

# Rapid Mixing of Several Markov Chains for a Hard-Core Model

Ravi Kannan<sup>1</sup>, Michael W. Mahoney<sup>2</sup>, and Ravi Montenegro<sup>3</sup>

<sup>1</sup> Department of Computer Science, Yale University, New Haven, CT 06520  
kannan@cs.yale.edu

<sup>2</sup> Department of Mathematics, Yale University, New Haven, CT 06520  
mahoney@cs.yale.edu

<sup>3</sup> School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332  
monteneg@math.gatech.edu

**Abstract.** The mixing properties of several Markov chains to sample from configurations of a hard-core model have been examined. The model is familiar in the statistical physics of the liquid state and consists of a set of  $n$  nonoverlapping particle balls of radius  $r^*$  in a  $d$ -dimensional hypercube. Starting from an initial configuration, standard Markov chain monte carlo methods may be employed to generate a configuration according to a probability distribution of interest by choosing a trial state and accepting or rejecting the trial state as the next configuration of the Markov chain according to the Metropolis filter. Procedures to generate a trial state include moving a single particle globally within the hypercube, moving a single particle locally, and moving multiple particles at once. We prove that (i) in a  $d$ -dimensional system a single-particle global-move Markov chain is rapidly mixing as long as the density is sufficiently low, (ii) in a one-dimensional system a single-particle local-move Markov chain is rapidly mixing for arbitrary density as long as the local moves are in a sufficiently small neighborhood of the original particle, and (iii) the one-dimensional system can be related to a convex body, thus establishing that certain multiple-particle local-move Markov chains mix rapidly. Difficulties extending this work are also discussed.

## 1 Introduction

A very simple model of particle interactions is the hard core model. Consider  $n$  points in a  $d$ -dimensional unit hypercube distributed uniformly subject to the condition that the distance  $r$  between any two particles is greater than some critical distance  $2r^*$ . Sampling uniformly from the set of such configurations is of interest in statistical physics and in two and three dimensions this hard core model is used as a simple model of liquid state systems.[1,2,20] The Markov chain monte carlo (MCMC) algorithmic procedure is well suited to perform such sampling.[9,10,18] Indeed, MCMC with the Metropolis filter was first applied to the hard core model in two dimensions.[16] Recent work examining the convergence rates of Markov chains has typically examined systems of computer scientific interest or simple models in statistical physics of solid state systems,

e.g., the Ising model.[6,10,11] In this work, we examine the convergence properties of several Markov chains for sampling from the set of configurations of the  $n$ -particle hard-core model of molecular liquids in  $d$  dimensions.

Starting from an initial configuration of particles, the MCMC procedure performs the sampling by constructing an irreducible aperiodic Markov chain that converges to the stationary distribution of interest.[9,10,16,18] Each step of the Markov chain consists of a trial move (for which there is much discretion in constructing) of one or more particles followed by an acceptance or rejection (according to the Metropolis rule) of the trial configuration as the next configuration of the Markov chain. One example of a trial move is a single-particle global move in which a single particle is selected and moved to a position that is chosen uniformly from anywhere within the hypercube. Another example is a single-particle local move in which a single particle is selected and moved to a position within the hypercube that is chosen according to a probability distribution that is localized around the original location of the particle. Another example of a trial move is a multiple-particle local move in which some or all of the particles are moved to a new set of positions chosen according to a probability distribution that is localized around the original locations of the particles. Of course, the multiple-particle global move in which every particle is independently randomly moved to any position in the box is a Markov chain which uses no information from the previous configuration.

Single-particle trial moves and multiple-particle trial moves are complementary in that in many cases the former empirically mix rapidly when the system is far from a critical point phase transition while the latter are often superior empirically when the system is near a critical point.[18] Similarly, global trial moves and local trial moves are complementary in that the former appear empirically to mix rapidly at extremely low densities and to mix very slowly at higher densities while the latter often appear empirically to mix rapidly at higher densities.[1] The intuition behind this is that when the density is high, typical configurations of the particles disallow (or more generally make extremely improbable) many of the possible global trial moves; thus these trial moves are typically rejected. On the other hand, by moving a particle to a nearby position, which is often free even at relatively high densities, the trial move is more likely to be accepted. The motivation behind these distinctions is that, as opposed to the Ising model where each spin may adopt only two possible values, for many models each particle may adopt a large number of possible configurations. Trial moves which are local use more information from the previous state of the Markov chain to construct the trial state. The heat-bath algorithm applied to the  $q$ -state Potts model is similar in spirit since it too uses local information to boost acceptance probabilities to accelerate the empirically observed mixing.[18] Note that single-particle local moves are often used in practice in statistical physics for the same reason.[1,14,16] See Figure 1 for a summary of the rapid mixing results that are empirically observed for moderately high density liquid state systems in one, two, and three dimensions away from phase transitions and for a summary of

Empirical Results for $d = 1, 2, 3$			Theoretical Results		
	Local	Global		Local	Global
Single	rapid	slow	Single	Thm. 3 ( $d = 1$ )	Thm. 2
Multiple	rarely used	N.A.	Multiple	Thm. 5 ( $d = 1$ )	N.A.

Fig. 1. Summary of empirically observed and our theoretical rapid mixing results.

our theoretical results;[1] results for both single-particle and multiple-particle local-move and global-move Markov chains are presented.

Despite the widespread use of these methods in statistical physics, dating back to the earliest days of the MCMC method,[16] very little has been rigorously known about the convergence properties of either local-move or global-move Markov chains for this hard-core model. After a review of the hard-core model and of rapidly mixing Markov chains in Section 2 we present several rapid mixing results for this model. In Section 3 we prove that a single-particle global-move Markov chain is rapidly mixing for all dimensions  $d$  if the density  $\rho$  is sufficiently low. In Section 4 we prove that a single-particle local-move Markov chain is rapidly mixing in one dimension up to density one as long as the moves are within a small enough region around the particles. In Section 5 we relate the  $n$ -particle one-dimensional system to a convex body in  $\mathbb{R}^n$  and as a corollary show that there exist rapidly mixing multiple-particle local-move Markov chains. In Section 6 we conclude with a discussion of difficulties extending this work to higher densities, other dimensions, and to other energy functions.

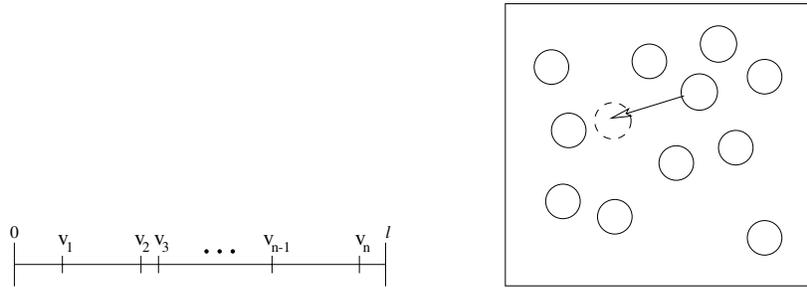
## 2 Background and Preliminaries

### 2.1 Background on the Hard Core Model

Let  $\Omega_{nd}$  be the set of all sets of  $n$  points in  $[0, \ell]^d$ , where we let  $\ell = 1$  without loss of generality. Thus, if  $X_i \in [0, 1]^d$  then  $X = (X_1, \dots, X_n) \in \Omega_{nd}$ . Let  $U(X)$  be the energy function of the configuration  $X$ , where  $U(X)$  has the form  $U(X) = \sum_{i \neq j} \phi(r_{ij})$ , where  $r_{ij} = |X_i - X_j|$  and  $\phi(r)$  is the two-particle energy function. A two-particle energy function of the form  $\phi(r) = 4\epsilon \left[ \left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6 \right]$ , where  $\epsilon$  and  $\sigma$  are parameters of the model, is known as the *Lennard-Jones energy function* and is widely used in applications.[1,14] A two-particle energy function of the form

$$\phi(r) = \begin{cases} \infty & \text{if } r \leq 2r^* \\ 0 & \text{if } r > 2r^* \end{cases} \tag{1}$$

is known as the *hard-core energy function* and is of interest in the present work.[1, 2,16,20] Note the similarity of this hard-core model to the hard-core Gibbs point process and of this work to related work on that process.[15,19] A two-particle energy function  $\phi(r)$  such that  $\phi(r) \geq 0$  for all  $r$ , e.g., the hard-core energy



**Fig. 2.** Illustration of the hard-core model in one and two dimensions

function, is said to be a *purely repulsive energy function*. In the hard-core case, each particle is modeled as a ball of radius  $r^*$  and the density of the system is defined to be  $\rho = nV_d(B_{r^*})$  where  $V_d(B_{r^*})$  is the volume of the  $d$ -dimensional ball of radius  $r^*$ . We focus attention on  $\rho$  (or  $r^*$ ),  $d$ , and  $n$ , and let  $n \rightarrow \infty$ , proving results for rapid mixing if  $\rho$  assumes certain values. Note that we will also consider the configuration space  $\Omega_{nd}^m$ , which is a discretized version of  $\Omega_{nd}$ ; more precisely, if  $X_i \in \Omega_{1d}^m = \frac{1}{m} \mathbb{Z}_m^d$  then  $X = (X_1, \dots, X_n) \in \Omega_{nd}^m$ . Sampling from  $\Omega_{nd}^m$  followed by a rejection sampling method is then sufficient to sample from  $\Omega_{nd}$ ; see [7]. Let  $\pi(X) = \frac{1}{Z} \exp(-\beta U(X))$  be the Boltzmann probability distribution on the configuration space  $\Omega$ ; in this expression  $\beta$  is a parameter known as the inverse temperature and  $Z$  is a normalization constant known as the partition function. In statistical mechanics it is often of interest to sample  $X \in \Omega$  according to the probability distribution  $\pi(X)$  and MCMC is often used as the method to do so. It should be clear that sampling from  $\Omega = \Omega_{nd}$  according to  $\pi$  for a system of hard-core particles corresponds to sampling from  $X \in \Omega_{nd}$  uniformly subject to the constraint that  $r_{ij} > 2r^*$  for all pairs of particles  $i$  and  $j$ . Particles are assumed to be indistinguishable, although for convenience we may assign them labels. In addition, we will assume either toroidal or truncated boundary conditions, as indicated; see [1] for more detail.

Figure 2 illustrates the system in one and two dimensions for the truncated boundary conditions; in the one-dimensional system the particles are represented as labeled points (with no radius shown) and in the two-dimensional system a single-particle trial move is illustrated.

### 2.2 Background on Rapid Mixing

Let  $\Omega$  be a set such as  $\Omega_{nd}$  or  $\Omega_{nd}^m$  and let  $p^{(0)}$  be an initial probability distribution on  $\Omega$ ,  $P$  be the transition kernel of a Markov chain  $\mathcal{M}$  on  $\Omega$ , and  $p^{(t)}$  be the  $t$ -step distribution; see [9,10,18]. The Markov chains we construct will be time-homogeneous, aperiodic and irreducible and the  $t$ -step distribution will converge to a stationary distribution  $p^{(\infty)}$ , i.e.,  $p^{(t)} \xrightarrow{t \rightarrow \infty} p^{(\infty)}$ . Let  $X^t = (X_1^t, \dots, X_n^t) \in \Omega$  be the state at the  $t$ -th step of a Markov chain and

let  $X_{TR}^{t+1}$  be the trial state chosen for the  $t + 1$ -st step. Since trial moves will be subjected to the Metropolis filter[9,16,18] the trial state is accepted and  $X^{t+1} = X_{TR}^{t+1}$  with probability  $\min\{1, \pi(X_{TR}^{t+1})/\pi(X^t)\}$ , where  $\pi$  is the Boltzmann distribution; otherwise the trial state is rejected and  $X^{t+1} = X^t$ . In the case of hard-core particles, the Metropolis rule leads to the rejection of any trial states in which two or more particles overlap and the acceptance of all other trial states. With this rule  $\pi = p^{(\infty)}$  and since the transition kernel is symmetric for the hard-core model, i.e.,  $P(A, B) = P(B, A)$  for all appropriately defined  $A, B \subset \Omega$ , the stationary distribution is uniform subject to the condition that no particles overlap. A trial move is a *single-particle trial move* if the trial state is of the form  $X_{TR}^{t+1} = (X_1^t, \dots, X_{\xi-1}^t, X_{\xi}^t, X_{\xi+1}^t, \dots, X_n^t)$  for some randomly chosen  $\xi$ , i.e., if the coordinates of only the single randomly chosen particle  $\xi$  are modified. A trial move is a *multiple particle trial move* if  $X_{TR}^{t+1} = (X_1^t, \dots, X_n^t)$ , where two or more of the  $X_i^t \neq X_i^t$ . A single-particle trial move is a *local trial move* if the chosen particle is moved to a trial state according to a probability distribution that is localized near the previous state and is a *global trial move* otherwise. A similar definition holds for multiple-particle local and global trial moves. In addition, all the Markov chains we consider will have a holding probability of 1/2, i.e., with probability 1/2 a trial move is not attempted and  $X^{t+1} = X^t$ ; this is to remove periodicity effects.

Given any two probability distributions  $\mu$  and  $\nu$  on some  $\Omega$ , the *variation distance* is a measure of how far  $\mu$  and  $\nu$  are from each other and is defined to be

$$\|\mu - \nu\|_{TVD} = \sup_{A \subseteq \Omega} |\mu(A) - \nu(A)|. \tag{2}$$

The *mixing time*  $\tau(\epsilon)$  is a measure how many steps it takes the Markov chain  $\mathcal{M}$  to come close to the uniform distribution and is defined as

$$\tau(\epsilon) = \sup_{p^{(0)}} \min \left\{ t : \|p^{(t)} - \pi\|_{TVD} \leq \epsilon \right\}. \tag{3}$$

We will define  $\tau = \tau(1/e)$ ; since  $\tau(\epsilon) \leq 2 \tau \log_2(1/\epsilon)$ , bounding  $\tau$  suffices to bound  $\tau(\epsilon)$ . [10]

There are several methods that may be used to bound the mixing time of a Markov chain. [10] One is the method of coupling; a coupling is a stochastic process  $(X_t, Y_t)_{t \in \mathbb{N}}$  on pairs of states in  $\Omega \times \Omega$  such that each of the  $(X_t)_{t \in \mathbb{N}}, (Y_t)_{t \in \mathbb{N}}$  considered independently is a faithful copy of the original Markov chain  $\mathcal{M}$ , and if  $X_t = Y_t$  then  $X_{t+1} = Y_{t+1}$ . A coupling may be used to bound the mixing time of a Markov chain via the coupling inequality, [10] which states that the worst variation distance is bounded by the coupling, i.e., that

$$\sup_{p^{(0)}} \|p^{(t)} - \pi\|_{TVD} \leq \sup_{X_0, Y_0} \mathbf{Pr} [X_t \neq Y_t]. \tag{4}$$

It is generally quite difficult to construct a coupling for complex Markov chains. The path coupling theorem of Bubley and Dyer simplifies and extends the applicability of coupling. [3] The path coupling theorem states that, given a neighborhood structure on the configuration space that connects every pair of states by

some path, it is sufficient to construct a coupling on pairs of neighboring states. We state the theorem in the simplest form necessary for our work.

**Theorem 1 (Bubley and Dyer[3]).** *Let  $\Omega$  be given and let  $X, Y \in \Omega$ . Define a metric  $\Phi : \Omega \times \Omega \rightarrow \{0, 1, \dots, D\}$  by letting  $\Phi(X, Y)$  be the length of the shortest path from  $X$  to  $Y$ . Suppose there exists a constant  $\zeta$  and a coupling  $(X_t, Y_t)$  of the Markov chain such that when  $\Phi(X_t, Y_t) = 1$  then  $\mathbf{E}[\Phi(X_{t+1}, Y_{t+1}) | X_t, Y_t] \leq \zeta \Phi(X_t, Y_t)$ . If  $\zeta < 1$  then the coupling can be extended to  $\Omega \times \Omega$  and the mixing time is bounded as  $\tau \leq \frac{\log(eD)}{1-\zeta}$ . If  $\zeta = 1$  and  $\Pr[\Delta\Phi_{t+1,t} \neq 0] \geq \alpha > 0$  for all  $X_t, Y_t \in \Omega$  then the mixing time is bounded as  $\tau \leq \frac{eD^2}{\alpha}$ .*

### 3 Single Particle Global Moves

In this section we prove that a single-particle global-move Markov chain is rapidly mixing for all dimensions  $d$  if the density is sufficiently low. We assume toroidal boundary conditions and have the following theorem.

**Theorem 2.** *Let  $\Omega_{nd}$  be the configuration space of  $n$  hard-core particles with radius  $r^*$  in dimension  $d$ . Let the density be  $\rho$  and the system have toroidal boundary conditions. Let  $\mathcal{M}$  be a Markov chain on  $\Omega_{nd}$  with the Metropolis filter in which trial moves are single particle global moves and let  $\gamma > 0$  be a constant. Then the Markov chain  $\mathcal{M}$  is rapidly mixing with mixing time bounded by*

$$\tau \leq \begin{cases} \frac{1+\gamma}{\gamma} 2n \log(2en) & \text{if } \rho \leq 1/(2^{d+1}(1 + \gamma)) \\ 2n^2 \log(2en) & \text{if } \rho = 1/(2^{d+1}) \end{cases} \tag{5}$$

*Proof.* The proof will be by path coupling and is a continuous analog of a hard-core result on graphs.[3] Since the  $n$  particles can be interpreted as  $n$  points (given by the particle centers) such that no two points can be within a distance  $2r^*$  of each other, each particle “blocks” other particles from being placed within a ball of radius  $2r^*$  around it. If this ball is denoted by  $B_{2r^*}$  then the total “blocked” volume is at most  $nV_d(B_{2r^*}) = 2^d nV_d(B_{r^*}) = 2^d \rho$ . Two  $n$ -particle configurations  $X$  and  $Y$  are defined to be neighbors if they differ by a single particle; denote the location of this particle in  $X$  by  $p_X$  and in  $Y$  by  $p_Y$ .

To bound the diameter, i.e., the maximal value  $D$  of the metric, it is necessary to construct paths between every pair of configurations  $X$  and  $Y$ . Number the particles  $1, 2, \dots, n$  in both configurations (any numbering will do). Pick up particle 1 in  $X$  and in  $Y$ , so that there are  $(n - 1)$  particles remaining in each. Each particle blocks a neighborhood of size  $V_d(B_{2r^*})$  from particle moves, so the  $2(n - 1)$  particles block at most  $2(n - 1)V_d(B_{2r^*}) = 2^{d+1}(n - 1)V_d(B_{r^*}) < 2^{d+1}\rho \leq 1$ . Since this is a strict inequality there is a spot on the torus that is blocked in neither  $X$  nor  $Y$  so we can place particle 1 at this spot in both configurations. The same procedure applies to the remaining particles  $2, 3, \dots, n$ , and gives a common “target” configuration. Thus it requires  $n$  steps from  $X$  to this target configuration, and  $n$  steps from there to return to  $Y$ , so the length of the longest path is  $D = 2n$ .

To define the path coupling, first label the particles in  $X$  and  $Y$  as follows. Label the particle where  $X$  and  $Y$  differ as particle 1; the rest of the particles can be labeled such that particle  $i > 1$  in  $X$  and particle  $i > 1$  in  $Y$  are at the same location. One step of the Markov chain proceeds as follows:

- Choose particle  $j$ .
- Choose trial location  $P_X$  for particle  $j$  in configuration  $X$ .
- If  $j = 1$  then chose trial location  $P_Y = P_X$  for  $j$  in  $Y$ .
- If  $j \neq 1$  chose:
  - If  $P_X \cap [B_{2r^*}(p_X) \cup B_{2r^*}(p_Y)] \neq \emptyset$  then choose  $P_Y = p_Y + (p_X - P_X)$ , i.e., if the new location overlaps with particle 1 in  $X$  or particle 1 in  $Y$  then reflect through the midpoint  $\frac{1}{2}(p_X + p_Y)$ .
  - Otherwise choose  $P_Y = P_X$ .
- Hold with probability  $1/2$ . Otherwise, attempt the trial move, accepting if it is a valid move.

To analyze the coupling, consider the case that particle  $j = 1$  is chosen in  $X$  and  $Y$ . The coupling was set up such that either the attempted move succeeds and couples ( $\Delta\Phi = -1$ ) in both  $X$  and  $Y$ , or fails in both in which case both particles return to their initial locations ( $\Delta\Phi = 0$ ). Each of the remaining  $(n-1)$  particles prevents a move into a radius  $2r^*$  region around them, for a total blocked volume of at most  $(n-1)V_d(B_{2r^*}) = 2^d(n-1)V_d(B_{r^*}) = 2^d(1-1/n)\rho$ . The probability of coupling is the probability of a successful move, which is thus at least  $\frac{1}{n}(1-2^d(1-1/n)\rho)$ . When a particle  $j > 1$  is chosen then there are several cases to consider. If  $P_X$  is near  $p_X$  then it fails to move in both configurations, so  $\Delta\Phi = 0$ . If  $P_X$  is near  $p_Y$  then it might make a move in only one of  $X$  or  $Y$  and the probability of distance increasing by one ( $\Delta\Phi = +1$ ) may be as high as  $V_d(B_{2r^*}) = 2^d\rho/n$ . Any other location of  $P_X$  leads to both configurations moving successfully or both failing to move, so  $\Delta\Phi = 0$ . This gives

$$\begin{aligned} \mathbf{E}[\Delta\Phi] &\leq -\frac{1}{2n} [1 - (1 - 1/n) 2^d \rho] + \frac{n-1}{2n} 2^d \rho/n \\ &= \frac{1}{2n} [-1 + 2(1 - 1/n) 2^d \rho] \end{aligned}$$

and the theorem follows because  $\mathbf{E}[\Delta\Phi] < 0$  when  $\rho \leq 1/2^{d+1}$ .

This theorem thus proves rapid mixing in 1-d for  $\rho \leq .25$ , in 2-d for  $\rho \leq .125$ , and in 3-d for  $\rho \leq .0625$ . The corresponding packing densities are  $1$ ,  $\frac{\pi}{2\sqrt{3}} \approx 0.9$ , and  $\frac{\pi}{3\sqrt{2}} \approx 0.74$ , respectively.[4,17] The slow empirical mixing indicated in Figure 1 for single-particle global-move Markov chains is for typical liquid state densities which are much closer to the packing density than are the densities for which our bounds hold. Although these bounds may thus seem rather weak, the packing density is not the best comparison; it is more appropriate to compare to the lowest density such that no global moves are possible, such as the 2-d square lattice with density  $\rho = \frac{\pi}{3\sqrt{3}} \approx 0.78$  and the 3-d cubic lattice with  $\rho = \frac{\pi}{6} \approx 0.52$ .

Alternatively, Rogers showed that  $\rho_{max} \geq d\zeta(d)/2^{d-1} e(1-e^{-d}) \xrightarrow{d \rightarrow \infty} d/2^{d-1} e$ ,

where  $\rho_{max}$  is the packing density and the Zeta function is  $\zeta(d) = \sum_{k=1}^{\infty} k^{-d}$ ; thus, Theorem 2 proves rapid mixing asymptotically up to a factor of  $(ed/4)$  of this lower bound on the packing density.[4,17] Note that even in two and three dimensions the density regime for which we have proven rapid mixing is significantly below densities typically of interest to practitioners simulating liquid state systems.[1]

## 4 Single Particle Local Moves

In this section we consider a single-particle local-move Markov chain on a discretized version of a one-dimensional  $n$ -particle hard-core system. Suppose one is interested in sampling configurations from  $\Omega_{n1}$  with toroidal boundary conditions, i.e., from on the circle of unit circumference with  $n$  particles. If this circle is discretized by placing on it  $m$  equally spaced grid points then an acceptable configuration is one in which particle centers are placed on grid points such that adjacent particles are at least a distance  $2r^*$  apart; equivalently, the particle centers are at least  $\lceil m\rho/n \rceil$  grid points apart. This resulting space  $\Omega_{n1}^m$  can be sampled with the *grid-walk Markov chain*, which is a single-particle local-move Markov chain in which trial moves are generated by choosing a particle uniformly at random and attempting to move it to either the right or the left one grid point, rejecting the move if this causes two particles to overlap and accepting it otherwise. The original space  $\Omega_{n1}$  can then be sampled by rejection sampling.[7] Our main result will be the following theorem.

**Theorem 3.** *Let  $\Omega_{n1}^m$  be the discretized version of the configuration space of  $n$  hard-core particles with radius  $r^*$  in dimension 1. Let the density be  $\rho$  and the system have toroidal boundary conditions. Let  $\mathcal{M}$  be a Markov chain on  $\Omega_{n1}^m$  with the Metropolis filter in which trial moves are the single-particle local-move grid walk. Then the Markov chain  $\mathcal{M}$  is rapidly mixing with mixing time bounded by*

$$\tau \leq \frac{4e}{3} n^3 m^2 (1 - \rho)^2. \quad (6)$$

*Proof.* The proof will be by path coupling. Let  $k = \lceil m\rho/n \rceil$  be the minimal number of grid points between adjacent particle centers. Denote some arbitrary (but fixed) grid point as the origin 0, and number the particles 1, 2, ...,  $n$  proceeding counterclockwise from 0. The location of the center of particle  $j$  in configuration  $X$  will be denoted  $p_j^X \in [0 \dots m - 1]$ , counting grid points counterclockwise from the origin 0, and likewise in  $Y$ . Two configurations  $X$  and  $Y$  will be neighbors if they differ by a single particle  $j$ , such that the positions of  $j$  ( $p_j^X$  in  $X$  and  $p_j^Y$  in  $Y$ ) are on adjacent grid points, i.e.,  $p_j^X - p_j^Y \equiv \pm 1 \pmod{m}$ .

To bound the diameter, first suppose that  $p_1^X < p_1^Y$ ; in this case, move particle 1 in configuration  $Y$  down by  $p_1^Y - p_1^X$  grid points to reach  $p_1^X$ . Likewise, if  $p_1^X > p_1^Y$  then move the particle in configuration  $X$  instead. Repeat this for the remaining particles, proceeding counterclockwise around the circle (i.e., by

increasing index  $j$ ). Then each particle was moved only once, in either  $X$  or  $Y$ , and moreover each was moved by at most  $m(1 - \rho)$  grid points. Therefore,  $D \leq nm(1 - \rho)$ .

To define the path coupling, let the origin 0 be set so that under the labeling given above the configurations differ only by particle 1, at locations  $p_1^X$  and  $p_1^Y \equiv p_X \pm 1$ ; without loss of generality assume  $p_1^X = p_1^Y + 1$ . Moves in the Markov chain are defined as follows. First choose particle  $j$  in  $X$  and trial move to location  $P_X \in p_j^X + \{-1, 0, 1\}$ . Choose the same particle  $j$  in  $Y$  and make a move as follows:

- If  $j > 1$  then  $P_Y = p_j^Y + (P_X - p_j^X)$  (i.e., make the same move in  $Y$  as in  $X$ )
- If  $j = 1$  then  $P_Y = p_1^Y + \delta$  where

$$\delta = \begin{cases} 0 & \text{if } P_X = p_1^X - 1 & \text{(couples and } \Delta\Phi = -1) \\ 0 & \text{with probability } 1/2 \text{ if } P_X = p_1^X & \text{(no change and } \Delta\Phi = 0) \\ 1 & \text{with probability } 1/2 \text{ if } P_X = p_1^X & \text{(couples and } \Delta\Phi = -1) \\ -1 & \text{if } P_X = p_1^X + 1 & \text{(doesn't couple and } \Delta\Phi = +2) \end{cases}$$

To analyze the coupling, first suppose that there are no particles adjacent to particle 1 in either  $X$  or  $Y$ . Then  $\Delta\Phi = 0$  when  $j \neq 1$  is chosen, and  $\mathbf{E}[\Delta\Phi] = 0$  when particle 1 is chosen. Overall,  $\mathbf{E}[\Delta\Phi] = 0$  and  $\alpha = 3/4n$ . Next, suppose that some particle, say particle 2, is adjacent to particle 1 in  $X$ , i.e.,  $p_2^X = p_1^X + k$ . Then, when particle 2 is chosen then  $\mathbf{E}[\Delta\Phi] = 1/4$  (when 2 tries to move to  $P_X = p_X - 1$ ), and when 1 is chosen then  $\mathbf{E}[\Delta\Phi] = -1/4$ . Overall, again  $\mathbf{E}[\Delta\Phi] = 0$  and  $\alpha = 3/4n$ . Likewise when particle 2 is adjacent to 1 in  $Y$  instead of in  $X$ . Thus,  $\mathbf{E}[\Delta\Phi] = 0$ ,  $\zeta = 1$  and  $\alpha = \mathbf{Pr}[\Delta\Phi \neq 0] = 3/4n$  in the path coupling theorem. Thus, the result follows from the path coupling theorem.

It might appear that the  $n^3$  term in the mixing time is weak and that one should be able to obtain a bound of  $O(n \log n)$ , since it takes  $O(n \log n)$  steps to choose all of the  $n$  particles and most of the particles may have to move  $O((m(1 - \rho))^2)$  steps. In fact, one can show that

$$\tau = O\left((m(1 - \rho))^2 n (\log n + \log \log m)\right)$$

by comparing this problem to an  $n$  particle exclusion process on the circle with  $m(1 - \rho)$  grid points and using a result of Diaconis and Saloff-Coste.[5] Aside from the  $\log \log m$  term, this bound is tight. For instance, consider the initial configuration where all the particles are at the “bottom” of the circle. In order to approach the stationary distribution, a significant fraction of the particles will need to become well distributed around the circle. It requires  $O((m(1 - \rho))^2)$  steps to move a particle a constant fraction around the circle (consider a random walk on a line). It requires  $n \log n$  steps to choose each particle once. Therefore it requires  $\Omega((m(1 - \rho))^2 n \log n)$  steps to distribute the particles around the circle. A similar lack of tightness between coupling and the optimal result is seen with a standard random walk on the grid  $[m]^n$ , where a coupling will yield  $\tau = O(m^2 n^3)$ , where  $D = mn$  and  $\alpha = 1/n$ , whereas one can obtain a bound of  $\tau = O(m^2 n \log n)$  via the method of log-Sobolev constants.[5]

## 5 Convex Bodies and Multiple Particle Local Moves

Insight into several aspects of the one-dimensional problem may be gained from the following theorem. It relates an  $n$ -particle one-dimensional local-moves problem to the problem of computing the volume of a convex body in  $n$  dimensions.

**Theorem 4.** *Let  $\Omega_{n1}$  be the configuration space of  $n$  hard-core particles with radius  $r^*$  in dimension 1. Let the density be  $\rho$ , and the system have truncated boundary conditions. Let  $\mathcal{M}$  be a Markov chain on  $\Omega_{n1}$  with the Metropolis filter in which trial moves are generated with the following multiple-particle local move procedure: for every particle, move that particle to a new position chosen uniformly from the 1-dimensional ball centered at the initial position of that particle and of radius less than  $4r^*$ . Then the set of states reachable by the Markov chain  $\mathcal{M}$  is a convex body in  $\mathbb{R}^n$ .*

*Proof.* Consider  $n$  particles of radius  $r^*$  positioned on a line with initial configuration  $0 \leq v_1 < v_2 < \dots < v_n \leq \ell = 1$  as in Figure 2. Then if the local moves are of size less than  $4r^*$  then the set of configurations reachable through local moves is exactly the polytope defined by the following equations.

$$\begin{aligned} v_1 &\geq 0 \\ v_n &\leq \ell = 1 \\ \forall i < n : v_{i+1} &\geq v_i + 2r^* \end{aligned} \tag{7}$$

To see this, observe that in the one-dimensional system the particles are ordered and this order is preserved if the local moves are less than  $4r^*$ . From the constraints due to the endpoints of the line and since the particles are hard balls we get (7). Conversely, any values of the  $v_i$  that satisfy the  $n + 1$  conditions (7) is a valid configuration. Thus, (7) exactly defines the set of valid configurations.

Thus, in one dimension sampling with a local-move Markov chain corresponds to choosing a random sample from the convex body defined by (7). Sampling with a global-move Markov chain also corresponds to choosing a random point from a convex set or one of a large number of (up to particle label) identical convex sets. This result makes available a great deal of work on various techniques for sampling from convex bodies.[6,12] As a corollary to Theorem 4, there exist rapidly mixing Markov chains for multi-particle continuous local-move Markov chains. For example, continuous moves can be implemented via the *ball walk* of [12] which in the present context is a multiple-particle local-move Markov chain in which particle moves are subject to certain conditions.

**Theorem 5.** *Let  $\Omega_{n1}$  be the configuration space of  $n$  hard-core particles with radius  $r^*$  in dimension 1. Let the density be  $\rho$ , and the system have truncated boundary conditions. Then there exists a multiple-particle local-move Markov chain  $\mathcal{M}$  on  $\Omega_{n1}$  with the Metropolis filter for which  $\tau \leq \text{poly}(n)$ .*

*Proof.* See [12] and note that the ball walk is a multiple-particle local-move algorithm.

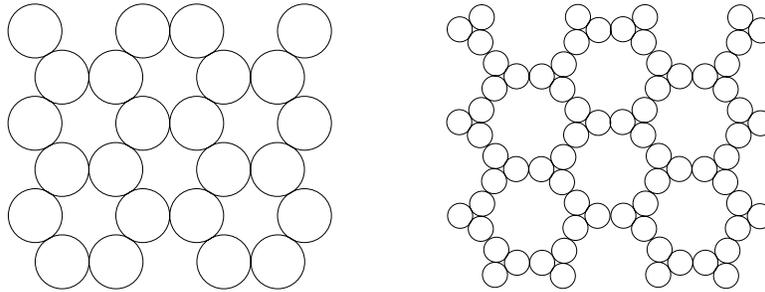
Due to the identification with the convex body problem, numerous other Markov chains can be seen to mix rapidly. For example, the *King's Moves* of [6] is a single-particle local-move Markov chain that mixes in polynomial time but that has more complex boundary conditions since steps are allowed slightly outside the convex body, i.e., since a small amount of particle overlap is allowed. One Markov chain that is of interest to practitioners involves performing single-particle continuous local moves. In this case the configuration space  $\Omega_{nd}$  is sampled (not by sampling from  $\Omega_{nd}^m$  followed by rejection sampling but instead) by performing trial moves in which a single particle is moved to a nearby location chosen from the continuous configuration space and no configurations in which there is any particle overlap are allowed, i.e., the Markov chain must not step outside of the convex body. This Markov chain is not rapidly mixing for general convex bodies, although this does not exclude the possibility that it is rapidly mixing for the convex body of Theorem 4.

## 6 Discussion and Conclusion

Although it is empirically observed that single-particle global-move Markov chains tend to work best at very low densities and that single-particle local-move Markov chains tend to work best at higher densities, the situation is more complicated for worst case analysis since the configuration space  $\Omega$  becomes disconnected or weakly connected under local and global moves at relatively low densities. For example, consider the two rigid packings shown in Figure 3. The first has density  $\rho \approx 0.6$  but both local and global trial moves will always be rejected; the second has density  $\rho \approx 0.39$  and, although global trial moves can be made relatively easily, local trial moves will always be rejected.[4,8] In three dimensions there is even a packing with density  $\rho \approx 0.0555$  such that no local trial moves will be accepted.[8] However know from Theorem 2 that a global-move Markov chain at this density mixes in time  $\tau \leq 18n \log(2en)$ . Note that for configurations such as these, if the particle radius  $r^*$  is decreased slightly then the resulting component of  $\Omega$  will have positive probability but will not communicate or will communicate only weakly, i.e, via a tight “bottleneck”, with the remainder of  $\Omega$ .

One method from physics to construct an initial state that mixes well empirically is to start with the  $n$  particles in a dense packing near the packing density and then decrease  $r^*$  until the density reaches the desired  $\rho$ ; this gives a configuration in which the particles are well separated. In one dimension this is equivalent to putting the  $n$  particles at locations  $\frac{i-0.5}{n} \ell$ ; from Theorem 4 we see that this is the center of the polytope of one-dimensional configurations. Although constructing such a starting configuration may be easy, determining the mixing rate within the main component is difficult since characterizing the main component is difficult and since methods for rapid mixing require rapid mixing from almost every starting point.

It would be nice to prove a rapid mixing result involving other purely repulsive energy functions or energy functions with both attractive and repulsive



**Fig. 3.** Two low density rigid packings

terms. One might think that if Theorem 2 holds for hard particles of radius  $r^*$ , then it should hold for at least some types of repulsive energy functions that are 0 outside radius  $r^*$ . However, even this seems difficult to show. To see why, consider the case in which the differing particle, #1, interacts with many particles in configuration  $X$  but with none in configuration  $Y$ . This will cause the particles that interact with #1 to have a higher probability of moving in  $X$  than in  $Y$ , and thus a higher chance of the distance increasing in the coupling. Hence the coupling approach fails. A more extreme example of this phenomenon may be seen by considering energy functions with an attractive term.

Other methods seem equally unlikely to yield positive results. Lovász et. al. have applied conductance methods to the ball walk and Jerrum et. al. have applied coupling and spectral methods to the ball walk.[12,10] These proofs rely on isoperimetric inequalities for convex bodies. By an argument similar to that in the proof of Theorem 4 one can show that if  $r^* = 0$  then  $\Omega_{nd}$  is a convex set. If  $r^* > 0$  then if particles  $i$  and  $j$  overlap then  $r_{ij} \leq 2r^*$ ; this constraint corresponds to the convex region in which one particle is placed anywhere, the second is placed a distance at most  $2r^*$  away from it, and the rest of the particles are placed anywhere. Then the configuration space  $\Omega_{nd}$  for  $r^* > 0$  is a convex set with a large number of convex regions removed from it. It does not seem possible to derive isoperimetric inequalities for the non-convex body formed by the configurations of the  $n$  particles if  $d > 1$ .

Related work on the hard-core Gibbs point process and on simulated tempering is of interest.[13,15,19] Simulated tempering is a method which avoids some of the hard-disk problems by running a Markov chain with multiple energy functions or at multiple temperatures. The Markov chain then consists of single-particle moves as well as temperature swaps; it is hoped that the softer energy functions will allow escape from bad configurations. However, analyzing the tempered chain will require knowledge of the mixing times of local-move algorithms with various energy functions and not just the hard-core energy function. In addition, the only mixing bounds currently known for tempered Markov chains establish that the tempered chain mixed no more slowly than the underlying chains; this would not be useful in the present context.

**Acknowledgments.** The authors sincerely thank Mark Jerrum, Santosh Vempala, and Dana Randall for numerous helpful discussions.

## References

1. M.P. ALLEN AND D.J. TILDESLEY, *Computer Simulation of Liquids*. (Oxford University Press, Oxford, 1987).
2. B.J. ALDER AND T.E. WAINWRIGHT, *Phase Transition in Elastic Disks*. *Physical Review*, 127:359–361, 1962.
3. R. BUBLEY AND M. DYER, *Path coupling: A technique for proving rapid mixing in Markov chains*. 38th Annual Symposium on Foundations of Computer Science, pp. 223–231, 1997.
4. J.H. CONWAY AND N.J.A. SLOANE, *Sphere Packings, Lattices and Groups*. (Springer-Verlag, New York, 1988).
5. P. DIACONIS AND L. SALOFF-COSTE, *Logarithmic Sobolev inequalities for finite Markov chains*. *Annals of Applied Probability*, 6:695–750, 1996.
6. M. DYER, A. FRIEZE AND R. KANNAN, *A random polynomial time algorithm for approximating the volume of convex bodies*. *Journal of the ACM*, 38:1–17, 1991.
7. G.S. FISHMAN, *Monte Carlo: Concepts, Algorithms, and Applications*. (Springer-Verlag, New York, 1996).
8. M. GARDNER, *Martin Gardner's New Mathematical Diversions from Scientific American*. (Simon and Shuster, CITY, 1966).
9. O. HÄGGSTRÖM, *Finite Markov Chains and Algorithmic Applications*. (Cambridge University Press, Cambridge, 2002).
10. M. JERRUM, *Counting, sampling and integrating: algorithms and complexity*. *Lectures in Mathematics*, ETH Zürich. (Birkhäuser Verlag, Basel, 2003).
11. M. JERRUM AND A. SINCLAIR, *Polynomial-time approximation algorithms for the Ising model*, *SIAM Journal on Computing*, 22:1087–1116, 1993.
12. L. LOVÁSZ AND R. KANNAN, *Faster mixing via average conductance*. 31st Annual ACM Symposium on Theory of Computing, pp. 282–287, 1999.
13. N. MADRAS AND D. RANDALL, *Markov chain decomposition for convergence rate analysis*. *Annals of Applied Probability*, 12:581–606, 2002.
14. M.W. MAHONEY AND W.L. JORGENSEN, *A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions*. *Journal of Chemical Physics*, 112:8910–8922, 2000.
15. S. MASE, J. MØLLER, D. STOYAN, R.P. WAAGEPETERSEN AND G. DÖGE, *Packing Densities and Simulated Tempering for Hard Core Gibbs Point Processes*. *Annals of the Institute of Statistical Mathematics*, 53:661–680, 2001.
16. N. METROPOLIS, A.W. ROSENBLUTH, M.N. ROSENBLUTH, A.H. TELLER AND E. TELLER, *Equation of state calculation by fast computing machines*. *Journal of Chemical Physics*, 21:1087–1092, 1953.
17. <http://mathworld.wolfram.com>.
18. M.E.J. NEWMAN AND G.T. BARKEMA, *Monte Carlo Methods in Statistical Physics*. (Oxford University Press, Oxford, 1999).
19. D. STOYAN, W.S. KENDALL AND J. MECKE, *Stochastic Geometry and Its Applications, 2nd ed.* (Wiley, New York, 1995).
20. L. TONKS, *The Complete Equation of State of One, Two and Three-Dimensional Gases of Hard Elastic Sphere*. *Physical Review*, 50:955–963, 1936.