13 December 1996

ELSEVIER

Chemical Physics Letters 263 (1996) 488–494

CHEMICAL PHYSICS LETTERS

Multigrid method versus staging algorithm

Wolfhard Janke [a], Tilman Sauer [b,c]

[a] *Institut für Physik, Johannes Gutenberg-Universität Mainz, 55099 Mainz, Germany*
[b] *Institut für Theoretische Physik, Freie Universität Berlin, 14195 Berlin, Germany*
[c] *Max-Planck-Institut für Wissenschaftsgeschichte, Wilhelmstrasse 44, 10117 Berlin, Germany*

Received 29 May 1996; in final form 15 October 1996

Abstract

We present a comparison of the performance of two non-local update algorithms for path integral Monte Carlo (PIMC) simulations, the multigrid Monte Carlo method and the staging algorithm. Looking at autocorrelation times for the internal energy we show that both refined algorithms beat the slowing down which is encountered for standard local update schemes in the continuum limit. We investigate the conditions under which the staging algorithm performs optimally and give a brief discussion of the mutual merits of the two algorithms.

1. Introduction

A well-known problem for path integral Monte Carlo simulations [1–4] using standard local update schemes such as the Metropolis algorithm is a severe slowing down in the continuum limit. By this one means that successively generated configurations in the Monte Carlo process are highly correlated, a phenomenon signalized by large autocorrelation times in the simulation. This slowing down problem is very similar to the critical slowing down encountered in simulations of statistical or lattice field theoretical systems near phase transitions of second order [5,6]. In both cases it is the diverging spatial correlations (in lattice units) which are the physical origin of the inefficiency of local update schemes.

In many of the applications in statistical physics and lattice field theory the critical slowing down problem can be overcome by the use of multigrid techniques [7–13]. These are non-local update schemes where updates are performed on a variety of length scales in order to sample most efficiently long wavelength fluctuations. In a recent Letter [14] we have shown that thanks to the generality of their definition multigrid techniques can be transferred to simulations of Euclidean path integrals. We explicitly demonstrated that also for these systems slowing down is almost completely reduced. Another advantage of their general definition is that multigrid techniques can also be combined with reweighting schemes such as multicanonical sampling in order to further reduce autocorrelation times in the presence of tunneling barriers [15,16].

For Monte Carlo simulations of path integrals another successful non-local update algorithm, which is somewhat similar in spirit but which works technically in a rather different manner, is known under the name of "staging" [17–20]. Although this algorithm also significantly reduces the slowing down of simulations in the continuum limit, to our knowledge no

PII S0009-2614(96)01231-6

detailed analysis of autocorrelation times for the staging algorithm exists in the literature. Also, we know of no work which gives a comparison of the two algorithms. In this Letter, we intend to fill this gap by reporting autocorrelation times for a standard energy estimator and two sample potentials, a convex one and a double well, employing both the staging algorithm and the multigrid method.

2. The algorithms

In the path integral Monte Carlo approach the quantum partition function $\mathcal{Z}$ at inverse temperature β is approximated as a discretized path integral consisting of L "beads" [21]

$$\mathcal{Z}_L(\beta) = \left[\prod_{i=1}^{L} \int \frac{\mathrm{d}x_i}{A}\right] \exp\{-\mathcal{A}_L\}, \tag{1}$$

with an action

$$\mathcal{A}_L = \epsilon \sum_{i=1}^{L} \left[\frac{1}{2}\left(\frac{x_i - x_{i-1}}{\epsilon}\right)^2 + V(x_i)\right], \tag{2}$$

where V is a potential to be specified below, $A = \sqrt{2\pi\epsilon}$, $\epsilon = \beta/L$, $x_0 = x_L$, and $\hbar = k_B = 1$. The original partition function $\mathcal{Z}$ is then recovered in the continuum limit $L \to \infty$ with $L\epsilon = \beta$ held fixed.

The basic idea of the multigrid approach [7–14] is to perform non-local updates of the x_i by working on a set of successively coarser discretizations of the time axis ("grids") in order to take into account long wavelength fluctuations of the paths more efficiently. The technical details of this algorithm have been described in detail elsewhere, see e.g. Ref. [14]. The algorithm requires the definition of a set of coarser grids and a prescription to set up coarsened actions on these grids. Given the grids and the corresponding actions the multigrid algorithm recursively defines a sequence in which the variables on the various grids are updated and interpolated back onto the original grid. Two of those sequences which have extensively been used and studied are known under the name of V-cycle and W-cycle.

The basic idea of the staging method is to rewrite the quantum statistical partition function in such a way that a sequence of j adjacent variables can be updated independently. Technical details can again be found in the literature [17–20]. The algorithm implies the random selection of the end points of some segment of the path of length j and to perform a change of variables which allows an elimination of the nearest neighbor coupling stemming from the kinetic energy. For the variables of the staging segment the partition function hence reduces to a collection of independent oscillators moving in an external potential which depends on the transformation of the variables. The staging variables may then be updated using Gaussian random numbers and a Metropolis like acceptance rule for the external potential. In contrast to the multigrid scheme the staging algorithm only allows for one single tunable parameter, namely the length j of the staging segment.

3. Results

As in Ref. [14] we studied two qualitatively different potential shapes, typical for a wide range of physical phenomena. A convex potential (CP), given by

$$V = 0.5x^2 + x^4,$$

is relevant for studying fluctuations around a unique minimum, and a double-well potential (DW), given by

$$V = -0.5x^2 + 0.04x^4,$$

is relevant for studying tunneling phenomena. We have simulated the path integral (1), (2) for grids of size $L = 2^3 = 8$ up to $2^{10} = 1024$, using both the multigrid scheme with Metropolis update and the staging algorithm. In all our simulations β was equal to 10.

An observable of central importance is the internal energy. For path integral Monte Carlo simulations two different estimators for this observable are well-known [22–24]. Straightforward application of the definition of the internal energy

$$U_L = -\partial \ln \mathcal{Z}_L / \partial\beta$$

leads to an estimator

$$U_L = U_L^{\mathrm{kin}} + \frac{1}{L}\sum_{i=1}^{L} \langle V(x_i)\rangle;$$

with

$$U_L^{\mathrm{kin}} = \frac{L}{2\beta} - \frac{1}{L}\sum_{i=1}^{L} \left\langle \frac{1}{2}\left(\frac{x_i - x_{i-1}}{\epsilon}\right)^2 \right\rangle.$$

We call U_L the "kinetic" estimator since it explicitly measures the kinetic part U_L^{kin} of the energy [22]. Another function of the Monte Carlo configurations is given by

$$\frac{1}{L}\sum_{i=1}^{L}\left[\tfrac{1}{2}x_i V'(x_i)+V(x_i)\right],$$

and its expectation value $\langle\ldots\rangle$ with respect to (1) also estimates the internal energy [23,24]. We call it the "virial" estimator since it can be obtained by invoking the virial theorem. As usual expectation values are approximated in the simulation by mean values of the estimators over the Monte Carlo sequence. For update schemes which reduce slowing down in the continuum limit the virial estimator is a priori superior in this limit since it has a constant variance whereas the variance of the kinetic estimators grows linearly with L. Clearly, the best estimator would be some linear combination of the two estimators which must take into account the individual variances *and* the covariance of the two estimators [25,26]. For our comparison of multigrid and staging update schemes we will here, however, only look at the virial estimator. A careful discussion of the problem of optimized energy estimation in path integral Monte Carlo simulations is beyond the scope of this Letter [26].

Since the main focus of this Letter is to present a comparison of the performance of the two different non-local update schemes we have taken care to measure precisely the autocorrelation times obtained for the multigrid and staging algorithms. Explicitly, the autocorrelation function $A(k)$ of an observable $\mathcal{O}$ is defined by [27]

$$A(k)=\frac{\langle\mathcal{O}_i\mathcal{O}_{i+k}\rangle-\langle\mathcal{O}_i\rangle^2}{\langle\mathcal{O}_i^2\rangle-\langle\mathcal{O}_i\rangle^2}, \tag{3}$$

where $\mathcal{O}_i$ stands short for the ith measurement of $\mathcal{O}$. The autocorrelation time τ_0 then follows from the asymptotic behaviour for large k, $A(k)\propto\exp(-k/\tau_0)$. The integrated autocorrelation time τ, defined by the area under the autocorrelation function of this observable,

$$\tau\equiv\frac{1}{2}+\sum_{k=1}^{\infty}A(k), \tag{4}$$

usually behaves qualitatively as τ_0. It can be shown to enter in the estimate for the statistical error of mean values as

$$\Delta\hat{\mathcal{O}}=\sqrt{2\tau}\sqrt{\sigma^2/N_{\mathrm{m}}},$$

where σ^2 is the observable's variance and N_{m} the number of measurements used to compute the mean value $\hat{\mathcal{O}}$. The effective statistics is thus reduced to $N_{\mathrm{eff}}=N_{\mathrm{m}}/2\tau$. Since $A(k)$ becomes very noisy for large k, the upper bound in (4) is usually cut off self-consistently at $f\tau$ with $f\approx 6\ldots 8$ [16,27]. In our analysis we used $f=8$, and all error bars for these autocorrelation times were obtained by jackkniving [28,29] the data with 100 blocks.

In our simulations we performed $N_{\mathrm{m}}=100\,000$ sweeps after discarding 5000 sweeps for thermalization. Measurements were taken after each sweep. In the case of the multigrid algorithm we understand a "sweep" to mean a complete V- or W-cycle respectively [14]. Similar to an ordinary Metropolis sweep, the computational work of a V-cycle is directly proportional to the number of variables L, while it grows slightly faster with a factor of $L\log L$ for a W-cycle on one-dimensional grids. In the case of the staging algorithm, a "sweep" means $\mathrm{int}[L/(j-1)]$ calls to the staging routine which moves $j-1$ adjacent variables at each call. Notice that the above definition of a staging sweep in general implies updates of less than L variables. We have therefore rescaled the measured autocorrelation times by a factor $\{\mathrm{int}[L/(j-1)]\}/[L/(j-1)]$.

For the staging algorithm the length j of the segment which is to be updated in the staging routine is the only parameter which may be tuned in order to optimize the performance. A rule of thumb here says that it should be set to such a value that the acceptance rate is 40% [20]. For the above definition of a staging pass the amount of numerical work to be done does not depend significantly on the parameter j. The only criterion for the optimal choice of j_{opt} therefore is the integrated autocorrelation time for the observable at hand. Fig. 1a shows the integrated autocorrelation times as a function of j for different values of L using the virial estimator for the internal energy of the convex potential. We notice that there certainly is an optimal value j_{opt} for each L even though the minima are quite shallow and autocorrela-

tion times do not differ very much for neighbouring values of j_{opt}. We also see that the autocorrelation times for j_{opt} stay roughly constant, independent of L. It is thus already seen qualitatively that the staging algorithm eliminates slowing down completely.

As discussed above, it is the divergence of the correlation length measured in lattice units which is the cause of the slowing down problem in the continuum limit. The staging algorithm overcomes this problem by updating a whole segment of the path in a completely decorrelating manner. One would hence expect that the efficiency of the staging algorithm depends on the ratio of the length of the staging segment to the correlation length along the path. Since the latter scales with L one would therefore expect that the optimal choice of j should also scale with L. In Fig. 1b we have therefore plotted the autocorrelation times with a rescaling of the x axis to j/L. Except for small values of L we notice that the curves do indeed collapse onto a common master curve under such a rescaling. Fig. 1b thus shows in particular that the optimal value j_{opt} scales with L. This observation shows that the optimal staging parameter j_{opt} can in practice easily be obtained by looking at the autocorrelation times for moderately coarse discretization. The optimal choices of j can then be obtained for any finer discretization by a simple rescaling.

For the virial estimator the rule of 40% acceptance probability turns out not to be too far off the mark. For the optimal j_{opt} we find an acceptance rate of roughly 55% independent of the grid size L for not too small grids (cf. Table 1). To give a numerical example, for the convex potential and a medium sized grid of $L = 256$ the optimal value is $j_{opt} = 44$. An acceptance of 40% on the other hand would be achieved for $j_{40\%} = 72$. For $j_{40\%}$ we then find an integrated autocorrelation time of 2.662(96) which is significantly larger than the minimal value of 2.119(67) (cf. Table 1). We also emphasize that the integrated autocorrelation times do depend on the observable and on its estimator. Thus the 40% rule may be rather misleading for a different observable or estimator respectively. In fact, we found that the acceptance probabilities in the staging algorithm for that $j_{opt,k}$ which optimizes the autocorrelation times for the kinetic estimator is roughly 90% [26]. For our comparison with the multigrid update schemes we have in any case used the value j_{opt} which minimized the integrated autocorrelation time.

Let us now turn to the comparison of the two update schemes. For the multigrid algorithm autocorrelation times for the moments $\langle x^n \rangle$, $n = 1, \ldots, 4$ were already reported in Ref. [14]. Here we look at the virial estimator for the internal energy but in order to facilitate comparison with our previous results we have used the same parameters as in Ref. [14] for the multigrid scheme. This means in particular that we performed only presweeps and no postsweeps [14]. The acceptance rates for the finest grid were adjusted to be roughly 50%. Table 1 lists

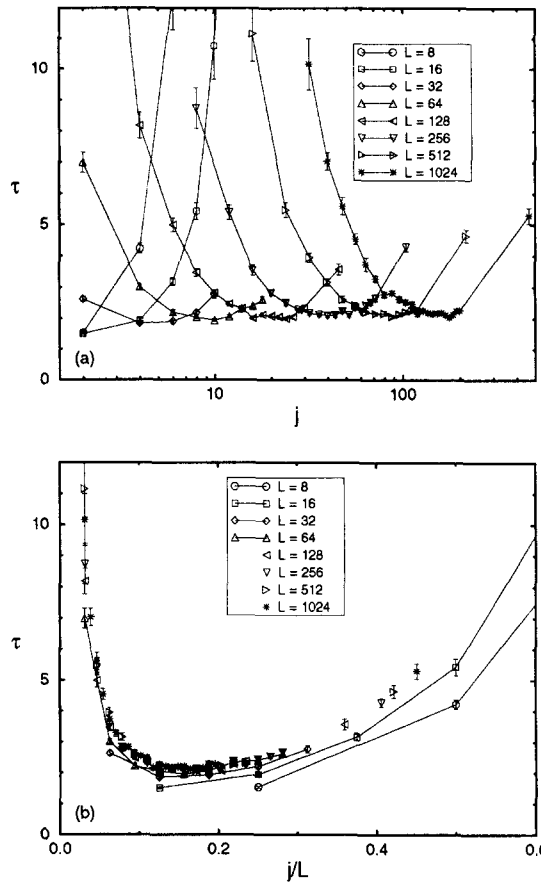

Fig. 1. (a) Integrated autocorrelation times for the virial estimator and the convex potential using the staging algorithm for various grid sizes L. The x axis is the length of the staging segment. (b) The same data as in (a), when plotted versus a rescaled x variable, collapse onto a common master curve.

for the virial estimator the integrated autocorrelation times τ for the V-cycle and the W-cycle as well as for the staging algorithm with j_{opt} as discussed above. These data are also plotted in Figs. 2a and 2b, together with fits to the data according to a power law of the form $\tau = \alpha L^z$. For the V-cycle these fits were done on the basis of the data for the three largest grids, for the W-cycle and the staging algorithm data for the four largest grids were used for the fits.

Since for a polynomial potential the virial estimator is a linear combination of the expectation values for the moments $\langle x^n \rangle$ one would expect that the autocorrelation times for the virial estimator would not differ too much from the autocorrelation times for the moments. A comparison with the data reported in Ref. [14] shows indeed a qualitative agreement. The multigrid V-cycle again reduces the L^2 divergence of autocorrelation times for standard local updates to a linear dependence with $z = 0.959(54)$ with a chi-square per degree of freedom of $\chi^2/\mathrm{d.o.f.} = 1.04$ for the convex potential (CP), and with $z = 0.808(42)$ with a $\chi^2/\mathrm{d.o.f.}$ of 0.58 for the double-well potential (DW). These exponents are indeed well comparable to the autocorrelation times for the even moment $\langle x^2 \rangle$ which are $z = 0.8356(92)$ (CP) and $z = 0.715(27)$ (DW) [14].

Also for the W-cycle the behaviour for the moments is qualitatively reproduced by the virial estimator. For the average path $\langle x \rangle$ slowing down was completely eliminated with values of z consistent with 0, while for $\langle x^2 \rangle$ the exponents were found to be $z = 0.1043(29)$ (CP) and $z = -0.015(11)$ (DW) [14]. For the virial estimator we find again an almost complete reduction of critical slowing down with exponents of $z = 0.128(12)$ with a $\chi^2/\mathrm{d.o.f.}$ of 0.20 (CP) and $z = 0.0467(86)$ with a $\chi^2/\mathrm{d.o.f.}$ of 0.49 (DW). The W-cycle thus almost completely eliminates slowing down in the continuum limit with absolute values of τ close to 0.5 which means complete decorrelation in between measurements.

Figs. 2a and 2b show that the staging algorithm also eliminates slowing down albeit with somewhat larger absolute values for τ. Judged from the exponents z the staging algorithm eliminates slowing down with exponents that are in fact perfectly consistent with 0 well within the statistical error bars. Here the fits give $z = 0.008(17)$ with a $\chi^2/\mathrm{d.o.f.}$ of 0.86 (CP) and $z = -0.005(20)$ with a $\chi^2/\mathrm{d.o.f.}$ of 0.33 (DW).

Table 1
Integrated autocorrelation times for the virial estimator of U using the staging algorithm (τ_s) and the multigrid algorithm with V-cycle (τ_v) and W-cycle (τ_w). The second and third columns give the optimal segment length and the acceptance rate in percent for the staging algorithm

L	j_{opt}	%	τ_s	τ_v	τ_w
convex potential (CP)					
8	2	64	1.545(32)	1.033(16)	0.851(16)
16	2	82	1.518(24)	0.912(14)	0.692(12
32	4	72	1.850(44)	0.909(18)	0.644(11)
64	10	68	1.959(44)	1.092(22)	0.6278(66)
128	24	55	2.012(49)	1.583(31)	0.660(14)
256	44	55	2.119(67)	2.701(82)	0.7119(93)
512	88	55	2.079(53)	5.07(21)	0.781(12)
1024	176	56	2.048(56)	10.80(94)	0.859(16)
double-well potential (DW)					
8	2	71	1.779(36)	0.752(18)	0.5478(76)
16	6	47	2.059(44)	0.746(12)	0.5121(79)
32	10	49	2.182(63)	0.784(12)	0.5128(79)
64	20	52	2.335(64)	0.945(16)	0.5291(75)
128	40	54	2.406(76)	1.338(25)	0.5377(72)
256	80	54	2.456(87)	2.297(61)	0.5479(72)
512	160	54	2.366(64)	3.92(15)	0.5650(95)
1024	320	54	2.406(69)	7.19(43)	0.5920(76)

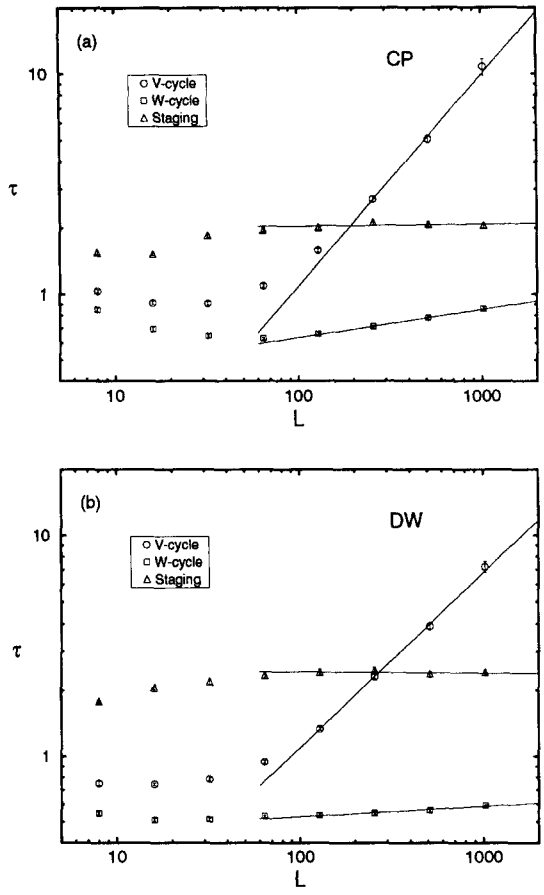

Fig. 2. (a) Double logarithmic plot of integrated autocorrelation times τ versus the grid size L for the virial estimator of the convex potential (CP) using the staging (S) algorithm and the multigrid V- and W-cycles. The straight lines are fits to the data according to $\tau = \alpha L^z$, yielding $z = 0.008(17)$ (S), $z = 0.959(54)$ (V), and $z = 0.128(12)$ (W). (b) The same plot as in (a) for the double-well potential (DW). Here the fits give $z = -0.005(20)$ (S), $z = 0.808(42)$ (V), and $z = 0.0467(86)$ (W).

Regarding the asymptotic behaviour the staging algorithm seems to be slightly superior to the W-cycle where a small L-dependence cannot be excluded from the data. One should also not forget that for one-dimensional systems the number of operations involved in the W-cycle scales with another log L dependence. For practical applications, however, it is also important to look at the absolute values of the autocorrelation times. These indeed turn out to be several times larger for the staging algorithm than for the W-cycle for grid sizes up to the largest one of $L = 1024$ considered in our investigations.

4. Discussion

Multigrid techniques and the staging algorithm provide two different but equally successful path integral Monte Carlo methods. An investigation of the integrated autocorrelation times for the virial estimator of the internal energy shows that both the staging algorithm and the multigrid schemes solve the slowing down problem of local update schemes in the continuum limit. For the staging algorithm this was demonstrated for the optimal choice of the parameter j which determines the length of the staging segment. This optimal choice differs notably from the one obtained following the common rule of achieving a certain acceptance probability. It scales with the number of lattice sites L in the same way as does the correlation length along the path measured in lattice units.

A comparison of the two update schemes from a practitioner's point of view shows that they both have their advantages and drawbacks. The staging algorithm completely beats slowing down with exponents which are even smaller than the ones for the multigrid W-cycle even though the absolute autocorrelation times are several times larger. One also has to take into account the number of operations per sweep which for one-dimensional systems grows proportional to L log L for the W-cycle. Also in more technical respects the staging algorithm is somewhat easier to implement for simple systems than the recursive multigrid scheme. Even though this is obviously hardware and platform dependent, one would probably find in most situations similar to the one investigated here that the staging algorithm will be preferable as far as CPU time is concerned. An advantage of the multigrid scheme is the generality of its definition and the fact that it is mathematically well defined and understood. Multigrid schemes are therefore quite easily generalizable to higher dimensions and quantum chains [25].

Another advantage may become crucial for the sampling of deep tunneling problems. High tunneling barriers typically result in exponentially diverging prefactors α in the power-law behaviour $\tau = \alpha L^z$. A similar problem is well-known in the simulation of phase transitions of first order where this problem can be overcome by using so-called "multicanonical" reweighting techniques [30]. As we have shown

elsewhere [4,15,16], the multicanonical method can successfully be adapted to deep tunneling problems and can furthermore quite generally be combined with multigrid update schemes. Using a combination of multicanonical reweighting with a multigrid W-cycle, we have demonstrated [15] that autocorrelation times can substantially be reduced even for very high tunneling barriers. Given the very good performance of the staging algorithm for simple systems it would be interesting to work out similar generalizations for this update scheme as well.

Acknowledgement

WJ thanks the Deutsche Forschungsgemeinschaft for a Heisenberg fellowship.

References

[1] B.J. Berne and D. Thirumalai, Ann. Rev. Phys. Chem. 37 (1986) 401.
[2] J.D. Doll, D.L. Freeman and T.L. Beck, Adv. Chem. Phys. 78 (1990) 61.
[3] N. Makri, Comp. Phys. Comm. 63 (1991) 389.
[4] W. Janke and T. Sauer, in: Path integrals from meV to MeV, Tutzing, 1992, eds. H. Grabert, A. Inomata, L. Schulman and U. Weiss (World Scientific, Singapore, 1993) p. 17.
[5] A.D. Sokal, in: Quantum fields on the computer, ed. M. Creutz (World Scientific, Singapore, 1992) p. 211.
[6] W. Janke, in: Computational physics: selected methods – simple exercises – serious applications, eds. K.H. Hoffmann and M. Schreiber (Springer, Berlin, 1996) p. 10.
[7] J. Goodman and A.D. Sokal, Phys. Rev. Lett. 56 (1986) 1015.
[8] J. Goodman and A.D. Sokal, Phys. Rev. D 40 (1989) 2035.
[9] G. Mack, in: Nonperturbative quantum field theory, Cargèse, 1987, eds. G. 't Hooft et al. (Plenum Press, New York, 1988) p. 309.
[10] G. Mack and S. Meyer, Nucl. Phys. B (Proc. Suppl.) 17 (1990) 293.
[11] D. Kandel, E. Domany, D. Ron, A. Brandt and E. Loh Jr., Phys. Rev. Lett. 60 (1988) 1591.
[12] D. Kandel, E. Domany and A. Brandt, Phys. Rev. B 40 (1989) 330.
[13] W. Hackbusch, Multi-grid methods and applications (Springer, Berlin, 1985).
[14] W. Janke and T. Sauer, Chem. Phys. Lett. 201 (1993) 499.
[15] W. Janke and T. Sauer, Phys. Rev. E 49 (1994) 3475.
[16] W. Janke and T. Sauer, J. Stat. Phys. 78 (1995) 759.
[17] E.L. Pollock and D.M. Ceperley, Phys. Rev. B 30 (1984) 2555.
[18] M. Sprik, M.L. Klein and D. Chandler, Phys. Rev. B 31 (1985) 4234.
[19] M. Sprik, M.L. Klein and D. Chandler, Phys. Rev. B 32 (1985) 545.
[20] M.E. Tuckerman, B.J. Berne, G.J. Martyna and M.L. Klein, J. Chem. Phys. 99 (1993) 2796.
[21] H. Kleinert, Path integrals in quantum mechanics, statistics and polymer physics, 2nd Ed. (World Scientific, Singapore, 1995).
[22] J.A. Barker, J. Chem. Phys. 70 (1979) 2914.
[23] M. Creutz and B. Freedman, Ann. Phys. 132 (1981) 427.
[24] M.F. Herman, E.J. Bruskin and B.J. Berne, J. Chem. Phys. 76 (1982) 5150.
[25] W. Janke and T. Sauer, Phys. Lett. A 197 (1995) 335.
[26] W. Janke and T. Sauer, to be published.
[27] N. Madras and A.D. Sokal, J. Stat. Phys. 50 (1988) 109.
[28] R.G. Miller, Biometrika 61 (1974) 1.
[29] B. Efron, The jackknife, the bootstrap and other resampling plans (SIAM, Philadelphia, PA, 1982).
[30] W. Janke, in: Computer simulations in condensed matter physics, Vol. 7, eds. D.P. Landau, K.K. Mon and H.-B. Schüttler (Springer, Berlin, 1994) p. 29, and references therein.