infty$, we see this clearly cannot hold since the fraction involving α approaches $1/2$. It is clear that no coupling can overcome this difficulty.

For our purposes in Section 6, we need to consider an exclusion process whose weights are defined by the Catalan numbers $C(n) = \frac{1}{n+1}\binom{2n}{n}$. This sequence is log–convex. We shall see that, unlike the previous example, log–convex sequences that satisfy a suitable restriction do permit a result analogous to Lemma 5.1. Hence, for sequences satisfying this restriction we also have analogues to Lemma 5.2 and Theorem 5.3, meaning the heat bath chain is rapidly mixing for such sequences.

One restriction for which the path coupling argument in the previous subsection will

work in the log–concave case is one ruling out cases where the sequence grows "too fast." Let us define $g_t = \sum_{i=1}^{t} f_{i-1} f_{t-i}$. As before, we consider the case when $\delta(\sigma_1, \sigma_2) = 1$ where the two configurations differ in the location of particle j. Referring back to Figure 3 as a representative case, we again let $m = \sigma_1(j) - \sigma_1(j-2) - 1$, for $i \in [m]$ $b_i = f_{i-1} f_{m-i}/Z_b$ is the probability to insert at position i in σ_1, and for $i \in [m+1]$ the probability to insert at position i in σ_2 is $a_i = f_{i-1} f_{m+1-i}/Z_a$ where $Z_b = g_m, Z_a = g_{m+1}$ are the respective normalizing constants. We first show a result similar to Lemma 5.1. In this case, the additional condition we impose for log-convex sequences makes the proof simpler than that for Lemma 5.1.

Lemma 5.4. *Suppose $f = \{f_i\}$ is a log–convex sequence and $\frac{f_t}{f_{t-1}} \leq \frac{g_{t+1}}{g_t}$ for all $t \geq 1$. Then for all $t \in [m]$ we have*

$$\sum_{i=1}^{t} a_i \leq \sum_{i=1}^{t} b_i.$$

Proof. We actually prove that $a_i \leq b_i$ for all $i \in [m]$. This set of inequalities implies those in the statement of the lemma.

The first inequality comes directly from the assumption relating f and g, i.e.,

$$\frac{f_m}{f_{m-1}} \leq \frac{g_{m+1}}{g_m}$$

implies that

$$a_1 = \frac{f_0 f_m}{g_{m+1}} \leq \frac{f_0 f_{m-1}}{g_m} = b_1.$$

Using log–convexity we have

$$\frac{f_{m-1}}{f_{m-2}} \leq \frac{f_m}{f_{m-1}} \leq \frac{g_{m+1}}{g_m},$$

so by rearranging this inequality and multiplying by f_1 we get

$$a_2 = \frac{f_1 f_{m-1}}{g_{m+1}} \leq \frac{f_1 f_{m-2}}{g_m} = b_2.$$

Similarly, using log–convexity, we have

$$\frac{f_{m-2}}{f_{m-3}} \leq \frac{g_{m+1}}{g_m},$$

so

$$a_3 = \frac{f_2 f_{m-2}}{g_{m+1}} \leq \frac{f_2 f_{m-3}}{g_m} = b_3.$$

Continuing in this manner, we obtain the inequalities $a_i \leq b_i$ for all $i \in [m]$. □

Using this lemma we also have results analogous to Lemmas 5.1 and 5.2 for log–convex sequences satisfying the same hypotheses, so we will not repeat them here. The proof of the following theorem is identical to that of Theorem 5.3.

Theorem 5.5. *Suppose $f = \{f_i\}$ is a log-convex sequence and $\frac{f_t}{f_{t-1}} \leq \frac{g_{t+1}}{g_t}$ for all $t \geq 1$. Then the heat bath Markov chain $\mathcal{M}_{HB}$ is rapidly mixing and:*

(a) $\tau_{\mathcal{M}_{HB}}(\varepsilon) = \mathrm{O}(n^5 \log(\frac{1}{\varepsilon}))$.
(b) $Gap(\mathcal{M}_{HB}) \geq \frac{c}{n^5}$ *for some constant* $c > 0$.

6. Sampling circuits in d-regular trees

We are now prepared to return to the decomposition method. The second application of decomposition we present is sampling circuits in d-regular trees. To do so, label the edges of the tree as a Cayley graph so that each vertex is adjacent to one edge with each of the d possible labels. As before, we wish to uniformly sample circuits of length $2n$ in the tree.

There is a connection between this problem and another of interest to the statistical physics community, that of sampling staircase walks according to a Gibbs distribution. Staircase walks are paths that join $(0,0)$ to $(2n,0)$ using diagonal steps by adding $(1,1)$ or $(1,-1)$ to the current location, and which never go below the x-axis. Letting $\lambda > 0$ be a fixed parameter, we assign a weight of λ^k to a staircase walk that hits the x-axis k times. The Gibbs distribution normalizes these weights, giving a probability distribution on the set of staircase walks. The key observation we use to sample circuits in a d-regular tree is a many-to-one mapping between the set of circuits on this tree to the set of staircase walks. A circuit of length $2n$, starting at the root, has n edges leading away from the root and n edges leading back towards the root. Whenever we are at the root we have d choices of labeled edges; whenever we are away from the root we have $d-1$ edges which move us farther away and a unique edge which will bring us closer to the root. Hence, there are $d^k(d-1)^{n-k}$ walks of length $2n$ that hit the root k times (including the initial time, but not the final). Using adsorbing staircase walks, sampling is easy: (1) Select a staircase walk of length $2n$ according to the Gibbs measure with $\lambda = \frac{d}{d-1}$. A staircase walk hitting the x-axis k times appears with probability proportional to $(\frac{d}{d-1})^k = \frac{d^k(d-1)^{(n-k)}}{(d-1)^n}$. The up edges in this walk correspond to steps in the tree that move away from the root, and the down edges are those that move back towards the root. (2) Assign labels to the up edges uniformly at random (from the set of d labels for edges starting from the x-axis, and from a suitable set of $d-1$ labels for edges above the x-axis), assigning labels to the down edges that equal the label of the most recent unpaired up edge preceding it. This gives a sequence of labeled edges corresponding to a labeled walk of length $2n$ in the d-regular tree.

Using this scheme, sampling circuits in the tree reduces to the problem of sampling staircase walks with the appropriate Gibbs measure, then labeling a staircase walk as outlined above. For this reason, we focus on analyzing an algorithm for sampling staircase walks. In Section 6.1 we define a Markov chain on staircase walks. In Section 6.2 we show that path coupling suffices to show rapid mixing when $\lambda \leq 1$, but it is precisely the case when $\lambda > 1$ that is necessary for sampling circuits in Cayley trees. In this case, path coupling fails since the Markov chain is not contracting on the state space.

Therefore, we apply the decomposition method to analyze the Markov chain in the

case when $\lambda > 1$. In fact, we actually apply decomposition twice to establish a bound on the mixing rate of this chain. Along the way, we make use of the analysis of exclusion processes found in Section 5, the comparison method (Theorem 2.3 in Section 2.2), and path coupling in some situations where it may successfully be applied.

6.1. The Markov chain on staircase walks

Fix some $n \geq 1$ and let $\mathcal{S}$ denote the set of all staircase walks with $2n$ edges. It is well known that $|\mathcal{S}| = C(n) = \frac{1}{n+1}\binom{2n}{n}$, the nth Catalan number (see [17]).

We define a natural Markov chain $\mathcal{M}$ on $\mathcal{S}$. This mountain/valley Markov chain has previously appeared in [8, 12, 18]. The transitions of the chain are *inversions* which replace local maxima with local minima, or vice-versa, by interchanging two edges along the walk. If the cth point on the path is $v_c = (x_c, y_c)$, we call it a mountain if $y_{c-1} = y_c - 1 = y_{c+1}$ and inverting it consists of setting $y'_c = y_c - 2$. Likewise, inverting a valley where $y_{c-1} = y_c + 1 = y_{c+1}$ consists of setting $y'_c = y_c + 2$. The Markov chain $\mathcal{M}$ iterates the following steps.

One step of Markov chain $\mathcal{M}$:

1. Pick c uniformly at random from $\{2, \ldots, 2n-2\}$, and let v denote the point on the walk whose x-coordinate is c.
2. - If v is the bottom of a valley lying on the x-axis, with probability $\frac{1}{2(1+\lambda)}$ set X_{t+1} equal to X_t inverted at v. Otherwise, set $X_{t+1} = X_t$.
   - If v is the top of a mountain, and inverting it will put it on the x-axis, with probability $\frac{\lambda}{2(1+\lambda)}$ set X_{t+1} equal to X_t inverted at v.
   - If v is the bottom of a valley not lying on the x-axis, or if v is the top of a mountain and inverting at v does not put it on the x-axis, with probability $\frac{1}{4}$ set X_{t+1} equal to X_t inverted at v.
   - In all other cases, set $X_{t+1} = X_t$.

Note that this Markov chain is aperiodic, reversible, and the stationary distribution is the Gibbs distribution, namely $\pi(\sigma) = \frac{\lambda^{k(\sigma)}}{Z}$, where $k(\sigma)$ is the number of times the walk σ touches the x-axis, and, as usual, $Z = \sum_{\sigma \in \mathcal{S}} \lambda^{k(\sigma)}$ is the normalization constant.

We first review the easier case when $\lambda \leq 1$.

6.2. Staircase walks with $\lambda \leq 1$

Path coupling suffices to show that the mountain/valley chain on staircase walks is rapidly mixing when $\lambda \leq 1$. We define our distance measure Φ to be one-half of the area between the configurations, i.e., drawing the configurations on the same set of axis bounds rectangular regions between the pair of walks. The distance between the two walks is one-half of the sum of the areas of these rectangular regions. For the coupling, we take the point c in step 1 of $\mathcal{M}$, and attempt to perform the same transition in each walk. To use path coupling, we must examine a pair of walks that differ solely by a single transition of the chain (a single square).

Lemma 6.1. *Let $X, Y \in \mathcal{S}$ with $\Phi(X, Y) = 1$. Provided that $\lambda \leq 1$, after one step of $\mathcal{M}$ we have $E(\Delta\Phi(X, Y)) \leq 0$.*

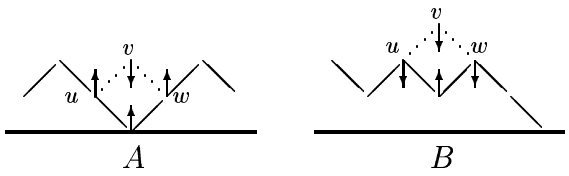

Figure 4. Typical situations for path coupling

Proof. Consider the configurations in Figure 4, which show pieces of walks that agree everywhere except at a single square, and let Y denote the "upper" walk. If the square is adjacent to the x-axis as in Figure 4.A, then there are two transitions that decrease the distance by one, inverting v in one of the walks so they now agree everywhere. Inverting u or w in Y increases the distance by one. Every other transition not involving u, v, or w does not change the distance between X and Y. Therefore, in this case we find $E(\Delta\Phi(X,Y)) \leq \frac{1}{2n}(\frac{1}{4} + \frac{1}{4} - \frac{1}{2(1+\lambda)} - \frac{\lambda}{2(1+\lambda)}) = 0$. (This is an inequality since, in some cases, one or both of the vertices u and w might not actually correspond to moves which increase the distance.) The second case is if the differences between the two walks occur a unit distance from the x-axis, as in Figure 4.B. There are also two good inversions at v, each with probability $\frac{1}{4}$, and those at u and w increase the distance between X and Y by one; all other moves preserve the distance between the pair. In this case, we have $E(\Delta\Phi(X,Y)) \leq \frac{1}{2n}(\frac{\lambda}{2(1+\lambda)} + \frac{\lambda}{2(1+\lambda)} - \frac{1}{4} - \frac{1}{4})$. This last expression is non-positive if $\lambda \leq 1$. Other situations where X and Y differ by a square that is far away from the x-axis are neutral; two good moves decrease the distance by one, and (at most) two bad moves increase the distance by one. Each of these moves occurs with equal probability, so in these cases we also have $E(\Delta\Phi(X,Y)) \leq 0$. □

An application of Theorem 2.2 gives a polynomial bound on the mixing rate. These details are left to the reader. We note in the case demonstrated in Figure 4.B that the distance will increase in expectation if $\lambda > 1$. We use the disjoint decomposition method to show rapid mixing in this case.

6.3. Staircase walks when $\lambda > 1$

Applying decomposition, we first partition $\mathcal{S}$ into sets $\mathcal{S}_k$ where each walk in $\mathcal{S}_k$ hits the x-axis exactly k times between the endpoints. The cardinalities $|\mathcal{S}_0|, |\mathcal{S}_1|, \ldots, |\mathcal{S}_{n-1}|$ form a log–concave sequence (Section 6.3.2) implying the first projection chain is rapidly mixing.

To examine the restricted chains on each of the sets $\mathcal{S}_k$, we apply decomposition a second time. We partition $\mathcal{S}$ by considering all walks that hit the x-axis at exactly the same locations. Within each $\mathcal{S}_k$ (and, hence, in each set after further partitioning $\mathcal{S}$) the distribution is uniform over walks. Rapid mixing of the new restrictions is now readily established using path coupling. The projection of the new partition is exactly the exclusion process of Section 5 where the location of the particles correspond to the places where the walks hit the x-axis. The probability of a particle configuration is proportional to a product of Catalan numbers, so we can use the analysis in Section 5.4 as the Catalan numbers are a log–convex sequence. Following a comparison of Markov chains, we finally establish rapid mixing of the second projection, hence rapid mixing of the restriction to

$\mathcal{S}_k$ via the new disjoint decomposition theorem. Applying that theorem again, together with the fact that the first projection was mixing because of the log–concavity of the sequence $|\mathcal{S}_i|$, we have rapid mixing on all of $\mathcal{S}$. Along the way in our analysis, we also apply the comparison method (Section 2.2). We formalize the analysis in what follows.

6.3.1. Decomposition of $\mathcal{S}$ The mountain/valley Markov chain $\mathcal{M}$ of Section 6.1 is insufficient for sampling from $\mathcal{S}_k$ since we will never be able to alter the places that a path hits the x-axis. For this reason, we need to introduce a slight variant $\widehat{\mathcal{M}}$ on $\mathcal{S}$ for the purposes of the analysis; the rapid mixing of the simpler chain $\mathcal{M}$ follows from the rapid mixing of $\widehat{\mathcal{M}}$ by a very simple application of the comparison method (Section 2.2).

In this new Markov chain $\widehat{\mathcal{M}}$ there are two basic types of moves. The first type of moves are inversions as described in Section 6.1. The second type of move consists of changing one "propeller-like" structure into its mirror image. Letting D denote a "down" edge and U an "up" edge, if there is a sequence of four edges $DUUD$, we can change it to the sequence $UDDU$, or vice-versa. These moves are only allowed when one point of the propeller touches the boundary. See Figure 5 for a pictorial depiction of this move. We call such a change a propeller move (centered) at v.

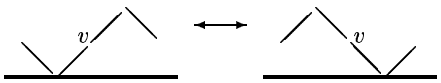

Figure 5. The propeller move

More formally, $\widehat{\mathcal{M}}$ iterates these steps.

One step of Markov chain $\widehat{\mathcal{M}}$:

1. Pick c uniformly at random from $\{2, \ldots, 2n-2\}$, and let v denote the point on the walk whose x-coordinate is c.
2. - If $c = 2$ and v is the bottom of a valley on the x-axis, with probability $\frac{1}{2(1+\lambda)}$ set X_{t+1} equal to X_t inverted at v.
   - If $c = 2$ and v is the top of a peak and inverting it will put it on the x-axis, with probability $\frac{\lambda}{2(1+\lambda)}$ set X_{t+1} equal to X_t inverted at v.
   - If $c \in \{3, 4, \ldots, 2n-2\}$, and v is the bottom of a valley not lying on the x-axis, or if v is the top of a peak and inverting at v does not put it onto the x-axis, with probability $\frac{1}{4}$ set X_{t+1} equal to X_t inverted at v.
   - If v is the central vertex of a propeller structure where the lowest point lies on the x-axis, with probability $\frac{1}{4}$ set X_{t+1} equal to X_t after performing a propeller move at v.
   - In all other cases, set $X_{t+1} = X_t$.

First note that $\widehat{\mathcal{M}}$ is aperiodic ($X_{t+1} = X_t$ with probability at least $\frac{1}{2}$). Second, the only time a transition is possible from a path hitting the boundary k times to one hitting the boundary $k+1$ (respectively $k-1$) times is when there is a peak (respectively valley) at the beginning of the walk, and we select that vertex in step 1 of the chain. All other moves of $\widehat{\mathcal{M}}$ preserve the weight of the walk.

Having described the Markov chain, we use it to define a metric d on $\mathcal{S}$. For any pair of states $X, Y \in \mathcal{S}$, if $P(X,Y) > 0$ (so X and Y are nearest neighbors), we define $d(X,Y)$ to equal one-half of the area of the symmetric difference of the two staircase walks. If $P(X,Y) = 0$ (i.e., moving from X to Y requires more than one move), first consider a *path of states* $X = \omega_0, \omega_1, \ldots, \omega_r = Y$ between X and Y, where $P(\omega_i, \omega_{i+1}) > 0$ for each $i = 0, \ldots r-1$; then define $d(X,Y) = \min \sum_{i=0}^{r-1} d(\omega_i, \omega_{i+1})$ where the minimum is taken over all paths joining X and Y. We call d the *transition metric*.

We apply the decomposition method of Section 3.2 to show rapid mixing of $\widehat{\mathcal{M}}$ on $\mathcal{S}$. To do so we need to examine the projection of the partition $\{\mathcal{S}_k\}$ and bound the spectral gap for the restriction to each of the subsets $\mathcal{S}_k$. We do this in the next sections following a brief combinatorial excursion.

6.3.2. A combinatorial look at $\mathcal{S}$ For this subsection, we let $\mathcal{S}^n$ denote the set of staircase walks with $2n$ edges and set $s^n = |\mathcal{S}^n|$. We use $\mathcal{S}_k^n$ to denote the subset of $\mathcal{S}^n$ containing those walks with k internal x-axis hits and let $s_k^n = |\mathcal{S}_k^n|$. These cardinalities can be shown to be log–concave, i.e., the sequence of numbers $s_0^n, s_1^n, s_2^n, \ldots, s_{n-1}^n$ is a log–concave sequence. This follows from two simple lemmas.

Lemma 6.2. *For $n \geq 3$ and $1 \leq k \leq n-2$, $s_k^n = s_{k-1}^{n-1} + s_{k+1}^n$.*

Proof. We partition $\mathcal{S}_k^n$ into two sets A_k^n, the subset of walks that begin with two edges UD, and B_k^n, the remaining walks in $\mathcal{S}_k^n$. We define two bijections. The first maps $\sigma \in A_k^n$ to an element in $\mathcal{S}_{k-1}^{n-1}$ by deleting the first two edges (the initial "bump" in the walk), and sliding the remaining piece of σ left two spaces. The second bijection is slightly more complicated; for $\sigma \in B_k^n$ let u denote the first point on the walk with y-coordinate equal to one (this is the point with x-coordinate equal to one). Let v be the first point to the right of u having y-coordinate one. Map σ to a walk in $\mathcal{S}_{k+1}^n$ by removing the first edge of σ, sliding the piece of σ between u and v to the left and down by one unit, adding an upwards edge at the end of this piece, and continuing the walk with the remainder of σ. See Figure 6 for an example of this mapping. It is easy to verify that each map is a bijection. □

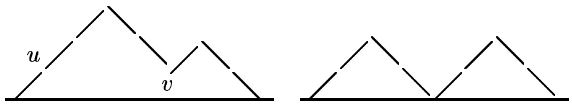

Figure 6. An example of the map from B_0^4 to $\mathcal{S}_1^4$

Lemma 6.3. *For $n \geq 3$ and $1 \leq k \leq n-2$,*

$$s_k^n = s_{k-1}^{n-1} + s_k^{n-1} + \cdots + s_{n-2}^{n-1}.$$

Also, for $n \geq 2$,

$$s_0^n = s^{n-1} = s_0^{n-1} + \cdots + s_{n-2}^{n-1} \qquad \textit{and}$$

$$s_0^n = s_1^n.$$

Proof. For the first part, we use Lemma 6.2 iteratively, so

$$\begin{aligned} s_k^n &= s_{k-1}^{n-1} + s_{k+1}^n \\ &= s_{k-1}^{n-1} + s_k^{n-1} + s_{k+2}^n = \cdots \end{aligned}$$

For the second, there is a bijection between $\mathcal{S}_0^n$ and $\mathcal{S}^{n-1}$ by taking $\sigma \in \mathcal{S}_0^n$, deleting the initial and terminal edges of the walk, and shifting the walk down and to the left by one to obtain a staircase walk joining $(0,0)$ to $(2n-2,0)$.

The above statement indicates that $s_0^n = C(n-1)$, so to prove the final claim of the lemma we want to show that $s_1^n = C(n-1)$ as well.

Recalling our notation that U represents an "up" edge and D represents a "down" edge, an element of $\mathcal{S}_1^n$ can be written as a string Uw_1DUw_2D, where w_1 is a staircase walk with $2j$ edges (j up edges and j down edges) and w_2 is a staircase walk with $2n-2j-4$ edges for some $j \in \{0,\ldots,n-2\}$. For each j, there are $C(j)$ possibilities for w_1 and $C(n-j-2)$ possibilities for w_2. Hence, we have

$$s_1^n = \sum_{j=0}^{n-2} C(j)C(n-j-2) = C(0)C(n-2) + C(1)C(n-3) + \ldots + C(n-2)C(0).$$

Checking the recurrence relation that defines the Catalan numbers (see [17]), we see this sum equals $C(n-1)$, as desired. □

Theorem 6.4. *For a fixed $n \geq 3$, s_k^n is log–concave. In particular, for $1 \leq k \leq n-2$,*

$$s_{k-1}^n \cdot s_{k+1}^n \leq (s_k^n)^2. \tag{6.1}$$

Proof. We use induction on n.

For $n = 3$, by a simple enumeration of the possibilities, we find that $s_0^3 = s_1^3 = 2$ and $s_2^3 = 1$, so that $s_0^3 \cdot s_2^3 \leq (s_1^3)^2$.

Now assume for some $n-1$ that s_k^{n-1} is log–concave. Also, assume first that $k \geq 2$. We want to show that (6.1) holds. To do this, it suffices to show the inequality

$$s_{k-2}^{n-1} \cdot [s_k^{n-1} + s_{k+1}^{n-1} + \cdots + s_{n-2}^{n-1}] \leq s_{k-1}^{n-1} \cdot [s_{k-1}^{n-1} + s_k^{n-1} + \cdots + s_{n-2}^{n-1}] \tag{6.2}$$

since Lemma 6.3 implies that (6.2) is equivalent to

$$s_{k-2}^{n-1} \cdot s_{k+1}^n \leq s_{k-1}^{n-1} \cdot s_k^n. \tag{6.3}$$

By adding $s_k^n \cdot s_{k+1}^n$ to both sides of (6.3), factoring, and applying Lemma 6.2, we get (6.1).

To show (6.2), it suffices to show the set of inequalities

$$s_{k-2}^{n-1} \cdot s_{k-1+i}^{n-1} \leq s_{k-1}^{n-1} \cdot s_{k-2+i}^{n-1}$$

for all $i \in [n-k-1]$. These inequalities all hold by our induction hypothesis that s_k^{n-1} is log–concave. Adding them, and the extra term $s_{k-1}^{n-1} \cdot s_{n-2}^{n-1}$ to the right hand side, gives us (6.2).

All that remains is the case $s_0^n \cdot s_2^n \leq (s_1^n)^2$ (when $k = 1$). We use that $s_0^n = s_1^n$ from Lemma 6.3 and, from Lemma 6.2, we see $s_2^n \leq s_1^n$. Therefore, $s_0^n \cdot s_2^n = s_1^n \cdot s_2^n \leq s_1^n \cdot s_1^n$. □

6.3.3. Projection 1: $\mathcal{S} = \dot{\cup}\mathcal{S}_k$ We bound the mixing rate of the projection by appealing to Theorem 6.4. The (disjoint) projection P_M of the partition $\mathcal{S} = \dot{\cup}_{k=0}^{n-1}\mathcal{S}_k$ is a random walk by a "hiker" on $\{0, \ldots, n-1\}$ with stationary probabilities $\rho(k) = \frac{\lambda^k|\mathcal{S}_k|}{Z}$, where $Z = \sum_{k=0}^{n-1}\lambda^k|\mathcal{S}_k|$, and Metropolis transitions with holding probability (at least) $1/2$, i.e., $P_M(k,j) = \frac{1}{4}\min\{1, \frac{\rho(j)}{\rho(k)}\}$ for $j \in \{k-1, k+1\}$, $P_M(k,k) = 1 - P_M(k-1) - P_M(k+1) \geq \frac{1}{2}$, and $P_M(k,j) = 0$ for $j \notin \{k-1, k, k+1\}$. Path coupling yields the following result:

Lemma 6.5. *The projection chain P_M is rapidly mixing and*

(a) $\tau_{P_M}(\varepsilon) = O(n^2 \log(\frac{1}{\varepsilon}))$.
(b) $Gap(P_M) \geq \frac{c_1}{n^2}$ *for some constant* c_1.

Proof. The distance between two configurations is simply the distance between the two hikers. Using path coupling we must consider a pair of states differing by unit distance as in Figure 7, where the hiker is at position k in the upper configuration, σ_1, and $k+1$ in the lower one σ_2.

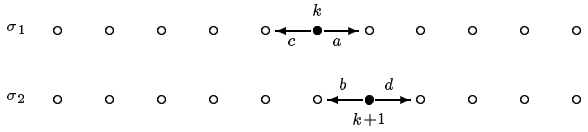

Figure 7. Typical situation for path coupling in P_M

We couple the moves so that the hiker moves in only one of the configurations during each time step of the coupled process, i.e., if we move the hiker left or right in σ_1, then we hold the hiker fixed in σ_2. The $1/2$ holding probabilities allow us to couple the moves in this fashion. Letting a, b, c, d denote the transition probabilities as labeled in Figure 7, the expected change in distance after one move in the coupled chain is $(-1)\cdot(a+b) + (1)\cdot(c+d)$. To apply Theorem 2.2 we need to show this quantity is less than or equal to zero, or, equivalently, show $a + b \geq c + d$. Consider first the case when $\lambda = \frac{|\mathcal{S}_k|}{|\mathcal{S}_{k+1}|}$. With this value, we have that $a = b = \frac{1}{4}$, $c = \frac{1}{4}\min\{1, \frac{|\mathcal{S}_{k+1}|\cdot|\mathcal{S}_{k-1}|}{|\mathcal{S}_k|^2}\} \leq \frac{1}{4}$, and $d = \frac{1}{4}\min\{1, \frac{|\mathcal{S}_k|\cdot|\mathcal{S}_{k+2}|}{|\mathcal{S}_{k+1}|^2}\} \leq \frac{1}{4}$. So $a + b \geq c + d$ as desired.

If $0 < \lambda < \frac{|\mathcal{S}_k|}{|\mathcal{S}_{k+1}|}$ we note that $a = \frac{1}{4}\frac{\lambda|\mathcal{S}_{k+1}|}{|\mathcal{S}_k|} \geq \frac{1}{4}\frac{\lambda|\mathcal{S}_{k+2}|}{|\mathcal{S}_{k+1}|} = d$, where the inequality follows using log–concavity of the sequence $|\mathcal{S}_k|$. Since $b = \frac{1}{4}\min\{1, \frac{|\mathcal{S}_k|}{\lambda|\mathcal{S}_{k+1}|}\} = \frac{1}{4} \geq c$ we again have $a + b \geq c + d$. The last case where $\lambda > \frac{|\mathcal{S}_k|}{|\mathcal{S}_{k+1}|}$ is similar, so that $a + b \geq c + d$ for all $\lambda > 0$.

Appealing to Theorem 2.2, noting that the maximum distance between two hikers is at most n and $\eta = \frac{1}{4}$ suffices in the hypothesis of the theorem (either $a = \frac{1}{4}$ or $b = \frac{1}{4}$ for

any value of λ) we have

$$\tau_{P_M}(\varepsilon) \leq \lceil \frac{en^2}{\frac{1}{4}} \rceil \lceil \log(\frac{1}{\varepsilon}) \rceil = O(n^2 \log(\frac{1}{\varepsilon})).$$

□

6.3.4. Restriction 1: Mixing on S_k By the disjoint decomposition theorem, it suffices to show that the restricted Markov chains ($\widehat{\mathcal{M}}$ restricted to S_k) are rapidly mixing in order to conclude that $\widehat{\mathcal{M}}$ is mixing on the whole state space $\mathcal{S}$. We show this in the next sections.

6.3.5. Decomposition of S_k In this section we show that $\widehat{\mathcal{M}}_{S_k}$, the Markov chain restricted to $\mathcal{S}_k$, is rapidly mixing. To do this we apply the decomposition method a second time. First we partition $\mathcal{S}_k$ and show that $\widehat{\mathcal{M}}_{S_k}$ is mixing when restricted to each of set of this partition. Following that, we show the projection is mixing using heat bath dynamics and the comparison theorem, setting the stage for Corollary 3.3.

Let T denote a subset of $\mathcal{S}_k$ where each walk touches the x-axis in the same k locations. For example, (in the case that $n \geq 6$) we can consider the set of walk that hit the x-axis at the points with x-coordinates $2, 6$, and 10 in the interior between the two endpoints. There are $\binom{n-1}{k}$ ways to specify the location of k internal hits, as the x-coordinate of each hit must be an even number. We write $\mathcal{S}_k = \dot{\cup}_T S_{k,T}$, where this union is over all $\binom{n-1}{k}$ ways of specifying the hits on the x-axis.

6.3.6. Restriction 2: Mixing of $\mathcal{S}_{k,T}$ Let $\widehat{\mathcal{M}}_{S_{k,T}}$ denote the restriction of $\widehat{\mathcal{M}}_{S_k}$ to the set T. We have the following result, whose proof is a simple application of path coupling:

Lemma 6.6. *Let T be a subset of $\mathcal{S}_k$ as above, and let $X, Y \in T$ with $d(X,Y) = 1$. After one step of the Markov chain, $\widehat{\mathcal{M}}_{S_{k,T}}$, on the set T, we have $E(\Delta d(X,Y)) \leq 0$.*

Proof. This proof is similar to Lemma 6.1, except that we need only consider the situation when the paths differ by a square that is at least distance one from the x-axis as in Figure 8. There are two good inversions at v that decrease the distance by one, and at most two inversions increasing the distance by one. Each of these inversions happens with equal probability, so $E(\Delta d(X,Y)) \leq 0$. □

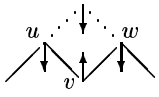

Figure 8. Possible transitions inside $\mathcal{S}_{k,T}$

Lemma 6.6 gives the first piece for the path coupling theorem. If $d(X,Y) > 0$, the probability of the distance changing in one step of $\widehat{\mathcal{M}}_{S_{k,T}}$ is at least $\frac{1}{4(n-k)}$, since when

we select a vertex on the walk we may avoid choosing one that lies on the x-axis, and its immediate neighbor to the right, as these vertices will never move.

For pairs of walks in $\mathcal{S}_k$, we have $0 \le d(X,Y) \le (n-k-1)^2 - (n-k-1)$. For $X, Y \in \mathcal{S}_{k,T}$ there is a sequence of $d(X,Y)$ inversions that will transform one walk into the other.

By a straightforward application of Theorem 2.2, we have

Lemma 6.7. *The mixing time of $\widehat{\mathcal{M}}_{S_k,T}$ satisfies*

$$\begin{aligned} \tau_{\widehat{\mathcal{M}}_{S_k,T}}(\varepsilon) &\le \lceil \frac{e((n-k-1)(n-k-2))^2}{\frac{1}{4(n-k)}} \log(\frac{1}{\varepsilon}) \rceil \\ &= O(n^5 \log(\frac{1}{\varepsilon})). \end{aligned}$$

6.3.7. Projection 2: $\mathcal{S}_k = \dot{\cup}_T \mathcal{S}_{k,T}$ The projection P_k of $\mathcal{S}_k = \dot{\cup}_T \mathcal{S}_{k,T}$ can be viewed as a particle process on $[n-1]$. The particles represent the places that a staircase walk hits the x-axis. The projection of the Markov chain can be viewed as nearest-neighbor dynamics on the set of particles which moves one particle to the left or right in each step. Interestingly, analyzing the mixing rate of this particle process cannot be done using a simple path coupling argument, which seems to isolate the difficulty with using path coupling on the original mountain/valley chain. However, we may use the analysis of exclusion processes from Section 5 to bound the mixing time of this projection.

In the framework of Section 5, we have a linear arrangement of $n-1$ sites, k of which contain particles. (There we have n sites, but the result is the same.) Denoting the collection of all $\binom{n-1}{k}$ arrangements by Ω_k, an element $\sigma \in \Omega_k$ corresponds to the set $\mathcal{S}_{k,T(\sigma)}$ of all staircase walks in $\mathcal{S}_k$ that have x-axis hits at the locations determined by the particles in σ. For example, if $k=3$ and σ is the configuration with particles at sites 2, 3, and 8, then $\mathcal{S}_{3,T(\sigma)}$ consists of all walks in $\mathcal{S}_3$ that hit the x-axis at coordinates 4, 6, and 16 (recalling that walks only hit the x-axis at even coordinates). The fixed particles added to each configuration at sites 0 and n correspond to the fixed endpoints of each staircase walk.

In this particle process the Catalan numbers are used to define weights on each configuration to count the number of paths in each set of the partition of $\mathcal{S}_k$. That is, the stationary probability $\rho(\sigma)$ of a configuration $\sigma \in \Omega_k$ is proportional to $C(x_1)C(x_2)\cdots C(x_{k+1})$, where x_j is the number of empty sites between particles j and $j+1$ in σ. The transitions of the projection are the single site dynamics described in Section 5.1 but, as mentioned, path coupling fails to directly show rapid mixing. However, taking $f_t = C(t)$, we check that the Catalan numbers satisfy the two conditions of Theorem 5.5, namely they are log–convex, and $\frac{C(t)}{C(t-1)} \le \frac{g_{t+1}}{g_t}$ (for $t \ge 1$) where $g_t = \sum_{i=1}^{t} C(i-1)C(t-i)$. Using that $C(t) = \frac{1}{t+1}\binom{2t}{t}$ and the recurrence relation that defines the Catalan numbers (see [17]), it is easy to show that they are log–convex and that $g_t = C(t)$. In fact $\frac{C(t)}{C(t-1)} \le \frac{g_{t+1}}{g_t} = \frac{C(t+1)}{C(t)}$ follows by the log–convexity of the Catalan numbers so Theorem 5.5 applies and the gap of the related heat bath chain $\tilde{P}_k$ has a lower bound of $\frac{c}{n^5}$, for some constant c. We then lower bound the gap of the projection P_k using the comparison theorem.

Lemma 6.8. *The Markov chain P_k on Ω_k with nearest-neighbor dynamics is mixing in polynomial time and $Gap(P_k) \geq \frac{c_2}{n^{19/2}}$, for some constant c_2.*

Proof. By part (a) of Theorem 2.3 it suffices to bound the parameter A for the chains P_k and $\tilde{P}_k$ to find a lower bound on $Gap(P_k)$.

A transition of the heat bath chain can be modeled by a sequence of nearest-neighbor transitions in an obvious manner, because a particle never moves over (or through) its neighboring particle(s). Thus, in the notation of Theorem 2.3, for any pair x, y with $(x, y) \in E(\tilde{P}_k)$ we see that $|\gamma_{xy}| \leq n - k$, and for any (nearest-neighbor) transition (z, w) of P_k, we have that $|\Gamma(z, w)| \leq n - k$.

We also note $\frac{\rho(x)}{\rho(z)P(z,w)} \leq c'\, k(n-k)^{3/2}$ for some constant $c' > 0$. Here we have used the fact that $C(t)$ is asymptotic to $4^t/(\sqrt{\pi}\ t^{3/2})$. The ratio $\frac{\rho(x)}{\rho(z)P(z,w)}$ has the form

$$4k \cdot \frac{C(i-1)C(m-i)}{C(j-1)C(m-j)} \tag{6.4}$$

for some $i, j \in \{1, \ldots m\}$ since x, y, z, w agree everywhere except at the position of a single particle. We maximize this expression by taking $i = 1$ and $j = \lfloor \frac{m}{2} \rfloor$. The asymptotic expression for the Catalan numbers, noting that $m \leq n - k$ for configurations having k particles, gives the stated bound on the ratio in (6.4).

Finally, we use the trivial bound $\tilde{P}_k(x, y) \leq \frac{1}{4k}$ (from the heat bath transition probabilities in equation (5.1) since $x \neq y$). We conclude that

$$A \quad \leq \quad c'\, k(n-k)^{3/2} \cdot (n-k) \cdot (n-k)^2 \cdot \frac{1}{4k} = \mathrm{O}(n^{9/2}).$$

The factor $(n-k)^2$ comes from the number of possible (x, y) pairs, based on the initial and final locations of the particle in question.

Combining this with the lower bound of $\frac{c}{n^5}$ for the eigenvalue gap of the heat bath chain we find that $Gap(P_k) \geq \frac{c_2}{n^{19/2}}$. □

We now have shown in Lemmas 6.7 and 6.8 that the restrictions defined by the decomposition $\mathcal{S}_k = \dot{\cup}_T \mathcal{S}_{k,T}$, as well as the projection, are all mixing in polynomial time. Appealing to Corollary 3.3 with $\beta = \frac{1}{8n}$ and $\gamma = \frac{1}{4}$, we find:

Lemma 6.9. *The collection of restricted Markov chains $\widehat{\mathcal{M}}_{\mathcal{S}_k}$ on the sets $\mathcal{S}_k$ are rapidly mixing and $Gap(\widehat{\mathcal{M}}_{\mathcal{S}_k}) \geq \frac{c_3}{n^{31/2}}$ for all $k \in \{0, \ldots, n-1\}$.*

Proof. When $k = 0$, the result of Lemma 6.7 directly applies to the mixing time of $\widehat{\mathcal{M}}_{\mathcal{S}_0}$ (as there is only one set in the "partition" of $\mathcal{S}_0$). This, together with Theorem 2.1(b), shows that

$$Gap(\widehat{\mathcal{M}}_{\mathcal{S}_0}) \geq \frac{r}{n^5}$$

for some constant $r > 0$.

Similarly, for $k \geq 1$, Lemma 6.7 also gives the same bound for the spectral gap of the

restrictions, namely

$$Gap(\widehat{\mathcal{M}}_{\mathcal{S}_k,T}) \geq \frac{r}{n^5}$$

for each set T in the partition of $\mathcal{S}_k$. Lemma 6.8 bounds the gap of the projection of the partition of $\mathcal{S}_k$, so using the disjoint decomposition method (Corollary 3.3), recalling that we take $\beta = \frac{1}{8n}$ and $\gamma = \frac{1}{4}$, we find

$$Gap(\widehat{\mathcal{M}}_{\mathcal{S}_k}) \geq \frac{1}{8n} \cdot \frac{1}{4} \cdot 2 \cdot \frac{c_2}{n^{19/2}} \cdot \frac{r}{n^5} = \frac{c_3}{n^{31/2}}.$$

Here we have used the fact that the maximum degree of any of the projection chains P_k is at least 2 (and this is achieved when $k = 1$). □

6.3.8. Mixing for $\mathcal{S}$: The final word A polynomial bound on the mixing time for $\widehat{\mathcal{M}}$ now follows from all of our previous work. By Lemma 6.9 the restrictions to each set $\mathcal{S}_k$ are all rapidly mixing, and the mixing time of the projection followed from the log–concavity of the sets $\mathcal{S}_k$ (Lemma 6.5). Using Corollary 3.3, we can bound the mixing time of the chain $\widehat{\mathcal{M}}$ on all of $\mathcal{S}$. Note that in this case we have $\beta = \frac{1}{4n(\lambda+1)}$ and $\gamma = \frac{1}{\lambda n}$. Also, since the projection chain is a simple random walk on $\{0, \ldots, n-1\}$ (recall Section 6.3.3), we have $\Delta = 2$ in this application of Corollary 3.3.

Theorem 6.10. *The Markov chain $\widehat{\mathcal{M}}$ on $\mathcal{S}$ is rapidly mixing and $Gap(\widehat{\mathcal{M}}) \geq \frac{c_4}{\lambda^2 n^{39/2}}$.*

Finally, a simple application of the comparison theorem establishes a polynomial bound for the mixing time of the original mountain/valley Markov chain $\mathcal{M}$ on $\mathcal{S}$.

Theorem 6.11. *The Markov chain $\mathcal{M}$ on $\mathcal{S}$ is rapidly mixing and*

(a) $Gap(\mathcal{M}) \geq \frac{c_5}{\lambda^3 n^{39/2}}$.
(b) $\tau_{\mathcal{M}}(\varepsilon) = O(\lambda^3 n^{41/2} \log(\frac{\lambda}{\varepsilon}))$.

Proof. The only interesting case to consider is a propeller move in $\widehat{\mathcal{M}}$. This move can be broken into a pair of inversions in $\mathcal{M}$. The propeller move occurs with probability $\frac{1}{4n}$, while each inversion happens with probability $\frac{\lambda}{(2n)2(1+\lambda)}$ or $\frac{1}{(2n)2(1+\lambda)}$. Since we consider the case with $\lambda > 1$ we have $A = 2(1+\lambda)$. From Theorem 2.3(a) we have $Gap(\mathcal{M}) \geq \frac{c_5}{\lambda^3 n^{39/2}}$. Using Theorem 2.1(a), observing that $\pi_* = \min_{x \in \mathcal{S}} \pi(x) \geq \frac{1}{4^n \lambda^n}$ so $\log(\frac{1}{\pi_* \varepsilon}) \leq n \log(\frac{4\lambda}{\varepsilon})$, we derive the stated bound on $\tau_{\mathcal{M}}(\varepsilon)$. □

7. Conclusions

As probably demonstrated by this last example on staircase walks, indirect methods such as decomposition and comparison should be used sparingly or the derived bounds on the mixing time can be incredibly weak. We feel it is necessary to emphasize that the utility of the decomposition method is primarily to establish whether a Markov chain converges in polynomial time. Moreover, analysis from the decomposition method can provide insights into why direct methods appear to fail, such as was the case with the particle processes

arising from the second projection. In fact, isolating such underlying structures can be instrumental for uncovering alternative direct proofs that would provide more reasonable bounds on the mixing time.

The major benefit of the decomposition method is for initially assessing the mixing rate of a chain by offering a top-down approach for systematically simplifying the chain into more manageable pieces. First, simpler arguments can often be used to bound the mixing rate of the pieces. Second, a hybrid approach is possible whereby the restrictions, for example, are analyzed through path coupling and the projection is analyzed by bounding some isoperimetric constant for the chain.

We leave as open problems finding rapidly mixing Markov chains for sampling circuits on other Cayley graphs. It would also be interesting to study whether decomposition can be used to show that chains are mixing on connected subsets of Cayley graphs, such as a subregion of $\mathbb{Z}^d$, since here alternative recursive methods would no longer be possible.

Acknowledgements

The authors are very grateful to Alan Sokal for observing that Theorem 3.1 could be used to greatly simplify the statement and proof of Theorem 3.2.

References

[1] R. Bubley and M. Dyer. Faster random generation of linear extensions. *Discrete Mathematics* **201** (1999), pp. 81–88.

[2] R. Bubley and M. Dyer. Path coupling: A technique for proving rapid mixing in Markov chains. *Proc. 38th Annual IEEE Symposium on Foundations of Computer Science (FOCS 1997)*, pp. 223–231.

[3] P. Diaconis and L. Saloff-Coste. Comparison theorems for reversible Markov chains. *Annals of Applied Probability* **3** (1993), pp. 696–730.

[4] M. Dyer and C. Greenhill. A more rapidly mixing Markov chain for graph colorings. *Random Structures and Algorithms* **13** (1998), pp. 285–317.

[5] M.R. Jerrum and A.J. Sinclair. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation* **82** (1989), pp. 93–133.

[6] M. Jerrum and J.-B. Son. Spectral gap and log-Sobolev constant for balanced matroids. *Proc. 43rd IEEE Symposium on Foundations of Computer Science (FOCS 2002)*, pp. 721–729.

[7] M. Jerrum, J.-B. Son, P. Tetali, and E. Vigoda. Elementary bounds on Poincaré and log-Sobolev constants for decomposable Markov chains. Isaac Newton Institute preprint NI03006-CMP, available at `http://www.newton.cam.ac.uk/preprints/NI03006.pdf`.

[8] M. Luby, D. Randall, and A.J. Sinclair. Markov chain algorithms for planar lattice structures. *SIAM Journal on Computing* **31** (2001), pp. 167–192.

[9] N. Madras and D. Randall. Factoring graphs to bound mixing rates. *Proc. 37th Annual IEEE Symposium on Foundations of Computer Science (FOCS 1996)*, pp. 194–203.

[10] N. Madras and D. Randall. Markov chain decomposition for convergence rate analysis. *Annals of Applied Probability* **12** (2002), pp. 581–606.

[11] R. Martin and D. Randall. Sampling adsorbing staircase walks using a new Markov chain decomposition method. *Proc. 41st Annual IEEE Symposium on Foundations of Computer Science (FOCS 2000)*, pp. 492–502.

[12] L. McShine and P. Tetali. On the mixing time of the triangulation walk and other Cata-

lan structures. *Randomization methods in algorithm design*, *DIMACS Series on Discrete Mathematics and Theoretical Computer Science* **43** (1999), pp. 147–160.

[13] D. Randall and P. Tetali. Analyzing Glauber dynamics by comparison of Markov chains. *Journal of Mathematical Physics* **41** (2000), pp. 1598–1615.

[14] B. van Rensburg. Adsorbing staircase walks and staircase polygons. *Annals of Combinatorics* **3** (1999), pp. 451–473.

[15] E.J.J. van Rensburg. Collapsing and adsorbing polygons. *Journal of Physics A: Mathematical and General* **31** (1998), pp. 8295–8306.

[16] A.J. Sinclair. *Algorithms for Random Generation & Counting: A Markov Chain Approach.* Birkhäuser, Boston, 1993.

[17] R. Stanley. *Enumerative Combinatorics*, Vol. II, Cambridge University Press, 1999.

[18] D.B. Wilson. Mixing times of lozenge tiling and card shuffling Markov chains. *Annals of Applied Probability*, to appear.