Cluster Algorithms for Discrete Models of Colloids with Bars

Sarah Miracle* Dana Randall† Amanda Pascoe Streib‡

Abstract

Colloids are mixtures of two different types of molecules. The model has a hard-core constraint forcing all the molecules to occupy non-overlapping positions, but there are no additional interactions between molecules; all non-overlapping arrangements are equally likely. It is believed that colloids undergo a phase transition whereby at low density the two types of molecules will be uniformly interspersed, while at high density large clusters will form and the two types of molecules will effectively separate. While local algorithms are not believed to work at or beyond the critical point, an algorithm due to Dress and Krauth [3] offers an alternative approach to sampling potentially beyond the critical point where clusters begin to form. We study the DK algorithm on a colloid model consisting of long bars and small diamonds on the periodic lattice $\mathbf{Z}_n^2$. We show that if we restrict the model to allow at most one bar in each column of the lattice region, then local algorithms are slow, but the DK algorithm is provably efficient (if the bars are long enough). However, we show that when we allow any number of bars per column, the DK algorithm also requires exponential time to reach equilibrium.

1 Introduction

Colloids are binary mixtures of substances that separate into clusters at sufficiently high density. When the density of each substance in the suspension is low, the mixture looks well-mixed and each substance looks evenly dispersed within the other. In contrast, when the density of each is high, the two substances separate whereby particles of one type join together to form large clusters sitting in a sea of molecules of the other type. Although this behavior is similar to other discrete models, such as the ferro-magnetic Ising model, here the two types of particles do not possess any *enthalpic* forces causing like particles to attract or disparate particles to repel. In contrast, the behavior of colloids is purely *entropic* — the only restriction is a "hard-core" constraint requiring objects to remain in non-overlapping positions, and clustering occurs at high density because the overwhelming majority of configurations in the stationary distribution are believed to exhibit such a separation.

While the emperical study of colloids is pervasive in surface chemistry, material science, physics, and nanotechnology, there has been little rigorous work explaining their behavior. In 1992, Frenkel and Louis [4] defined a discrete model that packs unit squares in a region of $\mathbf{Z}^2$ along with "specks" or diamonds of area $1/2$ that sit on lattice edges. They show that this colloid model corresponds precisely to the Ising model, where the density of squares fixes the magnetization of the Ising magnet and the density of specks determines the temperature. The colloid model then inherits the phase transition arising in the Ising model, showing that as the density of specks increases, effectively lowering the temperature in the Ising model, there will be a point beyond which the predicted clusters will form in the colloid model. However, this reduction does not readily generalize to any other models of colloids.

Given the dearth of rigorous analysis characterizing the distributions at high and low density, one would expect that sampling would be an invaluable experimental tool. However, designing rigorous sampling algorithms has proven equally challenging: intuitively, at low density standard approaches to sampling can be used to illustrate heterogeneous mixtures of particles, but sampling algorithms based on local updates are no longer efficient at high density – precisely because we expect large clusters reminiscent of the low temperature Ising model to form. In 1995 Dress and Krauth [3] introduced a clever non-local algorithm that they believed would be efficient at intermediate densities, even beyond the phase transition, and in 1998 Buhot and Krauth [2] provided simulations finally demonstrating heuristic evidence of the presence of two distinct phases in colloid models consisting of different sized squares.

Informally, the DK algorithm, which requires the colloid model to lie on a torus, operates as follows. In each step, a pivot on the torus is chosen and a dual image, formed by rotating the configuration 180°

*College of Computing, Georgia Institute of Technology, Atlanta, GA 30332-0765. Supported in part by NSF CCF-0830367 and a DOE Office of Science Graduate Fellowship.

†College of Computing, Georgia Institute of Technology, Atlanta, GA 30332-0765. Supported in part by NSF CCF-0830367 and CCF-0910584.

‡School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0280. Supported in part by the National Physical Sciences Consortium Fellowship, Georgia Institute of Technology ACO Fellowship, and NSF CCF-0910584.

around the pivot, is superimposed on top of the primal (original) configuration. For each component consisting of overlapping primal and dual clusters, one or the other is chosen; note that the complementary choice must be made for the corresponding (dual) component (see Figure 2). The DK algorithm potentially allows large parts of the configuration to be updated in one move which seems necessary at high density.

In fact, the DK algorithm can be generalized to a large class of hard-core models with symmetry, including independent sets on periodic lattices, although the efficiency of such an algorithm is not yet understood. This approach is compelling because various cluster algorithms have been used extensively to sample from the Gibbs (or Boltzmann) distribution for other fundamental models, such as the Ising and Potts models from statistical physics. While the Swendsen-Wang algorithm is used widely in practice to sample Ising configurations below the critical temperature when local algorithms fail to converge quickly, careful analysis of the algorithm revealed that it is not efficient at the critical point where it requires exponential time to reach equilibrium [5]. Likewise, approaches such as parallel and simulated tempering that try to circumvent obstacles causing slow mixing at low temperature by enlarging the state space are known to work in certain circumstances [1, 15], but further anlaysis has revealed that they can also require exponential time to converge [1, 14]. The rigorous analysis of these chains has helped us understand when these algorithms are sufficient for sampling and when alternative approaches are necessary.

1.1 Our results. We study the DK cluster algorithm on hard-core mixtures of "bars," or long, thin axis-aligned rectangles, and specks on the $n \times n$ torus T_n in Z^2. Bars have width 1 and length $L = c_1 n$ for any constant $0 < c_1 < 1$, so we should think of them as being microscopic in one dimension and macroscopic in the other. Specks are diamonds with height and width one that are bisected by edges of the lattice. When studying colloids, specks are a convenient choice because they cannot overlap; this is the same reason that Frenkel and Louis [4] use specks as the medium of suspension in their model that corresponds to the Ising model. We use bars for the second substance because models where both dimensions are microscopic include notoriously challenging sampling problems such as the Ising model (squares and specks) and independent sets (specks centered at odd vs. even lattice sites). By allowing one dimension to be large, we are able to derive rigorous bounds on the mixing time of the DK and local Markov chain algorithms at high density. In Section 3 we show that for the *restricted model*, when each column can have at most one bar, the DK algorithm is efficient. However, we show in Section 4 that for the *unrestricted model*, when a column can have any number of bars, the DK Markov chain can take exponential time to converge. In addition, we show that the local algorithm that only moves one object at a time is slow in both of the above settings. Thus, our results on the restricted model demonstrate that the DK cluster algorithm can be fast in the exact setting where the local algorithm is slow, as conjectured. However, for the unrestricted model we find that the DK algorithm is very sensitive to the specific problem being solved and can require exponential time with only minor changes to the definition of the state space.

1.2 Techniques. For all of the proofs we rely critically on a decomposition theorem for Markov chains [9, 10]. The theorem considers a partition of the state space Ω of any Markov chain into sets, and relates the mixing time of the original chain to the mixing time of the chain restricted to each set, as well as the mixing time of a projection chain capturing the flow between the sets. We define a partition according to the position of all of the bars (which we call the *bar structure*). The projection has state space $\widehat{\Omega}$ consisting of all possible bar structures. The stationary distribution of the projection is not uniform over $\widehat{\Omega}$, but rather each configuration σ has weight proportional to $\mu^{\kappa(\sigma)}$ where μ is a parameter related to the density and $\kappa(\sigma)$ is the total length of the perimeter of the bars in σ or the total contour length.

This transformation is significant because Markov chains on contour models have been extensively studied. For example, the Ising model can also be interpreted as a contour model where contours are edges in the dual, separating sites assigned + and -, and the weight of a configuration τ is proportional to $\mu^{\kappa(\tau)}$ where μ is a parameter related to temperature and $\kappa(\tau)$ is the total length of the contours. A special case of our contour model when there is one connected component where the first and last bars are at the same height is almost identical to the solid-on-solid model recently analyzed by Martinelli and Sinclair [11].

Bounding the convergence time of algorithms for sampling that make large changes to a configuration in one move is always challenging. Our analysis of the DK algorithm in the restricted setting relies on a complex canonical paths argument enabling us to encode information while preserving perimeter. The proof that the DK algorithm is slow in the unrestricted setting requires a combinatorial mapping argument to reveal a small cut in the state space.

2 Colloids and mixing times

We begin by formally defining our colloid model and the two Markov chains we will study. We also present some background on Markov chains and the main techniques used in our proofs.

2.1 The model. Consider the $n \times n$ discrete torus T_n and embedded graph $G = (V, E)$ with vertex set $V = \{(x, y) : x, y \in \mathbb{Z}^n\}$ and edge set E, where $((x_1, y_1), (x_2, y_2)) \in E$ if $|x_1 - x_2| + |y_1 - y_2| = 1$. We study the set of all non-overlapping packings of T_n of vertical bars on the faces of G and specks on the edges of G with a fixed density of each type of *tile* (bars and specks). Recall that when the density of bars and specks is high, we expect the bars to cluster together and the specks to be dispersed around them. One goal would be to sample from the uniform distribution over this set. We make a simplifying assumption allowing the number of specks to vary, and we sample from a weighted distribution where a configuration σ with d specks is given weight $\pi(\sigma) = \lambda^d/Z$ for constant $\lambda > 1$ and normalizing constant Z. The choice of λ effectively determines the density and this model can be shown to be stochastically equivalent to the original model where the density of both types of tiles is fixed. Moreover, by adding separate parameters for the density of bars and the expected density of specks, we can better understand the effect of each on the mixing time.

Given constants $\lambda > 1$, $0 < c_1 < 1$, and $0 < c_2 < 1$, where $c_1 n, c_2 n \in \mathbb{Z}$, let $L = c_1 n$ and $b = c_2 n$. Define $\Omega = \Omega(b, L, \lambda)$ to be the set of non-overlapping packings of T_n with b $(1 \times L)$ bars on faces and any number of specks (diamonds with height and width 1) on edges. We wish to sample according to the distribution

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

where $d(\sigma)$ is the number of specks in σ and $Z = \sum_{\sigma \in \Omega} \lambda^{d(\sigma)}$ is a normalizing constant. See Figure 1. Define $\Omega' \subset \Omega$ to be the set of configurations with the restriction that every column of T_n has at most one bar in it, and let π' be its restricted distribution with normalizing constant $Z' = \sum_{\sigma \in \Omega'} \lambda^{d(\sigma)}$. Define a *component* of σ to be a maximal set of adjacent columns each containing at least one bar. The requirement that $c_2 < 1$, which implies $b \leq n - 1$, is a useful simplification that prevents connected components of bars from wrapping around the torus.

2.2 Markov chains for colloids. As we are interested in sampling from the distribution π in order to determine whether clusters are forming, it is natural to design a Markov chain whose state space is Ω and whose stationary distribution is π (or Ω' and π'). There are many ways to do this. The simplest is the local algorithm that moves one speck or bar at a time. The DK algorithm is an alternative approach that has the same state space and limiting distribution, but potentially moves many tiles in one move. We will see rigorously how in some circumstances the DK algorithm is able to get around the obstacles to rapid mixing which the local algorithm encounters.

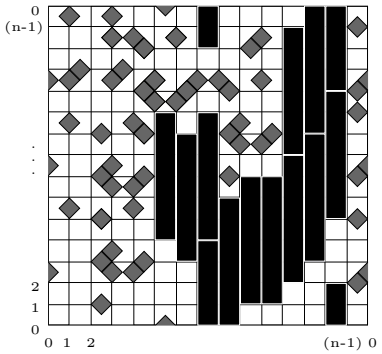

Figure 1: b bars on the $n \times n$ torus with specks.

The local Markov chain $\mathcal{M}_{Local}$

`Starting at any` σ_0`, iterate the following:`

- `With prob.` $1/2$`, pick a bar` t `and an offset` $u \in \{-1, 0, 1\} \times \{-1, 0, 1\}$ `u.a.r. (uniformly at random).`
  - `Shift the top left corner of` t `by` u `(mod` n`), if possible.`
- `With prob.` $1/2$`, pick an edge` $e \in E$ `u.a.r.`
  - `If` e `is not incident to a bar or a speck then add a speck at` e`.`
  - `Otherwise, if` e `is incident to a speck, then remove the speck with prob.` λ^{-1}`.`

This Markov chain clearly connects the state space, since it can remove all specks and then move the bars one at a time, e.g., to form a square in the top left corner. Since $\mathcal{M}_{Local}$ is also *aperiodic* (i.e., for all $\sigma, \rho \in \Omega$, g.c.d.$\{t \geq 1 : P^t(\sigma, \rho) > 0\} = 1$), $\mathcal{M}_{Local}$ converges to a unique stationary distribution (see, e.g., Chapter 2 of [13]). Moreover, since $\mathcal{M}_{Local}$ is *reversible* with respect to the distribution π over state space Ω (i.e., $\pi(\sigma)\mathcal{P}_{Local}(\sigma, \rho) = \pi(\rho)\mathcal{P}_{Local}(\rho, \sigma)$ for all $\sigma, \rho \in \Omega$), it follows that π is the unique stationary distribution of $\mathcal{M}_{Local}$ over Ω [13]. Analogously, π' is the stationary distribution of $\mathcal{M}_{Local}$ over Ω'. We will see in Section 3

that this Markov chain takes exponential time to converge to the stationary distribution π' since if $\mathcal{M}_{Local}$ tries to move a bar, it will fail whenever the new location contains any specks. The idea behind the DK Algorithm is to avoid simply rejecting these moves by allowing a bar to trigger a larger action causing everything it overlaps to move out of the way. This is achieved by pivoting the whole configuration σ around a pivot point p on the torus T_n to create a new configuration ρ, and superimposing σ and ρ to obtain ξ. See Figure 2. We say that two tiles are *DK-adjacent* if they overlap in this double configuration. These *DK-adjacencies* define *DK-components*, which are maximal *DK-connected* sets of tiles in ξ. For each DK-component C, there is a dual DK-component C' which is its 180° rotation around p (where C might equal C'). We will define an operator $\Lambda(\sigma, \rho)$ which uniformly at random selects from ξ one of its DK-components C that contains a bar, and moves all the tiles in C and C' to their ρ positions. If we restrict to Ω', where there is at most one bar per column, then $\mathcal{M}_{DK}$ is rapidly mixing. However, as we will see in Section 4, the Markov chain $\mathcal{M}_{DK}$ is inefficient on Ω.

We now define $\mathcal{P}_{DK}(\mathcal{P}'_{DK})$, the transition matrix of $\mathcal{M}_{DK}$ over the state space Ω (Ω'). Let H_n be the set of half-integers $\frac{k}{2}$, where $0 \leq k < 2n$ is an integer.

The DK Markov chain $\mathcal{M}_{DK}$

`Starting at any` σ_0`, iterate the following:`

- `With prob.` $1/8$`, pick a pivot position` $p \in H_n \times H_n$ `u.a.r.`
  - `Rotate` σ_t `by` 180° `around the pivot` p`. Call this configuration` ρ`.`
  - `Let` $\sigma_{t+1} = \Lambda(\sigma_t, \rho)$`.`
- `With prob.` $1/8$`, pick` $k \in H_n$ `u.a.r.`
  - `Reflect` σ_t `across the horizontal line` $y = k$`. Call this configuration` ρ`.`
  - `Let` $\sigma_{t+1} = \Lambda(\sigma_t, \rho)$`.`
- `With prob.` $1/8$`, pick an edge` $e \in E$ `u.a.r.`
  - `If` e `is not incident to a bar or a speck then add a speck at` e`.`
  - `Otherwise, if` e `contains a speck, then remove its speck with prob.` $1/\lambda$`.`
- `With prob.` $1/8$`, pick a position` $p \in V$ `u.a.r. Translate the entire configuration to set` $p = (0, 0)$`. Otherwise,` $\sigma_{t+1} = \sigma_t$`.`
- `Otherwise, do nothing.`

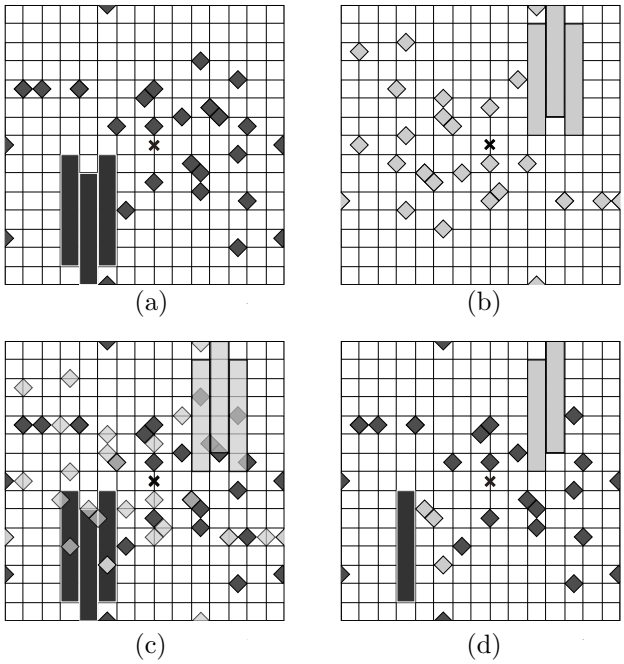

Figure 2: The DK Algorithm (a) $\sigma \in \Omega'$. (b) ρ obtained from σ by rotating around the marked pivot. (c) ξ obtained by superimposing σ and ρ. (d) $\Lambda(\sigma, \rho)$. In this case $\mathcal{M}_{DK}$ chose the component with two bars; notice that the leftmost bar of σ is not in this component since there is no speck from ρ along its shared border with the middle bar.

Since we allow $\mathcal{M}_{DK}$ to pivot on any half-integer, it can make any move that is possible using $\mathcal{M}_{Local}$; it follows that $\mathcal{M}_{DK}$ is also connected. Again, by connectivity and since $\mathcal{M}_{DK}$ is aperiodic and reversible with respect to π over Ω, π is the (unique) stationary distribution of $\mathcal{M}_{DK}$ (similarly for π' and Ω'). While this Markov chain achieves the desired stationary distribution, it is not useful if it converges slowly to equilibrium. This is especially relevant since this algorithm has been used in practice, where samples are taken after running the algorithm for a relatively short (polynomial) number of steps. If the algorithm does not converge quickly this data will not be chosen from the predicted distribution and conclusions are not reliable. Therefore, our goal for the remainder of the paper is to examine the efficiency of these algorithms, known as the mixing time.

2.3 Mixing time. The time a Markov chain takes to converge to its stationary distribution is measured in terms of the distance between the distribution at time t and the stationary distribution. The *total variation*

distance at time t is

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

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

$$\tau(\epsilon) = \min\{t : \|\mathcal{P}^{t'}, \pi\|_{tv} \le \epsilon, \forall t' \ge t\}.$$

We say that a Markov chain is *rapidly mixing* if the mixing time is bounded above by a polynomial in n and $\log(\epsilon^{-1})$, where n is the size of each configuration in Ω. It is well-known from probability theory that the eigenvalue gap of the transition matrix of a Markov chain provides a good bound on the mixing time of the chain. We let $Gap(\mathcal{P}) = 1 - |\lambda_1|$ denote the spectral gap, where $\lambda_0, \lambda_1, \ldots, \lambda_{|\Omega|-1}$ are the eigenvalues of the transition matrix $\mathcal{P}$ and $1 = \lambda_0 > |\lambda_1| \ge |\lambda_i|$ for all $i \ge 2$. The following result relates the spectral gap with the mixing time of the chain (see, e.g., [13]):

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

$$(a) \qquad \tau(\epsilon) \le \frac{1}{1-|\lambda_1|} \log\left(\frac{1}{\pi_*\epsilon}\right).$$

$$(b) \qquad \tau(\epsilon) \ge \frac{|\lambda_1|}{2(1-|\lambda_1|)} \log\left(\frac{1}{2\epsilon}\right).$$

An essential tool we will use to derive both upper and lower bounds on the mixing time of our Markov chains is conductance. The *conductance* of an ergodic Markov chain $\mathcal{M}$ with stationary distribution π is

$$\Phi_{\mathcal{M}} = \min_{\substack{S\subseteq\Omega \\ \pi(S)\le 1/2}} \phi_S,$$

where $\phi_S = \phi_S^{(\mathcal{M})}$ is the *conductance of a set $S \subset \Omega$*, defined by

$$\phi_S = \frac{1}{\pi(S)} \sum_{s_1\in S, s_2\in\overline{S}} \pi(s_1)\mathcal{P}(s_1,s_2).$$

The following relates conductance and spectral gap [6].

THEOREM 2.2. *For any Markov chain with conductance Φ and eigenvalue gap $Gap(\mathcal{P}) = 1-|\lambda_1|$, we have*

$$\frac{\Phi^2}{2} \;\le\; Gap(\mathcal{P}) \;\le\; 2\Phi.$$

It is sometimes convenient to analyze the square of a Markov chain $\mathcal{M}$ rather than the original. Each step (x,z) of $\mathcal{M}^2$ occurs with probability $\mathcal{P}^2(x,z) := \sum_{y\in\Omega} \mathcal{P}(x,y)\mathcal{P}(y,z)$. Not surprisingly, if $\mathcal{P}^2$ has large conductance, then so does $\mathcal{P}$.

LEMMA 2.1. *If $\mathcal{M}$ is ergodic and reversible, then $\Phi_{\mathcal{M}^2} \le 2\Phi_{\mathcal{M}}$.*

Proof. For any set $S \subset \Omega$,

$$\begin{aligned}
\phi_S^{(\mathcal{M}^2)}\pi(S) &= \sum_{\substack{x\in S \\ z\in\overline{S}}} \pi(x)\mathcal{P}^2(x,z) \\
&= \sum_{\substack{x,y\in S \\ z\in\overline{S}}} \pi(x)\mathcal{P}(x,y)\mathcal{P}(y,z) + \sum_{\substack{x\in S \\ y,z\in\overline{S}}} \pi(x)\mathcal{P}(x,y)\mathcal{P}(y,z) \\
&= \sum_{\substack{x,y\in S \\ z\in\overline{S}}} \pi(y)\mathcal{P}(y,x)\mathcal{P}(y,z) + \sum_{\substack{x\in S \\ y,z\in\overline{S}}} \pi(x)\mathcal{P}(x,y)\mathcal{P}(y,z) \\
&\le \sum_{\substack{y\in S \\ z\in\overline{S}}} \pi(y)\mathcal{P}(y,z) + \sum_{\substack{x\in S \\ y\in\overline{S}}} \pi(x)\mathcal{P}(x,y) \;=\; 2\phi_S^{\mathcal{M}}\pi(S).
\end{aligned}$$

Thus $\Phi_{\mathcal{M}^2} \le 2\Phi_{\mathcal{M}}$. □

Another tool used throughout our proofs is the decomposition method [9, 10]. Let $\Omega = \cup_{i=1}^m \Omega_i$ be a partition of the statespace into m disjoint pieces. For each $i = 1, \ldots, m$, define $\mathcal{P}_i = \mathcal{P}(\Omega_i)$ as the restriction of $\mathcal{P}$ to Ω_i which rejects moves that leave Ω_i. In particular, the restriction to Ω_i is a Markov chain, $\mathcal{M}_i$ with state space Ω_i, where the transition matrix $\mathcal{P}_i$ is defined as follows: If $x \ne y$ and $x, y \in \Omega_i$ then $\mathcal{P}_i(x,y) = \mathcal{P}(x,y)$; if $x \in \Omega_i$ then $\mathcal{P}_i(x,x) = 1 - \sum_{y\in\Omega_i, y\ne x} \mathcal{P}_i(x,y)$. Let π_i be the normalized restriction of π to Ω_i, i.e., $\pi_i(A) = \frac{\pi(A\cap\Omega_i)}{\pi(\Omega_i)}$. Define $\widehat{\mathcal{P}}$ to be the following aggregated transition matrix on the state space $\{1, \ldots, m\}$:

$$\widehat{\mathcal{P}}(i,j) = \frac{1}{\pi(\Omega_i)} \sum_{\substack{x\in\Omega_i, \\ y\in\Omega_j}} \pi(x)\mathcal{P}(x,y).$$

We then have the following [10].

THEOREM 2.3. *Let $\mathcal{P}_i = \mathcal{P}(\Omega_i)$ and $\widehat{\mathcal{P}}$ be as above. Then*

$$Gap(\mathcal{P}) \ge \frac{1}{2} Gap(\widehat{\mathcal{P}}) \min_{i=1,\ldots,m} Gap(\mathcal{P}_i).$$

In our case, each Ω_i will contain all elements of Ω with a particular bar structure X.

3 The restricted model (≤ 1 bar per column)

In this section we will see that on Ω', the restricted state space containing configurations with at most one bar per column, the local algorithm takes exponential time while the DK algorithm converges in polynomial time.

3.1 The local algorithm mixes slowly. We begin with a straightforward proof that any local Markov chain $\mathcal{M}_{Local}$ that moves a single bar or speck in each step will take exponential time to converge to stationarity. The idea is that in order to change which columns contain bars, the algorithm must take a bar and move it to a column i that currently has no bars. It can only do this if there is a consecutive set of $L+1$ (horizontal) edges in column i, that are all free of specks, but since $L = \Omega(n)$ this is exponentially unlikely.

THEOREM 3.1. *There exist constants c_3 and c_4 such that the mixing time of $\mathcal{M}_{Local}$ on Ω' satisfies*

$$\tau(\epsilon) \geq (c_3 \lambda^{c_4 n}/n^2)\log(1/2\epsilon).$$

Proof. Recall from Theorem 2.2 that to show the chain is slowly mixing, it is sufficient to demonstrate a cut in Ω' for which the conductance is exponentially small. Our cut will partition configurations according to whether or not they have a bar in the first column.

Define A to be the set of all configurations in Ω that do not have a bar in column 0. Next, define $K \subseteq A$ to be the set of configurations with $L+1$ consecutive (horizontal) edges in column 0 that do not have any specks present. Clearly if $\sigma \in A$, $\rho \in \overline{A}$ and $\mathcal{P}_{Local}(\sigma,\rho) > 0$, then $\sigma \in K$. First we will show that $\pi'(K)$ is exponentially smaller than $\pi'(A)$. Specifically,

$$\pi'(K) \leq \frac{n}{\lambda^{L+1}}\pi'(A). \tag{3.1}$$

We define a function $f : K \to A$ as follows. For $\sigma \in K$, let i be the smallest integer such that edges $((0,i),(1,i)),((0,i+1 \mod n),(1,i+1 \mod n)),\ldots,((0,i+L \mod n),(1,i+L \mod n))$ contain no specks. Add a speck to each of these edges, to create $f(\sigma)$. Since we have added $L+1$ specks, $\pi'(\sigma) = \lambda^{-L-1}\pi'(f(\sigma))$. For each configuration $\rho \in A$, there are at most n configurations $\sigma \in K$ such that $f(\sigma) = \rho$; corresponding to the (at most n) possible choices for the first (horizontal) edge in ρ to start removing specks. It follows that

$$\begin{aligned}\pi'(K) &= \sum_{\rho\in A}\sum_{\substack{\sigma\in K:\\ f(\sigma)=\rho}}\pi'(\sigma) = \sum_{\rho\in A}\sum_{\substack{\sigma\in K:\\ f(\sigma)=\rho}}\lambda^{-L-1}\pi'(\rho)\\ &\leq \sum_{\rho\in A}\frac{n}{\lambda^{L+1}}\pi'(\rho) = \frac{n}{\lambda^{L+1}}\pi'(A).\end{aligned}$$

Using Equation 3.1 we can upper bound the conductance of $\mathcal{M}_{Local}$ as follows. If $\pi'(A) \leq 1/2$,

$$\begin{aligned}\Phi_{\mathcal{M}_{Local}} &\leq \pi'(A)^{-1}\sum_{\sigma\in A,\rho\notin A}\pi'(\sigma)\mathcal{P}_{Local}(\sigma,\rho)\\ &= \pi'(A)^{-1}\sum_{\sigma\in K}\pi'(\sigma)\sum_{\rho\notin A}\mathcal{P}_{Local}(\sigma,\rho)\\ &\leq \pi'(A)^{-1}\sum_{\sigma\in K}\pi'(\sigma) \leq \frac{n}{\lambda^{L+1}}.\end{aligned}$$

Next, in order to handle the case where $\pi'(A) > 1/2$, we relate A and $\overline{A}$ by

$$\pi'(A) \leq n\pi'(\overline{A}). \tag{3.2}$$

We will define a function $g : A \to \overline{A}$. For any $\sigma \in A$, let i be the first column in σ that contains a bar. Translate the configuration so that column i maps onto column 0 to create $g(\sigma)$. Notice that for any configuration $\rho \in \overline{A}$, there are at most n configurations σ such that $g(\sigma) = \rho$. These n configurations correspond to the n possible ways to translate ρ. Combining these observations shows the following.

$$\begin{aligned}\pi'(A) &= \sum_{\rho\in\overline{A}}\sum_{\substack{\sigma\in A:\\ g(\sigma)=\rho}}\pi'(\sigma)\\ &= \sum_{\rho\in\overline{A}}\sum_{\substack{\sigma\in A:\\ g(\sigma)=\rho}}\pi'(\rho)\\ &\leq \sum_{\rho\in\overline{A}}n\pi'(\rho) = n\pi'(\overline{A}).\end{aligned}$$

If $\pi'(A) > 1/2$, by the reversibility of $\mathcal{M}_{Local}$ and Equation 3.2,

$$\begin{aligned}\Phi_{\mathcal{M}_{Local}} &\leq \pi'(\overline{A})^{-1}\sum_{\rho\in\overline{A},\sigma\in A}\pi'(\rho)\mathcal{P}_{Local}(\rho,\sigma)\\ &= \pi'(\overline{A})^{-1}\sum_{\rho\in\overline{A},\sigma\in A}\pi'(\sigma)\mathcal{P}_{Local}(\sigma,\rho)\\ &\leq \left(\frac{\pi'(A)}{\pi'(\overline{A})}\right)\left(\frac{n}{\lambda^{L+1}}\right) \leq \frac{n^2}{\lambda^{L+1}}.\end{aligned}$$

Applying Theorem 2.2 and Theorem 2.1 gives us that the mixing time of $\mathcal{M}_{Local}$ satisfies $\tau(\epsilon) \geq \left(\frac{\lambda^{c_1 n+1}}{4n^2} - \frac{1}{2}\right)\log\left(\frac{1}{2\epsilon}\right)$. □

REMARK 1. *Notice that this proof also holds for Ω, the unrestricted model. Moreover, even a non-local chain that is allowed to move a bar to any empty position in a single move will also be exponentially slow for the exact same reasons the local algorithm is slow.*

3.2 The DK algorithm mixes rapidly. The DK algorithm was introduced to allow moves even when there is a collision precisely to overcome the main obstacles causing the local algorithm to be slow. We will see that the DK Algorithm provides an exponential speedup in the mixing time over the local algorithm on Ω', where we restrict to having at most one bar per column. The DK Algorithm potentially makes huge global moves that can change a configuration significantly in a single step. Such Markov chains have proven to be difficult to analyze. However, in this case, we are able to exploit our decomposition of the state space Ω' to simplify the analysis. The projection chain arising emphasizes the dependence on total perimeter to determine the conditions under which the algorithm is efficient. If we restrict to having at most one bar per column of T, then $\mathcal{M}_{DK}$ is rapidly mixing. However, we will see in Section 4 that without this restriction the Markov chain is inefficient.

THEOREM 3.2. *If $\lambda \geq 4^{c_1^{-1}}$, then there exists a constant c_5 such that the mixing time of $\mathcal{M}_{DK}$ on Ω' satisfies*

$$\tau(\epsilon) \leq c_5 n^{14}(n^2 + \log \epsilon).$$

It will be useful to introduce some terminology. Given a configuration σ in Ω or Ω', define the *bar structure* $\Gamma(\sigma)$ as the configuration x obtained from σ by removing all of its specks. We consider the set of all such bar structures $\widehat{\Omega}'$ of b bars of length L. Let $\widehat{\pi}'$ be the induced distribution on $\widehat{\Omega}'$; that is, for $x \in \widehat{\Omega}'$, let

$$\widehat{\pi}'(x) = \sum_{\sigma \in \Gamma^{-1}(x)} \pi'(\sigma).$$

Let $e(\sigma)$ be the number of edges that are not incident to any bar in σ and define $\kappa(\sigma) = |\{e \in E : \text{ } e \text{ is incident with exactly one bar}\}|$, which we will call the *total perimeter* of the bars in σ. In equation 3.3, we find that each bar structure has weight proportional to $\mu^{\kappa(X)}$, where $\mu = (1+\lambda)^{-\frac{1}{2}}$. In other words, the total perimeter of the bar structure completely determines the probability that it will show up as the bar structure of some configuration in Ω'. Specifically,

$$\begin{aligned} \widehat{\pi}'(x) &= \sum_{k=0}^{e(x)} \frac{\lambda^k}{Z'} \binom{e(x)}{k} \\ &= \frac{1}{Z'}(1+\lambda)^{e(x)} \\ &= Z'^{-1}(1+\lambda)^{2n^2-2bL}\mu^{\kappa(x)}. \end{aligned} \tag{3.3}$$

In order to prove Theorem 3.2, we will explicitly use the projection $\widehat{\Omega}'$ as a decomposition of the state space to prove that $\mathcal{M}_{DK}$ is rapidly mixing. First, we have to show that the Markov chain is rapidly mixing within each set of configurations in Ω' that have the same bar structure (the restrictions); then we will show that the projection Markov chain that reduces each bar structure to a single configuration with the appropriate weighting is also rapidly mixing.

DEFINITION 3.1. *Define the projection Markov chain $\widehat{\mathcal{M}}_{DK}$ on $\widehat{\Omega}'$ by the following transitions: for $x, y \in \widehat{\Omega}'$,*

$$\widehat{\mathcal{P}}'_{DK}(x,y) = \frac{1}{\pi'(\Gamma^{-1}(x))} \sum_{\substack{\sigma \in \Gamma^{-1}(x) \\ \rho \in \Gamma^{-1}(y)}} \pi'(\sigma)\mathcal{P}'_{DK}(\sigma,\rho).$$

The main step of the proof of Theorem 3.2 is to show that the projection Markov chain $\widehat{\mathcal{M}}_{DK}$ is rapidly mixing. We achieve this using canonical paths; see, e.g., [6]. Between every pair $I, F \in \widehat{\Omega}'$ we define a canonical path $\gamma_{I,F}$ consisting of transitions of the Markov chain, ensuring that the weight of these paths is evenly distributed among edges of the Markov chain. Each path $\gamma_{I,F}$ routes a flow of $\widehat{\pi}'(I)\widehat{\pi}'(F)$ along its edges, ensuring that the total flow each edge (x_i, x_{i+1}) receives is within a polynomial factor of its capacity, which is defined to be $\widehat{\pi}'(x_i)\widehat{\mathcal{P}}'_{DK}(x_i, x_{i+1})$. Generally it is sufficient to define a complementary point $x' \in \widehat{\Omega}'$ for each edge $e = (x_i, x_{i+1})$ in $\gamma_{I,F}$ and to show that there is only a polynomial number of pairs $I, F \in \widehat{\Omega}'$ that use transition e with complementary point x'. We will identify this complementary point x' by describing the reverse path $\gamma_{F,I}$; for the ith edge (x_i, x_{i+1}) of the path $\gamma_{I,F}$, the complementary point is x'_i, where (x'_i, x'_{i+1}) is the ith edge of $\gamma_{F,I}$. The novelty of this proof lies in the choice of $\gamma_{I,F}$, which provides an encoding of the relative heights of the bars while maintaining the perimeter. This is important because the perimeter determines the weight of a configuration, and the edges of the path need to happen with high enough probability.

A key component of the proof is an operator called a bar shift, which we now explain. For $\sigma \in \widehat{\Omega}'$ and a bar a, let $h(a) = h_\sigma(a)$ denote the *height in the torus* of a in σ, defined as the y-coordinate of the top left corner of a. Assume that q and r are bars such that there are no bars in the columns immediately to the left of q and the right of r. Let j be the column containing r. For any integer $-n \leq d \leq n$, **bar shift** (q, r, d) is a transition of $\widehat{\mathcal{M}}^2_{DK}$ that moves bar q to column $j+1$ at height $h(r)+d$ with probability at least $(1/(64n^6)) \min\{1, \mu^{\Delta\kappa}\}$, where $\Delta\kappa$ is the change in perimeter of the move. Depending on the configuration, this may be achieved in a single step or a sequence of two steps.

For technical reasons the formal definition of the bar shift operator requires some case analysis to ensure that such a move is always possible with high enough

probability. Let q' be the bar to the right of q, if it exists, let $t = h(q) - h(q')$, and let l be the set of edges between columns j and $j+1$. If q' does not exist, then $\widehat{\mathcal{M}}_{DK}^2$ rotates q to height $h(r)+d$ in column $j+1$ in one step, and the next step does nothing. Otherwise, there are two cases to consider. Assume first that $t \geq 0$. Consider the move of $\widehat{\mathcal{M}}_{DK}$ that rotates q to column $j+1$ at height $h(r)+d$. In order for q to be in its own DK-component, there must be no specks on the vertical edges hit by the shared border of q and q', that is, on the edges in l in the interval $[h(r)+d-(L-t), h(r)+d]$. By the location of bar r, there must be no specks on l in the interval $[h(r)-L, h(r)]$. Thus, if $d \leq 0$ then the number of edges that may have specks is $\max\{0, -d-t\}$. Hence if $-d-t < 0$ then the move happens with probability at least $1/(8n^3)$ and otherwise with probability at least $(1+\lambda)^{d+t}/(8n^3) = \mu^{\Delta\kappa}/(8n^3)$. Thus in this case **bar shift** (q, r, d) performs this rotation and then does nothing in the second move. However, if $d \geq 0$, then the probability of this rotation succeeding (that is, q moving independently of q') depends on d, not on $\Delta\kappa$ as required, and indeed may be quite unlikely to happen. Instead, **bar shift** (q, r, d) first rotates q to column $j+1$ at height $h(r)-t$. Notice that when q and q' are rotated to this position, their shared border cannot hit any specks (as r completely overlaps q', and thus prevents specks in those positions). Thus q is in its own DK-component and so this move happens with probability at least $1/(8n^3)$. Then $\widehat{\mathcal{M}}_{DK}$ reflects q within column $j+1$ to height $h(r)+d$. In order for q to be in its own DK-component, there must be no specks on the edges in l in $[h(r)+d-L, h(r)+d-t]$. There are no specks on l in $[h(r)-L, h(r)]$ by the location of r; hence, the number of edges on which there may be specks is $\min(0, d-t)$. If $d < t$ then this move happens with probability at least $1/(8n^3)^2$. Otherwise, it happens with probability at least $(1+\lambda)^{-(d-t)} = \mu^{\Delta\kappa}/(64n^6)$. The other case, where $t < 0$ is similar. In this case, if $d < 0$ then **bar shift** rotates q first to height $h(r)-t$ and then reflects it to height $h(r)+d$. Otherwise, **bar shift** rotates q directly to height $h(r)+d$ and then does nothing.

We may now prove Theorem 3.2.

Proof of Theorem 3.2: As described above, we will prove that the restriction of $\mathcal{M}_{DK}$ to each Ω'_i and the projection chain $\widehat{\mathcal{M}}_{DK}$ are all rapidly mixing. For the restrictions the bar structure is fixed and we focus on moves that add or remove a speck (there may be other moves in the restrictions, but ignoring them can only decrease the spectral gap). Since each speck is independent of the other specks, this is equivalent to a biased walk on the hypercube on $e(x)$ vertices. Thus the spectral gap of the restriction with bar structure x is $1/e(x) \geq 1/(2n^2)$ (see, e.g., [8] p. 163).

To show that $\widehat{\mathcal{M}}_{DK}$ is rapidly mixing, we consider instead the square of this Markov chain $\widehat{\mathcal{M}}_{DK}^2$ and apply Lemma 2.1. We will use canonical paths to show that $\widehat{\mathcal{M}}_{DK}^2$ has high conductance, which will imply that $\widehat{\mathcal{M}}_{DK}$ does as well. Consider any set $S \subset \widehat{\Omega}'$. For any pair I, F of configurations in Ω' we will define a path $\gamma_{I,F} = \{I = x_0, x_1, x_2, \ldots x_t = F\}$ of transitions (x_i, x_{i+1}) of $\mathcal{M}_{DK}^2$ between them. For $I \in S$ and $F \in \overline{S}$, define $f(I,F)$ to be the first edge (x_i, x_{i+1}) of $\gamma_{I,F}$ that crosses the cut; that is, $x_i \in S$, $x_{i+1} \in \overline{S}$. We will show that for any $x \in S, y \in \overline{S}$,

$$\sum_{\substack{I \in S, F \in \overline{S}: \\ f(I,F)=(x,y)}} \widehat{\pi}'(I)\widehat{\pi}'(F) \;\leq\; 2^9 n^6 \widehat{\pi}'(x)\widehat{\mathcal{P}}'^2_{DK}(x,y), \tag{3.4}$$

which means that the weight over the paths $\gamma_{I,F}$ through (x,y) is not too high. This is the heart of the argument, and given (3.4), the rest is simple; it proceeds as follows. For every $S \subset \widehat{\Omega}'$ with $\widehat{\pi}'(S) \leq \frac{1}{2}$, we have

$$\begin{aligned}
\phi_S &= \frac{1}{\widehat{\pi}'(S)} \sum_{x \in S, y \in \overline{S}} \widehat{\pi}'(x)\widehat{\mathcal{P}}'^2_{DK}(x,y) \\
&\geq \frac{1}{2\widehat{\pi}'(S)\widehat{\pi}'(\overline{S})} \sum_{x \in S, y \in \overline{S}} \widehat{\pi}'(x)\widehat{\mathcal{P}}'^2_{DK}(x,y) \\
&\geq \frac{1}{2\widehat{\pi}'(S)\widehat{\pi}'(\overline{S})} \sum_{x \in S, y \in \overline{S}} \sum_{\substack{I \in S, F \in \overline{S}: \\ f(I,F)=(x,y)}} \frac{\widehat{\pi}'(I)\widehat{\pi}'(F)}{2^9 n^6} \\
&= \frac{1}{2^{10} n^6 \widehat{\pi}'(S)\widehat{\pi}'(\overline{S})} \sum_{I \in S, F \in \overline{S}} \widehat{\pi}'(I)\widehat{\pi}'(F) \\
&= 1 \;/\; (2^{10} n^6) \;.
\end{aligned}$$

Therefore the conductance of $\widehat{\mathcal{M}}_{DK}$ satisfies $\Phi_{\widehat{\mathcal{M}}_{DK}} \geq \frac{1}{2}\Phi_{\widehat{\mathcal{M}}_{DK}^2} \geq 1/(2^{11} n^6)$. By Theorem 2.2 the spectral gap satisfies $Gap(\widehat{\mathcal{P}}'_{DK}) \geq \frac{1}{2}\Phi^2_{\widehat{\mathcal{M}}_{DK}} \geq 1/(2^{21} n^{12})$. Finally, by the Decomposition Theorem 2.3,

$$Gap(\mathcal{P}'_{DK}) \geq \frac{1}{2^{22} n^{12}} \left(\frac{1}{2n^2}\right) \geq \frac{1}{2^{23} n^{14}}.$$

In order to apply Theorem 2.1, we note that the minimum weight configuration has no diamonds and weight $1/Z'$ and the normalizing constant $Z' \leq n^{2b}(1+\lambda)^{n^2}$, so for any $\epsilon > 0$, $\log(\frac{Z'}{\epsilon}) \leq 2\log(1+\lambda)n^2 - \log(\epsilon^{-1})$. Therefore by Theorem 2.1 we find that for all $\epsilon > 0$,

$$\tau(\epsilon) \leq 2^{23} n^{14} \left(2n^2 \log(1+\lambda) + \log \epsilon\right).$$

We now focus on establishing (3.4) for the remainder of the proof. First consider the case where I and F

each have a single component, as this is the key piece of the analysis. The general idea is that each edge of $\gamma_{I,F}$ will move a bar from the left side of I's component to the right side, lining up these bars with an offset that agrees with the offsets in F. Thus, after every bar has been wrapped around to the right, the resulting configuration F_T is just a translation of F. The final step of the path $\gamma_{I,F}$ is a translation that takes F_T to F.

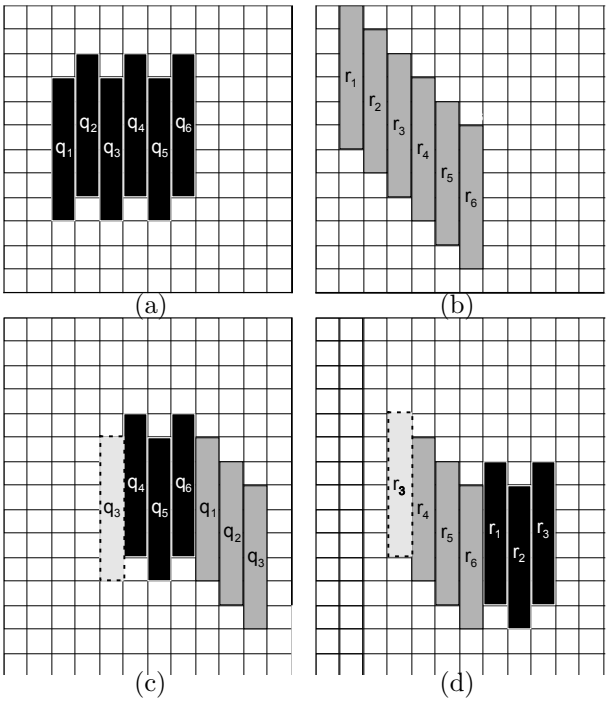

Figure 3: (a) Initial configuration I. (b) Final configuration F. (c) This transition (x, y) is the third step of $\gamma_{I,F}$ that takes bar q_3 to the right side at the same relative height with q_2 as r_4 was to r_3. Notice that q_4, q_5, q_6 remain unchanged from I; however, q_6, q_1, q_2, q_3 match the first four bars of F (up to translation). (d) (x', y') is the third step of $\gamma_{F,I}$; r_4, r_5, r_6 remain unchanged from F while r_6, r_1, r_2, r_3 record the first four bars of I.

Finally, we are ready to describe $\gamma_{I,F}$ when I and F each have a single component. Let C_1 be the unique component in I and let C_2 be the unique component in F. Label the bars of C_1 by $q_1, q_2, \ldots, q_b = q_0$ from left to right (beginning from the leftmost bar) and similarly label the bars of C_2 with $r_1, r_2, \ldots, r_b = r_0$. Assume without loss of generality that q_1 and r_1 are in column 1. In each step of $\gamma_{I,F}$, we wrap q_i to column $b + i$ at the same relative height to q_{i-1} as r_{i+1} is to r_i; meanwhile, each step of $\gamma_{F,I}$ moves r_i to column $b + i$ at the same relative height to r_{i-1} as q_{i+1} is to q_i. Let $t_i = h(q_i) - h(q_{i+1})$ and $t'_i = h(r_i) - h(r_{i+1})$. Then the ith step of $\gamma_{I,F}$ is (x, y) which performs **bar shift** $(q_i, q_{i-1}, -t'_i)$. Similarly, the ith step of $\gamma_{F,I}$, which is (x', y'), performs **bar shift** $(r_i, r_{i-1}, -t_i)$. See Figure 3.2. Notice that in x, the offsets of the bars that have been moved so far agree with F, and the offsets of the bars that have not moved yet agree with I. In x' the reverse is true. That is, the bars that have been moved so far agree with I, and the bars that have not moved yet agree with F. Thus, given (x, y) and x' we can recover I and F uniquely. Moreover, $\widehat{\pi}'(y)\widehat{\pi}'(y') = \widehat{\pi}'(x)\widehat{\pi}'(x') = \widehat{\pi}'(I)\widehat{\pi}'(F)$. Suppose that $\widehat{\pi}'(x) \leq \widehat{\pi}'(y)$; the other case is similar. In this case, $\widehat{\mathcal{P}}'^2_{DK}(x, y) \geq 1/(64n^6)$. Let S_1 (respectively $\overline{S}_1$) be the set of configurations in S ($\overline{S}$) with a single component. Then we have

$$\begin{aligned}\sum_{\substack{I\in S_1, F\in\overline{S}_1:\\ f(I,F)=(x,y)}} \widehat{\pi}'(I)\widehat{\pi}'(F) &= \sum_{\substack{I\in S_1, F\in\overline{S}_1\\ f(I,F)=(x,y)}} \widehat{\pi}'(x)\widehat{\pi}'(x') \\ &\leq \widehat{\pi}'(x) \sum_{x'\in\Omega'} \widehat{\pi}'(x') \\ &\leq \widehat{\pi}'(x) \;\leq\; \widehat{\pi}'(x)\widehat{\mathcal{P}}'^2_{DK}(x, y)64n^6. \qquad (3.5)\end{aligned}$$

The final edge of $\gamma_{I,F}$ is a simple translation that moves F_T to F. Recall F_T has a single component, and suppose its leftmost bar is in position (w, h). Suppose also that the leftmost bar of F is in position (w', h'). $\widehat{\mathcal{M}}_{DK}$ will perform a translation move that takes the component of F_T to position (w', h'). The next move will do nothing with probability at least $1/2$. This happens with probability $(\widehat{\mathcal{P}}'_{DK})^2(F_T, F) \geq 1/(16n^2)$. Clearly, given (F_T, F), we can identify F, so

$$\begin{aligned}\sum_{\substack{I\in S_1, F\in\overline{S}_1:\\ f(I,F)=(F_T,F)}} \widehat{\pi}'(I)\widehat{\pi}'(F) &\leq \sum_{I\in S_1} \widehat{\pi}'(I)\widehat{\pi}'(F) \\ &\leq \widehat{\pi}'(F) \\ &\leq 16n^2\widehat{\pi}'(F_T)\widehat{\mathcal{P}}'^2_{DK}(F_T, F). \qquad (3.6)\end{aligned}$$

For the case when I and/or F have multiple components we construct I' from I and F' from F so that I' and F' each have one component. The path $\gamma_{I,F}$ has three parts. In the first part we combine the components together until we reach I' which has a single component. If F contains several components, then the last part will be to break these components back up; that is, we will show how to take F and combine its components to create F' which has a single component, and the last part will be to undo these moves to go from F' back to F. The second part is precisely $\gamma_{I',F'}$. We will formally define each part of the path and bound for all transitions (x, y) the maximum number of paths $\gamma_{I,F}$ that can use (x, y) in each part. From the formal defintions of the path it will be clear that any transition (x, y) can be used in at most one part of

a path. Thus, it is sufficient to prove (3.4) for each part.

Part 1: Let C and C' be two components in I, where C' is the component immediately to the right of C. In this part, we combine C and C' by moving the bars of C', one at a time, to the right side of C, maintaining perimeter and relative heights of bars at every step until the last bar of C' is moved to fully align with the penultimate bar, for a loss in perimeter of $2L$. This loss in perimeter corresponds to an exponential gain in weight and there is only a polynomial number of places where C' could have been originally, so we can bound the weight of the configurations whose canonical path uses one of these transitions.

Let j_1 be the leftmost column of I containing a bar. If I has a single component, let $I' = I$ and proceed to Part 2. Otherwise, let C_1 be the component that begins in column j_1 and let r be the rightmost bar of C_1. Let C' be the first component to the right of C. Label the bars of C' by $q_1, q_2, \ldots, q_t$ from left to right. Suppose $1 \le i \le t-1$ (set $q_0 = r$) and we have already moved $q_1, q_2, \ldots, q_{i-1}$ over to C. Let $t_i = h(q_i) - h(q_{i+1})$. Move q_i to column $j+i$ to exactly align the shared border of q_i and q_{i+1} with the new shared border of q_{i-1} and q_i by performing **bar shift** $(q_i, q_{i-1}, -t_i)$. This results in no change in perimeter and encodes the relative height t_i. If $i = t$ then q_t is rotated over to fully align with q_{t-1}, for a loss of perimeter of $2L$. The second step of $\widehat{\mathcal{M}}^2_{DK}$ does nothing. These moves are performed with probability at least $1/(64n^6)$.

For a transition (x, y), how many paths $\gamma_{I,F}$ use (x, y) in Part 1? For this part we will let F be our complementary point x'. Given (x, y), in order to reconstruct I, we just need to remember where the other components used to be. Suppose that along $\gamma_{I,F}$ up to (x, y), $\widehat{\mathcal{M}}^2_{DK}$ has combined k components, so that x has k fewer components than I. Then there are at most n^{3k} ways to break these components back up, and $\widehat{\pi}'(x) \ge \widehat{\pi}'(I)\mu^{-2Lk}$. Hence for any $F \in \overline{S}$,

$$\sum_{\substack{I \in S: \\ f(I,F)=(x,y)}} \widehat{\pi}'(I) \;\le\; \sum_{k=0}^{n-2} n^{3k}\widehat{\pi}'(x)\mu^{2Lk} \;\le\; 2\widehat{\pi}'(x),$$

since, for $\lambda \ge 4^{c_1^{-1}}$ and any $n \ge 1$, we have

$$n^3\mu^{2L} \;=\; n^3(1+\lambda)^{-c_1 n} \;\le\; \frac{n^3}{4^n} \;\le\; \frac{1}{2} \tag{3.7}$$

We will return to this bound each time we need to push two components together. Finally,

$$\sum_{\substack{I \in S, F \in \overline{S}: \\ f(I,F)=(x,y)}} \widehat{\pi}'(I)\widehat{\pi}'(F) \le \sum_{F \in \overline{S}} \widehat{\pi}'(F) 2\widehat{\pi}'(x)$$
$$\le 2\widehat{\pi}'(x) \;\le\; 2^7 n^6 \widehat{\mathcal{P}}'^2_{DK}(x,y)\widehat{\pi}'(x).$$

This proves (3.4) for transitions (x, y) in Part 1.

Part 2: Let F' be obtained from F by the reverse of the process of Part 1. Given I' from Part 1 and F', let $\gamma_{I',F'}$ be defined as before for the case of a single component. Then for any edge (x, y) along this path,

$$\sum_{\substack{I \in S, F \in \overline{S}: \\ f(I,F)=(x,y)}} \widehat{\pi}'(I)\widehat{\pi}'(F)$$
$$\le \sum_{\substack{I' \in S_1, F' \in \overline{S}_1: \\ f(I,F)=(x,y)}} \widehat{\pi}'(I')\widehat{\pi}'(F') \sum_{k_1,k_2} \left(n^3\mu^{2L}\right)^{k_1+k_2}$$
$$\le 4 \sum_{\substack{I \in S_1, F \in \overline{S}_1 \\ f(I,F)=(x,y)}} \widehat{\pi}'(x)\widehat{\pi}'(x')$$
$$\le \widehat{\pi}'(x)\widehat{\mathcal{P}}'^2_{DK}(x,y)2^8 n^6,$$

where the second inequality follows from (3.7) and the last inequality follows from (3.5) and (3.6). This proves (3.4) for transitions (x, y) in Part 2.

Part 3: Let $\gamma_{F',F}$ be the reverse sequence of the path outlined in Part 1. For $(x, y) \in \gamma_{F',F}$, where y has k fewer components than F, there are at most n^{3k} ways to break these components up, and $\widehat{\pi}'(y) \ge \widehat{\pi}'(F)\mu^{-2Lk}$. Hence, as above, for any $I \in S$,

$$\sum_{\substack{F \in \overline{S}: \\ f(I,F)=(x,y)}} \widehat{\pi}'(F) \le 2\widehat{\pi}'(y).$$

Moreover,

$$\widehat{\pi}'(x)\widehat{\mathcal{P}}'^2_{DK}(x,y) \;=\; \widehat{\pi}'(y)\widehat{\mathcal{P}}'^2_{DK}(y,x) \;\ge\; \widehat{\pi}'(y)/(2^6 n^6).$$

Therefore

$$\begin{aligned}\sum_{\substack{I \in S, F \in \overline{S}: \\ f(I,F)=(x,y)}} \widehat{\pi}'(I)\widehat{\pi}'(F) \;&\le\; \sum_{I \in S} \widehat{\pi}'(I) 2\widehat{\pi}'(y) \\ &\le\; 2\widehat{\pi}'(y) \\ &\le\; 2^7 n^6 \widehat{\pi}'(x)\widehat{\mathcal{P}}'^2_{DK}(x,y).\end{aligned}$$

Given any transition (x, y), it is easy to see which part of a path it can be used in. Therefore we have proved (3.4) for all transitions (x, y). □

It is interesting to note that our canonical paths do not use any DK moves that involve more than a single bar. The only moves we require are essentially Glauber moves on the projection chain, which are possible since the DK algorithm moves the specks out of the way. Finally, we point out that the lower bound on λ that we require is just a consequence of the proof technique, and we believe that in Ω', $\mathcal{M}_{DK}$ should be rapidly mixing for all values of λ. Our canonical paths approach

starts by collapsing multiple components into one large component, and λ must be large enough to compensate for the loss of the original locations of the components. However, this requirement is reasonable, since our goal is to contrast the DK and local algorithms and this is the condition under which the local algorithm is slow.

Although we have framed this section with L linear in n, this bound on L is not necessary. In this proof, we use the length of L to ensure a sufficient decrease in perimeter when components are pushed together. In the slow mixing proof above, we also use the bound on L to argue that a bar will likely hit specks when moved to a different column. Notice that both of these proofs hold even when L is sublinear, as long as it is $\Omega(\log n)$.

4 The unrestricted model

While the results of Section 3 provide evidence that the DK algorithm can be useful, we now show that it is very sensitive to the setting in which it is used. We focus now on Ω where we remove the restriction on the number of bars per column. The proof of Theorem 3.1 can be extended to show $\mathcal{M}_{Local}$ also requires exponential time to converge on Ω. In fact, we will show that, on Ω, $\mathcal{M}_{DK}$ also takes exponential time to converge to the stationary distribution π. Specifically, we prove:

THEOREM 4.1. *Suppose $0 < c_1 \leq 1/4$ and $b \geq 8c_1 n/3$. If $\lambda \geq 8^{c_1^{-1}}$, then there exist constants c_6 and c_7 such that the mixing time of $\mathcal{M}_{DK}$ on Ω satisfies*

$$\tau(\epsilon) \geq (c_6(1+\lambda)^{c_7 n}/n^{10}) \log(1/2\epsilon).$$

As in our analysis of the local algorithm, the proof of this theorem again relies on conductance and identifies a bad cut in the state space. However, this slow mixing result for the DK algorithm requires more careful analysis than the local algorithm since there are many more moves crossing any particular cut.

Our proof again makes use of the decomposition method. Define $\widehat{\Omega}, \widehat{\pi}$, and $\widehat{\mathcal{P}}_{DK}$ analogously to $\widehat{\Omega}', \widehat{\pi}'$, and $\widehat{\mathcal{P}}'_{DK}$. Similarly, let $h(a) = h_x(a)$ denote the height in the torus of bar a in configuration $x \in \widehat{\Omega}$ analogous to the definition in Section 3.2. Throughout this section, we examine the projection Markov chain $\widehat{\mathcal{P}}_{DK}$, since an upper bound on the conductance of $\widehat{\mathcal{P}}_{DK}$ provides an immediate bound on $\mathcal{P}_{DK}$. Define $S \subset \widehat{\Omega}$ to be the set of configurations (bar structures) with at most one bar per column (thus $S = \Gamma(\Omega')$). We demonstrate that there is a bad cut in the projection chain $\widehat{\mathcal{M}}_{DK}$ between S and $\overline{S}$ as follows. Every path in $\widehat{\mathcal{M}}_{DK}$ between S and $\overline{S}$ must go through the set $K \subset \overline{S}$ of all configurations with at most 2 bars per column and at least one column with two bars. Define

$$Q(S,K) := \sum_{\sigma \in S, \rho \in K} \widehat{\pi}(\sigma)\widehat{\mathcal{P}}_{DK}(\sigma,\rho). \tag{4.8}$$

We find that $Q(S,K)$ is exponentially smaller than $\widehat{\pi}(S)$, which yields an upper bound on conductance. Estimating (4.8) is a challenging combinatorial problem; here we present a mapping argument between S and K that yields our bounds on (4.8), which relies critically on K having at most 2 bars per column. We believe that this map is interesting in its own right. To assist in explaining the mapping we define the *shared border length* $s(x,y)$ of two bars x and y to be the number of common edges of x and y. We say that x and y are *adjacent* if $s(x,y) \geq 0$.

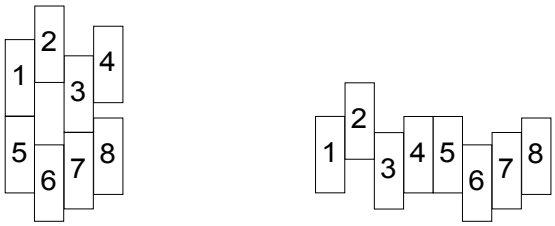

Figure 4: $y \in K$ and the corresponding $y' \in S$.

Let $y \in K$. Assume that y contains a single connected component of bars. The idea of the map is to remove one bar from each column in y, concatenating all of these bars on the right side of the component at the same relative heights that they had in y. Except for a decrease of $2L$ where the two rows meet, the perimeter remains essentially unchanged between y and y'. The crucial element of this map is that it maintains the perimeter, and yet we can still recover the original configuration from y'. The example shown in Figure 4 demonstrates the general approach. Notice that in y', bars 3 and 8 are no longer adjacent. In order to preserve the original perimeter we increased $s(3,4)$ by $s(3,8)$. However, knowing the original locations of bars 3 and 8 immediately tells us $s(3,8)$ which indicates that $s(3,4)$ in y' was increased by $s(3,8)$, allowing us to recover the original position of bar 4.

In order to use the conductance ϕ_S to provide the desired bound on the total conductance $\Phi_{\widehat{\mathcal{M}}_{DK}}$ we need that $\widehat{\pi}(S)$ is at most a polynomial times $\widehat{\pi}(\overline{S})$. When λ is large, this is indeed the case.

LEMMA 4.1. *Assume $0 < c_1 \leq 1/4$, $b \geq 8c_1 n/3$ and $\lambda \geq 4^{c_1^{-1}}$. Then $\widehat{\pi}(S) \leq 2\mu^{-1}\widehat{\pi}(\overline{S})$.*

Proof. Let S_k be the set of configurations in S with exactly k components. It is easy to show that $\widehat{\pi}(S) \leq$

$2\widehat{\pi}(S_1)$. Namely, we define a map $f : S \to S_1$ as follows. Given $x \in S$ with k components, numbered from the minimal leftmost component to the right, for $j = 2, 3, \ldots, k$, shift the entire jth component so that its leftmost bar aligns fully with the rightmost bar in the $j-1$st component. As in previous proofs, $\widehat{\pi}(x) = \mu^{2L(k-1)}\widehat{\pi}(f(x))$ and $|f^{-1}(f(x))| \leq n^{3(k-1)}$. Then for $\lambda \geq 4^{c_1^{-1}}$, by (3.7),

$$\begin{aligned}
\widehat{\pi}(S) &= \sum_{y\in S_1}\sum_{k=1}^{b}\sum_{\substack{x\in\\ S_k\cap f^{-1}(y)}} \mu^{2L(k-1)}\widehat{\pi}(y) \\
&\leq \sum_{y\in S_1}\sum_{k=1}^{b}\left(n^3\mu^{2L}\right)^{k-1}\widehat{\pi}(y) \\
&\leq 2\sum_{y\in S_1}\widehat{\pi}(y) = 2\widehat{\pi}(S_1). \qquad (4.9)
\end{aligned}$$

We will bound $\widehat{\pi}(S_1)$ by induction on b. For the purposes of this induction, we emphasize the effect of b by defining $S_1(b)$ to be the set S_1 under the condition that the model has b bars. Recall that the weight of $S_1 = S_1(b)$ in the projection is $\widehat{\pi}(S_1(b)) = Z^{-1}(1+\lambda)^{2n^2-2bL}\sum_{\sigma\in S_1(b)}\mu^{\kappa(\sigma)}$. We define $W(b) = \sum_{x\in S_1(b)}\mu^{\kappa(x)}$. We will show that for $b \geq 2$, $W(b) \leq \mu^{\frac{7}{4}b+2L}n^2$. Suppose there is just one bar; then $W(1) = \mu^{2L+2}n^2$, since there are n^2 places to put the first bar. Given a configuration with b bars, we can add another bar at the left side according to the following probabilities. If the bar fully aligns with the leftmost bar, then the change in perimeter is 2; on the other hand, there are 2 ways to have an shared border of i for each $i \in [L-1]$, each of which has a change in perimeter of $2L+2-2i$. There are $n-2L+1$ ways to have no shared border. Hence we have

$$\begin{aligned}
&W(b+1) \\
&= W(b)\left(\mu^2 + 2\sum_{i=1}^{L-1}\mu^{2L+2-2i} + (n-2L+1)\mu^{2L+2}\right) \\
&= W(b)\left(\mu^2 + 2\left(\frac{\mu^2-\mu^{2L}}{\mu^{-2}-1}\right) + (n-2L+1)\mu^{2L+2}\right) \\
&\leq W(b)\left(\mu^2 + \frac{2\mu^2}{\lambda} + \frac{\mu^2}{2}\right) \leq 2\mu^2 W(b).
\end{aligned}$$

Therefore $W(b) \leq (2\mu^2)^{b-1}\mu^{2L+2}n^2 \leq \mu^{\frac{7}{4}b+2L}n^2$, since $\lambda \geq 4^{1/c_1} \geq 4^4$, which implies $2^{b-1}\mu^{b/4} \leq 1$.

Consider $\overline{S}$. Let $\overline{S}_{2row}$ be the set of configurations with a single component of bars contained in a $2L \times \lfloor (b+1)/2 \rfloor$ rectangle. Then $\widehat{\pi}(\overline{S}_{2row}) \geq \mu^{4L+b+1}n^2$. Therefore since $b \geq 8L/3$,

$$\begin{aligned}
\frac{\widehat{\pi}(S)}{\widehat{\pi}(\overline{S})} &\leq \frac{2\widehat{\pi}(S_1)}{\widehat{\pi}(\overline{S}_{2row})} \leq \frac{2\mu^{\frac{7}{4}b+2L}n^2}{\mu^{4L+2b+1}n^2} \\
&\leq 2\mu^{\frac{3}{4}b-2L-1} \leq \frac{2}{\mu},
\end{aligned}$$

as claimed. □

REMARK 2. *The careful reader might notice that the recursion for $W(b)$ obtained in this proof, which exactly counts $\widehat{\pi}(S_1)$, provides a way to approximate $\pi'(\Omega')$ when λ is sufficiently large. In this case, the configurations in Ω' with more than one component are exponentially suppressed. Therefore this could be used to approximately sample from Ω' without the use of $\mathcal{M}_{DK}$. However, Ω' provides a nontrivial setting where $\mathcal{M}_{DK}$ yields an exponential speedup over the local algorithm.*

We may now prove Theorem 4.1 to show $\mathcal{M}_{DK}$ requires exponential time on Ω.

Proof of Theorem 4.1: Recall that our goal is to demonstrate that $(S, \overline{S})$ is a bad cut in the state space of the projection chain $\widehat{\mathcal{M}}_{DK}$. In particular, we show

$$\phi_S \leq 16n^{10}\mu^{2L}. \qquad (4.10)$$

If $\widehat{\pi}(S) \leq 1/2$ then this directly yields a bound on $\Phi_{\widehat{\mathcal{M}}_{DK}}$, since in this case $\Phi_{\widehat{\mathcal{M}}_{DK}} \leq \phi_S$. Suppose not. Then by the reversibility of $\widehat{\mathcal{M}}_{DK}$ and Lemma 4.1,

$$\begin{aligned}
\Phi_{\widehat{\mathcal{M}}_{DK}} &\leq \widehat{\pi}(\overline{S})^{-1}\sum_{x\in S, y\notin S}\widehat{\pi}(y)\widehat{\mathcal{P}}_{DK}(y,x) \\
&= \widehat{\pi}(\overline{S})^{-1}\sum_{x\in S, y\notin S}\widehat{\pi}(x)\widehat{\mathcal{P}}_{DK}(x,y) \\
&= \frac{\widehat{\pi}(S)}{\widehat{\pi}(\overline{S})}\phi_S \leq 2\mu^{-1}\phi_S.
\end{aligned}$$

Thus in either case, $\Phi_{\widehat{\mathcal{M}}_{DK}} \leq 32n^{10}\mu^{2L-1}$. It is easy to see from the definition of conductance and Definition 3.1 that $\Phi_{\mathcal{M}_{DK}} \leq \Phi_{\widehat{\mathcal{M}}_{DK}}$. Using Theorem 2.2, we find $Gap(\mathcal{P}_{DK}) \leq 64n^{10}\mu^{2L-1}$, and by Theorem 2.1,

$$\begin{aligned}
\tau(\epsilon) &\geq \left(\frac{1}{128n^{10}\mu^{2L-1}} - \frac{1}{2}\right)\log\left(\frac{1}{2\epsilon}\right) \\
&\geq \left(\frac{(1+\lambda)^{L-1}}{128n^{10}} - \frac{1}{2}\right)\log\left(\frac{1}{2\epsilon}\right),
\end{aligned}$$

which serves as an exponential lower bound, as desired. Thus it suffices to prove (4.10). Recall that if our Markov chain $\widehat{\mathcal{M}}_{DK}$ is currently at a configuration in S,

then it can only move to a configuration with at most two bars per column. Hence

$$\begin{aligned}\phi_S &= \widehat{\pi}(S)^{-1} \sum_{x\in S, y\notin S} \widehat{\pi}(x)\widehat{\mathcal{P}}_{DK}(x,y) \\ &= \widehat{\pi}(S)^{-1} \sum_{x\in S, y\in K} \widehat{\pi}(x)\widehat{\mathcal{P}}_{DK}(x,y) \\ &= \widehat{\pi}(S)^{-1} \sum_{x\in S, y\in K} \widehat{\pi}(y)\widehat{\mathcal{P}}_{DK}(y,x),\end{aligned}$$

where the last step follows from $\widehat{\mathcal{M}}_{DK}$ being reversible. For a configuration $y \in K$, define $c(y)$ to be the number of columns in y with two bars r_1, r_2 such that $s(r_1, r_2) \neq 0$. Define $\Psi(y) = \widehat{\pi}(y)\mu^{2c(y)}$ and $\Psi(K) = \sum_{y\in K} \Psi(y)$, which is the weight of K if we ignore the shared border of two bars in the same column.

We now bound the probability of moving from $y \in K$ into S. We will show that for all $y \in K$,

$$\sum_{x\in S} \widehat{\mathcal{P}}_{DK}(y,x) \leq n^2\mu^{2c(y)}. \tag{4.11}$$

The general idea is that in order to move from a configuration in K to a configuration in S using $\mathcal{M}_{DK}$, a pivot must be chosen that separates each pair of bars that are vertically adjacent by rotating their shared border to an edge without any specks. For a pivot point p the weight of the set of configurations in $\Gamma^{-1}(y)$ that have a possible transition to some σ' with bar structure in S is at most a $1/(1+\lambda)^{c(y)} = \mu^{2c(y)}$ fraction of the total weight of $\Gamma^{-1}(y)$.

For any $y \in K$ and any pivot p, define y_p to be the configuration obtained after rotating the bars in y 180° around p. Let $H_p(y) \subset E(G)$ be the set of horizontal edges in the lattice graph G that are incident to two bars in y_p, and let $\Gamma^{-1}_{H_p}(y) \subset \Gamma^{-1}(y)$ be the set of all configurations $\rho \in \Gamma^{-1}(y)$ that, after one move of $\mathcal{M}_{DK}$ with pivot p, can transition to a configuration whose bar structure is in S. In order for a configuration $\rho \in \Gamma^{-1}(y)$ to map to a configuration with at most one bar per column, no edge in $H_p(y)$ can be incident to a tile in y. If any $e \in H_p(y)$ is incident to a bar in y then regardless of where the specks are, such a configuration cannot transition to a configuration in S, and so $\Gamma^{-1}_{H_p}(y) = \emptyset$. Otherwise, if there is a speck in y on any edge $e \in H_p(y)$, then the two bars that are adjacent to e in y_p will be in the same DK-component and therefore stay in the same column after a transition of $\mathcal{M}_{DK}$ with pivot p. Therefore, in this case, $\Gamma^{-1}_{H_p}(y)$ is the set of all configurations in $\Gamma^{-1}(y)$ such that there are no specks on the edges of $H_p(y)$. Define $\Gamma^{-1}_H(y) = \cup_p \Gamma^{-1}_{H_p}(y)$. This implies that for any $y \in K$,

$$\begin{aligned}\sum_{x\in S} \widehat{\mathcal{P}}_{DK}(y,x) &= \sum_{x\in S} \frac{1}{\pi(\Gamma^{-1}(y))} \sum_{\substack{\sigma\in\Gamma^{-1}_H(y)\\ \rho\in\Gamma^{-1}(x)}} \pi(\sigma)\mathcal{P}_{DK}(\sigma,\rho) \\ &\leq \frac{1}{\pi(\Gamma^{-1}(y))} \sum_{\sigma\in\Gamma^{-1}_H(y)} \pi(\sigma) \;\leq\; \frac{\pi(\Gamma^{-1}_H(y))}{\pi(\Gamma^{-1}(y))}.\end{aligned}$$

For a configuration $y \in \widehat{\Omega}$, let $e(y)$ be the number of edges not incident to any bar in y. If $\Gamma^{-1}_{H_p}(y) \neq \emptyset$ for a pivot p, then $\pi(\Gamma^{-1}_{H_p}(y)) = Z^{-1}\sum_{i=0}^{t-i} \lambda^{e(y)-c(y)}\binom{e(y)-c(y)}{i} = Z^{-1}(1+\lambda)^{e(y)-c(y)}$. Thus

$$\begin{aligned}\sum_{x\in S} \widehat{\mathcal{P}}_{DK}(y,x) &\leq \frac{\pi(\Gamma^{-1}_H(y))}{\pi(\Gamma^{-1}(y))} \\ &= n^2 \frac{Z^{-1}(1+\lambda)^{e(y)-c(y)}}{Z^{-1}(1+\lambda)^{e(y)}} \;=\; n^2\mu^{2c(y)}.\end{aligned}$$

Using (4.11) we obtain

$$\begin{aligned}\Phi_{\widehat{\mathcal{M}}_{DK}} &\leq \widehat{\pi}(S)^{-1} \sum_{x\in S, y\in K} \widehat{\pi}(y)\widehat{\mathcal{P}}_{DK}(y,x) \\ &\leq \widehat{\pi}(S)^{-1} \sum_{y\in K} \widehat{\pi}(y) n^2\mu^{2c(y)} \\ &= n^2\widehat{\pi}(S)^{-1}\Psi(K). \qquad (4.12)\end{aligned}$$

Let $K_1 \subset K$ be the set of configurations in K with exactly one component. It is easy to show that since $\lambda \geq 4^{c_1^{-1}}$, $\Psi(K) \leq 2\Psi(K_1)$, by the same proof as in Lemma 4.1 used to show that $\widehat{\pi}(S) \leq 2\widehat{\pi}(S_1)$. Therefore it suffices to consider a single component in the analysis.

Finally, we proceed to the heart of the argument, where we show that $\frac{\Psi(K_1)}{\widehat{\pi}(S)} \leq 16n^8\mu^{2L}$. We define a map $g : K_1 \to S$ and use an information theoretic argument to show that for any $y' \in S$, the total weight of configurations y such that $g(y) = y'$ is exponentially small. Given $y \in K_1$, the following procedure iteratively creates $g(y) \in S$ while at the same time "encoding" sufficient information to recover y given $g(y)$. The encoded information gives a bound on the number of configurations in K_1 that could map to $y' \in S$, and the procedure carefully decreases the perimeter of the configurations when possible, thereby ensuring that the weight of y' is exponentially larger than the weight of any of its preimages. Together, this yields our bound.

For simplicity, we will assume that column 0 contains the leftmost bar in y. Let t be the number of columns in y which contain bars. For $i = 0, 1, \ldots, t-1$, we move a bar from column i to column $t+i$; label this bar q_i and the remaining bar in column i will be denoted

r_i, if it exists. We should think of this new configuration as embedded on a $2t \times n$ torus; it is possible that $2t > n$, in which case our configuration is not actually in $\widehat{\Omega}$. However, by the end of the procedure we will have created a configuration $g(y) \in \widehat{\Omega}$. Note that we may create empty columns through this process; say it results in k components. At the end we will shift the components together for a total loss in perimeter of $2L(k-1)$. We will show how to use this change in weight to bound the total weight of all possible preimages to $g(y)$.

First select any bar in column 0 to label q_0 and place it anywhere in column t. Encode the original height of q_0. Let $i \geq 0$ and suppose bars $q_0, \ldots, q_i$ have already been moved. There are several cases, depending on the location of the bars in y. For visual assistance, we've broken the cases down in terms of how many (0,1, or 2) bars there are in column $i+1$ that are adjacent to q_i. Cases 2.1, 2.3 and 3 are the most interesting, and in fact the other two cases are exponentially suppressed, which is precisely why we can use our information theoretic encoding to take care of them.

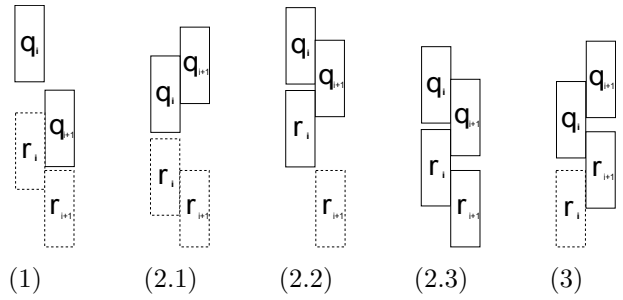

Figure 5: Example configurations for each case. The dotted lines represent bars that may not be present.

Case 1: *q_i is adjacent to 0 bars in column $i+1$.* Label any bar in column $i+1$ by q_{i+1}. Move q_{i+1} to column $t+i+1$ with height $h_{g(y)}(q_i)$. We can encode the column $i+1$ and the original height of q_{i+1}, since this move results in a loss in perimeter of $2L$, unless bar r_i exists. If r_i exists but not r_{i+1} then column i will be the rightmost column of a component after this move, so when we push the components together at the end, we will get a loss in perimeter of $2L$, as desired. Otherwise, if r_i and r_{i+1} both exist, shift every bar in columns $i+1$ through $t+i$ by the same vertical amount such that $s_{g(y)}(r_i, r_{i+1}) = L$, to get the loss in perimeter. Encode the original height of bar r_{i+1}.

Case 2.1: *q_i is adjacent to 1 bar in column $i+1$ (label this bar q_{i+1}) and either column i has only 1 bar or $s(r_i, q_{i+1}) = 0$.* Move q_{i+1} to column $t+i+1$ with height $h_{g(y)}(q_i) + h_y(q_i) - h_y(q_{i+1})$ (at the same relative height that they had previously).

Case 2.2: *q_i is adjacent to 1 bar in column $i+1$ (label this bar q_{i+1}), r_i exists, $s(r_i, q_{i+1}) \neq 0$, and if r_{i+1} exists, $s(r_i, r_{i+1}) = 0$.* Move q_{i+1} to column $t+i+1$ with height $h_{g(y)}(q_i)$. Encode the column $i+1$ and the original height of q_{i+1}. If bar r_{i+1} exists, shift every bar in columns $i+1$ through $t+i$ by the same vertical amount such that $s_{g(y)}(r_i, r_{i+1}) = L$. Encode the original height of bar r_{i+1}.

Case 2.3: *q_i is adjacent to 1 bar in column $i+1$ (label this bar q_{i+1}), r_i and r_{i+1} exist, $s(r_i, q_{i+1}) \neq 0$ and $s(r_i, r_{i+1}) \neq 0$.* Move q_{i+1} to column $t+i+1$ with height $h_{g(y)}(q_i) + h_y(q_i) - h_y(q_{i+1})$. Shift every bar in columns $i+1$ through $t+i$ by the same vertical amount such that $s_{g(y)}(r_i, r_{i+1})$ increases by $s_y(r_i, q_{i+1})$.

Case 3: *q_i is adjacent to 2 bars in column $i+1$.* Since $L \leq \frac{n}{4}$ there must be at least one bar in column $i+1$ that is only adjacent to q_i in column i (label this bar q_{i+1}). Move q_{i+1} to column $t+i+1$ at height $h_{g(y)}(q_i) + h_y(q_i) - h_y(q_{i+1})$. Shift q_{i+1} vertically to increase $s_{g(y)}(q_i, q_{i+1})$ by $s_y(q_i, r_{i+1})$.

After this procedure, the first t columns each contain at most 1 bar and the next t columns each contain exactly one bar. Assume there are k components, numbered from the minimal leftmost component to the right. For $j = 2, 3, \ldots, k$, shift the entire jth component so that its leftmost bar aligns fully with the rightmost bar in the $j-1$st component; using the $2L$ loss in perimeter, record where the components were originally. Next, let j be the largest j such that r_j exists. Shift all bars $q_0, \ldots q_t$ vertically by the same amount such that $s(r_j, q_0) = L$. Notice that we now have a single component with exactly one bar per column. Finally, shift the component horizontally so the first bar is in column 0 of the torus. Since $b < n$ there are at most $n-1$ columns with bars, implying $g(y)$ is a valid configuration in $\widehat{\Omega}$.

Next we will show that given the information encoded above and a configuration $y' \in S$ we can recover y. First, use the information encoded to recover the positions of the k components we pushed together. Recall that we also recorded the location of q_0 in y. Assume that we have already recovered the location of q_i and r_i; to recover the location of q_{i+1} and r_{i+1} we must determine which case was used to move q_{i+1}. If q_{i+1} was moved using Case 1 or Case 2.2, we would have encoded column i, the original location of q_{i+1}, and any amount r_{i+1} was shifted, so we can recover this information. Otherwise, if $s(q_i, r_{i+1}) \neq 0$ then Case 3 was used. This indicates that $s(q_{i+1}, q_i)$ was increased by $s(q_i, r_{i+1})$ in y' and so we can recover the original $s(q_{i+1}, q_i)$ and location. Finally, if none of these have occurred then we have used Case 2.1 or 2.3. In either of these cases the shared border between q_i and q_{i+1} was not changed so we can recover the original location of

q_{i+1}. If in addition $s(r_i, q_{i+1}) \neq 0$, then we used Case 2.3 where r_{i+1} and all bars from columns $i+1$ to $t+i$ were shifted to increase the shared border by $s(r_i, q_{i+1})$. Given this, we can undo the shift and recover the original location of r_i.

It is not difficult to see that the mapping g can be used to show that $\Psi(K_1)$ is exponentially smaller than $\widehat{\pi}(S)$; specifically, the bound obtained is

$$\frac{\Psi(K_1)}{\widehat{\pi}(S)} \leq 16n^8\mu^{2L}. \tag{4.13}$$

For a particular $y \in K_1$, consider the procedure that determines $g(y)$. Define $f_1(y)$ to be the number of times Case 1 is applied except when r_i exists but not r_{i+1}, $f_2(y)$ to be the number of times Case 2.2 is applied when r_{i+1} exists, $f_3(y)$ to be the number of times Case 2.2 or Case 1 is applied when r_{i+1} does not exist but r_i does, and $f_4(y)$ to be $k-1$ if r_j and q_0 are in adjacent columns before compression and $k-2$ otherwise. Define $f(y) = (f_1(y), f_2(y), f_3(y), f_4(y))$. Each time Case 1 or Case 2.2 is applied the column and the original height of one or two bars is encoded. During the final compression step for each component shifted, the amount shifted and the original height of each component is encoded. From these observations, for a particular y the total amount of information encoded is at most $n^{F(f_1(y),f_2(y),f_3(y),f_4(y))}$ where $F(i,j,k,l) = 2 + 3i + 3j + 2k + 3(l+1)$. The additional n^2 factor is needed to encode the original height and column of q_0. Since this information suffices to recover y such that $g(y) = y'$, it can be used to bound the number of preimages for a particular $y' \in S$. Each time Case 1 is applied (except when r_i exists but not r_{i+1}), Case 2.2 is applied when r_{i+1} exists and each time two components are joined together in the compression step, there is a decrease in perimeter of $2L$. In addition there is a decrease in perimeter of $2L$ between the two sets of bars q, r. Combining these observations shows that $\Psi(y) \leq \mu^{[2L(f_1(y)+f_2(y)+f_4(y)+1)]}\widehat{\pi}(g(y))$. Also notice that since $f_3(y)$ counts the number of indices i such that column i has 2 bars and column $i+1$ has 1 bar, and we remove one bar from each nonempty column, $f_3(y)$ corresponds to the number of components in the compression step. This implies $f_3(y) \leq f_4(y)+1$. Since $\lambda \geq 8^{c_1^{-1}}$ implies $n^5\mu^{2L} \leq 1/2$, we have

$$\begin{aligned}
\Psi(K_1) &= \sum_{y' \in S} \sum_{\substack{i,j,k,l \in \\ [0,n-1]}} \sum_{\substack{y \in K_1 : g(y) = y', \\ f(y) = (i,j,k,l)}} \Psi(y) \\
&\leq \sum_{y' \in S} \widehat{\pi}(y') \sum_{\substack{i,j,k,l \in \\ [0,n-1]}} n^{F(i,j,k,l)} \mu^{2L(i+j+l+1)} \\
&\leq n^5 \mu^{2L} \sum_{y \in S} \widehat{\pi}(y) \sum_{\substack{i,j,k,l \in \\ [0,n-1]}} n^{3i+3j+2(l+1)+3l} \mu^{2L(i+j+l)} \\
&\leq n^8 \mu^{2L} \sum_{y \in S} \widehat{\pi}(y) \sum_{\substack{i,j,l \in \\ [0,n-1]}} n^{3i+3j+5l} \mu^{2L(i+j+l)} \\
&\leq 16 n^8 \mu^{2L} \widehat{\pi}(S),
\end{aligned}$$

as desired. □

References

[1] N. Bhatnagar and D. Randall. Torpid mixing of simulated tempering on the Potts model. Proc. of the 15th Symp. of Discrete Algorithms (SODA), 478–487, 2004.

[2] A. Buhot and W. Krauth. Phase separation in two-dimensional additive mixtures. Phys. Rev. E **59:** 29392941, 1999.

[3] C. Dress and W. Krauth. Cluster algorithm for hard spheres and related systems. J. Phys. A: Math. Gen. **28:** L597, 1995.

[4] D. Frenkel and A. Louis. Phase separation in binary hard-core mixtures: an exact result. *Physical Review Letters,* **68:** 3363–3365, 1992.

[5] V. Gore and M. Jerrum. The Swendsen-Wang algorithm does not always mix rapidly. Proc. of the Twenty-Ninth ACM Symp. on the Theory of Computing, 674–681, 1997.

[6] M. Jerrum and A. Sinclair. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation,* **82:** 93133, 1989.

[7] W. Krauth. Cluster Monte Carlo algorithms. Chapter in *New Optimization Algorithms in Physics*, edited by A. K. Hartmann and H. Rieger, (Wiley-VCh), 2005.

[8] D. Levin, Y. Peres, and E. Wilmer. *Markov Chains and Mixing Times.* American Mathematical Society, Providence, 2009.

[9] N. Madras and D. Randall. Markov chain decomposition for convergence rate analysis. *Annals of Applied Probability*, **12**: 581-606, 2002.

[10] R. Martin and D. Randall. Disjoint Decomposition of Markov Chains and Sampling Circuits in Cayley graphs. Combinatorics, Probability and Computing, **15**: 411–448, 2006.

[11] F. Martinelli and A. Sinclair. Mixing time for the solid-on-solid model. *Proc. of the 41st ACM Symp. on the Theory of Computing (STOC) 2009*, 571-580, 2009.

[12] D. Randall. Slow mixing of Glauber dynamics via topological obstructions. Proc. of the 17th ACM-SIAM Symp. on Discrete Algorithms (SODA), 870–879, 2006.

[13] A.J. Sinclair. *Algorithms for random generation & counting: a Markov chain approach.* Birkhauser, Boston, 1993.

[14] D. B. Woodard, S. C. Schmidler and M. Huber. Conditions for torpid mixing of parallel and simulated tempering on multimodal distributions. Electronic Journal of Probability, **14**: 780804, 2009.

[15] D. B. Woodard, S. C. Schmidler and M. Huber. Conditions for Rapid Mixing of Parallel and Simulated Tempering on Multimodal Distributions. Annals of Applied Probability, **19**: 617640, 2009.