IMPROVED HEATBATH METHOD FOR MONTE CARLO CALCULATIONS IN LATTICE GAUGE THEORIES

A.D. KENNEDY

Institute for Theoretical Physics, University of California, Santa Barbara, CA 93106, USA

and

B.J. PENDLETON [1]

Physics Department, University of California, Santa Barbara, CA 93106, USA

Received 6 February 1985

Heatbath methods give more rapid equilibration and shorter correlation times for the Markov processes which generate lattice gauge configurations in Monte Carlo computations. We introduce an improved method for generating configurations according to the SU(2) heatbath distribution, which is also a central component of the SU(3) "quasi-heatbath" method of Cabibbo and Marinari. For problems of interest our method offers an improvement of a factor of about four.

1. Introduction. The numerical study of lattice gauge theories using Monte Carlo methods offers us hope of being able to calculate the properties of QCD for values of the renormalized coupling which are not small. In order to do so, however, it is becoming increasingly clear that we shall have to study very large lattices with very high statistics. As such computations are close to the bounds of what is possible using the current generation of computers, it is most important to find new methods which accelerate such Monte Carlo techniques.

The aim of Monte Carlo methods is to measure the average of some "physical" operator Ω over the space of gauge-field configurations $\{\mathcal{U}\}$ with respect to the measure given by the exponential of the action S:

$$\langle\Omega\rangle = Z^{-1}\int \mathrm{d}\mathcal{U}\,\exp[-\beta S(\mathcal{U})]\,\Omega(\mathcal{U})\,, \qquad Z = \int \mathrm{d}\mathcal{U}\,\exp[-\beta S(\mathcal{U})]\,. \tag{1}$$

For an SU(n) lattice gauge theory, the gauge configurations $\mathcal{U}$ are represented by an SU(n)-valued link variable U_ℓ on each link of a hypercubic lattice. Once this has been done, provided the lattice is large enough to preclude undesirable finite-size effects, and that the lattice spacing a is small enough that the system is "close to the continuum", we have a measurement to within known statistical errors of the expectation value $\langle\Omega\rangle$ for the continuum theory. Physically interesting operators Ω are, for example, the Wilson loop which measures the inter-quark potential; correlation functions whose decay determines hadron masses; or the Polyakov loop, which may be used as an order parameter to study the phase structure of the theory.

The number of dynamical variables U required to specify a lattice configuration $\mathcal{U}$ is large (typically of the order of 10^5) – the lattice is supposed to approximate a system with an infinite number of degrees of freedom – and thus the dimensionality of the space of configurations over which the functional integral of eq. (1) has to be taken is truly vast. Monte Carlo techniques are the only feasible means of performing such high dimensional integrals, which they do by selecting a sequence of configurations $\mathcal{U}$ with probability

[1] Present address: Physics Department, The University, King's Buildings, Edinburgh, EH9 3JZ, UK.

0370-2693/85/$ 03.30
(North-Holland Physics Publishing Division)

$$P_c(\mathcal{U})\, d\mathcal{U} = Z^{-1} \exp[-\beta S(\mathcal{U})]\, d\mathcal{U}\,, \tag{2}$$

and measuring the average $\bar{\Omega}$ over a sufficiently large number T of elements of this sequence. The statistical fluctuations in the mean value for Ω falls as $\sqrt{1/T}$ (T is the number of independent configurations generated) as $T \to \infty$. How large T has to be in practice depends upon the detailed dynamics of the theory being considered, and has to be determined empirically: it is found that at least tens of thousands of configurations have to be used in many cases.

The method used to generate this sequence of configurations is to construct a Markov process which has P_c as a fixed point and is ergodic; from these two facts it follows that the distribution of $\mathcal{U}$'s which are produced by this process will converge to P_c. The time required to relax to this distribution is system-dependent. It is usual to update the configurations by changing one link at a time, selecting the replacement link with probability $P_c(\mathcal{V})$ where $\mathcal{V}$ is the configuration obtained from $\mathcal{U}$ by changing the link $U_\ell \to V_\ell$. Although this single link update is not ergodic, after one complete sweep through the lattice during which every link has been updated at least once (in any order whatsoever) ergodicity is established.

The fundamental step in Monte Carlo computations is thus seen to be that of updating a single link. There are two common techniques of doing this, the Metropolis algorithm [1] and the heatbath algorithm. In the former case we choose the new link V from SU(n) subject to the conditions that V and $V^\dagger$ must be equiprobable, and that any SU(3) matrix can be reached after enough steps (ergodicity), and then accept it with probability [‡1]

$$\begin{aligned} P_a(V \leftarrow U) &= 1 && \text{if } P_c(V) \geqslant P_c(U)\,, \\ &= P_c(V)/P_c(U) && \text{if } P_c(V) < P_c(U)\,, \end{aligned} \tag{3}$$

and reject it otherwise. This procedure has P_c as a fixed point because it satisfies detailed balance, namely

$$P_a(V \leftarrow U)\, P_c(U) = P_a(U \leftarrow V)\, P_c(V)\,. \tag{4}$$

The heatbath algorithm is just a generic term for any method which selects the new link V with probability $P_c(V)$ completely independently of the old link U.

The advantage of the heatbath method is that the new configuration $\mathcal{V}$ is more independent of the old one $\mathcal{U}$ than is the case for the Metropolis algorithm. Of course, in both cases the correlations are very strong because $\mathcal{U}$ and $\mathcal{V}$ differ by only one link, but nevertheless the difference is significant. It is not hard to see why Metropolis-generated configurations are more correlated: for the heatbath V_ℓ depends on the links neighbouring ℓ, but not on U_ℓ. For the Metropolis case eq. (3) depends explicitly upon U; furthermore V is often chosen from a subset of SU(n) close to U_ℓ in order to increase the acceptance rate of the Metropolis algorithm.

The next question to address is whether we can construct a heatbath algorithm which can be implemented to run as quickly as the Metropolis algorithm. For the case of SU(2) this has been done by Creutz [2], and for SU(n) a method which we shall call the quasi-heatbath algorithm was introduced by Cabibbo and Marinari [3]. The quasi-heatbath method consists of performing heatbath updates on a sequence of SU(2) subgroups of SU(n): empirically it is observed that for the most interesting case of SU(3) the correlation time of quasi-heatbath generated configurations is about half that of ten-hit Metropolis ones [3,4]. When the gauge group is SU(3), as well as for SU(2) itself, it is thus most important to be able to generate SU(2) matrices according to a prescribed probability distribution P_c.

2. Mathematical preliminaries, introduction. The action $S(\mathcal{U})$ can take various forms, the simplest of which is Wilson's action

[‡1] We write $P_c(U)$ instead of $P_c(\mathcal{U})$ when we wish to concentrate on configurations which differ only a link ℓ.

$$S(\mathcal{U}) = \sum_{\square} \left(1 - n^{-1} \operatorname{Re} \operatorname{Tr} \prod_{\ell \in \partial \square} U_\ell\right), \tag{5}$$

where the sum extends over all the plaquettes $\square$ on the lattice, and the product over the U matrices is taken around the boundary $\partial\square$ of plaquette $\square$. The locality of this action means that as far as computing the relative probabilities of configurations differing by only one link ℓ is concerned, it suffices to compute only the contributions to S from those plaquettes in the coboundary of ℓ. The Wilson action can be cast into the form $\beta S(\mathcal{U})$ = constant $-$ β' Re Tr Σh, with β' a constant, Σ a complex 2×2 matrix made from the sum of products of link variables neighbouring ℓ, and h the SU(2) matrix associated with link ℓ which is to be updated. The problem, therefore, is to generate an SU(2) matrix according to the distribution

$$Q(h)(\mathrm{d}h) \propto \exp(\beta' \operatorname{Re} \operatorname{Tr} \Sigma h)(\mathrm{d}h)\,, \tag{6}$$

where $(\mathrm{d}h)$ is the SU(2)-invariant Haar measure.

We may simplify the problem by parametrizing the matrices in terms of Pauli σ matrices,

$$h = h_0 \mathbf{1} + \mathrm{i}\boldsymbol{h} \cdot \sigma \quad (h_\alpha \in \mathbf{R}, \quad h_0^2 + \boldsymbol{h} \cdot \boldsymbol{h} = 1)\,, \qquad \Sigma = \Sigma_0 \mathbf{1} + \mathrm{i}\boldsymbol{\Sigma} \cdot \sigma \quad (\Sigma_\alpha \in \mathbf{C})\,. \tag{7}$$

Now,

$$\operatorname{Re} \operatorname{Tr} \Sigma h = \operatorname{Re} \operatorname{Tr}[(\Sigma_0 \mathbf{1} + \mathrm{i}\boldsymbol{\Sigma} \cdot \sigma)(h_0 \mathbf{1} + \mathrm{i}\boldsymbol{h} \cdot \sigma)] = \operatorname{Re} \operatorname{Tr}[(\Sigma_0 h_0 - \boldsymbol{\Sigma} \cdot \boldsymbol{h})\,\mathbf{1} + (\ldots) \cdot \sigma] = \xi \operatorname{Tr} uh\,, \tag{8}$$

where $u \in \mathrm{SU}(2)$, and $\xi u_0 = \operatorname{Re} \Sigma_0$, $\xi \boldsymbol{u} = \operatorname{Re} \boldsymbol{\Sigma}$. Consider $\Sigma - \Sigma^\dagger = (\Sigma_0 - \Sigma_0^*)\,\mathbf{1} + \mathrm{i}(\boldsymbol{\Sigma} + \boldsymbol{\Sigma}^*) \cdot \sigma$ and $\operatorname{Tr} \Sigma^\dagger = 2\Sigma_0^*$, hence

$$\Sigma - \Sigma^\dagger + \mathbf{1} \operatorname{Tr} \Sigma^\dagger = 2 \operatorname{Re}(\Sigma_0)\,\mathbf{1} + 2\mathrm{i} \operatorname{Re}(\boldsymbol{\Sigma}) \cdot \sigma = 2\xi u_0 \mathbf{1} + 2\mathrm{i}\xi \boldsymbol{u} \cdot \sigma = 2\xi u\,, \tag{9}$$

so we have

$$4\xi^2 = \operatorname{Det}[\Sigma - \Sigma^\dagger + \mathbf{1} \operatorname{Tr} \Sigma^\dagger]\,, \quad u = (1/2\xi)(\Sigma - \Sigma^\dagger + \mathbf{1} \operatorname{Tr} \Sigma^\dagger)\,, \tag{10}$$

and from eq. (6) $Q(h)(\mathrm{d}h) \propto \exp(\beta' \xi \operatorname{Tr} uh)(\mathrm{d}h)$.

Writing $a \equiv uh \in \mathrm{SU}(2)$ we may parametrize a as $a = a_0 \mathbf{1} + \mathrm{i}\boldsymbol{a} \cdot \sigma$ with $(a_\alpha \in \mathbf{R}, a_0^2 + \boldsymbol{a} \cdot \boldsymbol{a} = 1)$, and Haar measure on SU(2) is $(\mathrm{d}a) = \mathrm{d}a_0\, \mathrm{d}^3 a_i\, \delta(1 - a_0^2 - \boldsymbol{a} \cdot \boldsymbol{a})$. In polar coordinates

$$(\mathrm{d}a) = \mathrm{d}a_0\, \mathrm{d}r\, \mathrm{d}\theta\, \mathrm{d}\varphi [\partial(a_1, a_2, a_3)/\partial(r, \theta, \varphi)]\, \delta(1 - a_0^2 - r^2) = \tfrac{1}{2}(1 - a_0^2)^{1/2}\, \mathrm{d}a_0\, \mathrm{d}r\, \mathrm{d}\theta\, \mathrm{d}\varphi \sin\theta\, \delta(r - (1 - a_0^2)^{1/2})\,, \tag{11}$$

which means we must generate a_0 with the distribution

$$P(a_0)\, \mathrm{d}a_0 \propto (1 - a_0^2)^{1/2} \exp(2\beta' \xi a_0)\, \mathrm{d}a_0\,, \tag{12}$$

and then generate a_i uniformly on a two-sphere of radius $(1 - a_0^2)^{1/2}$.

3. Heatbath algorithms

3.1. Creutz's algorithm [2]. This algorithm generates pseudo-random numbers a_0 distributed according to

$$P(a_0)\, \mathrm{d}a_0 = N^{-1}(1 - a_0^2)^{1/2} \exp(\alpha a_0)\, \mathrm{d}a_0 \quad (-1 \leqslant a_0 \leqslant 1, \quad \alpha > 0)\,, \tag{13}$$

where the normalization constant is

$$N = \int_{-1}^{1} \mathrm{d}a_0 (1 - a_0^2)^{1/2} \exp(\alpha a_0) = \int_0^\pi \mathrm{d}t \sin^2 t \exp(\alpha \cos t) = (\pi/\alpha) I_1(\alpha) \sim (\pi/2\alpha^3)^{1/3}\, \mathrm{e}^\alpha \quad (\alpha \to \infty)\,. \tag{14}$$

Creutz's method of generating such a_0 values is to generate them with an exponential distribution and then to fix-up the square root factor using an accept/reject procedure. Unfortunately, for cases of interest α is quite large

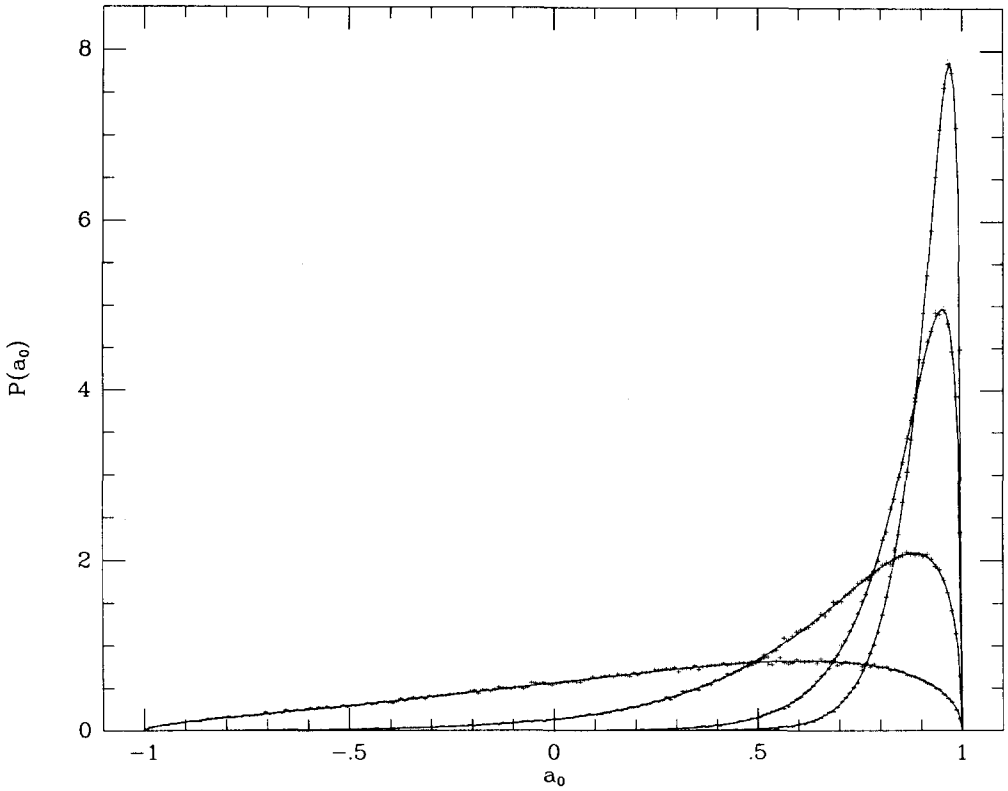

Fig. 1. Distribution of a_0 values for $\alpha = 1, 4, 10$, and 16. The solid curves show eq. (13), and the error bars show the measured distribution of a_0 values generated by the new algorithm.

(commonly of order 16), hence the exponentially distributed a_0 values cluster strongly about 1, which is where the prefactor is small, thus the rejection rate is high ($\approx 75\%$): this can easily be seen from fig. 1. We shall, therefore, introduce an algorithm which generates a good approximation to eq. (13) for $a_0 \approx 1$, and which will have a correspondingly high acceptance rate.

3.2. New algorithm. Let us first rewrite eq. (13) in terms of the more convenient variable $\delta = (1 - a_0)^{1/2}$:

$$P'(\delta)\,\mathrm{d}\delta = N'^{-1}(1 - \tfrac{1}{2}\delta^2)^{1/2}\,\delta^2 \exp(-\alpha\delta^2)\,\mathrm{d}\delta \quad (0 \leqslant \delta \leqslant \sqrt{2}), \tag{15}$$

which, for large α, is peaked near $\delta = 0$. We shall, therefore, generate the distribution

$$P''(\delta)\,\mathrm{d}\delta = N''^{-1}\,\delta^2 \exp(-\alpha\delta^2)\,\mathrm{d}\delta \quad (0 \leqslant \delta < \infty) \tag{16}$$

"analytically", and then impose the factor $(1 - \frac{1}{2}\delta^2)^{1/2}$ by an accept/reject test. The distribution eq. (16) is of the form of a gaussian weighted by a power of δ, and is readily generated by manipulating gaussian-like distributions. First note that if X is a uniformly distributed random variable lying in the unit interval, then the quantity $\xi = [-(\ln X)/\alpha]^{1/2}$ has the distribution

$$P_\xi(\xi) = \int_0^1 \mathrm{d}X\,\delta(\xi - [-(\ln X)/\alpha]^{1/2}) = 2\xi\alpha \exp(-\alpha\xi^2) \quad (0 \leqslant \xi \leqslant \infty). \tag{17}$$

Now that we can generate numbers according to eq. (17), we can also generate them with a gaussian distribution by the following trick. Take another uniform random value X' in $[0,1]$, and evaluate $\rho = \chi \cos 2\pi X'$, where χ is distributed according to P_ξ. The distribution of ρ is thus

$$P_\rho(\rho) = \int_0^\infty d\chi \int_0^1 dX' P_\xi(\chi)\, \delta(\rho - \chi \cos 2\pi X') = \int_0^\infty d\chi \int_0^{2\pi} d\theta (\chi\alpha/\pi) \exp(-\alpha\chi^2)\, \delta(\rho - \chi\cos\theta), \tag{18}$$

with $\theta = 2\pi X'$; viewing χ and θ as polar coordinates it is natural to rewrite the integral in cartesian form,

$$P_\rho(\rho) = \int_{-\infty}^\infty da \int_{-\infty}^\infty db (\alpha/\pi) \exp[-\alpha(a^2 + b^2)]\, \delta(\rho - a), \tag{19}$$

where $a = \chi\cos\theta$, $b = \chi \sin\theta$, and so

$$P_\rho(\rho) = \int_{-\infty}^\infty db(\alpha/\pi)\exp(-\alpha b^2)\exp(-\alpha\rho^2) = \sqrt{\alpha/\pi}\exp(-\alpha\rho^2) \quad (-\infty < \rho < \infty). \tag{20}$$

As an additional bonus, we see that $\rho' = \chi \sin 2\pi X'$ is also distributed according to P_ρ and is independent of ρ, so the computation of χ and X' gives us two gaussian random numbers ρ and ρ'.

If we combine one of these gaussian values with ξ from eq. (17) we can obtain a random number δ distributed according to eq. (16): Let $\delta = (\xi^2 + \rho^2)^{1/2}$, then

$$P''(\delta) = \int_0^\infty d\xi \int_{-\infty}^\infty d\rho\, P_\xi(\xi) P_\rho(\rho)\, \delta(\delta - (\xi^2 + \rho^2)^{1/2})$$

$$= \int_0^\infty d\xi \int_{-\infty}^\infty d\rho\, 2\alpha\xi\sqrt{\alpha/\pi}\exp[-\alpha(\xi^2 + \rho^2)]\, \delta(\delta - (\xi^2 + \rho^2)^{1/2}), \tag{21}$$

or in polar coordinates $r = (\xi^2 + \rho^2)^{1/2}$, $\phi = \tan^{-1}(\rho/\xi)$:

$$P''(\delta) = \int_0^\infty dr \int_{-\pi/2}^{\pi/2} d\phi\, 2\alpha r^2\sqrt{\alpha/\pi}\cos\phi\exp(-\alpha r^2)\,\delta(\delta - r)$$

$$= 2\alpha\sqrt{\alpha/\pi}\int_{-\pi/2}^{\pi/2} d\phi \cos\phi\, \delta^2 \exp(-\alpha\delta^2) = 4\alpha\delta^2\sqrt{\alpha/\pi}\exp(-\alpha\delta^2), \tag{22}$$

in accord with eq. (16).

To get from eq. (16) to eq. (15) we just generate δ' with probability $P''(\delta')$ and accept it with probability $(1 - \frac{1}{2}\delta'^2)^{1/2}$ if $\delta'^2 \leqslant 2$, and always reject it if $\delta'^2 \geqslant 2$. This process may be carried out by generating an extra uniform random number X'' between zero and one, and accepting δ' if $X''^2 \leqslant 1 - \delta'^2/2$. The validity of this procedure is easily verified:

$$P'(\delta) = \frac{\int_0^\infty d\delta' \int_0^1 dX''\, P''(\delta')\,\theta(1 - \frac{1}{2}\delta'^2 - X''^2)\,\delta(\delta - \delta')}{\int_0^\infty d\delta' \int_0^1 dX''\, P''(\delta')\,\theta(1 - \frac{1}{2}\delta'^2 - X''^2)}, \tag{23}$$

where the denominator corrects the normalization of $P'(\delta)$ because the procedure is repeated for new trial δ' until

one is accepted. Calling the denominator N''' we have

$$P'(\delta) = N'''^{-1} \int_0^1 dX'' P''(\delta)\, \theta(1 - \tfrac{1}{2}\delta^2 - X''^2) = N'''^{-1} \int_0^{(1-\delta^2/2)^{1/2}} dX' P''(\delta) \quad \text{for } \delta^2 < 2,$$

$$= 0 \quad \text{for } \delta^2 \geqslant 2,$$

$$= N'''^{-1}(1 - \tfrac{1}{2}\delta^2)^{1/2} P''(\delta) \quad (0 \leqslant \delta \leqslant \sqrt{2}), \tag{24}$$

in agreement with eq. (15). Finally, of course, we set $a_0 = 1 - \delta^2$.

4. Summary of the new algorithm.

A. Generate two uniformly distributed pseudo-random numbers R and R' in the unit interval;
B. Set $X \leftarrow -(\ln R)/\alpha$, $X' \leftarrow -(\ln R')/\alpha$;
C. Set $C \leftarrow \cos^2(2\pi R'')$, with R'' another uniform random number in $[0,1]$;
D. Let $A \leftarrow XC$;
E. Let $\bar{\delta} \leftarrow X' + A$;
F. If $R'''^2 > 1 - \frac{1}{2}\bar{\delta}$, for R''' pseudo-random and uniform in $(0,1]$, go back to step A;
G. Set $a_0 \leftarrow 1 - \bar{\delta}$.

Note that in step D setting $B \leftarrow X - A$ and using B in place of A in step E will generate a second independent a_0 value.

5. Acceptance probabilities. In order to make a quantitative comparison of the two algorithms, we must estimate their acceptance rates. Using Creutz's algorithm the probability of generating a candidate a_0 is

$$P_C(a_0) = N_C^{-1} \exp(\alpha a_0) \quad (-1 \leqslant a_0 \leqslant 1), \tag{25}$$

hence the probability of accepting a proffered a_0 is

$$P_C^{\text{acc}} = \int_{-1}^{1} da_0 \int_0^1 dR\, P_C(a_0)\, \theta(1 - a_0^2 - R^2) = N_C^{-1} \int_{-1}^{1} da_0 (1 - a_0^2)^{1/2} \exp(\alpha a_0) = N/N_C$$

$$= \pi I_1(\alpha)/(2 \sinh \alpha) \sim \sqrt{\pi/2\alpha} \quad (\alpha \to \infty), \tag{26}$$

with N from eq. (14).

For the new algorithm the acceptance probability is

$$P_{\text{new}}^{\text{acc}} = \int_0^\infty d\delta \int_0^1 dR\, P''(\delta)\, \theta(1 - \tfrac{1}{2}\delta^2 - R^2) = \sqrt{2\pi\alpha}\, \exp(-\alpha)\, I_1(\alpha)$$

$$\sim 1 - 3/8\alpha - 15/128\alpha^2 - 105/1024\alpha^3 + \ldots \quad (\alpha \to \infty). \tag{27}$$

For a typical value, $\alpha = 16$, we find $P_C^{\text{acc}} = 31\%$, whereas $P_{\text{new}}^{\text{acc}} = 97.6\%$. More generally the acceptance rates are shown in table 1.

We can see from table 1 that Creutz's algorithm does better for very small values of α; indeed, if we define the relative acceptance rate

$$R(\alpha) = P_{\text{new}}^{\text{acc}}(\alpha)/P_C^{\text{acc}}(\alpha) = \sqrt{2\alpha/\pi}\,[1 - \exp(-2\alpha)], \tag{28}$$

then $R(\alpha \to \infty) \to \infty$, $R(1) = 0.69$, $R(0) = 0$, and $R(1.6847) = 1$. The reason why Creutz's algorithm does better for

Table 1
Acceptance probabilities for the Creutz and new algorithms.

α	$e^{-\alpha}I_1(\alpha)$	$P_C^{acc}(\alpha)$	$P_{new}^{acc}(\alpha)$
1	0.207 910 415 4	0.7554	0.5212
4	0.178 750 839 4	0.5618	0.8961
10	0.121 262 681 4	0.3810	0.9612
16	0.097 349 614 7	0.3058	0.9761

$\alpha < 1.6847$ is that the new algorithm generates $0 < \xi, \rho < \infty$, rather than $0 < \xi, \rho \leqslant \sqrt{2}$, which would be more desirable. Because $I_1(z) \sim z/2$ as $z \to 0$ we have $P_{new}^{acc} \sim \sqrt{\pi/2}\alpha^{3/2}$ for $\alpha \to 0$, so the convolution of P_{new}^{acc} with reasonable distributions for α is likely to be non-zero (and large). If required, ξ and ρ may be generated in the desired range, although this is not entirely straightforward for a gaussian distribution.

Recent QCD Monte Carlo computations which attempt to probe the physically relevant continuum limit of lattice SU(3) gauge theory have been performed for couplings $g^2 \leqslant 1$, and in this regime the values of α obtained are typically 16 or more: therefore Creutz's algorithm is superior only for extremely strong coupling.

We would like to thank Julius Kuti, Doug Eardley, and Bon Sugar for their help, advice and support. This work was supported in part by the National Science Foundation under Grant No. PHY77-27084 and PHY83-13324, NASA, and the United Kingdom SERC.

References

[1] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, J. Chem. Phys. 21 (1953) 1087.
[2] M. Creutz, Phys. Rev. D21 (1980) 2308.
[3] N. Cabibbo and E. Marinari, Phys. Lett. 119B (1982) 387.
[4] K.C. Bowler and B.J. Pendleton, Nucl. Phys. B230 [FS10] (1984) 109.