Convergence and coupling for spin glasses and hard spheres

Cédric Chanal and Werner Krauth

Laboratoire de Physique Statistique, Ecole Normale Supérieure, CNRS, 24 rue Lhomond, 75231 Paris Cedex 05, France
(Received 15 October 2009; published 19 January 2010)

We discuss convergence and coupling of Markov chains, and present general relations between the transfer matrices describing these two processes. We then analyze a recently developed local-patch algorithm, which computes rigorous upper bound for the coupling time of a Markov chain for nontrivial statistical-mechanics models. Using the "coupling from the past" protocol, this allows one to exactly sample the underlying equilibrium distribution. For spin glasses in two and three spatial dimensions, the local-patch algorithm works at lower temperatures than previous exact-sampling methods. We discuss variants of the algorithm which might allow one to reach, in three dimensions, the spin-glass transition temperature. The algorithm can be adapted to hard-sphere models. For two-dimensional hard disks, the algorithm allows us to draw exact samples at higher densities than previously possible.

I. INTRODUCTION

The Monte Carlo method is a fundamental computational tool in science. Its goal is to sample configurations x in a given state space from a probability distribution $\pi(x)$. This can usually not be achieved directly for multidimensional distributions. Markov chain Monte Carlo methods [1,2] overcome this problem by generating configurations $x_0, x_1, x_2, \ldots$ starting from an initial configuration x_0, which belongs to a simpler distribution π^0 (often a fixed initial condition, or some *ad hoc* random choice). Configurations x_t are then generated from x_{t-1} according to a stochastic algorithm which guarantees, as time moves on, that π^t departs from the initial condition and converges for $t \to \infty$ toward the equilibrium distribution $\pi^\infty \equiv \pi$. The Markov chain approach can be implemented for arbitrary distributions π, using, for example, the Metropolis and the heat-bath algorithms. For many applications, enormous effort has gone into designing fast algorithms for which one reaches $\pi^t(x) \sim \pi(x)$ for reasonable running times t. In this paper, we are concerned with a related problem: rather than to find the fastest algorithm for a given problem, we are interested in quantifying the speed of a given Markov chain algorithm. This is, we want to prove after which time t the sample x_t is equilibrated. It then reflects the equilibrium distribution and no longer the initial configurations. In many practical applications, it is extremely difficult to decide from within the simulation whether it has indeed equilibrated [2–4]. Instead, one must validate the simulation results with other approaches, from exact solutions to experimental data. The correct characterization of the convergence toward equilibrium from within the simulation has remained a serious conceptual and practical problem of the Monte Carlo method.

From a fundamental viewpoint, the problem of rigorously proving convergence of a simulation was solved, at least in principle, through a paradigm called "exact sampling," which allows to generate, with Markov chains, samples x directly from the equilibrium distribution π without any influence of the initial configuration [5]. In practice, however, it has not been possible to implement exact sampling for many complicated problems, as, for example, disordered systems, for which standard methods of evaluating equilibration times fail.

The reason for this difficulty is as follows: exact sampling proves for a given Markov chain simulation that the correlation of the initial configuration with the configuration at time t strictly vanishes. This is done by showing explicitly that all possible initial configurations x_{t_0} yield the same output under coupled Monte Carlo dynamics. In many simple models, one can prove this coupling property indirectly. In general, however, one must indeed survey the entire configuration space. This is usually too complicated to be achieved.

We recently developed a local-patch algorithm [6], which indeed monitors the entire configuration space of complicated systems, even for very large sizes. The approach uses local information, concentrated on so-called "patches." The scale of these patches increases during the simulation. Information on patches can then be combined for the entire system to provide a crucial upper bound for the (global) coupling time, and to generate an exact sample. The algorithm was demonstrated to work for spin glasses at lower temperatures than previous methods [7,8], even though the physically interesting regime has still not been reached yet. The local-patch algorithm is quite general: in addition to spin glasses, we implement it in this paper for hard disks and improve on previous results [9,10]. The successful application of exact sampling to hard-sphere systems is remarkable because the configuration space is continuous so that, naively, its complete survey appears out of reach.

A. Transfer matrix

A Markov chain is fully characterized by the so-called "transfer matrix" of transition probabilities between any two configurations k and l. As will be illustrated shortly (Sec. II) in a specific example, the largest eigenvalue of the transfer matrix is $\lambda_1 = 1$ and the corresponding eigenvector Ψ_1 describes the equilibrium state. The convergence toward equilibrium is governed by the spectrum of the transfer matrix and by the overlap of the eigenvectors Ψ_k with the initial configuration

$$\pi^t(x) = \pi(x) + \sum_{\Psi_k, \lambda_k < 1} \langle \Psi_k | \pi^0 \rangle \Psi_k(x) \lambda_k^t. \qquad (1)$$

In the limit of infinite simulation time, the second-largest eigenvalue determines the exponential convergence of the

probability distribution toward equilibrium. This eigenvalue sets a time scale

$$\tau_{\text{corr}} = 1/|\log \lambda_2|,$$

and the convergence is as

$$\pi^t - \pi \propto \exp(-t/\tau_{\text{corr}}) \quad \text{for} \quad t \to \infty. \tag{2}$$

The rigorous determination of convergence properties of Markov chains has been undertaken in many cases, from urn models to card-shuffling (see [11]), diffusion processes, and many more (see [12]). Efficient algorithms, as, for example, the bunching method [2] are commonly used to perform an empirical error analysis of Monte Carlo data in more complicated cases, where rigorous calculations are out of the question. However, these methods are not failsafe. In practice, it is often difficult to extract τ_{corr} from the large number of physically relevant time scales. In disordered systems, for example, there is often no reliable way to ascertain that the simulation has run long enough, and τ_{corr} may be much larger than assumed (see, e.g., [2], Sec. 1.5).

B. Loss of correlation and exact sampling

In the limit of infinite times $t \to \infty$, the Markov chain converges toward the equilibrium distribution, and the positions x_t become independent of the initial condition. The loss of correlation with the initial condition is evident for Markov chains that couple, that is, which for each possible initial condition x_0 produce the same output x_t. In many cases of interest, this happens after a finite global "coupling time" $t \geq t_{\text{coup}}$, which depends on the realization of the Markov chain. Propp and Wilson [5] realized that this coupling property allows one to draw "exact" samples from the distribution π.

This approach, called "coupling from the past," eliminates the problem of analyzing the convergence properties. However, to establish that a Markov chain has coupled, the entire state space of the system must be supervised. This was believed infeasible except for special problems where the dynamics conserves a certain (partial) ordering relation on configurations. A partial order is conserved in heat-bath dynamics of the ferromagnetic Ising model, whereas the frustration in the spin-glass model foils this simplification.

II. COUPLING AND CONVERGENCE IN A ONE-DIMENSIONAL MODEL

We first discuss convergence and coupling for a Markov chain describing the hopping of a single particle on a simple N-site lattice with periodic boundary conditions (see Fig. 1). In one time step, the particle hops with probability $\frac{1}{3}$ from one site to its two neighbors

$$p_{k \to k+1} = p_{k \to k-1} = 1/3 \ (\text{if possible}). \tag{3}$$

In addition, we have $p_{1 \to N} = p_{N \to 1} = 1/3$.

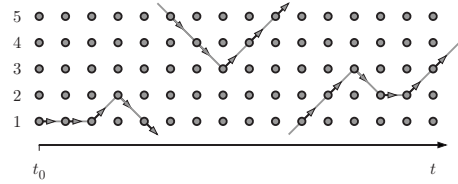The equal hopping probabilities imply via the detailed balance condition

FIG. 1. A Markov chain on a five-site lattice with periodic boundary conditions. The particle hops from a site k toward its neighbors with probability 1/3 each.

$$\pi_k p_{k \to l} = \pi_l p_{l \to k} \tag{4}$$

that the stationary probability distribution $\pi_k = 1/N$ of this problem is independent of k.

This system's Monte Carlo algorithm is encoded in the $N \times N$ transfer matrix $T^{1,1}$

$$T^{1,1} = \{p(i \to j)\} = \frac{1}{3} \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 & 1 \\ 1 & 1 & \ddots & & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & & \ddots & 1 & 1 \\ 1 & 0 & \cdots & 0 & 1 & 1 \end{pmatrix}. \tag{5}$$

The eigenvalues of $T^{1,1}$ are $\lambda_k^{1,1} = \frac{1}{3}(1 + 2 \cos\frac{2(k-1)\pi}{N}), k = 1, \ldots, \text{Int}[N/2] + 1$ (with multiplicities) that is, for $N = 5$, $\{1, \frac{1+\sqrt{5}}{6}, \frac{1-\sqrt{5}}{6}\}$. The largest eigenvalue, $\lambda_1^{1,1} = 1$ corresponds to the conservation of probabilities. By construction, it is associated with the equilibrium solution $(\pi_1, \ldots, \pi_N) = (\frac{1}{N}, \ldots, \frac{1}{N})$. The second-largest eigenvalue is $\lambda_2^{1,1}$. For $N = 5$ we have $\lambda_2^{1,1} = \frac{1+\sqrt{5}}{6} = 0.539$. This eigenvalue controls the long-time corrections to the stationary solution, which vanish as $[\lambda_2^{1,1}]^t = \exp[-t/\tau_{\text{corr}}]$, with

$$\tau_{\text{corr}} = 1/|\log \lambda_2^{1,1}|.$$

We note that the time scale τ_{corr} only describes the asymptotic behavior of the correlation. The calculation of the time t at which the probability distribution π^t itself is within a suitably chosen ϵ of the equilibrium distribution π is more involved (see, for example, [11,12],).

A. Coupling

As illustrated in Fig. 2, the Monte Carlo algorithm can be formulated in terms of random maps. In our example, this
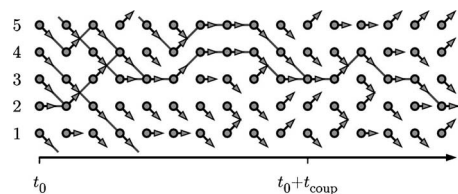

FIG. 2. Extended Monte Carlo simulation on $N = 5$ sites. Trajectories from all possible initial configurations at $t = t_0$ are indicated. They "couple" at $t = t_0 + t_{\text{coup}}$. The coupling time (here, $t_{\text{coup}} = 10$) depends on the realization of the Markov chain.

means that instead of prescribing one move per time step, as in Fig. 1, we now sample moves for all times t and all sites k, in such a way that the dynamics of a single particle again satisfies the detailed balance condition of Eq. (4). The most natural implementation of this approach is illustrated in Fig. 2: arrows are chosen independently for all times t and all sites k. At time t_0, for example, the particle should move down from sites 1, 3, 4, and 5 and straight from site 2. We can now check the outcome of the Monte Carlo calculation. In the example of Fig. 2, from time t_0+10 on, all initial configurations of the single particle yield the same output. This is remarkable because, evidently, at this time the initial conditions are completely forgotten.

The coupling time t_{coup} is a random variable ($t_{coup}=10$ in Fig. 2), which depends on the realization of the full Monte Carlo simulation from time t_0 onwards (until coupling has been reached). The independence of random maps on different time steps implies that the probability for not coupling vanishes at least exponentially fast in the limit $t \rightarrow \infty$.

Under the random-map dynamics, an initial state with N particles eventually evolves into a state with one particle (in later sections, spin-glass configurations will take the place of the single-particle positions). More generally, a state with k configurations can evolve at each time step into a state with $k' \leq k$ configurations. Figure 2 displays a sequence of random maps and illustrates the associated time-forward search of the coupling time.

This extended Monte Carlo dynamics on k-configuration states can again be described by a transfer matrix

$$T^{fw} = \begin{pmatrix} T^{1,1} & T^{2,1} & \cdots & & \cdots \\ 0 & T^{2,2} & T^{3,2} & & \cdots \\ 0 & 0 & T^{3,3} & & \cdots \\ \cdots & & & \ddots & \\ 0 & 0 & & & T^{N,N} \end{pmatrix}, \quad (6)$$

where the block $T^{k,l}$ [of sizes $\binom{N}{l} \times \binom{N}{k}$] concerns all the processes which lead from a state at time t with k configurations to a state with $l \leq k$ configurations at time $t+1$. The upper left block of this matrix, $T^{1,1}$, is the original matrix from Eq. (5). As an example, we find from Eq. (3) the following elements of this transfer matrix:

$$T^{fw}\{|\circ \circ \bullet \circ \bullet\rangle \rightarrow |\circ \bullet \circ \bullet \circ\rangle\} = 1/9,$$

$$T^{fw}\{|\circ \circ \bullet \circ \bullet\rangle \rightarrow |\circ \circ \bullet \circ \bullet\rangle\} = 1/9,$$

$$T^{fw}\{|\circ \circ \bullet \bullet \bullet\rangle \rightarrow |\circ \circ \circ \bullet \circ\rangle\} = 1/27,$$

etc. The matrix T^{fw} describes a physical system with variable particle number (from 1 to N) and a space comprising 2^N-1 states, the number of nonempty states in this new simulation [for a problem of N spins, the number of configurations is 2^N and the total number of k-configuration states (states with k configurations) is $2^{2^N}-1$].

The "forward" transfer matrix T^{fw} allows us to compute the coupling probabilities as a function of time in Fig. 3. The matrix T^{fw} is block-triangular in the number of particles (k,l), with the (1,1) block given by $T^{1,1}$. Therefore, all the
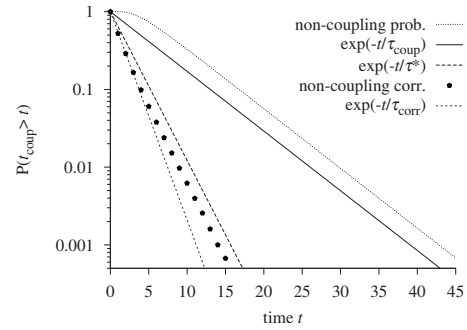

FIG. 3. The exact probability that the Markov chains have not coupled by time t (computed by repeated application of the forward transfer matrix), compared to the time-scales τ_{corr}, τ_{coup}, and τ^* (one-dimensional diffusion model with $N=5$).

eigenvalues of $T^{1,1}$ are also eigenvalues of T^{fw}. In particular, the largest eigenvalue of T^{fw} is again $\lambda_1^{fw}=1$, with corresponding right eigenvector $\frac{1}{5}(|\bullet\circ\circ\circ\circ\rangle+\cdots+|\circ\circ\circ\circ\bullet\rangle)$. The second-largest eigenvalue of T^{fw} belongs to the (2,2) block and leads to the time scale of the coupling, τ_{coup}. It is given by $\lambda_2^{fw}=0.838$, larger than $\lambda^{MC}\equiv\lambda_2^{1,1}$. This second-largest eigenvalue λ_2^{fw} governs the coupling probability $\mathcal{P}(t_{coup})$ for large times. It follows from the block-triangular form of the forward transfer matrix T^{fw} that the time scales satisfy $\tau_{coup} \geq \tau_{corr}$. A general argument allows us to better understand this result: for any running time t, we may separate all the Markov chains into those that have already coupled and those that have not. Only the noncoupling chains [whose number vanishes as $\exp(-t/\tau_{coup})$] contribute to connected correlation functions

$$\exp(-t/\tau_{corr}) \propto \langle \mathcal{O}(t)\mathcal{O}(0) \rangle$$
$$= \sum_{config.\sigma_t,\sigma_0} \mathcal{P}(t_{coup} > t,\sigma_t,\sigma_0)\mathcal{O}(\sigma_0)\mathcal{O}(\sigma_t)$$
$$+ \sum_{config.\sigma_0} \pi^0(\sigma_0)\mathcal{O}(\sigma_0) \sum_{config.\sigma_t} \mathcal{P}(t_{coup} \leq t,\sigma_t)\mathcal{O}(\sigma_t).$$

Here, $\mathcal{O}$ is an observable whose mean value is zero and σ_t the configuration of the system at time t. For the chains which have coupled by the time t, σ_t does not depend on σ_0, and the contribution to the correlation function vanishes. For the other chains, we find

$$\exp(-t/\tau_{corr}) = \sum_{config.\sigma_t,\sigma_0} \mathcal{P}(t_{coup} > t,\sigma_t,\sigma_0)\mathcal{O}(\sigma_0)\mathcal{O}(\sigma_t)$$
$$\propto \exp(-t/\tau_{coup})\exp(-t/\tau^*), \quad (7)$$

where we suppose that even the noncoupling chains converge toward equilibrium on a time scale τ^*, and use that the probability for a chain not to have coupled behaves as $\exp(-t/\tau_{coup})$ in the long-time limit. Equation (7) shows that the difference between τ_{coup} and τ_{corr} is caused by the convergence taking place within noncoupling chains.
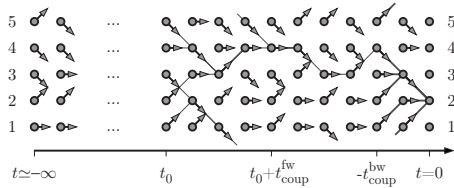
FIG. 4. Extended simulation on $N=5$ sites. The outcome of this simulation, from $t=-\infty$ up to $t=0$, is $k=2$. It can be obtained by backtracking from time $t=0$ to $-t_{\mathrm{coup}}^{\mathrm{bw}}$, or by forward simulation from any $t_0 \leq -t_{\mathrm{coup}}^{\mathrm{bw}}$, through the indicated trajectories. Backtracking from sites 1, 3, 4, and 5 leads to dead ends.

$$\frac{1}{\tau_{\mathrm{corr}}} = \frac{1}{\tau_{\mathrm{coup}}} + \frac{1}{\tau^*}. \tag{8}$$

This relation is illustrated in Fig. 3 for the one-dimensional diffusion model with $N=5$. A later figure, Fig. 8, will illustrate for the case of spin glasses the split between the general spin–spin correlation function and that same object computed for noncoupling chains only.

B. Forward and backward coupling

The probability distribution of coupling times in the forward direction can be obtained from the transfer matrix T^{fw} as we discussed in Sec. II A. Here, we analyze the distribution of coupling times in the backward direction for the application of the "coupling from the past" protocol which, as we will see, is the same as the one in the forward direction. The backward coupling process leads to a generalized transfer matrix, T^{bw}, which again describes an extended Monte Carlo simulation.

We consider a hypothetical simulation which has run since time $t=-\infty$ up to time $t=0$ (see Fig. 4). It follows from the discussion in Sec. II A that the simulation has coupled. Furthermore, because of the infinite separation between the infinitely remote initial condition and the final one, the resulting configuration (at $t=0$) is in equilibrium. But it remains to be seen which one of the five configurations at $t=0$ is generated. In the example of Fig. 4, a one-step backtrack to time $t=-1$ allows us to see that the configuration at $t=0$ can be neither $k=1$ nor $k=3$ nor $k=5$. Likewise, for $N-1$ output positions on the N-site ring this back-propagation leads to a dead end, and only a single position yields a full set of possibilities at some time $-t_{\mathrm{coup}}^{\mathrm{bw}}$. Thus, to find the output configuration of the simulation, one is interested in the first time in the past for which the simulation couples, that is one searches the "backward" coupling time. The implementation of this backward simulation, as defined for many particles, can again be described by a transfer matrix. For any distribution of arrows, any occupied site k at time t propagates its occupation back to all the sites at time $t-1$ which have arrows pointing toward k. The matrix element of T^{bw} between two states is given by the statistical weights of all the arrows connecting the two states. For example, we find for T^{bw}

$$T^{\mathrm{bw}}\{|\circ \circ \bullet \bullet \circ \circ\rangle \rightarrow |\circ \bullet \bullet \circ \circ\rangle\} = 8/81,$$

$$T^{\mathrm{bw}}\{|\bullet \bullet \circ \bullet \bullet\rangle \rightarrow |\bullet \circ \circ \bullet \bullet\rangle\} = 8/81,$$

$$T^{\mathrm{bw}}\{\underset{\text{time } t}{|\circ \circ \bullet \circ \bullet\rangle} \rightarrow \underset{\text{time } t+1}{|\circ \circ \circ \bullet \circ\rangle}\} = 2/27. \tag{9}$$

This is a nontrivial variant of the forward simulation as, for example, the matrix T^{bw} is not block-triangular as T^{fw}, but it is particle-hole symmetric (as we see in the above example).

Formally, a random-map f (here a set of arrows) associates configurations which are connected under the Monte Carlo dynamics. A k-configuration state $|x\rangle$ is by definition a set of configurations. At time t the state $|x\rangle$ can be associated with the state $|y\rangle$ at time $t+1$ via the forward matrix if and only if $|y\rangle$ is the (set) image of $|x\rangle$ by an allowed mapping f [i.e., $|y\rangle=f(|x\rangle)$]. The same holds for the backward matrix but $|x\rangle$ must be the reciprocal (set) image of $|y\rangle$ [i.e., $|x\rangle=f^{-1}(|y\rangle)$]. The backward transfer matrix T^{bw} manifestly differs from the forward matrix T^{fw}. However, we construct explicitly in Appendix A the similarity transform that maps T^{fw} onto T^{bw}. This means that

$$PT^{\mathrm{fw}} = T^{\mathrm{bw}}P$$

(see Appendix A). The similarity transformation P associates a k-configuration state $|x\rangle$ with the sum of states that share at least one configuration with $|x\rangle$, included itself. The spectrum of the backward transfer matrix thus agrees with the one of the forward matrix and the distribution of coupling times $\mathcal{P}(t_{\mathrm{coup}})$ is identical for backward and forward dynamics. This result is natural because the probabilities for not coupling for t time steps are identical in both forward and backward direction: $\mathcal{P}(t_{\mathrm{coup}}^{\mathrm{bw}} > t) = \mathcal{P}(t_{\mathrm{coup}}^{\mathrm{fw}} > t)$ (see [9]). The probability distribution $\mathcal{P}(t=t_{\mathrm{coup}}^{\mathrm{bw}})$ measures the weight of the configuration $|\bullet \bullet \bullet \bullet \bullet\rangle$ under repeated application of the backward transfer matrix from the configuration at $t=0$: $|\bullet \circ \circ \circ \circ\rangle + \cdots + |\circ \circ \circ \circ \bullet\rangle$.

C. Choice of random maps

Transition probabilities of the forward transfer matrix must satisfy the Markov chain transition probabilities for single particles, but the choice of random maps is otherwise unrestricted. The one-particle sector is trivially correct for independent moves as in our diffusion model of Sec. II A. We now discuss several alternative random maps for the one-dimensional diffusion, which may lower (or increase) the coupling time (with, however $\tau_{\mathrm{coup}} \geq \tau_{\mathrm{corr}}$) or achieve a rapid reduction in the number of configurations for smaller time scales.

A naive example for the one-dimensional diffusion example consists of arrows, such as in Fig. 2, but which for any given time t all point into the same direction, straight, up, and down, each with probability 1/3, so that single-particle moves satisfy the detailed balance condition. Evidently, this random map does not couple, and the noncoupling Markov chains, in Eq. (8), converge in a time $\tau^* = \tau_{\mathrm{corr}}$. We now modify this rigid algorithm by allowing arrows to change direction with probability ϵ. This makes the Markov chain couple on a time scale $\sim \log \frac{1}{\epsilon}$, much larger than the correlation time τ_{corr}, for small ϵ. The choice of independent random moves ($\epsilon=1$) is optimal in this class of maps, but it is
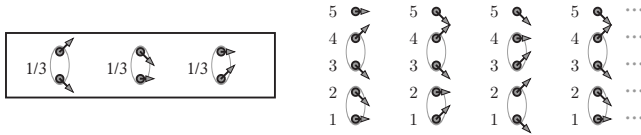
FIG. 5. Left: transition probabilities for pairs (1,2) and (3,4) for the correlated random map. Right: in the extended Monte Carlo simulation shown, the one-particle transition probabilities are as in Fig. 1, but nearest-neighbor coupling is favored.

not the choice minimizing τ_{coup} among all random maps. For example, we may choose correlated moves for selected neighboring pairs of sites say, for sites (1,2) and (3,4) and let the move from site 5 be independent (see Fig. 5).

Elements of $T^{fw,corr}$ are, for example,

$$T^{fw,corr}\{|\circ \circ \bullet \bullet \circ\rangle \rightarrow |\circ \circ \bullet \bullet \circ\rangle\} = 0,$$

$$T^{fw,corr}\{|\circ \circ \bullet \bullet \circ\rangle \rightarrow |\circ \circ \bullet \circ \circ\rangle\} = 1/3,$$

$$T^{fw,corr}\{|\circ \circ \bullet \bullet \circ\rangle \rightarrow |\circ \bullet \circ \circ \bullet\rangle\} = 1/3,$$

$$T^{fw,corr}\{|\circ \circ \bullet \bullet \circ\rangle \rightarrow |\circ \circ \bullet \circ \bullet\rangle\} = 0.$$

The single-particle sector of this algorithm is as before, but the second-largest eigenvalue of the transfer matrix T^{fw} with such correlated pair moves becomes smaller, indicating faster coupling

$$\lambda_2^{1,1} < \lambda_2^{fw,corr} = 0.777 < \lambda_2^{fw,indep} = 0.838.$$

Coupling times of both "independent-arrow" random mapping and "correlated-pair" random-mapping scale alike for large N. We note that in applications, as in our patch algorithm of Sec. V, it might be not so much of interest to speed up the coupling than to rapidly decrease the number of possible configurations at times $t < \tau_{coup}$. Therefore, one goal could be to decrease the eigenvalues of the matrix $T^{k,k}$, $k \gg 1$, whose time scales correspond to the rapid reduction in the number of configurations toward more manageable numbers.

D. Exact sampling, coupling from the past

As discussed in Sec. II B, the coupling of Markov chains allows one to produce exact samples of the equilibrium distribution: in the diffusion example, we were able to run the Monte Carlo simulation backward in time using T^{bw}, but this matrix can usually not be constructed. To find the sample at time $t=0$, one may tentatively set a time $t_0 < t$ and produce all the random maps between time t_0 and t. One can then check explicitly whether all the possible initial conditions at time t_0 have coupled, that is, for the diffusion problem, whether the initial N-particle configuration $|\bullet\bullet\bullet\bullet\bullet\rangle$ has yielded one of the one-particle configurations. If this goal has not been reached, one must complement the random maps already computed with random maps for earlier times (see Fig. 4). The one-dimensional diffusion problem without periodic boundary conditions illustrates an algorithm which determines the coupling time with much less effort. We con-
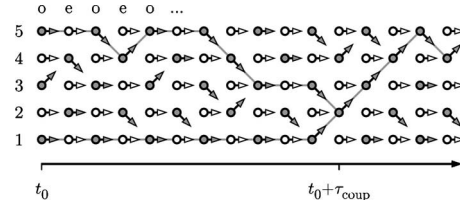

FIG. 6. Extended Monte Carlo simulation with odd (o) and even (e) time steps on a lattice without periodic boundary conditions. Trajectories cannot cross, and the coupling of the two extremal initial configurations (the simulations starting at time t_0 from sites 1 and 5) determines the coupling time.

sider odd times at which only sites 1, 3, and 5 and even time steps at which only sites 2 and 4 may flip (see Fig. 6). This preserves the correct stationary probability distribution, but the trajectories no longer cross each other (as at time t_0+1 in Fig. 2). As a consequence, it suffices to follow the two extremal configurations, which start at sites $k=1$ and $k=N$, from time t_0 on in order to determine the coupling time for a given full Monte Carlo simulation. The multiple-particle Monte Carlo simulation starts with the state $|\bullet\circ\circ\circ\bullet\rangle$ until it yields a single-particle state. The above strategy of following extremal configurations can be applied to the ferromagnetic Ising model (but not to spin glasses, see [2,5]). In this case, the two configurations with all spins up and all spins down, respectively, are extremal. This idea also holds for the heat-bath algorithm of two-dimensional directed polymers in a random medium.

III. COUPLING AND CONVERGENCE IN SPIN MODELS

We study the Edwards-Anderson $\pm J$ Ising spin glass on a d-dimensional square lattice. The energy of a configuration $\sigma = (\sigma_1, \ldots, \sigma_N)$ with $\sigma_k = \pm 1$ is

$$\mathcal{H}(\sigma) = -\sum_{\langle i,j \rangle} J_{ij}\sigma_i\sigma_j,$$

where $\langle i,j \rangle$ indicates the sum over the $2d$ nearest neighbors. This model has a phase transition at finite temperature in three dimensions and at zero temperature in two dimensions. Sampling spin-glass configurations with Markov chain algorithms is extremely difficult in $d=3$ dimensions below the critical temperature, but it is also nontrivial in the two-dimensional case [13]. We concentrate here on the study of a local heat-bath Monte Carlo algorithm, for which we apply the coupling-from-the-past protocol and obtain exact samples. We note that in two dimensions, the exact partition function of the Ising model on a finite lattice can be determined exactly for any choice of couplings [2,14]. This makes possible a direct-sampling algorithm, which is completely unrelated to the material presented here, but which we sketch, for the sake of completeness, in Appendix B.

The heat-bath Monte Carlo algorithm for spin models updates at each time step a randomly chosen site i of a spin configuration by comparing a function of the local field on site i with a uniform random number $\Upsilon_i(t) = \text{ran}[0,1]$
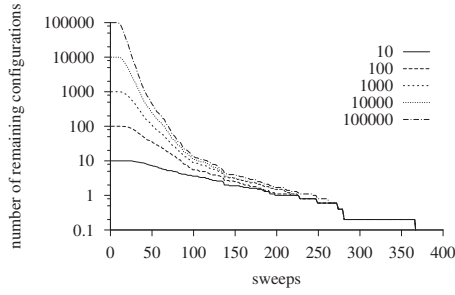
FIG. 7. Number of configurations of the partial-survey approximation with $\mathcal{N}(t_0)$ random initial configurations for the Ising spin glass on a 16×16 lattice at temperature $\beta = 0.5$. We average over 10 choices of J_{ij}, and use the same values of J_{ij} and the same random numbers Υ_i, i for all initial configurations.

$$\sigma_i(t+1) = \begin{cases} 1 & \text{if } \Upsilon_i(t) \leq [1 + e^{-2\beta h_i(t)}]^{-1} \\ -1 & \text{else} \end{cases}, \quad (10)$$

where $h_i(t) = \sum_j J_{ij}\sigma_j(t)$ is the local field on site i. A realization of the Markov chain corresponds to sampling the real-valued random numbers $\{\ldots, \Upsilon(t_0), \Upsilon(t_0+1), \ldots, \Upsilon(-1)\}$ and the random integers $\{\ldots, i(t_0), i(t_0+1), \ldots, i(-1)\}$. The unit of "physical" time (one "sweep") corresponds to N individual updates. The situation is now much more complicated than for the $1d$ diffusion, as the role of the five initial configurations in Fig. 2 is taken up by the 2^N possible spin configurations. To prove coupling one must show to which configuration they all converge at the coupling time. The state space is huge and one must find strategies to avoid enumerating and surveying 2^N configurations.

A. Partial-survey approximation

In [6], we presented an exact-sampling algorithm which works down to quite low temperatures in the two-dimensional Ising spin glass, and which is also operational in three dimensions. We found that practically the same results could be obtained by starting the simulation at time t_0 not from all the 2^N initial configurations, but from a more manageable number $\mathcal{N}(t_0)$ of randomly chosen configurations. We show in Fig. 7 that such "partial-survey" calculations yield useful lower bounds for the coupling time scale τ_{coup}. Each curve in the figure represents the mean number of distinct configurations remaining after coupled Monte Carlo simulations [that is, with the same random numbers (Υ, i) for all configurations] for different values of $\mathcal{N}(t_0)$. Increasing $\mathcal{N}(t_0)$ within this partial-survey approximation naturally improves the lower bound on the coupling time but, in practice, the value obtained saturates quite quickly.

B. Correlation functions

As discussed previously, τ_{coup} is always larger than τ_{corr} because only noncoupling chains contribute to correlation functions [see Eq. (8)]. To again illustrate the relation between coupling and convergence times, we separate in Fig. 8 noncoupling chains from the calculation of the spin-spin correlation function of a 8×8 spin glass at inverse temperature
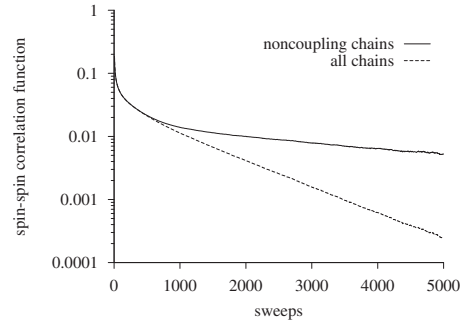

FIG. 8. Spin-spin correlation function for all Markov chains and for the noncoupling Markov chains only (with random initial conditions of the two-dimensional 8×8 spin glass at $\beta = 1$).

$\beta = 1$. Indeed, even if the chain has not coupled, the configurations σ_t may lose the dependence on the initial configuration σ_0.

We shall continue our study of spin glasses in Sec. V with an algorithm which rigorously establishes coupling down to a constant temperature for large systems.

IV. COUPLING FOR HARD-SPHERE SYSTEMS

In this section, we discuss the application of the "coupling from the past" protocol to hard-sphere systems. The study of Monte Carlo algorithms for hard-sphere systems goes back a long time, as the Metropolis algorithm was first implemented for hard disks, that is, two-dimensional spheres [1]. Even today, the physics of the hard-disk system is not well understood, and Monte Carlo algorithms have not been developed as successfully as, say, for the Ising model. In this very constrained system, the estimation of correlation times is quite controversial, especially at high densities [15], and rigorous results from exact-sampling approaches would be extremely welcome.

We first discuss the birth-death formulation of the Markov chain Monte Carlo algorithm for this system and then compute lower bounds on coupling times using the partial-survey algorithm. Its empirical coupling time saturates [for increasing $\mathcal{N}(t_0)$] to much smaller values than the coupling times obtained by the summary-state method [9,10]. This suggests that these previous algorithms are not optimal, an impression which is confirmed by our local-patch algorithm of Sec. V E.

The partition function of hard spheres in the grand-canonical ensemble, with fugacity λ, is given by a weighted sum over legal configurations of spheres

$$\mathcal{Z} = \sum_{N=0}^{+\infty} \int d^{2N}\sigma^{(N)} \lambda^N \Theta(\sigma^{(N)}). \quad (11)$$

Here, configurations of N spheres are written as

$$\sigma^{(N)} = \{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\},$$

where (x_k, y_k) denotes the centers of the spheres. In Eq. (11), $\Theta(\sigma^{(N)})$ equals one if spheres of the configuration $\sigma^{(N)}$ do not overlap and zero otherwise. We again use periodic boundary conditions.
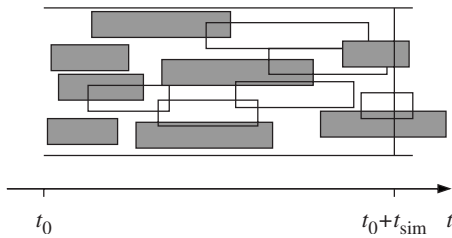
FIG. 9. Simulation of the birth-death algorithm for one-dimensional "spheres" in a box. The simulation starts at time t_0 and stops at t_0+t_{sim} with $N=2$ spheres. Transparent spheres are rejected because they overlap with spheres already present.

A. Birth-death algorithm for hard spheres

The spatial Poisson birth-death process allows us to apply coupling from the past to hard-sphere systems (see [9,10]): disks of radius r arrive ("are born") randomly on a two-dimensional unit square with constant rate λ. This transition from N to $N+1$ particles is described, in Eq. (12), by the probability $\mathcal{P}(\sigma^{(N)} \to \sigma^{(N+1)})$. Once born, they disappear ("die") with unit rate.

The probability for a disk to arrive within an infinitesimal time dt in a small box of area dS centered at (x,y) is $\lambda dt dS$. This disk is added to the configuration only if it overlaps with no other disk present. Each disk disappears with probability dt within the time interval dt. Sphere are added at point (x,y) or removed from the configuration $\sigma^{(N)}$ according to the detailed balance condition. With the notation $\sigma^{(N+1)} = \sigma^{(N)} \cup \{(x,y)\}$, we have

$$\mathcal{P}(\sigma^{(N)} \to \sigma^{(N+1)})\pi(\sigma^{(N)}) = \lambda\Theta(\sigma^{(N+1)})\pi(\sigma^{(N)}) = \pi(\sigma^{(N+1)}),$$

$$\mathcal{P}(\sigma^{(N+1)} \to \sigma^{(N)})\pi(\sigma^{(N+1)}) = 1 \times \pi(\sigma^{(N+1)}). \quad (12)$$

In Fig. 9, we illustrate the time evolution of accepted and rejected birth-death events on a one-dimensional hard-sphere problem, starting from an empty initial condition at time t_0.

In the hard-sphere algorithm, the probability distribution of time intervals between successive births is an exponential with parameter λ: $\mathcal{P}(\tau_b) = \lambda e^{-\lambda\tau_b}$. In Fig. 9, the lifetime of a sphere is represented by a horizontal extension of the box, irrespective of whether it has been accepted or not (the vertical dimension denotes the diameter). Lifetimes are exponentially distributed as well. For the exponential distribution, the time before the next death of a system of N spheres follows an exponential distribution with parameter N. Likewise the time before any event, (birth or death), follows an exponential distribution with parameter $\lambda+N$. The probability for the next event to be a birth is then $\frac{\lambda}{\lambda+N}$.

B. Coupling and partial-survey approximation

Coupling from the past applies to hard-sphere systems even though the space of configurations is continuous (unlike in lattice simulations). To apply the protocol, one considers a time evolution, as in Fig. 9, but stretching back to time $t=-\infty$. Two special aspects must now be handled.

First, we must determine which boxes (corresponding to spheres) are indeed placed ("True"), and which ones are re-
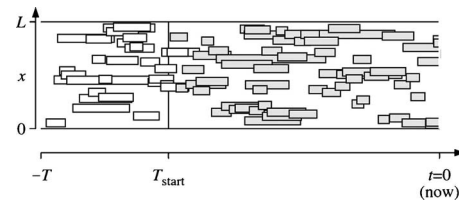

FIG. 10. Time evolution of a one-dimensional birth-death simulation in a box of size L. All spheres correspond to rectangles whose horizontal extension indicates their life time. From any possible cut of True boxes at T_{start} (seven boxes actually cut the line, so there are $\leq 2^7$ possible cuts) one can deduce the output at time $t=0$, as in Fig. 9.

jected ("False"). This is difficult to decide at high density. However, in the low-density case presented in Fig. 10, several spheres are True, simply because they do not overlap with already present True or False spheres. This allows the status of other spheres to be fixed and, finally, the configuration to be constructed. In the limit $N \to \infty$, the approach works up to a constant density [16]. This density is much higher than the density $\propto 1/N$ direct-sampling algorithm can achieve [2]. This approach [9,10] is equivalent to deciding whether a given spin is up or down in the "summary-state" algorithm for Ising systems [7,8], which, in the thermodynamic limit works down to a fixed constant temperature.

Second, one must fix the initial condition at time $-T$, because spheres born at times smaller than $-T$ may still be alive at time $-T$. This is solved through the sampling of a second time, T_{start}, after which we know that all spheres present at time $-T$ have disappeared. The time interval $T_{\text{start}}+T$ is sampled as the maximum of N_{max} lifetimes, where N_{max} is an upper bound on the number of spheres in a legal configuration.

Figure 10 sketches the time evolution of a Monte Carlo simulation for the one-dimensional hard-sphere problem, which has started at time $t=-\infty$. Boxes are drawn starting from time $-T$, but the simulation is picked up at time T_{start}. It is straightforward to complement the simulation shown (between times $-2T$ and $-T$, for example), in case it does not couple in the interval shown. However, we must show that it couples between $-2T$ and $-T$ or at least results in less than N_{max} spheres. In the Monte Carlo simulation in Fig. 10, the status of the boxes at later times can be easily decided, because at later times all spheres belong to clusters which are disjoint from the initial condition. However, this possibility disappears at higher densities. A simple example of this is shown in Fig. 11. As one cannot decide on the status of the initial sphere (which crosses the line at T_{start}), we should initialize the simulation with the two configurations, one corresponding to a True state and one to a False state. After several steps of the time evolution, we arrive in both cases at the same physical configuration (the two dark spheres, which are both True).

For all times $t > T_{\text{start}}$, we consider the set $\mathcal{C}(t)$ of all True or False spheres crossing the time line at t (see Fig. 10). From the set $\mathcal{C}$, one can in principle construct all the possible initial configurations, but their number remains huge. As in the spin-glass case, we may also select $\mathcal{N}(T_{\text{start}})$ among these
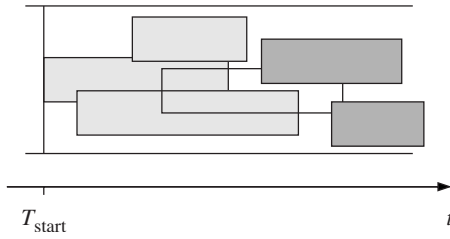
FIG. 11. Example of a time evolution of the one-dimensional birth-death simulation, which where no single sphere can be decided independently. Starting with all possible choices of the initial configuration at $t = T_{\text{start}}$ allows to prove coupling. (The two dark spheres are True, while the transparent sphere must be False.)

configurations, and propagate these. This is again the partial-survey approximation. In Fig. 12, we compare average lower bounds on the coupling time from this approximation with results from the summary-state algorithm [9]. In the time evolution of Fig. 10, one can determine the number of remaining configurations at any time $t > T_{\text{start}}$ and detect when exactly the coupling occurs.

V. LOCAL-PATCH ALGORITHM

In the present section, we discuss our local-patch algorithm, which performs the heat-bath dynamics for a general d-dimensional Ising spin glass on an N-site hypercubic lattice. This algorithm allows us to control all the 2^N initial configurations even for very large lattices and to eventually prove that the system has coupled. The PYTHON script implementing this algorithm has less than 300 lines. It is available electronically [17].

A. Patches

The (nonrigorous) partial-survey algorithm of Sec. III A determines the coupling time for a subset of all the configurations at time t. The (rigorous) patch algorithm, in contrast, works with a superset of all configurations at time t: by restricting the configurations to the smaller region of a patch, one severely limits their number, at the price of introducing
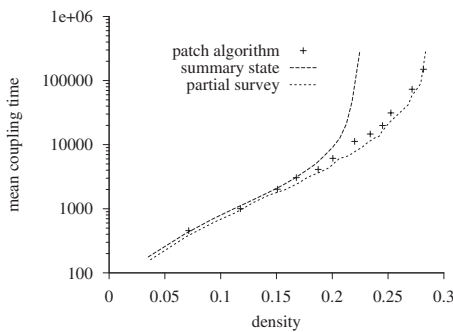

FIG. 12. Coupling times of the summary-state algorithm for hard disks (Refs. [9,10]) compared to the local-patch algorithm. Lower bounds are provided by the partial-survey approximation. The disks' radius is $r = 0.04$ in a unit square box with periodic boundary conditions, so that there are ~ 60 disks at density $\eta = 0.3$.
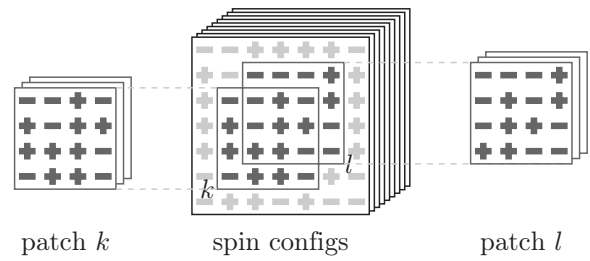

FIG. 13. Two patches, k and l, with a pair of compatible spin configurations.

compatibility problems between neighboring patches. For the two-dimensional spin glass, we use N rectangular patches of same shape and orientation, with M sites, and initially at $t = t_0$, we have 2^M spin configurations on each patch. Likewise, the set of global spin configurations is broken up into a list $[S_1(t), \ldots, S_N(t)]$ of sets $S_k(t)$ of spin configurations restricted to patches k. We can recover a superset $\Omega(t)$ of all relevant spin configurations from the direct product

$$\Omega(t) = S_1(t) \otimes S_2(t) \otimes \cdots \otimes S_N(t)/(\text{compat}). \quad (13)$$

Here, each configuration of $\Omega(t)$ is pieced together from configurations on all patches, with compatible spins on all lattice sites. (Two compatible spin configurations, on patches k and l, are shown in Fig. 13.) On large lattices, the direct product in Eq. (13) can be performed only if the number of spin configurations per patch is small. If there is only one configuration per patch, we can construct a unique global configuration on the whole lattice.

For each time step t of the heat-bath algorithm, we choose a random lattice site i and a random number $\Upsilon = \text{ran}[0, 1]$ and then update the spin σ_i for all configurations on all patches containing i. The site i may be in the center of a patch k (all the neighbors of site i also belong to k, see Fig. 13). In this case, each configuration of $S_k(t)$ yields one configuration of $S_k(t+1)$. Several configurations in $S_k(t)$ may yield the same configuration in $S_k(t+1)$, so that the size of S_k does not increase in this case. If the site i is on the boundary of a patch l, we only know upper and lower bounds for the field on the site i and, depending on the value of the random number Υ, may be unable to update σ_i. In this case, we add two configurations to the set $S_k(t+1)$, corresponding to $\sigma_i = -1$ as well as $\sigma_i = +1$. The set $S_k(t+1)$ may then contain more configurations than $S_k(t)$.

B. Compatibilities, pruning

Besides updating configurations on patches, we also perform a "pruning" operation: Fig. 13 presents two "compatible" configurations on patches k and l. These could possibly be pieced together into a global configuration, together with configurations on other patches. On the other hand, if the set S_l contains no configuration compatible with a configuration $\boldsymbol{\sigma}$ on patch k, we can eliminate (prune) $\boldsymbol{\sigma}$ from S_k. Pruning may be implemented through a dictionary (hash table), using as "key" the part of the patch configuration in the overlap region between k and l, and as "value" the list of patch configurations sharing this key (see Fig. 14). This is pro-
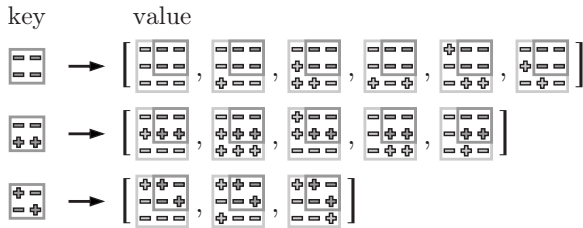
FIG. 14. A dictionary (hash table) associating keys (configurations in the overlap region between patches k and l) to values (lists of patch configurations with the given key in patch k).

grammed very easily in the PYTHON programming language [17].

Pruning can be iterated until all the sets $S_k(t)$ are pairwise compatible. To achieve this goal, it suffices to prune nearest-neighbor patches only. We have found it useful to perform one pruning operation for each pair of nearest-neighbor patches after a certain number of updates (see Fig. 15). The average number of configurations per patch saturates to a value that depends on the temperature and also the size of the patch. This is due to the balance between the decrease in the number of configurations induced by the coupling and its increase caused by the noise at the patch boundaries. This noise is reduced through the crucial pruning step of the algorithm.

C. Merging of patches

As illustrated in Fig. 15, the number of configurations per patch does not necessarily drop to one at large times, even if the underlying heat-bath dynamics couples. The entropy per spin is smaller for larger patches, because the influence of the boundaries is reduced. However, one cannot start the computation with large patches because of the large number of possible configurations. A merging procedure allows to increase the patch size in a rigorous way. Merging is implemented analogously to pruning: for overlapping patches k and l, dictionaries are again computed with the same keys

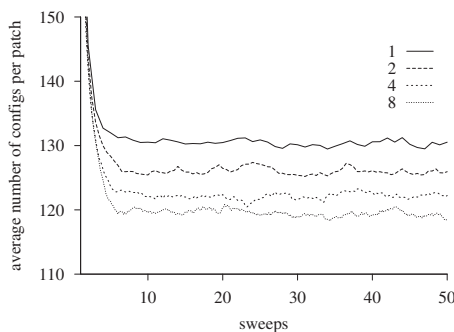

FIG. 15. Average number of configurations per patch vs time, at inverse temperature $\beta=1$ in the two-dimensional Ising spin glass on a 16×16 lattice at temperature $\beta=1$. For a given patch size (here, 3×3), the number of configurations per patch saturates to a value that depends on the number of prunings per sweep (here: from one to eight). Each pruning is done once for all nearest-neighbor patches.
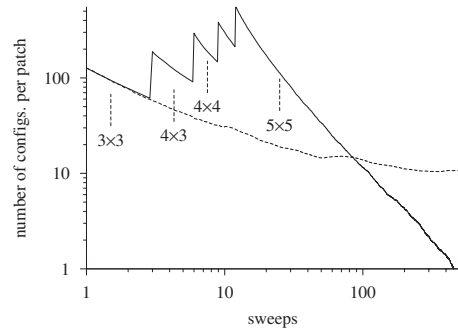

FIG. 16. Number of configurations per patch for the 32×32 spin glass at temperature $\beta=0.5$. A simulation with constant 3×3 patches is compared to the result of a jump-start procedure $3\times3\rightarrow4\times3\rightarrow4\times4\rightarrow,\ldots,\rightarrow5\times5$. All results are averaged over 100 samples.

and values (see again Fig. 13). For a given key configuration on the overlap region, we assemble the corresponding values in the dictionary of patch k with all corresponding values in the dictionary of patch l. All these couples of configuration must be taken into account for the larger patch $k\cup l$. The merging of the configurations on neighboring patches can be implemented very efficiently in the PYTHON programming language [17]. In our computations, we start with small square patches, say of size 3×3, and then pass to the size 4×3, after a few sweeps, then to size 4×4, etc. An analogous procedure is followed in higher dimensions. Results obtained with this "jump-start" approach are shown in Fig. 16 for the two-dimensional Ising spin glass at temperature $\beta=0.5$, with a disorder average performed over about 100 samples.

D. Memory of compatibilities, variants

In the patch algorithm, the pruning procedure detects inconsistent configurations in a particular patch. Two configurations on different patches are considered compatible if their spins match in the overlap region (at time t). More generally, we can keep track of the past evolution of patch configurations and may then declare them compatible only if they have matched for all times up to t. Otherwise, they cannot belong to a unique global spin configuration.

In [6], the bipartite nature of the square lattice was used to update one entire sublattice at a time. In this approach, only one sublattice is stored at a time. This allows one to start with larger patches, but the compatibilities between configurations are less well conserved. By contrast, in the present algorithm, we keep the information on both sublattices, and one of the sublattices is the past configuration. Two configurations are compatible in this modified version if they are compatible on both sub-lattices. Likewise one can use past values of a configuration $\sigma(t)$ to restrict its compatibilities with configurations on other patches.

Other generalizations are more straightforward, one can, for example, optimize the shape of patches in order to minimize the number of spins on the boundary, and work with more than N patches in order to increase the chance for detecting incompatibilities of configurations.
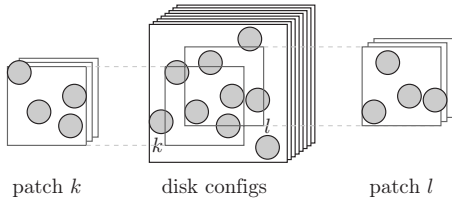
FIG. 17. Breaking up disk configurations into patches, with two patches k and l shown.

E. Exact sampling for hard spheres with local patches

In this section, we adapt the local-patch algorithm for the classical model of hard disks in a two-dimensional box with periodic boundary conditions. As mentioned, it works for large system sizes at higher density than previous method of exact sampling.

In Sec. IV B, we introduced the set $\mathcal{C}(t)$ of all disks the dynamics has tried to add in the box and which have not disappeared at time t. The coupled Monte Carlo simulations start with all possible configurations that are allowed by the set $\mathcal{C}$ at time T_{start}. Because of possible overlap between disks some of the $2^{\#\mathcal{C}}$ possible configurations are invalid but there may be far too many of them in practice. To reduce the number of configurations which must be handled, we introduce a regular square lattice with N sites covering the simulation box. Likewise, a superset of all feasible configurations on a patch at time T_{start} can be deduced from the set $\mathcal{C}$, restricted to disks (True or False) with centers inside the patch (see Fig. 17). From then on, whenever disks appear in the simulation box, we can decide whether they are accepted on a particular patch configuration by checking overlaps inside the patch only. Disks that disappear are simply removed from all the concerned patch configurations. At birth time, if the disk to be placed on a patch configuration may overlap with a disk outside the patch, the configuration is split into two: one configuration with the new disk and one without (as for spin systems). To detect and prune irrelevant configurations we check that the updated sets of configurations are compatible with other patches by a pruning procedure analogous to the one of Sec. V B. After several updates, the pruning is performed for most of overlapping patches several times.

For any pair of overlapping patches k and l, the pruning eliminates patch configurations with are inconsistent with all other configurations on a neighboring patch. As in the spin-glass case, this process can be implemented with dictionaries (hash tables) (see Fig. 18) and can be used to merge configurations on neighboring patches into larger local configurations.

Figure 12, which was first discussed in Sec. IV B, displays the mean coupling times of a hard-disk birth-death simulation for several choices of the fugacity λ. The radius of the disks is $r = 0.04$ in a unit square box with periodic boundary conditions. Results are compared to the partial-survey approximation algorithm, with $\mathcal{N}(T_{\text{start}}) = 1000$ initial configurations and to the results of the summary-state algorithm. We note that the lower bound on the coupling time (obtained by the partial-survey approximation) is just below
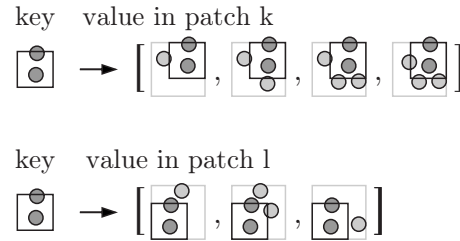

FIG. 18. The pruning of a pair of patches via the construction of dictionaries. Each dictionary associates keys (configurations in the overlap region) to values (lists of patch configurations with the given key) (compare with Fig. 14). The merging of patches k and l would lead in this example to $4 \times 3 = 12$ configurations on the combined patch.

its upper bound from the local-patch algorithm. This implies that we have indeed practically computed the true coupling time of the algorithm.

We concentrated on determining the coupling times of the birth–death dynamics for hard disks. However, the regime of operation of this algorithm is far in the liquid phase (see Fig. 12), and the physically interesting regime, around the liquid-solid transition density, $\eta \simeq 0.71$ [2] is still out of reach for exact-sampling methods. For hard disks, it remains a challenge to set up a working partial-survey algorithm with correlation times comparable to those of the usual Metropolis algorithm [15].

VI. CONCLUSION

In this paper, we have discussed exact-sampling algorithms that allow one to totally eliminate the influence of the initial condition from a Markov chain Monte Carlo simulation. This overcomes one of the main limitations of the method, namely, the rigorous estimation of the correlation time. We discussed central subjects, such as the relation between coupling times and convergence times, in a simple example of one-dimensional diffusion, before applying them to Ising spin glasses and to hard-sphere simulations. Algorithmically, the exact-sampling framework obliges one to follow the entire state space of a system. In the absence of simplifications, such as the half-order discussed in Sec. II, this can be done approximately through a partial survey of $\mathcal{N}(t_0)$ initial conditions. One can also restrict the configuration in size onto so-called patches, thereby restricting their number. A superset of the set of global configurations can in principle be reconstituted from the patches. This is easier when the patches are large, and we showed how pruning and merging operations allow one to increase the size of patches during the simulation and to finally prove coupling. Our exact-sampling algorithm works both for spin glasses and hard-disk systems, and we were able to go to lower temperatures, and higher densities than previous methods.

The partial-survey algorithm, which can be implemented easily, allowed us to prove that our local-patch algorithm is optimal for the local dynamics for both spin-glass and hard-disk systems. We have provided a number of ideas in order to allow exact-sampling methods to reach the phase transi-

tion of the three-dimensional Ising spin glass and the critical density of hard disks.

APPENDIX A: SIMILARITY OF FORWARD AND BACKWARD TRANSFER MATRICES

Forward and backward transfer matrices completely describe the coupling dynamic of general Markov chains (on a finite state space), that is their elements are the coupled transition probabilities between sets of configurations. Forward and backward matrices represent "extended" Monte Carlo dynamics, in the two time directions. These two formulations are equivalent, even though the forward dynamics, starting from $t=0$, does not generate exact samples. Here, we demonstrate similarity between forward and backward transfer matrices, and construct the similarity transformation between the two.

Let Ω be the finite space of configurations of the problem of interest. Ω may contain all N positions on the N-site diffusion problem or the 2^N configurations of a N-site spin system.

The k-configuration states build up an extended state space. They provide a natural basis for the forward and backward matrices. This basis is $2^\Omega - \varnothing$, the set of non trivial parts of Ω. For any state $I \in 2^\Omega - \varnothing$, we define $\tilde{I}$ as the set of states of the basis J that has at least one configuration in common with I ($J \in \tilde{I} \Leftrightarrow I \cap J \neq \varnothing$). The similarity matrix P is then defined as

$$P|I\rangle := \sum_{J \in \tilde{I}} |J\rangle. \tag{A1}$$

A random mapping—arrows for the case of one-dimensional diffusion—is a mapping on $\Omega : f : \Omega \to \Omega$. It defines a time step of the Markov chain for every configuration and satisfies $\mathcal{P}[f(i)=j]=p(i \to j)$ and its weight is noted $w(f)$. In the case of one-dimensional diffusion with independent arrows, or any "independent" random map in general, we naturally define the weight of the random map as a product of elements of the Monte Carlo transfer matrix as

$$w(f) = \Pi_{i \in \Omega} p[i \to f(i)].$$

The forward matrix associates any state I to all states that it is connected to by a mapping

$$T^{\mathrm{fw}}|I\rangle = \sum_{f \mathrm{rand.map}} w(f)|f(I)\rangle.$$

Using Eq. (A1), we find

$$PT^{\mathrm{fw}}|I\rangle = \sum_{f \mathrm{rand.map}} w(f) \sum_{J \in \widetilde{f(I)}} |J\rangle. \tag{A2}$$

The backward matrix T^{bw} has different rules but we will show that the similarity $PT^{\mathrm{fw}}=T^{\mathrm{bw}}P$ holds. Using a random-map f, a state J at time t evolves to another state K at time $t+1$ in the backward process if and only if $f^{-1}(K)=J$. For example, in the one-dimensional diffusion, a hole goes to a hole and a particle goes to a particle. Therefore,
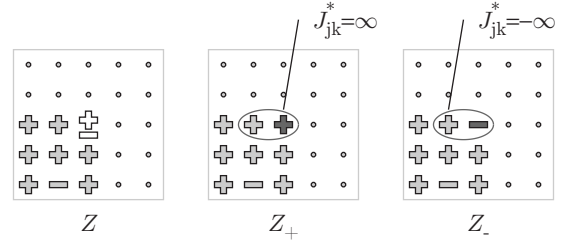

FIG. 19. One iteration in the direct-sampling algorithm for the two-dimensional Ising model. The probabilities $\pi(\sigma_k=+1)$ and $\pi(\sigma_k=-1)$ (with k the central spin) are obtained from the exact solution of the Ising model with fictitious couplings $J^*_{jk}=\pm\infty$.

$$T^{\mathrm{bw}}|J\rangle = \sum_{f \mathrm{rand.map}} w(f) \sum_{K, f^{-1}(K)=J} |K\rangle$$

and finally,

$$T^{\mathrm{bw}}P|I\rangle = \sum_{f \mathrm{rand.map}} w(f) \sum_{J \in \tilde{I}} \sum_{K, f^{-1}(K)=J} |K\rangle$$

$$= \sum_{f \mathrm{rand.map}} w(f) \sum_{K, f^{-1}(K) \in \tilde{I}} |K\rangle. \tag{A3}$$

In fact Eqs. (A3) and (A2) are equivalent because $f^{-1}(K)$ overlaps I if and only if K overlaps $f(I)$ [$f^{-1}(K) \cap I \neq \varnothing \Leftrightarrow K \cap f(I) \neq \varnothing$]. This proves the similarity of the backward and forward matrices.

APPENDIX B: EXACT SAMPLING OF TWO-DIMENSIONAL SPIN GLASS USING ANALYTIC SOLUTION

In this appendix, we sketch for completeness an unrelated direct-sampling algorithm for the two-dimensional Ising spin glass. To generate exact samples, this algorithm does not use Markov chains. It rather relies on the fact that the partition function of the two-dimensional Ising model or of one sample of the spin glass on a planar lattice with N sites can be expressed as the square root of the determinant of one $4N \times 4N$ matrix (for open boundary conditions) or of four such matrices (for periodic boundary conditions) [14]. This relation has been much used in the recent literature, in order to study the physics of the two-dimensional Ising spin glass at low temperature [18,19]. The partition function yields the thermodynamics of the system, but the knowledge of entire configurations gives, for example, access to complicated spatial configuration functions.

The sampling algorithm for two-dimensional spin-glass configurations constructs the sample one site after another. Let us suppose that the gray spins in the left panel of Fig. 19 are already fixed, as shown. We can now set a fictitious coupling J^*_{ll} either to $-\infty$ or to $+\infty$ and recalculate the partition function $Z_\pm$ with both choices. The statistical weight of all configurations in the original partition function with spin "+" is then given by

$$\pi_+ = \frac{Z_+ \exp(\beta J_{kl})}{Z_+ \exp(\beta J_{kl}) + Z_- \exp(-\beta J_{kl})} \tag{B1}$$

and this two-valued distribution can be sampled with one random number. Equation (B1) resembles the heat-bath algorithm of Eq. (10), but it is not part of a Markov chain: after obtaining the value of the spin on site k, we keep the fictitious coupling, and add more sites. Going over all sites, we can generate direct spin-glass samples at any temperature. We note that this algorithm is polynomial, and the effort is basically temperature-independent, both for the two-dimensional Ising model and the Ising spin glass (see also [20]).

APPENDIX C: PYTHON CODE

The PYTHON 2.5 code, which has produced all the spin-glass data in this article, contains less than 300 lines. It has been deposited as supplementary material [17] and may assist in understanding our patch algorithm. An analogous program was used for the hard-disk system.

[1] N. Metropolis *et al.*, J. Chem. Phys. **21**, 1087 (1953).

[2] W. Krauth, *Statistical Mechanics: Algorithms and Computations* (Oxford University Press, Oxford, U.K., 2006).

[3] R. N. Bhatt and A. P. Young, Phys. Rev. B **37**, 5606 (1988).

[4] R. N. Bhatt and A. P. Young, Phys. Rev. Lett. **54**, 924 (1985).

[5] J. G. Propp and D. B. Wilson, Random Struct. Algorithms **9**, 223 (1996).

[6] C. Chanal and W. Krauth, Phys. Rev. Lett. **100**, 060601 (2008).

[7] M. Huber, Random Struct. Algorithms **22**, 43 (2003).

[8] A. M. Childs, R. B. Patterson, and D. J. C. MacKay, Phys. Rev. E **63**, 036113 (2001).

[9] D. B. Wilson, Random Struct. Algorithms **16**, 85 (2000).

[10] W. S. Kendall and J. Møller, Adv. Appl. Probab. **32**, 844 (2000).

[11] D. Bayer and P. Diaconis, Ann. Appl. Probab. **2**, 294 (1992).

[12] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov Chains and Mixing Times* (American Mathematical Society, Providence, 2009).

[13] J. S. Wang and R. H. Swendsen, Phys. Rev. B **38**, 4840 (1988).

[14] L. Saul and M. Kardar, Phys. Rev. E **48**, R3221 (1993).

[15] E. P. Bernard, W. Krauth, and D. B. Wilson, Phys. Rev. E **80**, 056704 (2009).

[16] M. Luby and E. Vigoda, Random Struct. Algorithms **15**, 229 (1999).

[17] See supplementary material at http://link.aps.org/supplemental/10.1103/PhysRevE.81.016705

[18] A. Galluccio, M. Loebl, and J. Vondrak, Phys. Rev. Lett. **84**, 5924 (2000).

[19] J. Lukic, A. Galluccio, E. Marinari, O. C. Martin, and G. Rinaldi, Phys. Rev. Lett. **92**, 117202 (2004).

[20] W. Krauth, *Les Houches Lectures* (Oxford University Press, New York, in press).