Physica A 418 (2015) 88–93

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Monte Carlo methods beyond detailed balance

Raoul D. Schram [a,*], Gerard T. Barkema [b,a]

[a] Instituut-Lorentz, Leiden University, P.O. Box 9506, 2300 RA Leiden, The Netherlands
[b] Institute for Theoretical Physics, Utrecht University, P.O. Box 80195, 3508 TD Utrecht, The Netherlands

HIGHLIGHTS

- We propose a Monte Carlo approach, which is not restricted by detailed balance.
- The high efficiency of this new approach is demonstrated in a simple model.
- Its general applicability is demonstrated by applying it to the Ising model.

ARTICLE INFO

Article history:
Available online 19 June 2014

Keywords:
Monte Carlo simulations
Detailed balance
Ising model

ABSTRACT

Monte Carlo algorithms are nearly always based on the concept of detailed balance and ergodicity. In this paper we focus on algorithms that do not satisfy detailed balance. We introduce a general method for designing non-detailed balance algorithms, starting from a conventional algorithm satisfying detailed balance. This approach is first applied to a very simple model, which shows the basic viability of the method. Then we apply it to the Ising model, where we find that the method is an improvement compared to the standard Metropolis algorithm, be it with a modest gain of a factor 2.3.

1. Introduction

To study equilibrium properties of models in statistical physics by means of computer simulations, Monte Carlo methods are often used (see [1] for a more thorough description). Starting from some initial configuration C_0, the basic idea of most Monte Carlo simulations is to iteratively propose a small random change in a configuration C_i, resulting in the trial configuration C^t_{i+1}. Next, the trial configuration is either accepted, i.e. $C_{i+1} = C^t_{i+1}$, or rejected, i.e. $C_{i+1} = C_i$. The resulting set of configurations for $i = 1 \ldots M$ is known as a Markov chain. If the proposition and acceptance probabilities are well chosen, the probability that a configuration A is sampled by the Markov process (after thermalization) is equal to its Boltzmann weight:

$$P_A \sim \exp(-\beta E_A), \tag{1}$$

in which E_A is the energy of configuration A and β is the inverse temperature, defined by $\beta \equiv 1/(k_B T)$ with temperature T and the Boltzmann constant k_B. A proposed change in the configuration is usually referred to as a Monte Carlo move.

The key question in Monte Carlo algorithms is obviously which small changes one should propose, and how large the acceptance probabilities should be. The first constraint is ergodicity: starting from any configuration C_0 with nonzero Boltzmann weight, any other configuration with nonzero Boltzmann weight should be reachable through a finite number of well-chosen Monte Carlo moves. Apart from a very small number of exceptions, a second constraint to guarantee its

* Corresponding author. Tel.: +31 642562091.
E-mail addresses: schram@lorentz.leidenuniv.nl (R.D. Schram), g.t.barkema@uu.nl (G.T. Barkema).

http://dx.doi.org/10.1016/j.physa.2014.06.015
0378-4371/

correctness is known as the condition of detailed balance. If the system is in state A, and a single Monte Carlo move brings it to state B, then detailed balance requires

$$P_A \cdot T(A \to B) \cdot A(A \to B) = P_B \cdot T(B \to A) \cdot A(B \to A), \tag{2}$$

in which $T(A \to B)$ is the probability that, given the system is in state A, a move towards state B is proposed, and $A(A \to B)$ is the probability that this proposed move is accepted. The combination of ergodicity and detailed balance assures a correct algorithm, i.e., given a long enough time, the Boltzmann distribution is sampled. Though this combination is sufficient, it is not a necessary condition. In fact, the slithering snake algorithm [2] from polymer physics is an example of an algorithm violating detailed balance. One often-used approach to realize detailed balance is to randomly propose a small change in state A, resulting in another state B, in such a way that the reverse process (starting in B and then proposing a small change that results in A) is equally likely. More formally, a process in which the condition $T(A \to B) = T(B \to A)$ holds for all pairs of states $\{A, B\}$. If that is the case, then detailed balance can be obtained by the so-called Metropolis algorithm [3], in which the acceptance probability is given by

$$A_{\text{met}}(A \to B) = \min\left[1, \exp(-\beta(E_B - E_A))\right]. \tag{3}$$

Thus, a proposed move which does not raise the total energy is always accepted, but one resulting in higher energy is accepted with a probability that decreases with increasing energy.

For our purposes further on, it is convenient to introduce also the 'unconditional transition probability', defined as

$$U(A \to B) = P_A \cdot T(A \to B) \cdot A(A \to B). \tag{4}$$

Expressed in this quantity, detailed balance is simply the condition $U(A \to B) = U(B \to A)$ for all pairs of states $\{A, B\}$, and the Metropolis algorithm is obtained by $T(A \to B) = T(B \to A)$ combined with the requirement that the maximum of $A(A \to B)$ and $A(B \to A)$ is unity.

Recently, there has been a flurry of efforts to explore dynamics which violates detailed balance [4–10]. In particular, the approach presented by Turitsyn et al. [6] is closely related to our approach. Here, we follow a more generic description. Additionally, to explore the efficiency of this approach in a more relevant model, we apply it to the Ising model in two dimensions.

2. Beyond detailed balance: a one-dimensional model

We first introduce a very simple model, to illustrate a drawback of standard Monte Carlo algorithms. The phase space consists of states $i = 1 \ldots N$ which are placed on a one-dimensional ring. Each state i has an energy E_i. Transitions occur from state i to $i \pm 1$, modulo N. At high temperatures, all states are more or less equally likely, as well as all transitions. In that regime, it is reasonable to expect that the correlation time of the Monte Carlo algorithm, measured in the number of transitions, scales as $\tau_c \sim N^2$, as the exploration of phase space occurs by diffusion.

A more efficient exploration of phase space takes place if the dynamics would resemble molecular dynamics with inertia, in which a transition to the left (right) is predominantly followed by another transition in the same direction. In principle, this allows for asymptotically faster exploration during the time span where the direction of the transitions is correlated. How can this be achieved? One way to do this, is as follows. We double the phase space, into states $i = 1^+ \cdots N^+$ (the forward states) and states $i = 1^- \cdots N^-$ (the backward states), while maintaining that all states should be visited with their Boltzmann probabilities:

$$P(i^+) = P(i^-) \sim \exp(-\beta E_i). \tag{5}$$

As in the standard Monte Carlo, we randomly propose moves in either direction, but now we also add transitions between the forward and backward states:

$$\begin{aligned} T(i^+ \to (i+1)^+) &= T(i^- \to (i+1)^-) = T((i+1)^+ \to i^+) = T((i+1)^- \to i^-) \\ &= T(i^+ \to i^-) = T(i^- \to i^+) = 1/3. \end{aligned} \tag{6}$$

We do, however, reduce the acceptance probabilities in the undesirable directions as much as possible. Along each fourfold loop visiting the states i^+, $(i+1)^+$, $(i+1)^-$ and i^-, we modify the unconditional transition probability by subtracting a cycle with probability

$$U_{\text{loop}}(i + 1/2) = p_i \cdot A_{\text{met}}(i \to i+1) = p_{i+1} \cdot A_{\text{met}}(i+1 \to i) \tag{7}$$

so that the resulting acceptance probabilities are:

$$\begin{aligned} A(i^+ \to (i+1)^+) &= A_{\text{met}}(i \to i+1) \\ A((i+1)^- \to i^-) &= A_{\text{met}}(i+1 \to i) \\ A((i+1)^+ \to i^+) &= A_{\text{met}}(i+1 \to i) - P_{i+1}^{-1} \cdot U_{\text{loop}}(i+1/2) \\ A(i^- \to (i+1)^-) &= A_{\text{met}}(i \to i+1) - P_i^{-1} U_{\text{loop}}(i+1/2) \\ A(i^+ \to i^-) &= P_i^{-1} \cdot \text{Max}\left[0, U_{\text{loop}}(i+1/2) - U_{\text{loop}}(i-1/2)\right] \\ A(i^- \to i^+) &= P_i^{-1} \cdot \text{Max}\left[0, U_{\text{loop}}(i-1/2) - U_{\text{loop}}(i+1/2)\right]. \end{aligned} \tag{8}$$

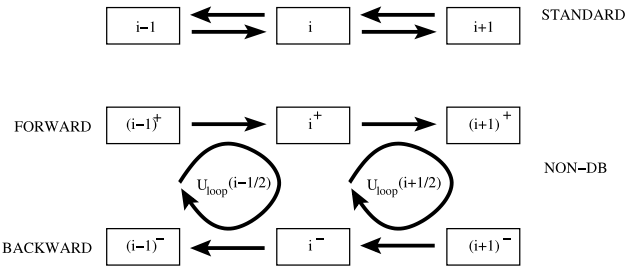

Fig. 1. A schematic depiction of the difference between standard Monte Carlo and non-detailed balance Monte Carlo. The upper part shows the algorithm satisfying detailed balance. The bottom part shows the alternative algorithm violating detailed balance. A particle (or system) going in the 'forward' (backward) direction will predominantly continue in the same direction, with a small chance of reversing direction. This is done by doubling the phase space and adding cycles with well chosen unconditional transition probabilities.

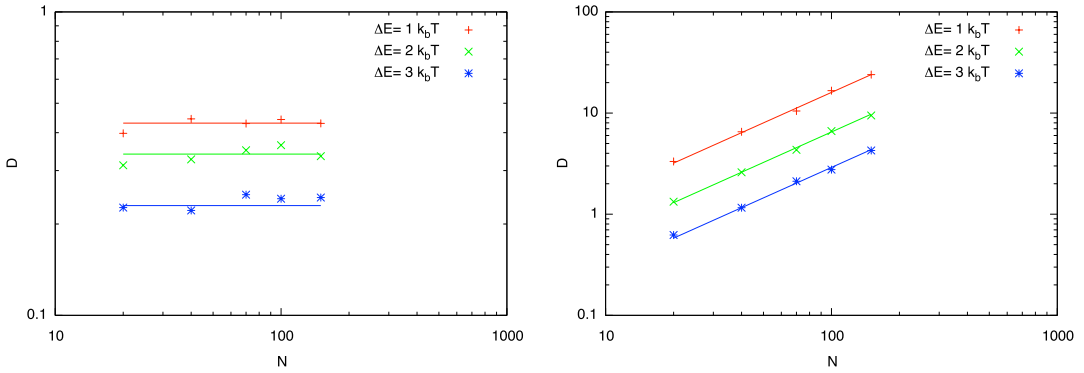

Fig. 2. The diffusion coefficient D as a function of the system size N, for a standard MC algorithm (left) and the non-detailed balance algorithm (right). The model consists of a single particle in a (single) "saw-tooth" potential where the difference between the minimum and maximum energy is ΔE.

Thus, we maintain the transitions from i^+ to $(i+1)^+$, as well as the transitions from i^- to $(i-1)^-$, while at the same time transitions from i^+ to $(i-1)^+$ and from i^- to $(i+1)^-$ are suppressed. To restore a correct sampling, we introduce transitions from i^+ to i^- or reverse, with rates as small as possible. The consequence is that in the upper states, there is a trend to move up (to states with a higher index), while in the lower states, there is an opposite trend. This new approach is illustrated in Fig. 1.

Irrespective of the values of $U_{\text{loop}}(i+1/2)$, the Boltzmann distribution is sampled. If $U_{\text{loop}}(i+1/2)$ is set to zero, then the standard Metropolis algorithm is retrieved. On the other hand, the choice in Eq. (7) is the maximal value that keeps all acceptance probabilities in the interval [0,1], while minimizing the number of reversals. In this case the backward transition probabilities in the forward states become exactly zero, and vice versa.

We have implemented the algorithm defined by Eq. (8), for a model in which the energy of state i equals $E_i = i/N$. Thus, the energy landscape resembles a saw-tooth. The efficiency of the exploration of phase space can be characterized by measuring the diffusion coefficient, i.e. the mean squared distance per unit of time, measured over times in which a typical simulation crosses the periodic boundary many times. For a precise definition of this effective diffusion coefficient, we introduce the coordinate x which is incremented by 1 for every step in the positive direction, and decremented by 1 for every step in the negative direction. Note that x is not limited to the interval $[0 \ldots N-1]$. The diffusion is then defined as

$$D = \lim_{t\to\infty} \frac{\left\langle (x_t - x_0)^2 \right\rangle}{2t}, \tag{9}$$

where t is the number of Monte Carlo iterations. Fig. 2 shows this diffusion coefficient for a number of system sizes N and temperatures T, for the standard Monte Carlo (left) as well as the new approach (right). Analysis of these data shows that with the standard Monte Carlo the diffusion coefficient is independent of the system size, while in case of the new algorithm, the diffusion coefficient increases linearly with the system size. Thus, breaking detailed balance, we achieve a faster exploration of phase space, by reducing the rates of steps that undo the previous step.

3. Beyond detailed balance: a general formulation

For a more general formulation of our new approach, we switch to continuous time simulations with rejection-free sampling, as explained in the previous paper. Given that we are in a state C_i, we list the collection of states $B \in \{C_i^{\text{trial}}\}$ reachable from C_i with nonzero probability, for which we define the rate $r(C_i \to B)$:

$$r(C_i \to B) = T(C_i \to B) \cdot A(C_i \to B). \tag{10}$$

Next, we determine the total rate $R(C_i)$ out of state C_i, using the definition

$$R(A) = \sum_B r(A \to B) = \sum_{B \in \{C_i^{\text{trial}}\}} r(A \to B), \tag{11}$$

where the rightmost expression is more suited for implementation, since it leaves out all terms that are zero because they are never proposed. The algorithm now proceeds by iterating two steps:

(i) increment the time scale: $t_{i+1} = t_i + \Delta t_i$ with $\Delta t_i = 1/R(C_i)$.
(ii) from the complete set of states $B \in C_i^{\text{trial}}$, select one with a conditional probability proportional to the rate towards it, i.e.

$$P(C_i \to B) = r(C_i \to B)/R(C_i). \tag{12}$$

The time-averaged value of some observable Q as obtained from this sampling equals the ensemble-averaged value of it; as the time increments are not equal, each value should be weighted with the time it lasts. Thus:

$$\langle Q \rangle = \frac{\sum_i \Delta t_i Q_i}{\sum_i \Delta t_i}, \tag{13}$$

in which Q_i is the value of the observable in state C_i. To complete the picture we introduce the equivalent of the unconditional transition probability, which is the flux (compare Eq. (4))

$$\phi(A \to B) = P(A) \cdot r(A \to B). \tag{14}$$

Instead of detailed balance, we can now write down the generalized balance equations:

$$\forall A \in \mathcal{C} :: \sum_B [\phi(A \to B) - \phi(B \to A)] = 0. \tag{15}$$

In the case of detailed balance the constraint is more strict: $\phi(A \to B) = \phi(B \to A)$. To allow for more general Monte Carlo simulations, the flux network can also be constructed using cycles. This was done exactly in the previous section with U_{loop}.

In continuous time simulations, the *time-averaged* probability $P(A)$ of being in state A is no longer equal to the frequency of occurrence $\rho(A)$ of state A in the sequence of states visited by the Markov process, as the time spent in states will generally differ. Apart from an overall normalization constant, the two are related by

$$\rho(A) \sim P(A)/\Delta t(A). \tag{16}$$

At first sight, one might think that a large spread in the times $\Delta t(A)$ should always be avoided: if only few states have very long residual times, then the sampling will be poor as those few states will dominate the expectation values of the observables of interest, and the contribution of states with a very short residual time to the observable of interest is negligible and should therefore be avoided. However, this is not entirely true. For instance, in a model in which deep-lying minima are separated by huge energy barriers, it might be beneficial to spend a significant number of Monte Carlo steps to cross these barriers (and thereby visiting the intermediate high-energy states) for a faster overall exploration of phase space.

In statistical physics, one-dimensional models are an exception, more than the rule. Our MC beyond detailed balance needs a local distinction between forward positive and backward negative directions, however. A convenient choice for this distinction turns out to be provided by the energy. As a result, in the spirit of our simulation approach for the one-dimensional model, once we move up in energy, we continue to do so for a large number of steps, until we switch the direction and make many steps that decrease the energy. In case of a model with discrete energies, a tie-breaker might be introduced, to achieve that all pairs of states before and after a single move have a well-defined direction between them. One possible set of fluxes satisfying the balance equations (15) (and thus sampling the Boltzmann distribution), while having a small spread in rates $r(A \to B)$, and at the same reducing the number of reversals, is given by

$$\begin{aligned}
\phi(A^+ \to B^+) &= \exp(-\beta E_B) \\
\phi(A^- \to C^-) &= \exp(-\beta E_A) \\
\phi(A^+ \to A^-) &= \text{Max}\left[0, \sum_C \phi(A^- \to C^-) - \sum_B \phi(A^+ \to B^+)\right] \\
\phi(A^- \to A^+) &= \text{Max}\left[0, \sum_C \phi(A^- \to C^-) - \sum_B \phi(A^+ \to B^+)\right]
\end{aligned} \tag{17}$$

where $\{B\}$ and $\{C\}$ are the sets of states with a higher, resp. lower energies; the superscript distinguishes the forward and the backward states, as before. Note that from now on, we assume that a tie-breaker rule is enforced which avoids ambiguities.

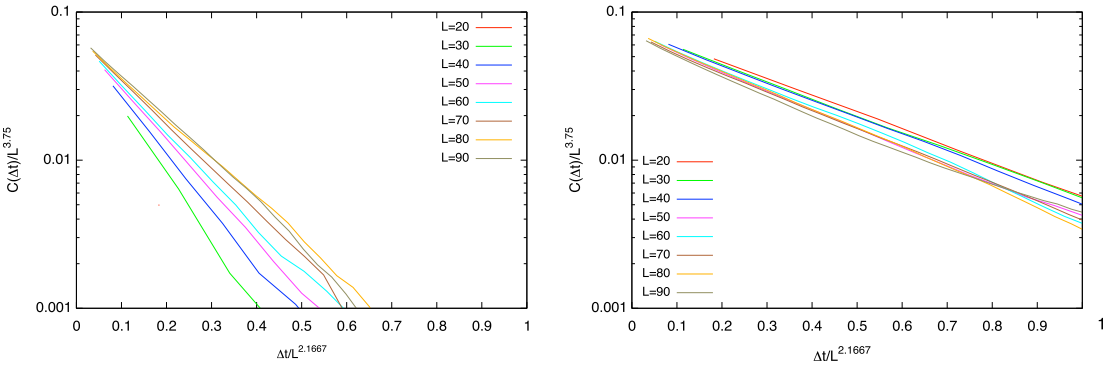

Fig. 3. Simulations of the Ising model, using the traditional Metropolis algorithm on the right, and the new non-detailed balance algorithm on the left. The lattices are squares and range from 20×20 to 90×90. The simulations are done at the known critical temperature with $\beta = \log(1+\sqrt{2})/2$. The x-axis is rescaled as $\tilde{dt} = t_a dt/L^z$, with $z = 2.1665(12)$ [11]. The y-axis is rescaled as $\tilde{C}(M, \tilde{dt}) = C(M, \tilde{dt})/L^{\gamma}$, with $C(M, dt) = \langle M(t) * M(t+dt) - \langle M(t)\rangle^2\rangle$ and $\gamma = 3.75$. For the non-detailed balance algorithm, the best fit is given by $\tilde{C}(M, \tilde{dt}) = a * \exp(-\tilde{dt}/\tau)$, with $a = 0.07$ and $\tau = 0.15$. In the case of the detailed balance algorithm, the best parameters were found to be $a = 0.07$ and $\tau = 0.35$. Thus, the non-detailed balance algorithm is about 2.3 times faster than its standard counterpart.

4. Results: two-dimensional Ising model

To test the efficiency of the non-detailed balance algorithm, we compared the autocorrelation of the magnetization M with the Metropolis algorithm. Different variants of the non-detailed balance algorithm were tested. The most effective variant was found to be one that uses the magnetization as the distinction between forward and backward moves. An attempt was made using the energy as the distinction between forward and backward directions, but to our surprise, the algorithm was even slower than the standard Metropolis algorithm. For the comparison both algorithms use rejection free sampling. The autocorrelation functions are shown in Fig. 3.

Though the dynamic exponent z is equal for the two algorithms, the prefactor is better for the non-detailed balance algorithm. The difference is a factor of approximately 2.3, which shows that the introduction of the new Monte Carlo approach can improve simulations for less trivial problems as well. The number of spin flips per second is roughly the same for the rejection-free implementations of the algorithm with detailed balance and the one in which detailed balance is violated. Interestingly, the figure also shows that the non-detailed balance algorithm is more advantageous for smaller system sizes, before converging to a constant factor at larger system sizes. We currently do not have a coherent explanation for this phenomenon.

5. Conclusion, discussion and outlook

Algorithms violating detailed balance are currently very rarely used. We show, that despite low development efforts, such algorithms have the potential to improve the performance significantly. In a simple one-dimensional model, the non-detailed balance algorithm was asymptotically faster than the detailed balance algorithm. In the case of the Ising model, we found that a simple algorithm using the magnetization as a distinction between forward and backward moves was about 2.3 times faster than an otherwise identical algorithm with detailed balance. Thus, for this model, the increase in efficiency cannot compete with the much faster cluster algorithms [12,13] at the critical temperature. However, applied to other models, where cluster algorithms are not available, the new approach may very well advance Monte Carlo computation.

In future research we will continue along the direction explained in this paper, to apply the method to other models in statistical physics, such as spin-glasses and lattice polymer models with geometric constraints. Preliminary results are encouraging.

Acknowledgments

This work is part of the research programme of the Foundation for Fundamental Research on Matter (FOM), which is financially supported by the Netherlands Organisation for Scientific Research (NWO), grant L2604M and also part of the D-ITP consortium, a program of the NWO that is funded by the Dutch Ministry of Education, Culture and Science (OCW).

References

[1] M.E.J. Newman, G.T. Barkema, Monte Carlo Methods in Statistical Physics, Oxford University Press, 1998.
[2] F.T. Wall, F. Mandel, Macromolecular dimensions obtained by an efficient Monte Carlo method without sample attrition, J. Chem. Phys. 63 (1975) 4592.
[3] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. 21 (1953) 1087.

[4] P. Diaconis, H. Susan, R.M. Neal, Analysis of a Non-Reversible Markov Chain Sampler, Cornell University Technical Report, BU-1385-M, 1997.
[5] H. Suwa, S. Todo, Markov chain Monte Carlo method without detailed balance, Phys. Rev. Lett. 105 (2011) 120603.
[6] K.S. Turitsyn, M. Chertkov, M. Vucelja, Irreversible Monte Carlo algorithms for efficient sampling, Physica D 240 (2011) 410–414.
[7] T. Lelièvre, F. Nier, G.A. Pavliotis, Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion, preprint arXiv:1212.0876v1, 2011.
[8] A. Ichiki, M. Ohzeki, Violation of detailed balance accelerates relaxation, Phys. Rev. E 88 (2013) 020101.
[9] S. Todo, H. Suwa, Geometric allocation approaches in Markov chain Monte Carlo, preprint arXiv:1310.6615v11, 2013.
[10] Y. Sakai, K. Hukushima, Dynamics of one-dimensional Ising model without detailed balance condition, J. Phys. Soc. Japan. 82 (2013) 064003.
[11] M.P. Nightingale, H.W.J. Blöte, The dynamic exponent of the two-dimensional Ising model and Monte Carlo computation of the sub-dominant eigenvalue of the Stochastic matrix, Phys. Rev. Lett. 76 (1996) 4548.
[12] R.H. Swendsen, J.S. Wang, Nonuniversal critical dynamics in Monte Carlo simulations, Phys. Rev. Lett. 58 (1987) 86.
[13] U. Wolff, Collective Monte Carlo updating for spin systems, Phys. Rev. Lett. 62 (1989) 361.