Analyzing Glauber Dynamics by Comparison of Markov Chains*

Dana Randall† Prasad Tetali‡

November 12, 1999

Abstract

A popular technique for studying random properties of a combinatorial set is to design a Markov chain Monte Carlo algorithm. For many problems there are natural Markov chains connecting the set of allowable configurations which are based on local moves, or "Glauber dynamics." Typically these single site update algorithms are difficult to analyze, so often the Markov chain is modified to update several sites simultaneously. Recently there has been progress in analyzing these more complicated algorithms for several important combinatorial problems.

In this work we use the comparison technique of Diaconis and Saloff-Coste to show that several of the natural single point update algorithms are efficient. The strategy is to relate the mixing rate of these algorithms to the corresponding non-local algorithms which have already been analyzed. This allows us to give polynomial time bounds for single point update algorithms for problems such as generating planar tilings and random triangulations of convex polygons. We also survey several other comparison techniques, along with specific applications, which have been used in the context of estimating mixing rates of Markov chains.

1 Introduction

Random sampling of combinatorial structures such as tilings, colorings and independent sets of a graph has attracted the attention of researchers in combinatorics, theoretical computer science and statistical physics in recent years. The Markov chain Monte Carlo method has played a crucial role in establishing efficient algorithms for almost uniform sampling of such structures and in yielding fully polynomial randomized approximation schemes for the corresponding counting and sampling problems (see e.g. [12], [17], and [42] for general surveys).

Establishing such rigorous bounds for Markov chain Monte Carlo algorithms has proven a compelling challenge. In many cases, very simple algorithms based on local updates suggest themselves, and they have been widely used in practice to study various physical

*A preliminary version of this paper appeared in *Springer Lecture Notes in Computer Science,* Vol. 1380, 1998, pp. 292-304.

†College of Computing and School of Mathematics, Georgia Institute of Technology, Atlanta GA 30332, randall@math.gatech.edu. Research supported by NSF Grant No. CCR-9703206.

‡School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, tetali@math.gatech.edu. Research supported by NSF Grant No. CCR-9503952.

systems. Typically it is straightforward to design a Markov chain which connects the state space and which converges to the desired distribution over configurations (e.g., the Gibbs or the uniform distribution). These simple chains are commonly referred to as *Glauber dynamics*. The widespread use of these algorithms is based on a belief that the algorithms converge quickly to their stationary distribution so that sampling after a small number of steps is representative of this limiting distribution. However, typically a rigorous analysis is omitted due to the difficulty in establishing such bounds.

Recently there has been great progress in developing analytical techniques for bounding the convergence rates of such chains where the underlying state space represents configurations in a physical system or, more generally, elements of a combinatorial set. The first notable achievement is the method for bounding the *conductance* (or discrete Cheeger constant) due to Jerrum and Sinclair [34], which has been used to develop fully polynomial randomized approximation schemes for the partition function associated with many dimer systems [15, 19] and the Ising model [14]. The second significant advance is the coupling technique refined by Aldous [1], which has led to efficient approximation algorithms for colorings [13, 40], independent sets [24], and planar dimer models [23], as well as problems of a more combinatorial nature (e.g., [3, 18, 29]). Further refinement, via *path coupling*, due to Bubley and Dyer [3] has made the coupling technique by far the most successful as a way to prove the rapid mixing of Markov chains.

While this explosion of results has been encouraging, many natural Markov chains which are believed to be rapidly converging to stationarity continue to resist analysis. In fact, many of the solutions cited above have involved clever manipulations of the simplest Markov chains so that they still have the desired properties (i.e., they are easy to implement and have the desired stationary distributions) but they are enhanced with additional moves which enables a simple analysis. These enhanced chains are inspired by heat bath algorithms, including the "tower moves" in the case of tilings [23], "edge moves" in the case of independent sets [24,9] and "Kempe-chain moves" for colorings [40].

In this work we derive rigorous bounds for the Glauber dynamics for these problems by comparing these simpler chains to the enhanced chains which have already been shown to converge rapidly. We focus on two applications which best demonstrate the versatility of this method: generating lozenge tilings on the triangular lattice (a planar dimer model) and generating random triangulations of a convex polygon. Further applications of this method to the case of domino tilings, colorings and independent sets can be found in [32] (see also [40]).

The proof technique uses a comparison theorem due to Diaconis and Saloff-Coste [4]. Their theorem yields a geometric comparison inequality that gives bounds on the eigenvalues of a reversible Markov chain in terms of the eigenvalues of a second chain. The main application in [4] was to give a sharp upper bound on the second eigenvalue of the symmetric exclusion process on a graph. The symmetric exclusion process on a graph is a certain generalization of a simple random walk on a graph and can be described as follows. Start with an arbitrary placement of r particles on r vertices of a graph. At each discrete time step, a particle is chosen at random, and then one of its neighboring vertices is chosen at random. If the neighbor is unoccupied (by a particle) then the chosen particle is moved there, otherwise the system stays as it was. The special case of $r = 1$ corresponds to the simple random walk on a graph. Diaconis and Saloff-Coste bound the second eigenvalue of this chain by comparing it to a well-studied chain (the Bernoulli–Laplace model for

diffusion) whose eigenvalues are known.

Our approach is somewhat different in this paper since the Markov chains we work with are much more combinatorial in nature; in particular, it is very hard to determine the second eigenvalue of these chains or of related chains with the same stationary distributions. However, the "known" chains in our applications are chains whose mixing times are known. Using the existing literature on relating the time to reach equilibrium and the second eigenvalue (e.g. [8, 34]), together with the comparison theorem, we derive an inequality relating the mixing times of two chains. This allows us to estimate the rate of mixing of the "unknown" chains (based on single site updates) mentioned above. Direct analysis of any of these chains seems challenging and might yield tighter bounds on the mixing times.

After describing our two main applications, we devote a section to survey other comparison techniques which have found useful applications in this topic. Perhaps the first person to use a comparison technique in the context of Markov chain Monte Carlo was Holley [11]. Holley's hypothesis however, as we shall see in Section 5, is too stringent for the kind of applications we mention in this paper. Diaconis and Saloff-Coste also mention a few variants of the comparison theorem one of which, using the notion of *flows*, is slightly more general than the version we use in our two main applications here. Very recently Vigoda managed to make use of this generality in the context of sampling k-colorings of a graph. We describe this result briefly, while referring the reader to [40] for the full details. We then describe the role of the logarithmic Sobolev constant in bounding mixing times, and the corresponding comparison techniques, following [6]. Recently Dyer and Greenhill also used a different, simpler comparison technique to argue that the Glauber dynamics for independent sets (i.e. the hardcore (lattice) gas model on arbitrary graphs) is efficient [9]. Their basic idea is nice, and is also a statement about comparison of eigenvalues (see Section 5 for details). However, their method is not as widely applicable and is only effective when the enhanced moves can be implemented using a (very small) bounded number of Glauber moves; we will see that this is not the case for the two main applications given in this paper. Finally, there have been several results relating block dynamics and Glauber dynamics for certain problems, due to Martinelli [28], van den Berg and Brouwer [2], and (indirectly) Madras and Randall [26, 27].

In Section 2 we describe relevant results from the theory of rapidly mixing Markov chains, including relations between mixing times and eigenvalues, and comparison inequalities. In Sections 3 and 4 we describe in detail the application to lozenge tilings and triangulations. In Section 5 we describe briefly several other related comparison techniques.

2 The comparison theorem and mixing rates

Let (Ω, P, π) denote an ergodic (i.e. irreducible and aperiodic) Markov chain with finite state space Ω, transition probability matrix P, and stationary distribution π. Furthermore, we assume that the chain is reversible, i.e. that we have the detailed–balance conditions, $\pi(x)P(x,y) = \pi(y)P(y,x)$, for all $x, y \in \Omega$. Assuming we are dealing with discrete-time Markov chains, for $x, y \in \Omega$, $t \in \mathcal{Z}^+$, let $P^t(x,y)$ denote the t-step probability of going from x to y. Then the time a Markov chain takes to be close to equilibrium can be measured

using the *variation distance* between P^t and π, where the variation distance is given by

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

We also denote by $\Delta(t)$ the variation distance starting from the worst state, i.e. $\Delta(t) = \max_{x \in \Omega} \Delta_x(t)$.

• mixing time and the second eigenvalue

For $\epsilon > 0$, the *mixing time*, starting from state x, is defined by

$$\tau_x(\epsilon) = \min\{t : \Delta_x(t') \leq \epsilon, \qquad \forall t' \geq t\}.$$

Once again we denote by $\tau(\epsilon)$, the mixing time starting from the worst state, i.e. $\tau(\epsilon) = \max_{x \in \Omega} \tau_x(\epsilon)$. For the rest of the paper, when we refer to *mixing time*, we always mean $\tau(\epsilon)$.

Let $1 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{|\Omega|-1} > -1$ denote the eigenvalues of P. The following result of Sinclair [35] (which is an extension of a key result from [34]; see also [8]) shows the relationship between mixing times and the maximum eigenvalues. Strictly speaking, λ_1 in the following theorem should be replaced by $\lambda_{max} = \max(\lambda_1, |\lambda_{|\Omega|-1}|)$, but in all our applications below we make sure that $\lambda_1 > |\lambda_{|\Omega|-1}| > 0$ by adding self-loops with weight 1/2.

Theorem 1 *For $\epsilon > 0$, we have*
(i) for all $x \in \Omega$, $\tau_x(\epsilon) \leq \frac{1}{1-\lambda_1} \log\left(\frac{1}{\pi(x)\epsilon}\right)$;
(ii) $\max_x \tau_x(\epsilon) \geq \frac{\lambda_1}{2(1-\lambda_1)} \log\left(\frac{1}{2\epsilon}\right)$.

• mixing time and coupling time

Another method for bounding the mixing time $\tau(\epsilon)$ is to construct a *coupling* for the Markov chain. A coupling is a new Markov chain on the state space $\Omega \times \Omega$ (where Ω is the original state space, e.g., the set of 3-colorings) with the following properties. Rather than updating two configurations independently, the coupled process correlates the random coin flips while maintaining that each configuration, when observed in isolation, is just performing transitions of the original Markov chain. In addition, we need that if the two configurations agree, the coupled process will force them to agree at all future times. Coupling is a crucial ingredient in all of the applications in section 3.

The following theorem states that the coupling time, which is the expected time it takes for two configurations to meet starting from the worst starting point, provides a good bound on the mixing time. More formally, let x and y be the starting configurations. Then

$$T^{x,y} = \min\{t : X_t = Y_t \mid X_0 = x,\, Y_0 = y\},$$

and define the *coupling time* to be $T = \max_{x,y} \mathrm{E}T^{x,y}$. The following result relates the mixing time and the coupling time (see [1]).

Theorem 2 $\tau(\epsilon) \leq 6T(1 + \ln \epsilon^{-1})$.

• **comparison of eigenvalues (via Dirichlet forms)**

Let $\tilde{P}$ and P denote two reversible Markov chains on the same state space Ω with the same stationary distribution π. Then Diaconis and Saloff-Coste (see [4]) provide the following geometric bound between the two eigenvalues $\lambda_1(\tilde{P})$ and $\lambda_1(P)$. Strictly speaking, the result in [4] compares the Dirichlet forms associated with $\tilde{P}$ and P thus yielding the following comparison result between *all* nontrivial eigenvalues, and not just the second eigenvalue. (Also, the assumption that the stationary distributions be identical can be weakened to the stationary distributions be comparable.)

First we need some more notation. As we shall see, in applications, $\tilde{P}$ is the chain with known eigenvalues (or known mixing time), and P is the chain whose mixing time we would like to bound by comparing with $\tilde{P}$. Let $E(P) = \{(x,y) : P(x,y) > 0\}$ and $E(\tilde{P}) = \{(x,y) : \tilde{P}(x,y) > 0\}$ denote the sets of edges of the two chains, viewed as directed graphs. For each x, y with $\tilde{P}(x,y) > 0$, define a *path* γ_{xy} using a fixed sequence of states, $x_0 = x, x_1, \ldots, x_{k-1}, x_k = y$ with $P(x_i, x_{i+1}) > 0$. The length $(= k)$ of such a path will be denoted by $|\gamma_{xy}|$. Further let

$$\Gamma(z,w) = \left\{(x,y) \in E(\tilde{P}) \text{ such that } (z,w) \in \gamma_{xy}\right\}$$

denote the set of paths which use the transition (z,w).

Theorem 3 *With the above notation, we have*

$$(1 - \lambda_1(P)) \geq \frac{1}{A}(1 - \lambda_1(\tilde{P})),$$

where

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

It is worth noting that the quantity A above depends on our choice of paths $\{\gamma_{xy}\}$; thus these paths play a role akin to that of the *canonical paths* (introduced by Jerrum and Sinclair [15, 34]) in bounding the *conductance* of a Markov chain. However, the crucial difference, as pointed in [4], is that we need only define these paths between pairs of states which are adjacent in the known chain.

In the following our strategy is as follows. We begin with a bound on the mixing time of a chain established, say, via the coupling method and Theorem 2. We then use part (ii) of Theorem 1 above to lower bound the spectral gap of such a chain. Next we use the comparison theorem (Theorem 3) to lower bound the spectral gap of an unknown chain by carefully bounding the parameter A. This in turn provides us with a bound on the mixing time of the unknown chain in view of part(i) of Theorem 1. The following technical proposition makes precise the aforementioned strategy, and is thus crucial to our results of the next section.

Let $\tilde{\tau}(\epsilon)$ and $\tau(\epsilon)$ denote the mixing times of $\tilde{P}$ and P respectively. Then with A as in Theorem 3, we have the following comparison result relating the mixing times. Let π_* denote $\min_{x\in\Omega} \pi(x)$.

Proposition 4 *For $0 < \epsilon < 1$, and for all $x \in \Omega$, we have,*

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

Proof. For $0 < \epsilon < 1$, from part (ii) of Theorem 1, we have,

$$\tilde{\tau}(\epsilon) \geq \frac{\lambda_1(\tilde{P})}{2(1-\lambda_1(\tilde{P}))} \log\left(\frac{1}{2\epsilon}\right).$$

This implies that

$$(1-\lambda_1(\tilde{P})) \geq \frac{1}{4\tilde{\tau}(\epsilon)} \log\left(\frac{1}{2\epsilon}\right),$$

wherein we also used the trivial bound, $\lambda_1(\tilde{P}) \geq 1/2$. Now using the comparison theorem, we get that

$$1-\lambda_1(P) \geq \frac{1}{A}(1-\lambda_1(\tilde{P})) \geq \frac{1}{A}\frac{1}{4\tilde{\tau}(\epsilon)} \log\left(\frac{1}{2\epsilon}\right).$$

Finally, using part(i) of Theorem 1 we can bound the mixing time of P, starting from any state x,

$$\tau_x(\epsilon) \leq \frac{4\log(1/(\epsilon\pi(x)))}{\log(1/2\epsilon)} A\tilde{\tau}(\epsilon),$$

completing the proof of the proposition. □

Remark 1. The above proposition illustrates the fact that the comparison argument is effective as long as we can control the factor A, which depends on the choice of paths in the unknown chain. The dependence on π_*, albeit not as crucial, can affect the mixing time by another factor involving the size of the input, since $1/\pi_*$ in most cases is at most exponential in the size of the input.

Remark 2. Note that Theorem 3 also holds with spectral gaps replaced by the *logarithmic Sobolev* constants of the Markov chains, since the definitions of both the spectral gap and the log-Sobolev constant use the (same) Dirichlet form (see Section 5 and also [6]). Although the log-Sobolev constant offers a tighter upper bound on the mixing time (as defined here), due to the lack of an appropriate lower bound, we are not able to make use of the comparison of log-Sobolev constants. However, this is certainly a useful avenue, if one is able to bound other (e.g. $L_2(\pi)$) notions of mixing time for the chain that one is comparing to (see comments following Corollary 10 in Section 5).

3 Lozenge tilings

Let R be a region of the triangular lattice. A lozenge tiling of R is a covering of the region with lozenges tiles, where each lozenge covers two adjacent cells in R and no two lozenges overlap. Just looking at a lozenge tiling causes a three dimensional surface to appear – in fact the set of lozenge tilings correspond bijectively with the surfaces formed by placing unit cubes in a larger three-dimensional frame such that each cube is supported on its back three sides. The shape of the frame is uniquely determined from the region R (see Fig. 1).

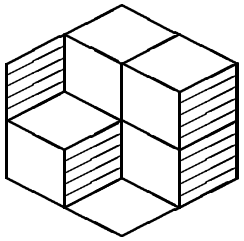

Figure 1: A lozenge tiling viewed as a surface

Given this equivalence, there is an obvious Markov chain $\mathcal{M}_{ssu}$ for generating tilings. Namely, connect any two tilings whose surfaces differ by the addition or removal of a single cube. In the two dimensional picture of a tiling this corresponds to chosing a hexagonal window and if it is comprised of three tiles, rotate them by 60^o (see Fig. 2). More precisely, the transition matrix $P(\cdot\,,\cdot)$ of $\mathcal{M}_{ssu}$ is defined as

$$P(x,y) = \begin{cases} 1/2N, & \text{if } x \oplus y \text{ is a cube (or hexagon)} \\ 1 - \sum_{z \neq x} \tilde{P}(x,z), & \text{if } y = x, \end{cases}$$

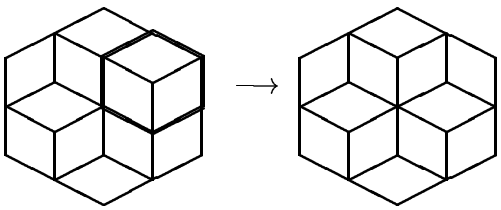

Figure 2: A move in the Markov chain $\mathcal{M}_{ssu}$

In [23] a modified algorithm based on tower moves was analyzed. Again the state space is the set of all lozenge tilings. Two tilings differ by a tower of height k if they differ by the addition or removal of a $1 \times 1 \times k$ vertical column of cubes. Let x and y be lozenge tilings of the region R. Let $\mathcal{M}_{loz}$ represent the Markov chain in which there is a move from x to y if and only if the symmetric difference of the edges of x and y is a tower. Recall that the transition probabilities $\tilde{P}(\cdot\,,\cdot)$ of $\mathcal{M}_{loz}$ are defined by

$$\tilde{P}(x,y) = \begin{cases} 1/2Nh, & \text{if } x \oplus y \text{ is a} \\ & \text{tower of height } h; \\ 1 - \sum_{z \neq x} \tilde{P}(x,z), & \text{if } y = x, \end{cases}$$

where N is the area of the region being tiled. Note that both $\mathcal{M}_{ssu}$ and $\mathcal{M}_{loz}$ have the uniform distribution as the stationary distribution. Wilson [43] improves the analysis given in [23] to show that the Markov chain based on tower moves mixes in time $\mathrm{O}(W^2 N \log N \log(2/\epsilon))$, where W is the width and N is the area of the triangular lattice region to be tiled. However, neither approach gives the mixing time of $\mathcal{M}_{ssu}$.

3.1 Comparing Markov chains for tilings

We now define a set of paths and then bound A corresponding to these paths. For each (x, y) which differ by a tower move of height h_{xy} (and are thus adjacent in the known chain), there is a unique minimum length sequence of single site update moves of length h_{xy} which transforms x into y. Such a sequence defines a path γ_{xy}, in a natural way, using transitions of $P(\cdot, \cdot)$. Note that the length of the path is h_{xy} and $\tilde{P}(x, y) = 1/2Nh_{xy}$. Consider an arbitrary (z, w) where z and w are lozenge tilings which differ by a single cube. Note that $P(z, w) = 1/2N$. Furthermore, for a given (z, w), the number of (x, y) such that the path γ_{xy} uses (z, w) is at most H^2, where H is the maximum height of a tower. (This is because the bottom and the top of any tower containing a particular single site can be chosen in at most H ways.) Thus yielding the following bound for the quantity A from the comparison theorem.

$$\begin{aligned} A &= \max_{(z,w)\in E(P)} \left\{ \frac{1}{\pi(z)P(z,w)} \sum_{\Gamma(z,w)} |\gamma_{xy}|\pi(x)\tilde{P}(x,y) \right\} \\ &= \max_{(z,w)\in E(P)} \left\{ \frac{2N}{\pi(z)} \sum_{\Gamma(z,w)} h_{xy}\pi(x)(1/2Nh_{xy}) \right\} \\ &\leq \sum_{\Gamma(z,w)} 1 \leq H^2. \end{aligned}$$

Theorem 5 *Let R be a region in the triangular lattice whose convex hull has area N. Then the mixing time of $\mathcal{M}_{ssu}$ for generating a lozenge tiling of R is given by*

$$\tau_{ssu} = \mathrm{O}\left(N^4 \log N + N^3 \log N \log(1/\epsilon)\right).$$

Proof. Clearly the number of lozenge tilings of a region of size N is at most 3^N (since we can overcount by replacing each triangle in the underlying region with a triangle with one side identified – those configurations where identified edges line up are the set of valid tilings). The bound on the mixing time of $\mathcal{M}_{loz}$ given by Wilson [43] is $\tau(\epsilon) = \mathrm{O}(W^2 N \log N \log(1/\epsilon))$, where W is the width of the region. Therefore, by proposition 4,

$$\begin{aligned} \tau_{ssu} &\leq \frac{\log(1/(\epsilon\pi_*))}{\log(2/\epsilon)}(H^2)(W^2 N \log N \log(1/\epsilon)) \\ &= \mathrm{O}\left(N^4 \log N + N^3 \log N \log(1/\epsilon)\right), \end{aligned}$$

since $H * W = \mathrm{O}(N)$. □

3.2 Domino tilings and 3-colorings

The single site algorithm for domino tilings and 3-colorings (on regions with fixed boundary conditions) follows exactly the same analysis. Starting from any domino tiling, e.g., choose a 2×2 window; if there are two parallel dominoes, rotate them by 90^o. This simple

algorithm is motivated by the linear time tiling algorithm of Thurston [39] for generating a single tiling. In [23] a tower algorithm is presented which achieves a mixing time of $O(N^{3.5}(1 + \log(1/\epsilon)))$, where N is the area of the Cartesian lattice region to be tiled. We can show that $A \leq H^2$, where H is the size of the maximal tower; for square regions this is $O(N^{1/2})$ and in general is at most N. In addition, the number of domino tilings is trivially bounded by 4^N. Thus, the comparison theorem establishes the efficiency of this local algorithm. The analysis for 3-colorings (also referred to as Eulerian orientations or the ice model) is completely analogous.

4 Triangulations of convex polygons

The set of triangulations of a convex n-gon is a well-known characterization of the Catalan numbers $\{c_n\}$. Two other common representations of c_n which will be useful in this discussion are *Dyck paths* and *binary trees* (see [38] for a general survey):

- A Dyck path from $(0, 0)$ to $(2n, 0)$ is a lattice path with steps $(1, 1)$ and $(1, -1)$ never falling below the x-axis.
- A binary tree of size n is a rooted tree with n *internal nodes* (those with two descendants) and $n + 1$ *external nodes* or *leaves* (those with no descendants).

Each of these representations offers its own Glauber dynamics, which we refer to as the "interchange graphs." The interchange graphs for the three representations are defined as follows (see Fig. 3):

- **Triangulations:** The set of triangulations forms the vertex set of the interchange graph, and two triangulations are adjacent if one can be obtained from the other by a diagonal flip, as described in [37]. Every diagonal in a triangulation of a convex polygon defines a quadrilateral. A diagonal flip replaces that diagonal with the other diagonal of the same quadrilateral. Sleator et al. [37] show this move connects the state space and they obtained tight upper and lower bounds (of $2n-6$) on the diameter of this interchange graph and other results on triangulations of the sphere (see [20] for a simpler proof).
- **Binary trees:** Two binary trees with n internal nodes are adjacent if one can be transformed into the other by applying the rotation operation. A rotation at a node is defined as shown in Fig. 5. Sleator et al. also showed that this graph is isomorphic to the previous one on triangulations of a convex $(n + 2)$-gon.
- **Dyck paths:** Similarly, in the collection of Dyck paths of length $2n$, two elements are adjacent if one may be changed into the other by flipping a peak into a valley (that is, changing $(1, 1), (1, -1)$ to $(1, -1), (1, 1)$) or a valley into a peak (that is, changing $(1, -1), (1, 1)$ to $(1, 1), (1, -1)$). It is easy to see that the diameter of this graph is precisely $n(n - 1)/2$.

The interchange graph on triangulations is of particular interest because it suggests a Markov chain on a general planar point set, a problem of central interest in computer graphics and computational geometry. We will review the use of the comparison technique

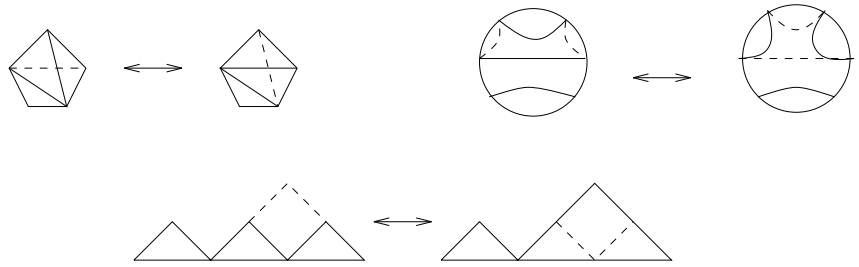

Figure 3: Local moves defining interchange graphs

analyzed for Markov chain in the convex case by McShine and Tetali [29] which shows that $O(n^5 \log(n/\epsilon))$ are sufficient to get close (within ϵ in variation distance) from the stationary distribution, which in this case is uniform over triangulations. Previously Molloy et. al. [31] showed that $O(n^{23} \log(n/\epsilon))$ steps are sufficient using a conductance argument. Their lower bound determining $\omega(n^{3/2})$ suggests that there is still room for improving the upper bound on the convergence time.

Naturally there are known bijections between these representations for the Catalan numbers. It was shown in [37] that the interchange graphs for triangulations and binary trees are isomorphic. Thus, it follows that it is sufficient to bound the mixing rate of a Markov chain based on the interchange graph on binary trees (to establish the corresponding bound on a Markov chain on triangulations). This is our first step.

The bijection between binary trees and Dyck paths is more interesting, and it is the relation between their interchange graphs which underlies our application of the comparison theorem of Diaconis and Saloff-Coste. That is, we use the bound on the mixing time of the chain on Dyck paths as established by Wilson [43] and then use this to bound the mixing time of a chain on binary trees.

4.1 Comparing chains based on interchange graphs

Let $RG(n)$ denote the interchange graph on binary trees, where two trees are connected if they differ by a single rotation. The transition probabilities of the Markov chain on $RG(n)$ are defined as follows. For two distinct binary trees z and w,

$$\begin{aligned} P(z,w) &= 1/[2(n-1)], \quad \text{if } (z,w) \in E(RG(n)) \\ P(z,z) &= 1/2. \end{aligned}$$

The transition probabilities of the Markov chain on $DG(n)$, the interchange graph for Dyck paths of length $2n$, are defined as follows. For two distinct Dyck paths x and y,

$$\begin{aligned} \tilde{P}(x,y) &= 1/[2(2n-3)] \quad \text{if } (x,y) \in E(DG(n)) \\ \tilde{P}(x,x) &= 1 - \textstyle\sum_{y \sim x} P(x,y) \geq 1/2 \end{aligned}$$

It is easy to verify that the two Markov chains as defined above do indeed satisfy the reversibility (i.e. detailed-balance) condition and also that they share the uniform distribution as the stationary distribution π. So for $x \in \tilde{\Omega}$ and $z \in \Omega$, we have $\tilde{\pi}(x) = \pi(z) = 1/c_n$.

Let $\tau(\epsilon)$ and $\tilde{\tau}(\epsilon)$ denote the mixing times of the Markov chains on $RG(n)$ and $DG(n)$ respectively. First we may deduce from [43] that the Markov chain on $DG(n)$ has mixing time

$$\tilde{\tau}(\epsilon) = O(n^3(\log n + \log 1/\epsilon)) \ , \ 0 < \epsilon < 1.$$

Our known chain, $\tilde{P}$, is the chain with the set of Dyck paths of length $2n$ as the state space, denoted $\tilde{\Omega}$; our unknown chain, P, is the chain with the set of binary trees with n internal nodes as the state space, denoted Ω. We would like to use Proposition 4 to get a bound on the mixing time $\tau(\epsilon)$. First note that Proposition 4 (and 2) require the state spaces $\tilde{\Omega}$ and Ω of $\tilde{P}$ and P, respectively, to be the same. In the present case, although they are not the same, the propositions are still applicable in view of the fact that we are able to define a bijection $f : \tilde{\Omega} \to \Omega$ between them.

The bijection between binary trees and Dyck paths is easiest to describe through a bijection from each to another Catalan structure – the set of binary strings of length $2n$ with equal number of 1's and 0's, wherein the number of 1's in each string is always greater than or equal to the number of 0's as we count from left to right in the string (see Fig. 4). Given a Dyck path of length $2n$, a segment of slope $+1$ corresponds to a 1 and a segment of slope -1 corresponds to a 0. Given a binary string on n internal nodes, label the *left edges* (edges leading to left descendants) with a 1 and the *right edges* with a 0. Now the corresponding binary string is the one obtained by reading the labels as the tree is traversed, recursively, starting from the root, first visiting the left subtree and then the right subtree. We leave it to the reader to verify that this is indeed a bijection.

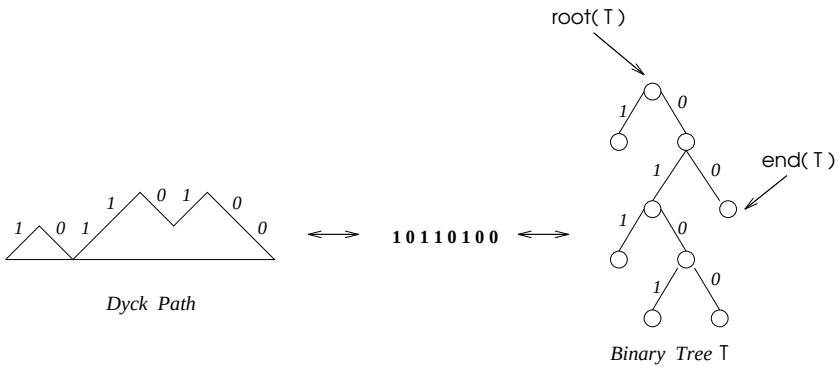

Figure 4: Bijection via binary strings

The canonical path γ_{xy} of the comparison technique, is now a path in P, which can be described as a sequence of states, $f(x) = z_0, z_1, \ldots, z_{k-1}, z_k = f(y)$, for $(x, y) \in E(DG(n))$. The description of the paths will be simplified by using the above binary string representation of Dyck paths from now on.

Following [37], we will state some definitions. A *subtree* of a binary tree is either a single node or a binary tree with at least one internal node. Subtrees will be denoted by T_i, and this will stand both for the subtree and the binary string representation of T_i. The *depth* of a node in a binary tree is the length of the shortest path from the root to that node. It is convenient to view these binary trees as binary search trees with labels on the nodes, namely, with the property that the label of a node is bigger than the labels of all the

nodes in its left subtree, and smaller than the labels of the nodes in its right subtree. This gives a natural ordering on the nodes of a tree. The rotation operation mentioned earlier preserves this node ordering, not surprisingly, since the rotations were invented as a way of restructuring binary search trees. Specifically, the rotation operation is defined as shown in Fig. 5. Subtrees T_i consist of single nodes or larger binary trees. After a rotation at Y, the parent of Y becomes the parent of X; if Y is the root of the tree then X becomes the new root after the rotation.

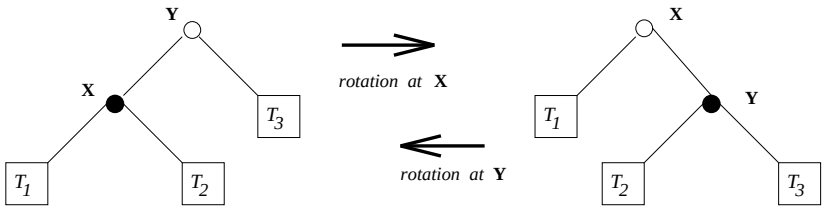

Figure 5: Local moves on binary trees

Rotations do not change the number of internal nodes, and for a tree with n internal nodes, there are $n-1$ rotations possible at any time, one for each internal node, except for the root. The rotation at a node brings the node one step closer to the root, thus decreasing the depth of that node by exactly one.

Referring to Fig. 5 above, a rotation at X (called an X-rotation) decreases the depth of nodes in subtree T_1 by 1, increases the depth of nodes in subtree T_3 by 1, and leaves the depth of all other nodes the same. In the same way, a rotation at Y (called a Y-rotation) decreases the depth of nodes in subtree T_3 by 1, increases the depth of nodes in subtree T_1 by 1, and leaves the depth of all other nodes the same.

4.2 Canonical paths

We need to define $\Gamma = \{\gamma_{xy}\}$, for $(x, y) \in E(\tilde{P})$. This can be done in a natural way, once we analyze a transition from x to y in $DG(n)$, according to whether it is a $01 \to 10$ flip or a $10 \to 01$ flip, and interpret the flip in terms of the corresponding binary trees, $f(x)$ and $f(y)$.

The easy case is if $f(x)$ and $f(y)$ differ by a single rotation – if $f(x) \sim f(y)$ in P, then γ_{xy} is simply the edge $(f(x), f(y)) \in E(P)$. But in general, $(x, y) \in E(\tilde{P})$ does not imply $(f(x), f(y)) \in E(P)$. Fig. 6 characterizes the differences in two binary trees, which have adjacent representations as Dyck paths. In such a case, we will define a unique sequence of rotations which transforms $f(x)$ into $f(y)$, and the corresponding sequence of edges in P which forms the corresponding canonical path γ_{xy}. We shall do this for the case when $x \to y$ is a $01 \to 10$ flip, and in the other case the path is just the reverse of the path in this case; we are justified in doing this since the interchange graphs can be viewed as *undirected* graphs.

To simplify the discussion, we will now introduce some new terms which are illustrated in Fig. 3. The *root* of a subtree T_i, *root*(T_i), is the top vertex of that subtree. The *end* of a subtree T_i, *end*(T_i), is the rightmost vertex of T_i.

The following observation is key to understanding Γ and to bounding $A(\Gamma)$. A $01 \to 10$ flip moves a particular left subtree (T_3 in Fig. 6) hanging from the right child of some

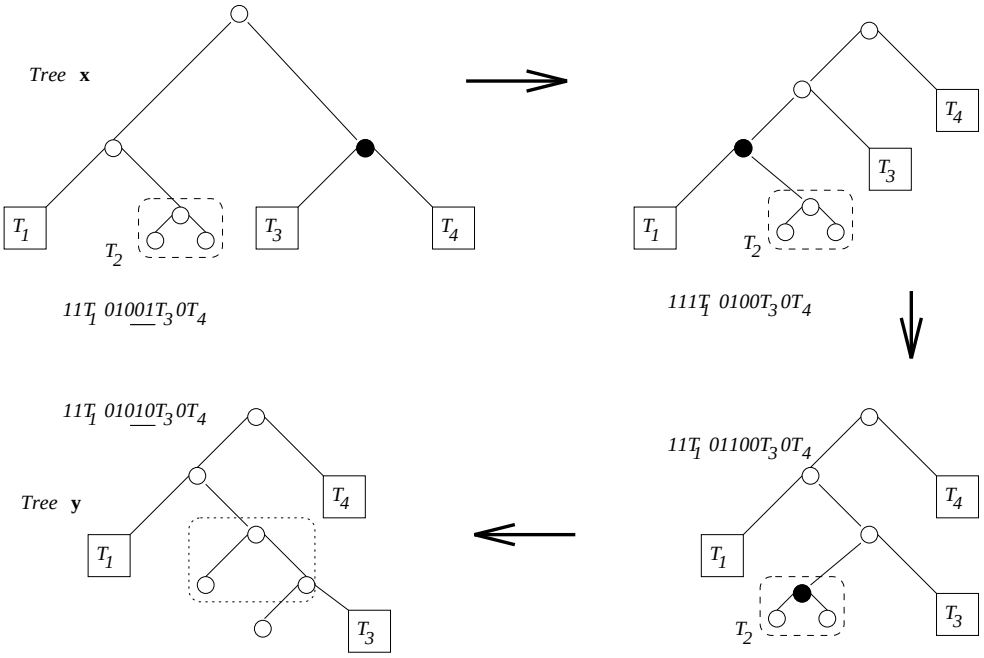

Figure 6: Canonical path from **x** to **y**, differing in a 01–10 flip

node N to being the rightmost subtree of the left child of N. This is easy to verify by considering the binary strings corresponding to x and y, which differ in a $01 \to 10$ flip, and then by constructing $f(x)$ and $f(y)$. We call such a subtree the *characteristic subtree* of that particular $01 \to 10$ flip, since it uniquely identifies the edge $(x, y) \in E(\tilde{P})$. In Fig. 6, T_3 is the characteristic subtree of the flip which takes x to y. Note that such a subtree can also be a single node.

The canonical path γ_{xy} is the unique sequence of rotations which transforms the binary tree $f(x)$ into $f(y)$ – the first rotation is performed at the parent of the root of the characteristic subtree, and then every subsequent rotation is at the sibling of the root of the characteristic subtree. (In Fig. 6 the black nodes denote the nodes at which rotations are performed; also for convenience, the trees are called **x** and **y**, rather than $f(x)$ and $f(y)$.)

Claim. With Γ being the set of canonical paths as described above, the comparison factor $A(\Gamma) = O(n)$.

Proof. We will show that $A(\Gamma)$ is at most $O(n)$, by arguing that the length of a canonical path is at most n, and that each rotation in a binary tree is used by at most one canonical path γ_{xy}, where $x \to y$ is a $01 \to 10$ flip, and by at most one canonical path corresponding to a $10 \to 01$ flip.

First notice that in a path corresponding to a $01 \to 10$ flip, the depth of the root of the characteristic subtree remains the same after the first rotation, but increases by exactly one with every subsequent rotation. At most $n-2$ internal nodes can participate (by being siblings of the root of the characteristic subtree) in increasing the depth – the grandparent of the characteristic subtree and any nodes in the right subtree of the right child of the grandparent are the nodes where a rotation is not performed in such a canonical path.

This shows that the length of a canonical path can be at most $n-1$. (The argument for a $10 \to 01$ flip is analogous.)

Secondly, consider an arbitrary rotation $(z,w) \in E(P)$. Whether it is an X-rotation or a Y-rotation, there are always at most two choices for a subtree to play the role of a characteristic subtree. Referring to Fig. 5, if (z,w) is an X-rotation, then either T_3 is the characteristic subtree of a $01 \to 10$ flip or T_2 is the characteristic subtree of a $10 \to 01$ flip. The rotation (z,w) and the choice of either T_2 or T_3 as the characteristic subtree, uniquely identifies the pair (x,y) such that $(z,w) \in \gamma_{xy}$. Referring once again to Fig. 5, if (z,w) is an X-rotation, then T_3 will eventually end up as the right subtree of $end(T_2)$, giving us y. Knowing y and the fact that T_3 is the characteristic subtree uniquely determines x. If on the other hand, T_2 were to be the characteristic subtree, then w is in fact y, and x can be uniquely determined given that $x \to y$ is now a $10 \to 01$ flip.

Thus, for a fixed (z,w), $|\Gamma(z,w)| \leq 2$, and when $(z,w) \in \gamma_{xy}$, $|\gamma_{xy}| \leq n-1$.
We have $\pi(x) = \pi(z) = (n+1)/\binom{2n}{n}$. Also, $P(z,w) = 1/(2n-2)$, for all $(z,w) \in E(P)$, and $\tilde{P}(x,y) = 1/(4n-6)$, for all $(x,y) \in E(\tilde{P})$. Thus,

$$\begin{aligned} A(\Gamma) &= \max_{(z,w)\in E(P)} \left\{ \frac{1}{\pi(z)P(z,w)} \sum_{\Gamma(z,w)} |\gamma_{xy}| \pi(x) \tilde{P}(x,y) \right\} \\ &\leq \frac{n-1}{2n-3}[2(n-1)] = O(n). \end{aligned}$$

□

Applying Proposition 4, with the known bound on the mixing time of $\tilde{P}$, we can now bound the mixing time of P, the Markov chain on binary trees:

$$\begin{aligned} \tau(\epsilon) &\leq \frac{4\log\left(\frac{1}{\epsilon\pi_*}\right)}{\log\left(\frac{1}{2\epsilon}\right)} n(n^3(\log n + \log 1/\epsilon)) \\ &= \frac{4\left(\log(1/\epsilon) + \log(c_n)\right) n^4(\log n + \log 1/\epsilon)}{\log(1/2) + \log(1/\epsilon)} \\ &= O\left(\frac{n^5 \log n}{\log(1/\epsilon)} + n^4 \log n + n^5 + n^5 \log n\right) \\ &= O(n^5((\log n + \log 1/\epsilon)), \end{aligned}$$

thus establishing the following theorem. (See comments following Corollary 10 for a possible way to improve this bound.)

Theorem 6 *The mixing time of the Markov chain on triangulations of a convex $(n+2)$-gon (equivalently, on binary trees with n internal nodes) satisfies, for $0 < \epsilon < 1$,*

$$\tau(\epsilon) = O\left(n^5 \log(n/\epsilon)\right).$$

5 Other comparison techniques

• **Holley's result**. In his well-known paper [11] on possible rates of convergence in finite range attractive spin systems, Holley proved the following comparison type result as well. Let S denote the finite set of spins and let $\Omega = S^{\mathcal{Z}^d}$ denote the state space corresponding to S-valued configurations on $\mathcal{Z}^d$. Holley considers finite range, translation invariant, attractive jump rates on the infinite volume, with the jump rates denoted by $c(k, \eta, s)$ for site k with the spin η at time t to switch to spin s. Then Theorem 0.2 of [11] states the following.

Theorem 7 *Let $\{c(k, \cdot, \cdot) : k \in \mathcal{Z}^d\}$ be the finite range, translation invariant, attractive jump rates for an infinite system of interacting processes and suppose that the corresponding semi-group converges exponentially fast to its equilibrium. Then there is a $\delta > 0$ such that if $\{\bar{c}(k, \cdot, \cdot) : k \in \mathcal{Z}^d\}$ is a similar collection of jump rates and*

$$\sup_{\eta \in \Omega} \sum_{s \in S} |c(0, \eta, s) - \bar{c}(0, \eta, s)| < \delta,$$

then the semi-group generated by the jump rates $\bar{c}$ also converges to equilibrium exponentially fast.

Note that due to the translation invariance, it suffices to consider the jump rates at a fixed site such as 0. Holley uses this theorem in conjunction with the main theorem in his paper to conclude that for the so-called contact process (see [10]) *at* the critical value of the parameter, either the process is not ergodic or it converges to the state $\mathbf{0}$ at a rate which is at most t^{-d}. (When dimension $d = 1$, this result was known due to Griffeath [10].)

Loosely paraphrasing Holley's result in our present context of discrete-time finite state Markov chains, would be saying: if the dynamics generated by P is rapidly mixing, then a related dynamics $\tilde{P}$ is also rapidly mixing, provided

$$\max_x \sum_y |P(x, y) - \tilde{P}(x, y)| \leq \frac{\delta}{n},$$

for some fixed $\delta > 0$, where n is the size of the underlying input parameter (e.g. the number of sites in the system or the number of vertices in a graph), introduced by the sequentiality of the Glauber dynamics considered in this paper. It can be checked that the above, although being a useful result in Holley's considerations, demands in its hypothesis too stringent a condition to be satisfied by the applications mentioned in the preceding sections.

• **Diaconis–Saloff-Coste comparison using flows and Vigoda's result**. Going back to the framework of Theorem 3, in applications sometimes one finds that there is more than one path $x = x_0, x_1, \ldots, x_k = y$ with $P(x_i, x_{i+1}) > 0$ (i.e. $(x_i, x_{i+1}) \in E(P)$) between pairs x, y such that $\tilde{P}(x, y) > 0$ (i.e. $(x, y) \in E(\tilde{P})$). In such situations one can reduce the comparison factor $A(\Gamma)$ by "distributing the weight" (or routing the flow) over *a set of paths*, rather than accounting for each nonlocal move $\tilde{P}(x, y)$ via a single (canonical) path. This is captured precisely in the following theorem of Diaconis and Saloff-Coste [4]. The following version is simpler, due to our added assumption of identical stationary

distributions. Let Γ_{xy} denote the set of all simple paths connecting x and y as before and set $\Gamma = \cup_{(x,y)\in E(\tilde{P})}\Gamma_{xy}$. For $\gamma \in \Gamma$, let $|\gamma|$ denote its length. A function $f : \Gamma \to \mathbb{R}^+$ is called a $(P, \tilde{P})$ flow if for every $(x, y) \in E(\tilde{P})$,

$$\sum_{\gamma\in\Gamma_{xy}} f(\gamma) = 1.$$

Then Theorem 2.3 of [4] yields the following.

Theorem 8 *With the above notation, we have*

$$(1 - \lambda_1(P)) \geq \frac{1}{A(f,\Gamma)}(1 - \lambda_1(\tilde{P})),$$

where

$$A(f,\Gamma) = \max_{(z,w)\in E(P)} \left\{ \frac{1}{\pi(z)P(z,w)} \sum_{\gamma\in\Gamma_{xy}:(z,w)\in\gamma} |\gamma| f(\gamma)\pi(x)\tilde{P}(x,y) \right\}.$$

It is easy to see that Theorem 3 is a special case of the above theorem – simply route all the flow through the chosen (canonical) path γ_{xy} for each pair $(x, y) \in E(\tilde{P})$. In [4] this theorem was used in analyzing the exclusion process. Of course, Proposition 4 also holds with A replaced by $A(f, \Gamma)$.

Very recently, Vigoda found a nice application of this theorem in showing that the Glauber dynamics for sampling k-colorings of an n-vertex graph G of maximum degree $\Delta(G)$ is at most $O((k \log k)n^2 \log n)$ as long as $k > (11/6)\Delta(G)$. (Vigoda also shows that the dynamics is rapidly mixing when $k = (11/6)\Delta$, but the bound on the mixing time is worse.) In terms of rapid mixing, the previously best known bound on the number of colors was $k \geq 2\Delta(G)$, as shown by Jerrum [13].) Recall that the Glauber dynamics for sampling colorings picks a vertex and a color uniformly at random and attempts to color the vertex with the chosen color (i.e. if none of the neighbors of the vertex have the chosen color). Direct analysis of Glauber for this problem for k significantly less than 2Δ (and still above, say $\Delta+2$) seems an extremely difficult task, and Vigoda succeeded in obtaining the modest improvement by using the so-called Kempe-chain moves: the transitions consist of flipping two-colored clusters in the following sense. Given the current coloring σ, pick a vertex v and a color $\sigma(v)$ uniformly at random; consider the maximal *cluster* of vertices which contain v and have color c or $\sigma(v)$. (By a cluster, we mean a connected induced subgraph.) With a carefully chosen probability, flip this cluster by interchanging colors c and $\sigma(v)$ on it. In his algorithm only clusters of size 6 or less are ever flipped. For this chain, Vigoda shows a mixing time of $O(kn \log n)$ whenever $k > (11/6)\Delta$.

To bound the mixing time of Glauber dynamics, Vigoda uses the availability of extra colors (since $k > \Delta$) to advantage and defines multiple paths, and manages to show that $A(f, \Gamma) = O(1)$, where the flows f are simply chosen to be uniformly distributed over the available paths. The details can be found in [40]. Do note that flows help here, since a straightforward application of Theorem 3 only gives $A(\Gamma) = O(k\Delta^2)$ or so. However, the curse of $\log(1/\pi_*)$ can not be avoided, which in this application is $O(n \log k)$ (since there can be $\Omega(k^n)$ proper colorings of G), thus giving the aforementioned bound of $O((k \log k)n^2 \log n)$, via Proposition 4.

• **Comparison using the log-Sobolev constant.** It turns out that the extra factor $\log(1/\pi_*)$ can be avoided if one manages to bound the $L_2(\pi)$ distance (of the known chain) from stationarity rather than the total variation distance. In short this is because the log-Sobolev constant captures the time to make the L_2 distance small in a sharp way – up to a factor of $\log\log(1/\pi_*)$.

First recall a few definitions. We let (Ω, P, π) denote the standard triple from Section 2. The logarithmic Sobolev constant $\rho = \rho(P) > 0$ is the optimal constant in the functional inequality: for all $f : \Omega \to \mathbb{R}$,

$$\rho \mathrm{Ent}_\pi \mathrm{f}^2 \leq \frac{1}{2} \sum_{\mathrm{x,y}} (\mathrm{f(x)} - \mathrm{f(y)})^2 \, \mathrm{P(x,y)} \pi(\mathrm{x})$$

where $\mathrm{Ent}_\pi \mathrm{f} := \mathrm{E}_\pi \mathrm{f} \log \mathrm{f} - \mathrm{E}_\pi \mathrm{f} (\log \mathrm{E}_\pi \mathrm{f})$. (In some of the literature ρ^{-1} is referred to as the log-Sobolev constant instead.) Also recall that the spectral gap $\lambda(P) := 1 - \lambda_{\max}$ has the variational (functional) characterization as the optimal constant $\lambda > 0$ in

$$\lambda \mathrm{Var}_\pi \mathrm{f} \leq \frac{1}{2} \sum_{\mathrm{x,y}} (\mathrm{f(x)} - \mathrm{f(y)})^2 \, \mathrm{P(x,y)} \pi(\mathrm{x}),$$

where $\mathrm{Var}_\pi \mathrm{f}$ denotes the variance $E_\pi f^2 - (E_\pi f)^2$. Given that the comparison theorems of Diaconis–Saloff-Coste actually compare the Dirichlet forms of P and $\tilde{P}$ (i.e. the right hand sides of the above definitions), it should come as no surprise that the analogs of Theorems 3 and 8 hold with the spectral gaps replaced by the log-Sobolev constants of P and $\tilde{P}$.

The other key ingredient in the comparison method we have used was Theorem 1, and we need an analog in terms of ρ rather than λ. Towards this, following [6], it is also useful to define the L_p-time, for $1 \leq p \leq \infty$, which captures the approach to stationarity using the $L_p(\pi)$ norm. For $1 \leq p < \infty$, let

$$\left\| \frac{\mu}{\pi} - 1 \right\|_{p,\pi} := \left(\sum_x \left| \frac{\mu(x)}{\pi(x)} - 1 \right|^p \pi(x) \right)^{1/p}$$

Also, let

$$\left\| \frac{\mu}{\pi} - 1 \right\|_{\infty,\pi} := \left(\max_x \left| \frac{\mu(x)}{\pi(x)} - 1 \right| \right)$$

For $\epsilon > 0$, and $x \in \Omega$, let

$$T_p(\epsilon) = \max_x \min \left\{ \hat{t} > 0 : \left\| \frac{P^t(x,\cdot)}{\pi(\cdot)} - 1 \right\|_{p,\pi} \leq \epsilon, \ \ \forall \, t \geq \hat{t} \right\}$$

Note that the total variation norm is one-half of the $L_1(\pi)$ norm, and so $\tau(\epsilon) = T_1(2\epsilon)$. For convenience, we let $\epsilon = 1/e$, and let $T_p(1/e) = T_p$. Then the analog of Theorem 1 asserts the following (see [6, 33] for a proof in continuous time, and [30] for a discrete-time version).

Theorem 9 *Let (Ω, P, π) be as above. Then for $2 \leq p \leq \infty$,*

$$\frac{1}{2\rho} \leq T_p \leq \frac{1}{2\rho} \left(c + \log\log(1/\pi_*) \right),$$

where $c > 0$ is a small absolute constant.

Remark 3. For $p = 1$, a case of special interest and relevance, the lower bound in Theorem 9 is unfortunately not known to be true, even with a worse absolute constant. The upper bound is indeed true, with $p = 1$ and has an elementary proof, but is not directly useful for comparison of total variation mixing times.

It should be easy to see that an analog of Proposition 4 is possible relating the L_p-times of P and $\tilde{P}$ which in addition to the comparison factor of $A(\Gamma)$ would only have a $\log\log(1/\pi_*)$ factor. Let T_p and $\tilde{T}_p$ denote the L_p-times of P and $\tilde{P}$, which are both reversible Markov chains. Then the analog of Theorem 3 for the log-Sobolev constant and Theorem 9 together yield,

Corollary 10 *Given* (Ω, P, π), $(\Omega, \tilde{P}, \pi)$, *and* $A(\Gamma)$ *as defined in Theorem 3, for* $2 \leq p \leq \infty$, *we have*

$$T_p \leq (c + \log\log(1/\pi_*))\, A(\Gamma)\tilde{T}_p,$$

where $c > 0$ *is a small absolute constant.*

The obvious difficulty in making direct use of the above theorem for comparison purposes is that one first needs a "known chain" $\tilde{P}$ for which either $\rho(\tilde{P})$ or T_p, for $p \geq 2$, should be known. We briefly mention two examples where the log-Sobolev comparison theorem has yielded improved bounds on mixing rates. Using Fourier techniques, Diaconis and Shashahani [7] computed all the eigenvalues of the Markov chain arising from random transposition shuffle. In [5] such precise information of eigenvalues was used in the comparison of L_2-times of random walks on groups. In particular, T_2 of the (random) adjacent transposition shuffle was shown to be $O(n^3 \log n)$. It is very likely that one can use this information together with the connection between permutations and *threshold functions* (see [43]) to show that T_2 of Dyck paths is $O(n^3 \log n)$. In turn, this would yield for the T_2 (and hence T_1) of the chain on triangulations an improved bound of $O(n^4 \log^2 n)$, via Corollary 10.

• **Dyer-Greenhill result**. Consider as in the previous sections, two (reversible) Markov chains with the same state space Ω and the same stationary distribution π. Then under the further assumption of a certain *linear* relationship between the entries of the two transition probability matrices (P and $\tilde{P}$) of the chains, a corresponding linear relationship between the spectral gaps of the two matrices is derived in [9]. Once this is established, using Proposition 4 or a similar one (as done in [9]), the mixing rates of the two Markov chains can also be related. The precise statement of the Dyer-Greenhill result is as follows. Let $\pi^* = \max_{x \in \Omega} \sqrt{(1 - \pi(x))/\pi(x)}$.

Theorem 11 *Suppose that there exists an* α *such that* $0 < \alpha < 1$ *and* $P \geq \alpha(\tilde{P})$, *then*
(i) $(1 - \lambda_1(P)) \geq \alpha(1 - \lambda_1(\tilde{P}))$, *and*
(ii) $\tau(\epsilon) \leq 2\log(\pi^*/2\epsilon)\alpha^{-1}\tilde{\tau}(1/e)$.

Notice that we can achieve a very similar result by taking trivial canonical paths consisting of single moves in the known chain $\tilde{P}$ and applying Proposition 4. However, the proof in [9], which we omit here, offers a different and worthwhile perspective on the comparison technique. The basic idea behind (i) is to use the min-max characterization of eigenvalues in showing that the second largest eigenvalue of a *convex combination* of two symmetric

matrices is bounded from above by the convex combination of the second largest eigenvalues of the two matrices.

The main application of this technique in [9] was to show that the Glauber dynamics for the hardcore lattice gas model with activity λ on *arbitrary* graphs on n vertices with maximum degree Δ mixes rapidly in $O(n^2 \log n)$ time as long as $\lambda < 2/(\Delta - 2)$. In reporting these bounds, we suppress the dependence on ϵ, Δ and λ. Recall that in the hardcore model a valid configuration prohibits two particles to reside on neighboring vertices, and each such (valid) configuration I has the Gibbs measure proportional to $\lambda^{|I|}$, where $|I|$ denotes the number of particles in I.

To achieve this result, they first introduce a slightly different dynamics $\mathcal{M}_2$ which, besides the Glauber moves, also allows "sliding" the particle at an occupied site to a neighboring unoccupied site, whenever the resulting configuration is a valid hardcore configuration. Using the path coupling technique, they derive bounds on the mixing time of this new dynamics, and then use their comparison technique to bound the Glauber dynamics. An interesting subtlety here is that the Glauber dynamics does not quite dominate the new dynamics in the sense of the linear relationship specified by Theorem 11; however, they replace Glauber $\mathcal{M}$ (yet once more) by an equally fast dynamics $\mathcal{M}'$ which runs two moves of Glauber at each step and also has a holding probability of one-half; $\mathcal{M}'$ has the same mixing time as $\mathcal{M}$, but crucially, $\mathcal{M}'$ does dominate $\mathcal{M}$, allowing them to use Theorem 11.

More recently, Luby and Vigoda [25] have shown that the Glauber dynamics for the hardcore model mixes in $O(n^2 \log n)$ time as long as $\lambda < 2/(\Delta - 2)$, and $O(n^2 \log n)$ time when $\lambda = 2/(\Delta - 2)$. Luby and Vigoda achieve this improvement through a *direct* analysis of the Glauber dynamics using a more subtle coupling argument.

• **Other related methods**. We merely mention here some other results related to the theme of this paper. The first two results are in the framework of relating certain block dynamics to Glauber dynamics, both with the same Gibbs distribution as the invariant distribution on some Ising-type configurations. Towards this consider $D = \{V_1, V_2, \ldots, V_m\}$, an arbitrary collection of finite sets $V_i \in \mathcal{F}$, and let $V = \cup_i V_i$. Then by *block dynamics* with blocks $\{V_1, V_2, \ldots, V_m\}$ one typically means the continuous-time Markov chain in which each block waits an exponential time of mean one and then the configuration inside the block is replaced by a new configuration distributed according to the Gibbs measure of the block, given the previous configuration outside the block. (Note that the tower moves of Section 3 *do not* fall under this category.)

The first result is mentioned as Proposition 3.4 in Martinelli's lecture notes [28]. This result asserts that the spectral gap of the (single-site) Glauber dynamics is lower bounded by the spectral gap of the smallest of the spectral gaps of the same dynamics restricted to each of the blocks of some block dynamics *times* the spectral gap of the block dynamics itself. Note that if the size of the blocks is chosen to be at most a constant, indepedent of the size of the system (V), then this guarantees that the spectral gap of the Glauber dynamics is at least a certain constant (strictly between 0 and 1) times the spectral gap of the block dynamics. The proof is short and uses once again the variational characterization of the spectral gap. We refer the reader to [28] for the details.

The second one is fairly natural, although it appears in a rather specialized context. It is due to van den Berg and Brouwer [2] in the context of random sampling for the monomer-dimer model on the d-dimensional torus. These authors first use certain block dynamics

to sample matchings (dimers) of the d-dimensional torus of "side length" n (viewed as a graph, this corresponds to the Cartesian product of d copies of an n-cycle). They first show that the block dynamics has mixing time $O(n \log n)$ (fixing the monomer-dimer parameter $\lambda > 0$, dimension d, and ϵ in the definition of mixing time), through the analysis of spatial dependencies of the monomer-dimer model. Now to get a bound on the original Glauber dynamics, they essentially simulate the block dynamics using several steps of the Glauber dynamics – each step of the block dynamics is replaced by a number of "micro steps", where each micro step is a Glauber move inside the block, *fixing* the configuration on the boundary of the block. The number of such micro steps is chosen so that at the end of such a run of micro steps, the configuration inside the block is distributed *approximately* (rather than exactly) according to the correct distribution – the one that is used to run the block dynamics. The authors only pay a penalty of an extra $O(\log n)$ factor in the mixing time in going from the block dynamics to the Glauber dynamics; most of their analysis is carried out via suitable coupling arguments.

Finally, we mention a decomposition technique due to Madras and Randall [26, 27] which considers a decomposition of the state space of a Markov chain into overlapping pieces. The Markov chain is compared to a family of "restricted Markov chains," each representing the original Markov chain restricted to a piece of the state space, as well as a global Markov chain capturing how well the pieces are interconnected. The spectral gap of the original Markov chain can be bounded by a product of minimum of the spectral gaps of the restricted Markov chains times the spectral gap of the global (interconnection) Markov chain. They apply this technique in the context of sampling independent sets according to their Gibbs measure in the special case when the size of the independent set is bounded by $|V|/(2\Delta+1)$, where $|V|$ is the number of sites and Δ is the maximum degree in the graph [27]. Although this result is of a different flavor from the comparison techniques mentioned here, it offers an alternative way to relate block and Glauber dynamics. Namely, we can decompose a chain defining Glauber dynamics into pieces corresponding to block moves by only allowing those Glauber moves restricted to that block. The decomposition theorem implies that if the Glauber moves are mixing within each block (i.e., the restricted Markov chains quickly converge to the uniform distribution within blocks) and the pieces are sufficiently well connected (i.e., the block dynamics are rapidly mixing) then we can conclude that the Glauber dynamics is mixing as well. This is an alternative formulation of Martinelli's result.

6 Conclusions

In this paper we address the issue that although two Markov chains appear quite similar, it is often the case that only one admits a simple analysis using currently available tools. We envision that there are many other applications for comparison techniques.

For instance, the results on independent sets in [32] can be extended to relate the mixing rate of the single site updates to the mixing rate of a dynamics which would update all the sites in a rectangle of fixed size $a \times b$. This so-called "heat bath" algorithm is used experimentally in statistical physics to study the uniqueness of the Gibbs state of the hard-core lattice gas model. Using the method given here, we derive a bound on the mixing rate of this new Markov chain which introduces a factor which depends exponentially on $\min(a, b)$.

Needless to say, direct analysis of several of the Glauber dynamics mentioned in this paper remains elusive, and should lead to tighter bounds.

Acknowledgment. The authors enjoyed and benefitted from discussions on this topic with J. van den Berg, Jim Fill, Lisa McShine, and Eric Vigoda.

References

[1] Aldous, D. Random walks on finite groups and rapidly mixing Markov chains. *Séminaire de Probabilités* XVII, 1981/82, Springer Lecture Notes in Mathematics 986, pp. 243–297.

[2] van den Berg, J. and Brouwer, R. Random Sampling for the monomer-dimer model on a lattice. *In this special issue of the J. Math. Phys.* 1999.

[3] Bubley, R. and Dyer, M. Path coupling: A technique for proving rapid mixing in Markov chains. *38th IEEE Symposium on Foundations of Computer Science* 1997.

[4] Diaconis, P. and Saloff-Coste, L. Comparison theorems for reversible Markov chains. *Ann. Appl. Probability* **3**, 1993, pp. 696-730.

[5] Diaconis, P. and Saloff-Coste, L. Comparison techniques for random walks on finite groups. *Ann. Probability* **21**, 1993, pp. 2131-2156.

[6] Diaconis, P. and Saloff-Coste, L. Logarithmic Sobolev Inequalities for Finite Markov Chains. *Ann. of Appl. Probab.* **6**, 1996, pp. 695-750.

[7] Diaconis, P. and Shashahani, M. Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Geb.* **57** 1981, 159-179.

[8] Diaconis, P. and Stroock, D. Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probability* **1**, 1991, pp. 36-61.

[9] Dyer, M. and Greenhill, C. On Markov chains for independent sets. Preprint (1997).

[10] Griffeath, D. The basic contact processes. *Stochastic Proc. Appl* **11** (1981), 151-185.

[11] Holley, R. Possible rates of convergence in finite range, attractive spin systems. in Particle systems, Random media, and Large deviations. *Contemp. Math.* **41** 1985, pp. 215-234.

[12] Jerrum, M. Mathematical Foundations of the Markov Chain Monte Carlo Method. *Probabilistic Methods for Algorithmic Discrete Mathematics* ed. by Habib et al., Springer-Verlag (1998), pp. 116-165.

[13] Jerrum, M. A very simple algorithm for estimating the number of k-colorings of a low-degree graph. *Random Structures and Algorithms* **7**, 1995, pp. 157-165.

[14] Jerrum, M. R. and Sinclair, A. J. Polynomial-time approximation algorithms for the Ising model. *SIAM Journal on Computing* **22** (1993), pp. 1087-1116..

[15] Jerrum, M.R. and Sinclair, A.J. Approximating the permanent. *SIAM Journal on Computing* **18** (1989), pp. 1149–1178.

[16] Jerrum, M., Valiant, L. and Vazirani, V. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science* **43** (1986), pp. 169–188.

[17] Kannan, R. Markov chains and polynomial time algorithms. *Proc. 35th IEEE Symposium on Foundations of Computer Science,* 1994, pp. 656–671.

[18] Kannan, R., Tetali, and P., Vempala, S. Simple Markov chain algorithms for generating bipartite graphs and tournaments. *Proc. of the 8th ACM-SIAM Symp. on Discrete Algorithms* January 1997.

[19] Kenyon, C., Randall, D. and Sinclair, A. Approximating the number of dimer coverings of a lattice. *Journal of Statistical Physics* **83** (1996), pp. 637–659.

[20] Kenyon, C. and Thurston, W. Rotation distance using flows. Preprint (1992).

[21] Lieb, E.H. Residual entropy of square ice. *Physical Review* **162**, 1967, pp. 162-172.

[22] Lubin, M. and Sokal A.D. Comment on "Antiferromagnetic Potts Model". *Phys. Rev. Lett.* **71**, 1993, pp. 1778.

[23] Luby, M., Randall, D. and Sinclair, A. Markov Chain Algorithms for Planar Lattice Structures. *Proc. 36th IEEE Symposium on Foundations of Computing* (1995), pp. 150-159.

[24] Luby, M., Vigoda, E. Approximately counting up to four. *Proc. 29th ACM Symposium on Theory of Computing* (1997), pp. 150-159.

[25] Luby, M., Vigoda, E. Fast Convergence of the Glauber dynamics for sampling independent sets.. *Random Structures and Algorithms* **15**, 1999, pp. 229-241.

[26] Madras, N. and Randall, D. Factoring Markov chains to bound mixing rates. *Proc. 37th IEEE Symposium on Foundations of Computing* (1996).

[27] Madras, N. and Randall, D. Decomposing Markov chains. Preprint (1999).

[28] Martinelli, F. Lectures on Glauber dynamics for discrete spin models. Preprint (1997).

[29] McShine, L. and Tetali, P. On the mixing time of the triangulation walk and other Catalan structures. *Randomization methods in Algorithm Design* DIMACS-AMS vol. **43** (1998), pp. 147-160.

[30] Miclo, S. *Lecture Notes in Math.* Springer **1655** (1995), pp. 136-167.

[31] Molloy, M., Reed, B. and Steiger, W. On the mixing rate of the triangulation walk. Preprint (1997).

[32] Randall, D. and Tetali, P. Analyzing Glauber dynamics by comparisons of Markov chains. *Proc. 3rd Latin American Theoretical Informatics Symposium,* Springer Lecture Notes in Computer Science Vol. **1380**, 1998, pp. 292-304.

[33] Saloff-Coste, L. Lectures on Finite Markov chains. *Lecture Notes in Math.* **1665** Springer (1997), 301–413.

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

[35] Sinclair, A.J. *Algorithms for random generation & counting : a Markov chain approach.* Birkhäuser, Boston, 1993, pp. 47-48.

[36] Sinclair, A.J. *Improved bounds for mixing rates of Markov chains and multicommodity flow.* Combinatorics, Probability, & Computing. **1** (1992), pp. 351-370.

[37] Sleator, D., Tarjan, R. and Thurston, W. Rotation distance, triangulations, and hyperbolic geometry. *J.AMS.* **1**, 1988.

[38] Stanley, R. *Enumerative Combinatorics, Volume II.* Version of 12 November 1997, to be published by Cambridge University Press.

[39] Thurston, W. Conway's tiling groups. *American Mathematical Monthly* **97**, 1990, pp. 757-773.

[40] Vigoda, E. Improved bounds for sampling colorings. *IEEE Foundations of Computer Science* (to appear 1999).

[41] Welsh, D.J.A. *The computational complexity of some classical problems from statistical physics.* In Disorder in Physical Systems, (G. Grimmett and D. Welsh eds.). Claredon Press, Oxford, 1990, pp. 323-335.

[42] Welsh, D.J.A. *Approximate counting.* In Surveys in Combinatorics, (R.A. Bailey, ed.). Cambridge University Press, London Math Society Lecture Notes 241, 1997, pp. 287-317.

[43] Wilson, D.B. Mixing times of lozenge tiling and card shuffling Markov chains. Preprint (1997).