M

MARKOV CHAIN MONTE CARLO SIMULATIONS

Markov Chain Monte Carlo (MCMC) simulations are widely used in many branches of science. They are nearly as old as computers themselves, since they started in earnest with a 1953 paper by Nicholas Metropolis, Arianna Rosenbluth, Marshall Rosenbluth, Augusta Teller, and Edward Teller (1) at the Los Alamos National Laboratory, New Mexico.These authors invented what is nowadays called the Metropolis algorithm. Various applications and the history of the basic ideas are reviewed in the proceedings (2) of the 2003 Los Alamos conference, which celebrated the fiftieth anniversary of the Metropolis algorithm.

OVERVIEW

The Monte Carlo method to compute integrals (3) amounts to approximate an integral $\int_\Omega d\mu(X)\mathcal{O}(X)$, inside a volume Ω, where $d\mu(X)$ is some integration measure and $\int_\Omega d\mu(X) < \infty$, by sampling. Without loss of generality, one can assume that $\int_\Omega d\mu(X) = 1$, and consider accordingly $d\mu(X)$ as a probability measure. Drawing independently N_{sample} values of $X \in \Omega$ according to this probability measure, one has

$$\int_\Omega d\mu(X)\mathcal{O}(X) \approx \frac{1}{N_{\text{samples}}} \sum_{s=1}^{N_{\text{samples}}} \mathcal{O}(X_s) \tag{1}$$

where X_s is the s'th random variable drawn. The right hand side of Equation 1 is an unbiased estimator of the integral, namely it is exact in average, for any N_{sample}. It converges toward the correct value when $N_{\text{samples}} \to \infty$, with a $1/\sqrt{N_{\text{samples}}}$ leading correction. This method is of practical use if one can easily draw, in a computer program, random values $X_s \in \Omega$ according to the measure $d\mu(X)$. This is the case in particular if one integrates inside an N- dimensional hypercube with the flat measure $d\mu^{\text{flat}}(X) \propto \prod_{k=1}^{N} dX^{(k)}$, where the $X^{(k)}$'s are the Cartesian components of X, or when the integral reduces to a finite sum. This is not the case in the situation where this simple measure is multiplied by a nontrivial function $\omega(X)$ of the components $X^{(k)}$. If $\omega(X)$ is a smooth function in Ω with a limited range of variations, one can still draw values of X according to the simple measure $d\mu^{\text{flat}}$ and write (assuming without loss of generality that $\int_\Omega d\mu(X) = 1$ and $\int_\Omega d\mu^{\text{flat}}(X) = 1$)

$$\begin{aligned}\int_\Omega d\mu(X)\mathcal{O}(X) &= \int_\Omega d\mu^{\text{flat}}(X)\omega(X)\mathcal{O}(X) \\ &\approx \frac{1}{N_{\text{samples}}} \sum_{s=1}^{N_{\text{sample}}} \omega(X_s)\mathcal{O}(X_s)\end{aligned} \tag{2}$$

If the function $\omega(X)$ has a very large range of variations with one or several sharp peaks, the sum in Equation 2 is dominated by a few rare configurations, and the Monte Carlo method does not work (most samples are drawn in vain). This is the case of the problem considered in Ref. 1, where $\omega(X)$ is a Boltzmann weight, the exponential of N times a function of order one, with N the number of molecules in a box, which is a notoriously large number.

The Markov chain Monte Carlo method allows overcoming this problem by generating a set of random $X \in \Omega$, distributed according to the full measure $d\mu(X)$, using an auxiliary Markov chain. Note that often, in particular in the physics literature, Monte Carlo is used as a short name for Markov chain Monte Carlo (MCMC). MCMC is sometimes called dynamic Monte Carlo, in order to distinguish it from the usual, "static," Monte Carlo. Many probability distributions, which cannot be sampled directly, allow for MCMC sampling. From now on we write a formula for a discrete space Ω with K_{st} states, although the results can be generalized. A Markov chain (4–7) is a sequence of random variables $X_1, \quad X_2, \quad X_3, \ldots$, that can be viewed as the successive states of a system as a function of a discrete time t, with a transition probability $\mathbf{P}(X_{t+1} = r | X_t = s) = W_{r,s}$ that is a function of r and s only. The next future state X_{t+1} is a function of the current state X_t alone. Andrey Markov (8) was the first to analyze these processes. In order to analyze Markov chains, one introduces an ensemble of chains with an initial probability distribution $\mathbf{P}(X_0 = s)$. By multiplying this vector by the transition matrix repeatedly, one obtains $\mathbf{P}(X_1 = s)$, and then $\mathbf{P}(X_2 = s), \ldots,$ successively. The natural question is whether this sequence converges. One says that a probability distribution $w = \{w_s\}_{s \in \Omega}$ is an equilibrium distribution of a Markov chain if it is let invariant by the chain, namely if

$$\sum_{s=1}^{K_{st}} W_{r,s}\omega_s = \omega_r \tag{3}$$

This condition is called balance in the physics literature. A Markov chain is said to be ergodic (irreducible and aperiodic in the mathematical literature) if for all states $r, s \in \Omega$ there is a $N_{r,s}$ such that for all $t > N_{r,s}$ the probability $(W^t)_{s,r}$ to go from r to s in t steps is nonzero. If an equilibrium distribution exists and if the chain is irreducible and aperiodic, one can show (4–7) that, starting from any distribution $\mathbf{P}(X_0 = s)$, the distribution after t steps $\mathbf{P}(X_t = s)$ converges when $t \to \infty$ toward the equilibrium configuration ω_s.

The Metropolis algorithm (see the next section) offers a practical way to generate a Markov chain with a desired equilibrium distribution, on a computer using a pseudorandom numbers generator. Starting from an initial configuration X_0, successive configurations are generated. In most cases, the convergence of $\mathbf{P}(X_t = s)$ toward ω_s is exponential in t and one can safely assume that, after some number of steps t_{eq}, "equilibrium is reached" and that the

Wiley Encyclopedia of Computer Science and Engineering, edited by Benjamin Wah.

configurations $X_{t_{eq}+1}, X_{t_{eq}+2}, \ldots$ are distributed according to $\omega(X)$.

Whereas the random samples used in a conventional Monte Carlo (MC) integration are obviously statistically independent, those used in MCMC are correlated. The effective number of independent events generated by an equilibrated Markov chain is given by the number of steps done divided by a quantity called "integrated autocorrelation time" τ_{int}. In most cases, τ_{int} does not depend on the quantity $\mathcal{O}$ measured, but there are exceptions like, for example the vicinity of a second-order (continuous) phase transition. The algorithm will fail if τ_{int} is too big. In practice it can be quite difficult to estimate τ_{int} reliably and there can also be a residual effect of the starting position. More sophisticated MCMC-based algorithms such as "coupling from the past" (9,10) produce independent samples rigorously but at the cost of additional computation and an unbounded (although finite on average) running time.

The method was originally developed to investigate a statistical physics problem. Suitable applications arise as well in computational biology, chemistry, physics, economics, and other sciences. MCMC calculations have also revolutionized the field of Bayesian statistics.

Most concepts of modern MCMC simulation were originally developed by physicists and chemists, who still seem to be at the cutting edge of new innovative developments into which mathematician, computer scientists, and statisticians have joined. Their interest developed mainly after a paper by Hastings (11), who generalized the accept/reject step of the Metropolis method. Unfortunately a language barrier developed that inhibits cross-fertilization. For instance the "heat bath" algorithm, which may be applied when the conditional distributions can be sampled exactly, was introduced by physicist Refs. (12) and (13). Then it was rediscovered in the context of Bayesian restoration of images under the name "Gibbs sampler" (14). Another example is the "umbrella sampling" algorithm (15) that was invented in chemical physics and later rediscovered and improved by physicists. This is just two of many examples of how different names for the same method emerged. The reader should be aware that this article was written by physicists, so that their notations and views are dominant in this article. The book by Liu (16) tries to some extent to bridge the language barrier. Other textbooks include those by Robert and Casella (17) for the more mathematically minded reader, Landau and Binder (18) for statistical physicists, Berg (19) from a physicist point of view, but also covering statistics and providing extensive computer code (in Fortran). Kendall, et al. (20) edited a book that combines expositions from physicists, statisticians, and computer scientists.

In the following discussion, we will first explain the basic method and illustrate it for a simple statistical physics system, the Ising ferromagnet in two dimensions, followed by a few remarks on autocorrelations and cluster algorithms. An overview of the MCMC updating scheme is subsequently given. The final section of this article focuses on so-called generalized ensemble algorithms.

MCMC AND STATISTICAL PHYSICS

As mentioned, the MCMC algorithm was invented to investigate problems in statistical physics. The aim of statistical physics is to predict the average macroscopic properties of systems made of many constituents, which can be, for example, molecules in a box (as in the original article of Metropolis et al.), magnetic moments (called spins) at fixed locations in a piece of material, or polymer chains, in a solvent. If one considers the so-called canonical ensemble, where the system has a fixed temperature T, one knows that the probability to observe a given microscopic configuration (or micro state) s (defined in the three examples given above by the positions and velocities of all molecules; the orientations of all the spins; and the exact configuration of all the polymer chains, respectively) is proportional to the Boltzmann weight $\exp(-E_s/k_BT) = \exp(-\beta E_s)$, where E_s is the energy of the configuration s, k_B is the Boltzmann constant, and T is the temperature. (In the following discussion, we will use a unit of temperature such that $k_B = 1$). Let $\mathcal{O}_s$ be the value of $\mathcal{O}$ computed in configuration s. The mean value of any macroscopic observable $\mathcal{O}$ (e.g. the energy) is given by the average of $\mathcal{O}_s$ over all possible configurations s, weighted by the Boltzmann weight of the configuration. This sum (or integral if one has continuous variables) is to be normalized by the sum over all configurations of the Boltzmann weight [the so-called partition function $Z(T)$], namely

$$\hat{O} = \hat{O}(T) = \langle O \rangle = Z^{-1}(T) \sum_{s=1}^{K_{st}} O_s e^{-E_s/T} \tag{4}$$

where $Z(T) = \sum_{s=1}^{K_{st}} \exp(-E_s/T)$. The index $s = 1, \ldots, K_{st}$ labels the configuration (states) of the system.

A particularly simple system is the Ising model for which the energy is given by

$$E = \sum_{<ij>} s_i s_j \tag{5}$$

Here the sum is over the nearest-neighbor sites of a hypercubic D-dimensional lattice and the Ising spins take the values $s_i = \pm 1$, $i = 1, \ldots, N$ for a system of $N = \prod_{i=1}^{D} L_i$ spins. Periodic boundary conditions are used in most simulations. The energy per spin is $e = E/N$. The model describes a ferromagnet for which the magnetic moments are simplified to ± 1 spins at the sites of the lattice. In the $N \to \infty$ limit (and for $D > 1$), this model has two phases separated by a phase transition (a singularity) at the critical temperature $T = T_c$. This is a second-order phase transition, which is continuous in the energy, but not in specific heat. This model can be solved analytically when $D = 2$. (This means that one can obtain exact analytical expressions for the thermodynamic quantities, e.g., the energy per spin, as a function of temperature, in the $N \to \infty$ limit.) This makes the two-dimensional (2-D) Ising model an ideal testbed for MCMC algorithms.

The number of configurations of the Ising model is $K_{st} = 2^N$, because each spin can occur in two states (up or down). Already for a moderately sized lattice, say linear dimension $L_1 = L_2 = 100$ in $D = 2$, K_{st} is a tremendously large number. With the exception of very small systems, it is therefore impossible to do the sums in Equation 4 explicitly. Instead, the large value of K_{st} suggests a statistical evaluation.

IMPORTANCE SAMPLING THROUGH MCMC

As explained in the previous section, one needs a procedure that generates configurations s with the Boltzmann probability

$$P_{B,s} = c_B\, w_{B,s} = c_B e^{-\beta E_s} \tag{6}$$

where the constant c_B is determined by the condition $\sum_s P_{B,s} = 1$. The vector $P_B := \{P_{B,s}\}$ is called the Boltzmann state. When configurations are generated with the probability $P_{B,s}$, the expectation values [Equation (4)] are estimated by the arithmetic averages:

$$\langle O \rangle = \lim_{N_{\text{samples}} \to \infty} \frac{1}{N_{\text{samples}}} \sum_{s=1}^{N_{\text{samples}}} O_s \tag{7}$$

With the MCMC algorithm, this is obtained from a Markov chain with the following properties:

1. *Irreducibility and aperiodicity (ergodicity in the physics literature)*: For any two configurations r and s, an integer $N_{r,s}$ exists such that for all $n > N_{r,s}$, the probability $(W^n)_{r,s}$ is nonzero.
2. *Normalization*: $\sum_r W_{r,s} = 1$.
3. *Stationarity of the Boltzmann distribution (balance in the physics literature)*: The Boltzmann state [Equation (6)] is an equilibrium distribution, namely a right eigenvector of the matrix W with eigenvalue one, i.e., $\sum_s W_{r,s} e^{-\beta E_s} = e^{-\beta E_r}$ holds.

There are many ways to construct a Markov process satisfying 1, 2, and 3. In practice, MCMC algorithms are often based on a stronger condition than 3, namely,

3. *Detailed balance*:

$$W_{r,s} e^{-\beta E_s} = W_{s,r} e^{-\beta E_r} \quad \forall\, r, s. \tag{8}$$

METROPOLIS ALGORITHM AND ILLUSTRATION FOR THE ISING MODEL

Detailed balance obviously does not uniquely fix the transition probabilities $W_{r,s}$. The original Metropolis algorithm (1) has remained a popular choice because of its generality (it can be applied straightforwardly to any statistical physics model) and its computational simplicity. The original formulation generates configurations with the Boltzmann weights. It generalizes immediately to arbitrary weights (see the last section of this article). Given a configuration s, the algorithm proposes a new configuration r with a priori probabilities $f(r,s)$. This new configuration r is accepted with probability

$$a^{r,s} = \min\left[1, \frac{P_{B,r}}{P_{B,s}}\right] = \begin{cases} 1 \text{ for } E_r < E_s \\ e^{-\beta(E_r - E_s)} \text{ for } E_r > E_s \end{cases} \tag{9}$$

If the new configuration is rejected, the old configuration is kept and counted again. For such decisions one uses normally a pseudorandom number generator, which delivers uniformly distributed pseudorandom numbers in the range $[0, 1)$; see Refs. (18) and (19) for examples, and Ref. (21) for the mathematics involved. The Metropolis algorithm gives then rise to the transition probabilities

$$W_{r,s} = f(r,s) a_{r,s} \quad r \neq s \tag{10}$$

and

$$W_{s,s} = f(s,s) + \sum_{r \neq s} f(r,s)(1 - a_{r,s}) \tag{11}$$

Therefore, the ratio $W_{r,s}/W_{s,r}$ satisfies the detailed balance condition [Equation (8)] if

$$f(r,s) = f(s,r) \tag{12}$$

holds. This condition can be generalized when also the acceptance probability is changed (11). In particular such "biased" Metropolis algorithms allow for approaching the heat-bath updating scheme (22).

For the Ising model, the new putative configuration differs from the old one by the flip of a single spin. Such a "single spin update" requires a number of operations of order one and leads to a ratio $P_{B,r}/P_{B,s}$ in Equation (9). This is fundamental in order to have an efficient MCMC algorithm. The spin itself may be chosen at random, although some sequential procedure is normally more efficient (19). The latter procedure violates the detailed balance condition but fulfills the balance condition, and thus, it does lead to the probability of states approaching the Boltzmann distribution [Equation (6)]. The unit of Monte Carlo time is usually defined by N single spin updates, aka a lattice sweep or "an update per spin." Many ways to generate initial configurations exist. Two easy to implement choices are as follows:

1. Use random sampling to generate a configuration of ± 1 spins.
2. Generate a completely *ordered* configuration, either all spin $+1$ or -1.

Figure 1 shows two Metropolis energy time series (the successive values of the energy as a function of the discrete Monte Carlo time) of 6000 updates per spin for a 2D Ising model on a 100×100 lattice at $\beta = 0.44$, which is close to the (infinite volume) phase transition temperature of this model ($\beta_c = \ln(1 + \sqrt{2})/2 = 0.44068\ldots$). One can compare with the exact value for e (23) on this system size, which is indicated by the straight line in Fig. 1. The time series

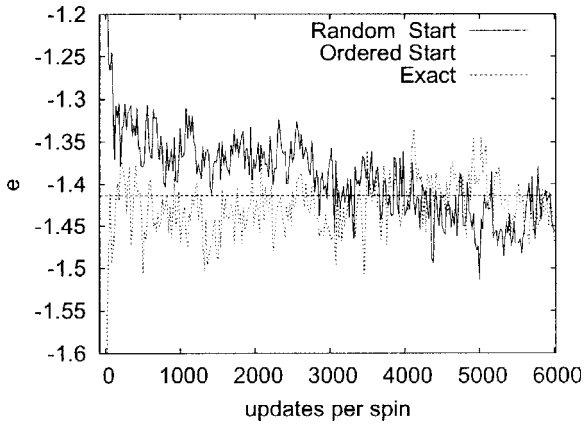

Figure 1. Two-dimensional Ising model: Two initial Metropolis time series for the energy per spin e.

corresponding to the ordered start begins at $e = -2$ and approaches the exact value from below, whereas the other time series begins (up to statistical fluctuation) at $e = 0$ and approaches the exact value from above. It takes a rather long MCMC time of about 3000 to 4000 updates per spin until the two time series start to mix. For estimating equilibrium expectation values, measurements should only be counted from there on. A serious error analysis for the subsequent equilibrium simulation finds an integrated autocorrelation time of $\tau_{\text{int}} \approx 1700$ updates per spin. This long autocorrelation time is related to the proximity of the phase transition (this is the so-called critical slowing down of the dynamics). One can show that, at the transition point, τ_{int} diverges like a power of N.

For this model and a number of other systems with second-order phase transitions, cluster algorithms (24,25) are known, which have much shorter autocorrelation times, in simulations close to the phase transition point. For the simulation of Fig. 1, $\tau_{\text{int}} \approx 5$ updates per spin instead of 1700. Furthermore, for cluster algorithms, τ_{int} grows much slower with the system size N than for the Metropolis algorithm. This happens because these algorithms allow for the instantaneous flip of large clusters of spins (still with order one acceptance probability) in contrast to the local updates of single spins done in the Metropolis and heat-bath-type algorithms. Unfortunately such nonlocal updating algorithms have remained confined to special situations. The interested reader can reproduce these Ising model simulations discussed here using the code that comes with Ref. 19.

UPDATING SCHEMES

We give an overview of MCMC updating schemes that is, because of space limitations, not complete at all.

1. We have already discussed the Metropolis scheme (1) and its generalization by Hastings (7). Variables of the system are locally updated.
2. Within such local schemes, the heat-bath method (12-14) is usually the most efficient. In practice it can be approximated by tabulating Metropolis–Hastings probabilities (22).
3. For simulations at (very) low temperatures, event-driven simulations (26,27), also known as the "N-fold way," are most efficient. They are based on Metropolis or heat-bath schemes.
4. As mentioned before, for a number of models with second order phase transitions the MCMC efficiency is greatly improved by using nonlocal cluster updating (24).
5. Molecular dynamics (MD) moves can be used as proposals in MCMC updating (28,29), a scheme called "hybrid MC", see Ref. (30) for a review of MD simulations.

More examples can be found in the W. Krauth contribution in Ref. (31) and A. D. Sokal in Ref. (32).

GENERALIZED ENSEMBLES FOR MCMC SIMULATIONS

The MCMC method, which we discussed for the Ising model, generates configurations distributed according to the Boltzmann–Gibbs canonical ensemble, with weights $P_{B,s}$. Mean values of physical observables at the temperature chosen are obtained as arithmetic averages of the measurements made [Equation 7]. There are, however, in statistical physics, circumstances where another ensemble (other weights $P_{NB,s}$) can be more convenient. One case is the computation of the partition function $Z(T)$ as a function of temperature. Another is the investigation of configurations of physical interest that are rare in the canonical ensemble. Finally the efficiency of the Markov process, i.e., the computer time needed to obtain a desired accuracy, can depend greatly on the ensemble in which the simulations are performed. This is the case, for example when taking into account the Boltzmann weights, the phase space separates, loosely speaking, into several populated regions separated by mostly vacant regions, creating so-called free energy barriers for the Markov chain.

A first attempt to calculate the partition function by MCMC simulations dates back to a 1959 paper by Salsburg et al. Ref. (33). As noticed by the authors, their method is restricted to very small lattices. The reason is that their approach relies on what is called in the modern language "reweighting." It evaluates results at a given temperature from data taken at another temperature. The reweighting method has a long history. McDonald and Singer (34) were the first to use it to evaluate physical quantities over a range of temperatures from a simulation done at a single temperature. Thereafter dormant, the method was rediscovered in an article (35) focused on calculating complex zeros of the partition function. It remained to Ferrenberg and Swendsen (36), to formulate a crystal clear picture for what the method is particularly good, and for what it is not: The reweighting method allows for focusing on maxima of appropriate observables, but it does not allow for covering a finite temperature range in the infinite volume limit.

To estimate the partition function over a finite energy density range Δe, i.e., $\Delta E \sim N$, one can patch the histograms

from simulations at several temperatures. Such multi-histogram methods also have a long tradition. In 1972 Valleau and Card (37) proposed the use of overlapping bridging distributions and called their method "multistage sampling." Free energy and entropy calculations become possible when one can link the temperature region of interest with a range for which exact values of these quantities are known. Modern work (38,39) developed efficient techniques to combine the overlapping distributions into one estimate of the spectral density $\rho(E)$ [with $Z(T) = \int dE\rho(E)\exp(-\beta E)$] and to control the statistical errors of the estimate. However, the patching of histograms from canonical simulations faces several limitations:

1. The number of canonical simulations needed diverges like $\sqrt{N}$ when one wants to cover a finite, noncritical range of the energy density.
2. At a first-order phase transition point, the canonical probability of configurations with an interface decreases exponentially with N.

One can cope with the difficulties of multi-histogram methods by allowing arbitrary sampling distributions instead of just the Boltzmann–Gibbs ensemble. This was first recognized by Torrie and Valleau (15) when they introduced umbrella sampling. However, for the next 13 years, the potentially very broad range of applications of the basic idea remained unrecognized. A major barrier, which prevented researchers from trying such extensions, was certainly the apparent lack of direct and straightforward ways of determining suitable weighting functions $P_{NB,s}$ for problems at hand. In the words of Li and Scheraga (40): *The difficulty of finding such weighting factors has prevented wide applications of the umbrella sampling method to many physical systems*.

This changed with the introduction of the multicanonical ensemble multicanonical ensemble (41), which focuses on well-defined weight functions and offers a variety of methods to find a "working approximation." Here a working approximation is defined as being accurate enough, so that the desired energy range will indeed be covered after the weight factors are fixed. A similar approach can also be constructed for cluster algorithms (42). A typical simulation consists then of three parts:

1. Construct a working approximation of the weight function P_{muca}. The Wang–Landau recursion (43) is an efficient approach. See Ref. (44) for a comparison with other methods.
2. Perform a conventional MCMC simulation with these weight factors.
3. Reweight the data to the desired ensemble: $\langle \mathcal{O} \rangle = \lim_{N_{\text{samples}} \to \infty} \frac{1}{N_{\text{samples}}} \sum_{s=1}^{N_{\text{samples}}} \mathcal{O}_s P_{B,s}/P_{\text{muca,s}}$ Details can be found in Ref. 19.

Another class of algorithms appeared in a couple of years around 1991 in several papers (45–50), which all aimed at improving MCMC calculations by extending the confines of the canonical ensemble. Practically most important has been the replica exchange method, which is also known under the names parallel tempering and multiple Markov chains. In the context of spin glass simulations, an exchange of partial lattice configurations at different temperatures was proposed by Swendsen and Wang (45). But it was only recognized later (46,50) that the special case for which entire configurations are exchanged is of utmost importance, see Ref. 19 for more details. Closely related is the Jump Walker (J-Walker) approach (51), which feeds replica from a higher temperature into a simulation at a lower temperature instead of exchanging them. But in contrast to the replica exchange method this procedure disturbs equilibrium to some extent. Finally, and perhaps most importantly, from about 1992 on, applications of generalized ensemble methods diversified tremendously as documented in a number of reviews (52–54).

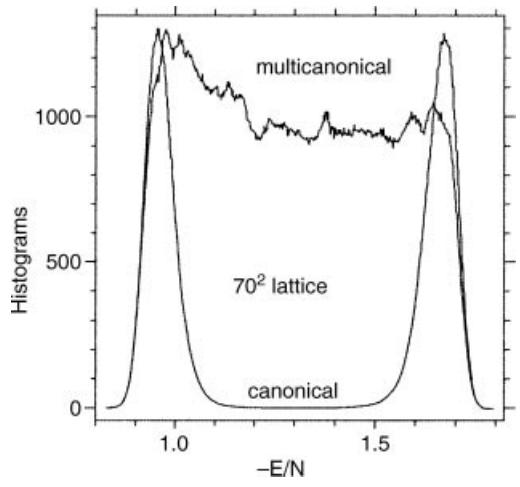

Figure 2. Multicanonical $P_{\text{muca}}(E)$ together with canonical $P(E)$ energy distribution as obtained in Ref. 41 for the $2d$ 10-state Potts model on a 70×70 lattice. In the multicanonical ensemble, the gap between the two peaks present in $P(E)$ is filled up, accelerating considerably the dynamics of the Markov Chain.

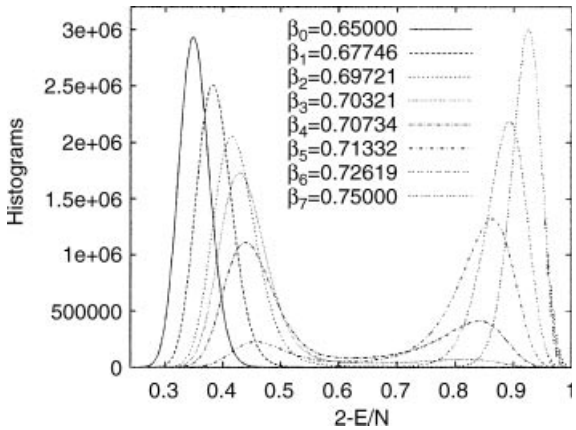

Figure 3. Canonical energy distributions $P(E)$ from a parallel tempering simulation with eight processes for the $2d$ 10-state Potts model on 20×20 lattices (Fig. 6.2 in Ref. 19).

The basic mechanisms for overcoming energy barriers with the multicanonical algorithm are best illustrated for first-order phase transitions (namely a transition with a nonzero latent heat, like the ice–water transition), where one deals with a single barrier. For a finite system the temperature can be fine-tuned to a pseudocritical value, which is defined so that the energy density exhibits two peaks of equal heights. To give an example, Fig. 2 shows for the $2d$ 10-state Potts (55) the canonical energy histogram at a pseudocritical temperature versus the energy histogram of a multicanonical simulation (41). The same barrier can also be overcome by a parallel tempering simulation but in a quite different way. Figure 3 shows the histograms from a parallel tempering simulation with eight processes on 20×20 lattices. The barrier can be "jumped" when there are on both sides temperatures in the ensemble, which are sufficiently close to a pseudocritical temperature for which the two peaks of the histogram are of competitive height. In complex systems with a rugged free energy landscape (spin glasses, biomolecules, ...), the barriers can no longer be explicitly controlled. Nevertheless it has turned out that switching to the discussed ensembles can greatly enhance the MCMC efficiency (53,54). For a recent discussion of ensemble optimization techniques, see Ref. (56) and references given therein.

BIBLIOGRAPHY

1. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines. *J. Chem. Phys.*, **21**: 1087–1092, 1953.
2. J. Gubernatis (Editor). The Monte Carlo method in the physical sciences: Celebrating the 50th anniversary of the Metropolis algorithm, *AIP Conference Proc.* Vol 690, Melville, NY, 2003.
3. N. Metropolis and S. Ulam, The Monte Carlo method. *J. Am. Stat. Assoc.*, **44**: 335–341, 1949.
4. J. G. Kemeny and J. L. Snell, *Finite Markov Chains*. New York: Springer, 1976.
5. M. Iosifescu, *Finite Markov Processes and Their Applications*. Chichester: Wiley, 1980.
6. K. L. Chung, *Markov Chains with Stationary Transition Probabilities*, 2nd ed., New York: Springer, 1967.
7. E. Nummelin, *General Irreducible Markov Chains and Non-Negative Operators*. Cambridge: Cambridge Univ. Press, 1984.
8. A. A. Markov, Rasprostranenie zakona bol'shih chisel na velichiny, zavisyaschie drug ot druga. Izvestiya Fiziko-matematicheskogo obschestva pri Kazanskom Universitete, 2-ya seriya, tom 15: 135–156, 1906.
9. J. G. Propp and D. B. Wilson, Exact sampling with coupled Markov chains and applications in statistical mechanics. *Random Structures and Algorithms*, **9**: 223–252, 1996.
10. J. G. Proof and D. B. Wilson, Coupling from the Past User's Guide. DIMACS Series in Discrete Mathematics and Theoretical Computer Science (AMS), **41**: 181–192, 1998.
11. W. K. Hastings, Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**: 97–109, 1970.
12. R. J. Glauber, Time-dependent statistics of the Ising model. *J. Math. Phys.*, **4**: 294–307, 1963.
13. M. Creutz, Monte Carlo study of quantized $SU(2)$ gauge theory. *Phys. Rev. D.*, **21**: 2308–2315, 1980.
14. S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intelli.* **6**: 721–741, 1984.
15. G. M. Torrie and J. P. Valleau, Nonphysical sampling distributions in Monte Carlo free energy estimation: Umbrella sampling. *J. Comp. Phys.*, **23**: 187–199, 1977.
16. J. S. Liu, Monte Carlo strategies in scientific computing. New York: Springer, 2001.
17. C. P. Robert and G. Casella, Monte Carlo statistical methods (2nd ed.), New York: Springer, 2005.
18. D. P. Landau and K. Binder, A guide to Monte Carlo simulations in statistical physics. Cambridge: Cambridge University Press, 2000.
19. B. A. Berg, Markov chain Monte Carlo simulations and their statistical analysis. Singapore: World Scientific, 2004.
20. W. S. Kendall, F. Liang, and J.-S. Wang (Eds), Markov Chain Monte Carlo: Innovations and applications (Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore). Singapore: World Scientific, 2005.
21. D. Knuth, The art of computer programming, Vol 2: Semi numerical algorithms, Third Edition. Reading, MA: Addison-Wesley, 1997, pp. 1193
22. A. Bazavov and B. A. Berg, Heat bath efficiency with Metropolis-type updating. *Phys. Rev. D.*, **71**: 114506, 2005.
23. A. E. Ferdinand and M. E. Fisher, Bounded and inhomogeneous Ising models. I. Specific-heat anomaly of a finite lattice. *Phys. Rev.*, **185**: 832–846, 1969.
24. R. H. Swendsen and J.-S. Wang, Non-universal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.*, **58**: 86–88, 1987.
25. U. Wolff, Collective Monte Carlo updating for spin systems. *Phys. Rev. Lett.*, **62**: 361–363, 1989.
26. A. B. Bortz, M. H. Kalos, and J. L. Lebowitz, A new algorithm for Monte Carlo simulation of Ising spin systems. *J. Comp. Phys.*, **17**: 10–18, 1975.
27. M. A. Novotny, A tutorial on advanced dynamic Monte carlo methods for systems with discrete state spaces. *Ann. Rev. Comp. Phys.*, **9**: 153–210, 2001.
28. S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, Hybrid Monte Carlo. *Phys. Lett. B.*, **195**: 216–222, 1987.
29. B. Mehlig, D. W. Heermann, and B. M. Forrest, Hybrid Monte Carlo Methods for condensed-matter systems. *Phys. Rev. B.*, **45**: 679–685, 1992.
30. D. Frenkel and B. Smit, Understanding molecular simulation. San Diego CA: Academic Press, 1996.
31. J. Kertesz and I. Kondor (Ed.), Advances in Computer Simulation. *Lecture Notes in Physics*, Heidelberg: Springer Verlag, 1998.
32. K. Binder (Ed.), Monte Carlo and molecular dynamics simulations in polymer science, Oxford: Oxford University Press, 1996.
33. Z. W. Salsburg, J. D. Jacobson, W. S. Fickett, and W. W. Wood. Applications of the Monte Carlo method to the lattice-gas model. I. Two-dimensional triangular lattice. *J. Chem. Phys.*, **30**: 65–72, 1959.
34. I. R. McDonald and K. Singer, Calculation of thermodynamic properties of liquid argon from Lennard-Jones parameters by a Monte Carlo Method. *Discussions Faraday Soc.*, **43**: 40–49, 1967.
35. M. Falcioni, E. Marinari, L. Paciello, G. Parisi, and B. Taglienti, Complex zeros in the partition function of four-dimensional $SU(2)$ lattice gauge model. *Phys. Lett. B.*, **108**: 331–332, 1982.

36. A. M. Ferrenberg and R. H. Swendsen, New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett.*, **61**: 2635–2638, 1988; **63**: 1658, 1989.
37. J. P. Valleau and D. N. Card, Monte Carlo estimation of the free energy by multistage sampling. *J. Chem. Phys.*, **37**: 5457–5462, 1972.
38. A. M. Ferrenberg and R. H. Swendsen, Optimized Monte Carlo data analysis. *Phys. Rev. Lett.*, **63**: 1195–1198, 1989.
39. N. A. Alves, B. A. Berg, and R. Villanova, Ising-Model Monte Carlo simulations: Density of states and mass gap. *Phys. Rev. B.*, **41**: 383–394, 1990.
40. Z. Li and H. A. Scheraga, Structure and free energy of complex thermodynamic systems. *J. Mol. Struct. (Theochem)*, **179**: 333–352, 1988.
41. B. A. Berg and T. Neuhaus, Multicanonical ensemble: A new approach to simulate first-order phase transitions. *Phys. Rev. Lett.*, **68**: 9–12, 1992.
42. W. Janke and S. Kappler, Multibondic cluster algorithm for Monte Carlo simulations of first-order phase transitions. *Phys. Rev. Lett.*, **74**: 212–215, 1985.
43. F. Wang and D. P. Landau, Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.*, **86**: 2050–2053, 2001.
44. Y. Okamoto, Metropolis algorithms in generalized ensemble. In Ref. (2), pp. 248–260. On the web at http://arxiv.org/abs/cond-mat/0308119.
45. R. H. Swendsen and J.-S. Wang, Replica Monte Carlo simulations of spin glasses. *Phys. Rev. Lett.*, **57**: 2607–2609, 1986.
46. C. J. Geyer, Markov Chain Monte Carlo maximum likelihood, in E. M. Keramidas (ed.), *Computing Science and Statistics, Proc. of the 23rd Symposium on the Interface*. Fairfax, VA, Interface Foundation, 1991, pp. 156–163.
47. C. J. Geyer and E. A. Thompson, Annealing Markov chain Monte Carlo with applications to ancestral inference. *J. Am. Stat. Ass.*, **90**: 909–920, 1995.
48. E. Marinari and G. Parisi, Simulated tempering: A new Monte Carlo scheme. *Europhys. Lett.*, **19**: 451–458, 1992.
49. A. P. Lyubartsev, A. A. Martsinovski, S. V. Shevkanov, and P. N. Vorontsov-Velyaminov, New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles. *J. Chem. Phys.*, **96**: 1776–1783, 1992.
50. K. Hukusima and K. Nemoto, Exchange Monte Carlo method and applications to spin glass simulations. *J. Phys. Soc. Japan*, **65**: 1604–1608, 1996.
51. D. D. Frantz, D. L. Freemann, and J. D. Doll, Reducing quasi-ergodic behavior in Monte Carlo simulations by J-walking: Applications to atomic clusters. *J. Chem. Phys.*, **93**: 2769–2784, 1990.
52. W. Janke, Multicanonical Monte Carlo simulations. *Physica A.*, **254**: 164–178, 1998.
53. U. H. Hansmann and Y. Okamoto, The generalized-ensemble approach for protein folding simulations. *Ann. Rev. Comp. Phys.*, **6**: 129–157, 1999.
54. A. Mitsutake, Y. Sugita, and Y. Okamoto, Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers (Peptide Science)*, **60**: 96–123, 2001.
55. F. Y. Wu, The Potts model. *Rev. Mod. Phys.*, **54**: 235–268, 1982.
56. S. Trebst, D. A. Huse, E. Gull, H. G. Katzgraber, U. H. E. Hansmann, and M. Troyer, Ensemble optimization techniques for the simulation of slowly equilibrating systems, Invited talk at the*19th Annual Workshop on Computer Simulation Studies in Condensed Matter Physics*, in D. P. Landau, S. P. Lewis, and H.-B. Schuettler (eds.), Athens, GA, 20–24 February 2006; Springer Proceedings in Physics, Vol. 115, 2007. Available on the internet at: http://arxiv.org/abs/cond-mat/0606006.

BERND A. BERG
Florida State University
Tallahassee, Florida
ALAIN BILLOIRE
Service de Physique Théorique
CEA Saclay
Gif-sur-Yvette, France