Noisy Monte Carlo Algorithm *

T. Bakeyev[a] † and Ph. de Forcrand [b] [c]

[a]Joint Inst. for Nuclear Research, 141980 Dubna, Russia

[b]Inst. für Theoretische Physik, ETH-Hönggerberg, CH-8093 Zürich, Switzerland

[c]CERN, Theory Division, CH-1211 Geneva 23, Switzerland

We present an exact Monte Carlo algorithm designed to sample theories where the energy is a sum of many couplings of decreasing strength. The algorithm avoids the computation of almost all non-leading terms. Its use is illustrated by simulating SU(2) lattice gauge theory with a 5-loop improved action. A new approach for dynamical fermion simulations is proposed.

When sampling the partition function $Z = \int \prod dU \;\; e^{-H(\{U\})}$, the most common algorithm is that of Metropolis: at each step, starting from the configuration $\{U\}$, a candidate configuration $\{U'\}$ is proposed, and it is accepted with probability $P_{acc} = \min(1, \mathrm{e}^{-(\mathrm{H}(\{\mathrm{U}'\})-\mathrm{H}(\{\mathrm{U}\}))})$. This acceptance test is realized by comparing P_{acc} to a random number uniformly distributed in $[0,1]$. This seems like an excess of information: why compute $H(\{U'\})$ *exactly*, then compare it with a random number? It should be sufficient to *estimate* it. This logical proposition was studied long ago [1,2]. The main difficulty which prevented the construction of an efficient algorithm at that stage was the existence of probability bound violations for the noisy estimator of P_{acc}, which caused intolerable systematic errors. This difficulty was overcome in [3]. Ref.[3], however, introduces an infinite number of auxiliary variables, and tests of the method are performed on a toy model with 5 degrees of freedom only. Here, we simplify the method of [3], by introducing only 1 auxiliary variable per term in H. Moreover, we separate H into a leading part to be calculated exactly, and a sum of correction terms, which we treat stochastically.

Consider a generic Hamiltonian of the type $H = \sum_{k=0}^{m} c_k W_k$, where as k increases, $|c_k|$ decreases. It is often the case that one would like to study a Hamiltonian of such type resulting from an expansion, be it perturbative [4], non-perturbative [5], or based on the fixed point of a renormalization group transformation [6]. As k increases, the number of geometrically equivalent terms grouped into W_k increases exponentially. This combinatoric explosion normally makes the simulation of extended Hamiltonians prohibitively expensive. However, in most cases the couplings c_k decrease exponentially with k, so that the overall Hamiltonian is dominated by W_0. By making use of stochastic methods to estimate the correction terms $W_k, k \geq 1$, we aim at postponing the combinatoric explosion of the simulation costs incurred when including higher terms W_k. This opens the possibility of studying numerically much more complicated Hamiltonians including higher-order correction terms.

By shifting the W_k's by irrelevant constants, one can arrange that the terms $c_k W_k$ be nonpositive starting from $k = 1$: $c_k W_k(U) \leq 0 \;\; \forall U$. The contribution of the terms $W_k(U), k \geq 1$ is estimated stochastically by introducing auxiliary fields σ_k, which can take two values: 0 and 1. We represent the probability e^{-H} in the form $P_0[U] * P_1[U,\sigma]$, where

$$P_0[U] = e^{-c_0 W_0(U)}$$

*A more detailed description of the NMC algorithm can be found in Ref.[11]

†Talk presented by T.B. at Lattice 2000, Bangalore, India

$$P_1[U,\sigma] = \prod_{k=1}^{m} \sum_{\sigma_k=0,1} [\delta_{\sigma_k,0} + \delta_{\sigma_k,1}(e^{-c_k W_k} - 1)] \quad (1)$$

$P_0[U] * P_1[U,\sigma]$ can be interpreted as the joint probability distribution for the original fields of the model and the new σ fields.

One can easily see why the introduction of auxiliary σ fields can be useful. Starting from the configuration $\{U_1, \sigma\}$, a candidate configuration $\{U_2\}$ distributed with the weight $P_0[U_2]$ is proposed, and accepted with probability

$$P_{acc} = \min\Big(1, \frac{\mathrm{P_1[U_2,\sigma]}}{\mathrm{P_1[U_1,\sigma]}}\Big) =$$
$$= \min\Big(1, \prod_{\mathrm{k\,:\,\sigma_k=1}} \frac{\mathrm{e^{-c_k W_k(U_2)} - 1}}{\mathrm{e^{-c_k W_k(U_1)} - 1}}\Big) \quad (2)$$

Since the terms $c_k W_k(U)$ contribute in P_{acc} only if $\sigma_k = 1$, the amount of computational work is greatly reduced if the configurations with $\sigma_k = 0$ are dominating. That is certainly the case when the coefficients $|c_k|$ are small: the probabilities for σ_k to be unity are negligible then: $p_{\sigma_k=1} = 1 - e^{c_k W_k(U)} \approx |c_k W_k(U)| \approx 0 \quad \text{if} \quad c_k \approx 0$.

We illustrate these ideas on a 5-loop perturbatively improved SU(2) gauge model in $4d$:

$$S = \sum_{i=1}^{5} c_i \frac{1}{m_i^2 n_i^2} S_{m_i,n_i}$$
$$S_{m_i,n_i} =$$
$$\sum_{x,\mu,\nu} \Big(-2\mathrm{sgn}\ (\mathrm{c_i}) - \frac{\mathrm{Tr}}{2}\ (\nu \square_{x\ \ \mu} + \nu \square_{x\ \ \mu}\ ) \Big) \quad (3)$$

where $(m_i, n_i) = (1,1), (2,2), (1,2), (1,3), (3,3)$ for $i = 1, \ldots, 5$ denote the planar, fundamental loops of size $m \times n$. The Gibbs factor is $\exp(-\frac{\beta}{2} S)$. One can construct a one-parameter set of actions which have no $\mathcal{O}(a^2)$ and $\mathcal{O}(a^4)$ corrections [8]:

$$c_1 = (19 - 55\ c_5)/9, \quad c_2 = (1 - 64\ c_5)/9,$$
$$c_3 = (-64 + 640\ c_5)/45, \quad c_4 = 1/5 - 2\ c_5 \quad (4)$$

Here we take $c_5 = 1/20$ and $\beta = 2.4$. We estimate the contribution of all loops except the plaquette stochastically. For each loop l of sort $2 \le i \le 5$ we introduce the auxiliary variable $\sigma_i(l) = 0, 1$; and rewrite the contribution of this loop to the Gibbs factor in the form

$$e^{-\frac{\beta}{2} S_{i,l}} = \sum_{\sigma_i(l)=0,1} [\delta_{\sigma_i(l),0} + \delta_{\sigma_i(l),1}(e^{-\frac{\beta}{2} S_{i,l}} - 1)] \quad (5)$$

After each N_i updates of the fields U we update the σ fields of sort i. We measure the average values of σ_i, as listed in Table 1. They are quite small, so one avoids the computation of almost all of the extended "staples" in the U update.

Table 1
Average value of σ_i field for each loop of sort i.

loop	1x2	1x3	2x2	3x3
$\langle\sigma\rangle$	0.0753	0.0199	0.0202	0.0018

The efficiency of NMC is estimated by comparing it with the non-noisy updating procedures (heatbath, overrelaxation) which are commonly used for the simulation of actions like (3).We label these usually applied techniques with the collective name "Usual Monte Carlo" (UMC), to contrast with NMC.

For NMC the average computational cost per one update of the U fields on the entire lattice in units of matrix (link) multiplications is

$$t_{tot}^{NMC} = t_U^{NMC} + \sum_{i=2}^{5} \frac{t_{\sigma_i}}{N_i} =$$
$$= t_U^{pl} + 6V \sum_{i=2}^{5} P_i s_i (\frac{1}{N_i} + P_i \langle\sigma_i\rangle) \quad (6)$$

where t_U^{pl} is the update cost for the elementary plaquette action, P_i is the perimeter of loop i, s_i is a symmetry factor: $s_i = 1$ for square loops and $s_i = 2$ for rectangular loops (see Ref.[11] for details). The computational cost for UMC per U update is approximately equal to the r.h.s. of expression (6) in the limit $N_i \to \infty$ and $\langle\sigma_i\rangle \to 1$:

$$t_{tot}^{UMC} = t_U^{pl} + 6V \sum_{i=2}^{5} P_i^2 s_i \quad (7)$$

The naive gain in efficiency from using NMC is given by the ratio between the costs (7) and (6):

$$r_{gain}^{naive} = \frac{t_U^{pl} + 6V \sum_{i=2}^{5} P_i^2 s_i}{t_U^{pl} + 6V \sum_{i=2}^{5} P_i s_i (\frac{1}{N_i} + P_i \langle\sigma_i\rangle)} \quad (8)$$

Table 2
Integrated autocorrelation times for average loop traces in units of U updates for the UMC algorithm (first column), and the NMC algorithm with different frequencies of σ updates for 1x2, 1x3, 2x2 and 3x3 loops (other columns). The last row presents the naive gain for the NMC algorithm (8). The real gain is given by eq.(9) and depends on the observable. From the values of τ_{int} below, the real gain is $\mathcal{O}(4-6)$.

number of U updates per 1 σ update	UMC no σ	1 for all	5 for all	5 for 1x2 15: 1x3,2x2 105 for 3x3	10 for all	10 for 1x2 30: 1x3,2x2 210 for 3x3	20 for all	30 for all	40 for all
τ_{int} (1x1)	0.7(1)	1.9(2)	2.3(1)	2.5(2)	3.1(2)	3.2(2)	4.3(4)	4.5(4)	3.8(2)
τ_{int} (1x2)	0.8(1)	2.6(3)	2.8(2)	3.2(2)	4.3(4)	3.9(3)	5.2(4)	5.6(5)	5.7(4)
τ_{int} (1x3)	0.8(1)	2.7(3)	2.8(2)	3.2(2)	4.3(4)	3.9(3)	5.1(4)	5.4(5)	5.4(4)
τ_{int} (2x2)	1.0(1)	3.4(5)	3.3(3)	3.7(3)	4.7(4)	4.7(3)	5.3(4)	5.8(6)	5.7(4)
τ_{int} (2x3)	1.4(3)	4.2(7)	3.8(3)	4.2(3)	5.5(5)	5.7(4)	5.7(5)	6.2(6)	6.3(5)
τ_{int} (3x3)	1.8(4)	5.0(8)	4.5(5)	5.0(5)	5.8(6)	6.4(5)	5.9(5)	6.3(6)	6.3(5)
r_{gain}^{naive}	1	7.2	14.8	18.3	17.9	20.6	18.3	20.8	21.1

One should also take into account the increase of autocorrelation times coming from the introduction of variables σ in the NMC algorithm, so the real gain is

$$r_{gain}^{real} \equiv r_{gain}^{naive} * \frac{\tau_{int}^{UMC}}{\tau_{int}^{NMC}} \quad . \tag{9}$$

where the second factor depends on the observable under consideration.

It is not practical to keep the same updating frequencies $1/N_i$ for all sorts i of loops. In order for the work in the σ and in the U updates coming from loops of sort i to remain comparable, one should keep the updating frequencies $1/N_i$ proportional to $\langle\sigma_i\rangle$:

$$\frac{1}{N_i} \;\sim\; P_i\langle\sigma_i\rangle \tag{10}$$

Due to the small influence of weakly coupled terms on the dynamics of the system, one can expect only insignificant changes in the autocorrelation behavior as N_i increases.

Table 2 gives an impressive demonstration of the benefits which come from using the NMC algorithm. For the runs where the updating frequencies for σ fields are adjusted as per eq.(10) we infer that the 'real gain' in computer time is of order $\mathcal{O}(4-6)$. One can expect a much greater gain for more complicated highly-improved actions.

Let us now speculate on possibilities to use our algorithm to simulate a Hamiltonian with a very large number of terms. A specific example we have in mind is the case of full QCD, where the measure is, for 2 flavors of Wilson quarks:

$$\frac{1}{Z} e^{-S_g(U)} \;\; det^2(\mathbf{1} - \kappa M(U)) \tag{11}$$

where S_g is the local gauge action, $M(U)$ is a hopping matrix connecting nearest neighbours on a $4d$ hypercubic grid, and Z normalizes the distribution. The determinant can be turned into $exp(Tr(Log(\mathbf{1} - \kappa M(U))))$, then the logarithm expanded around 1, giving the loop expansion of the measure above:

$$\frac{1}{Z} e^{-S_g(U) - 2\sum_{l=4}^{\infty} \frac{\kappa^l}{l} TrM(U)^l} \quad . \tag{12}$$

$TrM(U)^l$ can be represented as a sum over all closed non-backtracking loops of length l on the $4d$ hypercubic lattice. The number of types of contributing loops n_l is bounded by 7^l. Although this upper bound is not saturated, it is clear that the multiplicity of terms of a given length l grows exponentially:

$$n_l \sim F_1(l)\, \alpha^l; \qquad \alpha < 7 \tag{13}$$

where $F_1(l)$ is a rational function of l, and α^l is the leading exponential behavior of the number of loops of length l in the limit of large l. At first sight, it seems that sampling numerically the distribution (12) is a disastrous idea: the action

contains an infinite number of terms, of exponentially growing multiplicity. Instead, other strategies are being used, based on the transformation of the determinant (11) into a Gaussian integral.

Nevertheless, the coupling $\frac{\kappa^l}{l}$ decreases exponentially as l increases. Therefore, the auxiliary variables σ_l associated in our approach with various loops of length l will take value 0 almost always. One gets:

$$\langle \sigma_l \rangle \ \sim c_l \sim F_2(l)\ k^l \gamma^l \tag{14}$$

where the exponentially growing factor γ^l comes from the average trace of Dirac matrices along the loops of length l.

If one arranges the updating frequencies for each loop i as per eq.(10), one can expect that the average computer time needed for estimating the contribution of all loops of length l behaves as

$$t_l \sim \ n_l l^2 \langle \sigma_l \rangle \ \sim F(l) * (\alpha\gamma\kappa)^l \tag{15}$$

For $\kappa < \kappa_{ca} = \frac{1}{\alpha\gamma}$ the computational cost t_l decreases exponentially with l and the total computational cost of the algorithm $t = \sum_l t_l$ converges to some finite value. Our rough estimation from fitting n_l and average trace of Dirac matrices in the interval $4 \leq l \leq 12$ gives $\alpha \approx 5.4$; $\gamma \approx 1.4$, and therefore $\kappa_{ca} \approx 0.13$.

In the regime $\kappa < \kappa_{ca}$ we are in an interesting situation where the influence of very large loops is negligible because their associated coupling in the effective action is extremely small. Therefore, truncating the loop expansion above a certain order will introduce a statistically unobservable bias. Equivalently, one can freeze the associated σ variables at the value zero, or update them with arbitrarily low frequency. In spite of this extremely (or infinitely) slow dynamic mode of the σ's, the dynamics of the gauge fields are not affected. Note that the cost of our algorithm grows linearly with the volume V of the system. This is better than alternative approaches to the simulation of full QCD: Hybrid Monte Carlo (cost $\propto V^{5/4}$) and MultiBoson (cost $\propto V(LogV)^2$) [9]. In addition the stepsize, or typical change at each update of a gauge link U, does not seem restricted a priori for small quark mass, unlike in the two alternative approaches above.

A less speculative use of our algorithm for full QCD consists of truncating the loop expansion eq.(12) to some order l_{max}, and representing the higher orders with the MultiBoson approach [10]. This strategy, called "UV-filtered MultiBoson", has already been used successfully [7]. However, in Ref.[7] the loop expansion is truncated to its lowest term $l = 4$, because the exact evaluation of larger loops is too time-consuming. With our stochastic approach, these larger loops can be estimated at low cost. We expect this composite strategy to be particularly efficient.

Acknowledgments: T.B. was supported by INTAS 96-0370 and RBRF 99-01-00190.

REFERENCES

1. A.D. Kennedy, J. Kuti, Phys.Rev.Lett. **54** (1985) 2473-2476.
2. G. Bhanot, A.D. Kennedy, Phys.Lett. **B157** (1985) 70.
3. L. Lin, K.F. Liu and J. Sloan, Phys. Rev. **D61** (2000) 074505, hep-lat/9905033.
4. K. Symanzik, Nucl.Phys. **B226** (1983) 187.
5. M. Alford, W. Dimm, G.P. Lepage, G. Hockney, P.B. Mackenzie, Phys.Lett. **B361** (1995) 87-94, hep-lat/9507010.
6. P. Hasenfratz, F. Niedermayer, Nucl.Phys. **B414** (1994) 785-814, hep-lat/9308004.
7. Ph. de Forcrand, Nucl. Phys. B (Proc. Suppl.) **73** (1999) 822, hep-lat/9809145; C. Alexandrou, Ph. de Forcrand, M. D'Elia and H. Panagopoulos, Phys. Rev. **D61** (2000) 74503, hep-lat/9906029.
8. M. García Pérez, A. Gonzalez-Arroyo, J. Snippe and P. van Baal, Nucl. Phys. **B413** (1994) 535.
9. Ph. de Forcrand, hep-lat/9702009, Proceedings of the "International conference on multiscale phenomena and their simulation", (Oct. 1996, Bielefeld, Germany), World Scientific pub., p.169-178.
10. M. Lüscher, Nucl.Phys. **B418** (1994) 637-648, hep-lat/9311007;
    A. Slavnov, hep-th/9611154.
11. T. Bakeyev, Ph. de Forcrand, hep-lat/0008006, to appear in Phys. Rev. D