International Workshop on

QUANTUM SIMULATIONS
OF CONDENSED MATTER PHENOMENA

8 – 11 August 1989 Los Alamos, NM, USA

Editors

J.D. Doll
J.E. Gubernatis

World Scientific
Singapore • New Jersey • London • Hong Kong

Antithetic Langevin variables : a way to take care of the fermionic sign problem?

F.F. Assaad and Ph. de Forcrand
IPS, ETHZ, Zürich, Switzerland

Abstract

We propose to reduce the uncertainty on any observable $< W >$ estimated during a Langevin simulation, by measuring $(W_1 + W_2)/2$, where W_1 and W_2 correspond to 2 Langevin processes using anti-correlated noise. The Langevin noise of the first process is used in the second process after a rotation chosen to encourage opposite deviations from the average $< W >$. This variance reduction is illustrated for simple and realistic models. The variance is reduced by a factor $\mathcal{O}(40)$ or even becomes exactly zero in symmetric cases. This method is put to best use when W is the sign of a fermionic determinant. Tests on a one-dimensional model show a dramatic improvement in the determination of the average sign. All these improvements result from the anti-correlation of the noises, and require no additional computing effort.

1 Introduction

The simulation of systems with many Fermions poses a notorious challenge. The "minus-sign problem" comes from the anti-commuting properties of the fermion fields. It is present in Green's function Monte Carlo methods. There, the ground-state wave-function alternates in sign. Methods are still being developed [1, 2] to generate self-consistently the nodal surface. In path-integral Monte Carlo, paths can have positive or negative weights. Only in $(1 + 1)$ dimension can one find a formulation which does not exhibit a sign problem [5].

Here we want to address the general case of sampling the partition function Z. Quadratic fermion interactions can be integrated, and give rise to a determinant, so that

$$Z = \int [d\phi] e^{-S(\phi)} det(M(\phi)) \tag{1}$$

where ϕ are bosonic degrees of freedom. Since the determinant can become negative, Monte Carlo simulations sample the modified partition function

$$\tilde{Z} = \int [d\phi] e^{-S(\phi)} \, |det(M(\phi))| \tag{2}$$

The missing sign is accounted for in the averaging. For any observable W :

$$< W >= \frac{\int [d\phi] e^{-S(\phi)}\, |det(M(\phi))|\, sign(det) W(\phi)}{\int [d\phi] e^{-S(\phi)}\, |det(M(\phi))|\, sign(det)} = \frac{< W sign >_{\tilde{Z}}}{< sign >_{\tilde{Z}}} \quad (3)$$

The denominator in (3), $< sign >_{\tilde{Z}} = < sign >_{Z}^{-1}$ is called the average sign. It results from the subtraction of two large numbers, and is consequently very poorly determined by Monte Carlo sampling. This uncertainty affects in turn every observable W. The problem gets worse as $< sign >_{\tilde{Z}}$ decreases. Unfortunately the physically interesting region often corresponds to an almost complete cancellation. Such cases include the Hubbard model away from half- filling [6], QCD with an odd number of Wilson fermionic flavors [7] , or supersymmetric models [8].

Ergodicity already represents a first challenge, since the effective action contains a term $Ln\,(|det|)$. Infinite potential barriers separate regions of opposite signs in phase space. Fortunately it was shown in [16] that the Langevin equation still maintains a finite tunneling rate as the step size is reduced. We therefore use the Langevin approach, and try to reduce the variance of the denominator in (3).

We use two Langevin simulations, and consider the average of pairs of measurements performed at equal Langevin times on both simulations. By anti-correlating the two Langevin noises, we favor configuration pairs of opposite signs, thereby reducing the variance of the sign. An analogous idea was introduced by Hammersley and Morton [3] for Monte Carlo integration under the name of antithetic variates. Parisi [4] also suggested to correlate Langevin noises among two simulations in order to measure a response function.

If the average sign takes the value s, uncorrelated simulations will yield probabilities P_{++}, P_{--}, P_{+-} of forming pairs with sign combinations $++, --, +-$ respectively equal to $\left(\frac{1+s}{2}\right)^2, \left(\frac{1-s}{2}\right)^2, \frac{1-s^2}{4}$. The variance of the estimator of s will be $v = P_{++} + P_{--} - s^2 = \frac{1-s^2}{2}$.

Let us now correlate the two simulations, while maintaining proper sampling of Z, so that one still has $P_{++} - P_{--} = s$. The variance can be minimized if $P_{--} = 0$ (assuming $s > 0$; otherwise $P_{++} = 0$). Then $P_{++} = s$, and the new variance will be

$$v' = s - s^2 \quad (4)$$

The variance is thus reduced by a factor $\frac{v}{v'} = \frac{1+1/s}{2}$, which represents the effective gain in the computer time required to achieve a given accuracy on the measurement of any observable. It is unbounded as s goes to zero. v' itself goes to zero with s.

To achieve the above redistribution of probabilities, we use the freedom available in choosing the Langevin noise of simulation 2 as a transform of the noise of simulation 1 which preserves its Gaussian properties. The method is described in Section 2. It is by no means restricted to the evaluation of the average sign. The variance of *any* observable can be reduced by appropriate noise correlations. Examples are provided in Section 3, for the average of x in a one-dimension harmonic potential, for the magnetization in the 2d half-filled Hubbard model, and for the sign in a one-dimensional potential. The conclusion tries to evaluate the limits of applicability of our method.

2 The method

We wish to evaluate by means of the Monte Carlo method the quantity:

$$< W >= \frac{\int dx e^{-S(x)} W(x)}{\int dx e^{-S(x)}} \quad (5)$$

The uncertainty on the Monte Carlo average $< W >$ is $\frac{\sigma(W)}{\sqrt{N}}$, where N is the number of independent values of $W(x)$ along the Monte Carlo path, and the variance is :

$$\sigma^2(W) = < W^2 > - < W >^2 \quad (6)$$

In order to reduce $\sigma^2(W)$, we carry out two simulations W_1, W_2 which we correlate so as to minimize the fluctuations of the quantity $W = (W_1 + W_2)/2$. The variance of W is now given by :

$$\sigma^2(W) = \frac{1}{4}\sigma^2(W_1) + \frac{1}{4}\sigma^2(W_2) + \frac{1}{2}C(W_1, W_2) \quad (7)$$

where $C(W_1, W_2) = < (W_1 - < W_1 >)(W_2 - < W_2 >) >$. Whenever one is able to correlate the two Monte Carlo simulations so as to obtain $C < 0$, one reduces the variance. In the case of complete failure at correlating W_1 and W_2 (ie. $C = 0$), there still is no loss in computer time : the two simulations combine to reduce the variance by a factor 1/2, and yield the same uncertainty on $< W >$ as a single simulation with 2 N measurements.

We sample the partition function with discretized Langevin dynamics:

$$x_{1,2}(t + \delta t) = x_{1,2}(t) - \delta t \left.\frac{\partial S(x)}{\partial x}\right|_{x = x_{1,2}(t)} + \sqrt{2\delta t}\, \eta_{1,2}(t) \quad (8)$$

where $< \eta(t)_{1,2} \eta(t')_{1,2} > = \delta_{t,t'}$. Observables are measured using:

$$< W > = \lim_{M \to \infty} \frac{1}{M} \sum_{m=1}^{M} W(x(m\delta t)) \,. \quad (9)$$

The above discretization of the Langevin equation introduces a systematic error of order δt. In the limit $\delta t \to 0$, the random noise (of order $\sqrt{\delta t}$) completely dominates the drift term (of order δt), and thus provides a means to correlate the two Langevin simulations. The constraint is to preserve the Gaussian properties of the random noise in both simulations. The goal is to minimize C.

We denote by $P_1^t(W_1)$ the probability distribution at Langevin time t of W_1. The probability distribution of W_2 is given by:

$$P_2^t(W_2) dW_2 = dW_2 \int dW_1 P_1^t(W_1) Q(W_2 \mid W_1) \quad (10)$$

where $Q(W_2 \mid W_1)$ is the conditional probability of W_2 given W_1. Perfect correlation between the two simulations ($\sigma^2(W) = 0$ in (6)) requires:

$$Q(W_2 \mid W_1) = \delta\,(W_1 + W_2 - 2 < W >) \quad (11)$$

Assuming that at time t we have perfect correlation, we may preserve it as long as:

$$W_1(t+\delta t) \;+\; W_2(t+\delta t) \;=\; 2 < W > \tag{12}$$

where

$$W_{1,2}(t+\delta t) \;=\; W_{1,2}(t) + \left.\frac{\partial W}{\partial x}\right|_{x=x_{1,2}(t)} \left\{ -\left.\frac{\partial S}{\partial x}\right|_{x=x_{1,2}(t)} \delta t \;+\; \sqrt{2\delta t}\,\eta_{1,2}(t) \right\}$$

so that

$$\sqrt{2\delta t}\left\{ \left.\frac{\partial W}{\partial x}\right|_{x=x_1(t)} \eta_1(t) \;+\; \left.\frac{\partial W}{\partial x}\right|_{x=x_2(t)} \eta_2(t) \right\} \;+\; \mathcal{O}(\delta t) \;=\; 0.$$

The antithetic condition thus requires:

$$\left.\vec{\nabla} W\right|_{\vec{x}=\vec{x}_1(t)} \cdot \vec{\eta}_1(t) \;=\; -\left.\vec{\nabla} W\right|_{\vec{x}=\vec{x}_2(t)} \cdot \vec{\eta}_2(t) \tag{13}$$

where we have generalized our result to several dimensions. If the random noise $\vec{\eta}_1(t)$ increases the value of $W_1(t)$, the random noise $\vec{\eta}_2(t)$ has to be chosen so as to decrease the value of $W_2(t)$ by the same quantity.

The above choice of the conditional probability (11) implies a symmetric probability distribution $P_1(W)$ around $< W >$. Only if this is true, can one achieve perfect anti-correlation. (See example 1.) Otherwise, the above antithetic condition needs to be relaxed to :

$$\left(\left.\vec{\nabla} W\right|_{\vec{x}=\vec{x}_1(t)} \cdot \vec{\eta}_1(t) \right) \left(\left.\vec{\nabla} W\right|_{\vec{x}=\vec{x}_2(t)} \cdot \vec{\eta}_2(t) \right) < 0\,. \tag{14}$$

2.1 One-dimensional systems.

The random noise $\eta_2(t)$ is built from the random noise $\eta_1(t)$ through the relation:

$$\begin{aligned} \eta_2(t) &= \epsilon(t)\,\eta_1(t), \\ \epsilon(t) &= -sign\left(\left.\frac{\partial W}{\partial x}\right|_{x=x_1(t)} \left.\frac{\partial W}{\partial x}\right|_{x=x_2(t)} \right). \end{aligned} \tag{15}$$

The above construction clearly satisfies the antithetic condition (14). Since $\epsilon(t)$ depends on all the random noises $\eta_1(t')$, $t' < t$ but not on $\eta_1(t)$, the white noise properties of $\eta_1(t)$ are transmitted to $\eta_2(t)$. This is shown below.

$$\begin{aligned} &a)\quad t = t', && \epsilon^2(t) = 1 \\ & && < \eta_2(t)\eta_2(t) > = < \eta_1(t)\eta_1(t) > = 1 \\ &b)\quad t \neq t', && t' < t \\ & && < \eta_2(t')\eta_2(t) > \\ & && = < \epsilon(t')\epsilon(t)\eta_1(t')\eta_1(t) > \\ & && = < \epsilon(t')\epsilon(t)\eta_1(t') >_{\eta_1(t=0)\ldots\eta_1(t-1)} < \eta_1(t) >_{\eta_1(t)} \\ & && = 0 \end{aligned} \tag{16}$$

2.2 Multi-dimensional systems.

We propose two correlation schemes.

• As in the one-dimensional case, each component of the random vector $\vec{\eta}_2(t)$ is built from the corresponding component of the random vector $\vec{\eta}_1(t)$:

$$\eta_{i,2}(t) \;=\; \epsilon_i(t)\,\eta_{i,1}(t)\;,\;\; \epsilon_i(t) \;=\; -sign\left(\left.\frac{\partial W}{\partial x_i}\right|_{x_i=x_{i,1}(t)} \left.\frac{\partial W}{\partial x_i}\right|_{x_i=x_{i,2}(t)} \right). \tag{17}$$

In analogy with the one-dimensional case, one sees that the white noise properties of $\vec{\eta}_1$ are transmitted to $\vec{\eta}_2$. On closer inspection, however, this correlation scheme does not necessarily satisfy the antithetic condition (14), because of the cross-terms $\frac{\partial W}{\partial x}_{1,i}\eta_{1,i}\frac{\partial W}{\partial x}_{2,j}\eta_{2,j}$ in (14). We thus propose below another way of introducing the correlation which satisfies the antithetic condition but is somewhat more complicated.

• Denote by P the plane defined by the two drift terms $\vec{\nabla}_1 W$ and $\vec{\nabla}_2 W$, and by α their angle. The random noise $\vec{\eta}_1$ is decomposed into a vector $\vec{\eta}_{1,\perp}$ perpendicular to P and a vector $\vec{\eta}_{1,P}$ in P. We rotate $\vec{\eta}_{1,P}$ in the plane P by an angle $\beta = \pi - \alpha$, obtaining $R(\beta)\vec{\eta}_{1,P}$. The random noise for the second simulation is then given by:

$$\vec{\eta}_2 \;=\; \vec{\eta}_{1,\perp} \;+\; R(\beta)\,\vec{\eta}_{1,P} \tag{18}$$

where

$$\vec{\eta}_{1,P} \;=\; \left(\vec{\eta}_1 \cdot \frac{\vec{\nabla} W_1}{|\vec{\nabla} W_1|} \right) \frac{\vec{\nabla} W_1}{|\vec{\nabla} W_1|} \;+\; \left(\vec{\eta}_1 \cdot \frac{\vec{\nabla} W_{\perp,1}}{|\vec{\nabla} W_{\perp,1}|} \right) \frac{\vec{\nabla} W_{\perp,1}}{|\vec{\nabla} W_{\perp,1}|}$$

and

$$\vec{\eta}_{1,\perp} \;=\; \vec{\eta}_1 \;-\; \vec{\eta}_{1,P}\,.$$

Here, we have defined:

$$\vec{\nabla} W_{\perp,1} \;=\; \vec{\nabla} W_2 \;-\; \left(\vec{\nabla} W_2 \cdot \frac{\vec{\nabla} W_1}{|\vec{\nabla} W_1|} \right) \frac{\vec{\nabla} W_1}{|\vec{\nabla} W_1|}\,.$$

The above correlation scheme preserves the white noise properties of $\vec{\eta}_2$ since the latter is built by rotating the projection in the plane P of $\vec{\eta}_1$. (Note that the rotation is independent of $\vec{\eta}_1$.) Furthermore, the antithetic condition (14) is satisfied.

In practice, the rotation in equation (18) is realized through:

$$R(\beta)\vec{\eta}_{1,P} \;=\; r\frac{\vec{\nabla} W_{\perp,2}}{|\vec{\nabla} W_{\perp,2}|} \;-\; \left(\vec{\eta}_{1,P} \cdot \frac{\vec{\nabla} W_1}{|\vec{\nabla} W_1|} \right) \frac{\vec{\nabla} W_2}{|\vec{\nabla} W_2|} \tag{19}$$

where

$$\vec{\nabla} W_{\perp,2} \;=\; \vec{\nabla} W_1 \;-\; \left(\vec{\nabla} W_1 \cdot \frac{\vec{\nabla} W_2}{|\vec{\nabla} W_2|} \right) \frac{\vec{\nabla} W_2}{|\vec{\nabla} W_2|}$$

and

$$r \;=\; \pm\left| \vec{\eta}_{1,P} \;-\; \left(\vec{\eta}_{1,P} \cdot \frac{\vec{\nabla} W_1}{|\vec{\nabla} W_1|} \right) \frac{\vec{\nabla} W_1}{|\vec{\nabla} W_1|} \right|.$$

3 Examples

3.1 One-dimensional harmonic potential.

We want to reduce the variance of

$$< x > = \frac{\int dx\, e^{-x^2}\, x}{\int dx\, e^{-x^2}} \tag{20}$$

The antithetic condition (14) requires:

$$\eta_2(t) \;=\; -\,\eta_1(t) \tag{21}$$

We start the simulation in the unfavorable situation $x_1(0) = x_2(0)$. The state $x_1(t) = -x_2(t)$ is an attractor of the dynamics, and one obtains $< x > = 0$ with *zero variance* after thermalization (See Fig. 1). This zero variance result occurs whenever the Boltzmann distribution of W is symmetric around $< W >$, so that (13) can be satisfied exactly at every step.

3.2 Two-dimensional Hubbard model at half filling.

The simulation of the two-dimensional Hubbard model is carried out with a projected trial wave function technique. The sampling is done with Langevin dynamics [11, 9]. We wish to evaluate the magnetic quantities:

$$< W_M > \equiv \sum_i < S_{z,i} > = \frac{1}{\sqrt{\delta\tau U}} \sum_i < \sigma_i > \tag{22}$$

as well as

$$< W_{SM} > \equiv \sum_i (-1)^{x_i+y_i} < S_{z,i} > = \frac{1}{\sqrt{\delta\tau U}} \sum_i (-1)^{x_i+y_i} < \sigma_i > . \tag{23}$$

Here $S_{z,i}$ denotes the z component of the spin on lattice site i, U the on-site Coulomb repulsion, σ the Hubbard-Stratonovitch fields and $\delta\tau$ the imaginary time interval introduced through the Trotter decomposition. We have chosen to correlate the two random noises component by component, following (17). The linearity of both observables in the σ's results in the simple prescription :

$$\eta_{i,2} \;=\; -\,\eta_{i,1} \tag{24}$$

and guarantees that the antithetic condition (14) will be satisfied. Figures 2 to 7 show the Langevin histories of W_M and W_{SM} for individual and combined simulations. The equilibrium distribution of W_M is symmetric around 0, so that again one achieves zero variance for the combined simulation. For W_{SM}, the variance is reduced by $\mathcal{O}(40)$.

3.3 Fermionic sign.

In order to minimize the fluctuations of the sign we choose to correlate the two simulations with respect to $W = det(M(\phi))$. Optimally, when the average sign vanishes, the antithetic simulations will select exlusively $(+,-)$ pairs thus reducing the error on $< sign_1 + sign_2 > /2$ to zero.

In order to test the above correlation scheme we have modified a toy model proposed in [16]. The effective action is given by (see Fig. 8) :

$$S_{eff}(\phi) \;=\; \phi^2 \;-\; ln \;|\; \phi^4 \;-\; 3\phi^2 \;+\; 0.75 \;| \,. \tag{25}$$

Here, $(\phi^4 - 3\phi^2 + 0.75)$ plays the role of the determinant and the parameters have been chosen so that the average sign is zero. In this example, ergodicity is essential since there are several disconnected regions of same sign. The results (see Fig. 10) show that after very long runs, results of the individual simulations are still wrong. On the other hand the quantity $< sign_1 + sign_2 > /2$ rapidly converges to the right value. This is due to the anti- correlation of ϕ_1 and ϕ_2, as shown in Fig. 9. At the same time, fluctuations are reduced by a factor 2.

4 Conclusion

We have explained how to reduce the variance of any observable W by correlating the noises of 2 Langevin simulations and measuring $\frac{W_1+W_2}{2}$. The method has been illustrated on both simple (1d harmonic potential) and complex (2d Hubbard model) systems. When the system has an n-fold symmetry, zero variance can be achieved by correlating n simulations. This variance reduction is most crucial in measurements of the average sign of the fermionic determinant. Simulation of a one-dimensional model shows a dramatic improvement.

Work is under way to assess the variance reduction achievable for the sign of a 2d Hubbard determinant [9], and for Wilson and Polyakov loops in QCD [10]. Preliminary results indicate that the Langevin time steps commonly used in Hubbard simulations [11] are large compared to the width of the potential barrier separating regions of opposite signs. There is no precursory sign announcing the proximity of the barrier. This is very good from the point of view of ergodicity, but makes any correlation scheme rather inefficient. An attractive feature of our approach however is that, even in the case of complete failure, the 2 sets of results are uncorrelated and can still be averaged together, yielding the variance of a normal simulation at no extra cost of computer time.

Finally it is easy to generalize the idea to antithetic Metropolis steps [12], and to the integration of complex functions. The latter adaptation might help in the evaluation of fast oscillating integrals which appear in one reformulation of the Hubbard model [13], and more generally in the presence of a chemical potential [14, 15].

Acknowledgements

We are grateful for conversations with A.I. Bochkarev, W.P. Petersen and D. Würtz. F.F. Assaad acknowledges the support of the Zentenarfonds of ETHZ, and Ph. de Forcrand the support of the Cray Research Foundation.

References

[1] V. Elser, Phys. Rev. A34 (1986) 2293.

[2] J.W. Negele and K.D. Brommer, Ann. Phys. 192 (1989) 119.

[3] J.M. Hammersley and K.W. Morton, Proc. Camb. Phil. Soc. 52 (1956) 449.

[4] G. Parisi, Nucl. Phys. B180 (FS2) (1981) 378 ; B205 (FS5) (1982) 337.

[5] J.E. Hirsch, R.L. Sugar, D.J. Scalapino and R. Blankenbecler, Phys. Rev. B26 (1982) 5033.

[6] E.Y. Loh, J.E. Gubernatis, R.T. Scalettar, S.R. White, D.J. Scalapino and R.L. Sugar, The sign problem in the numerical simulation of many electron systems, preprint UCSBTH-89-21.

[7] Ph. de Forcrand and I.O. Stamatescu, Nucl. Phys. B (Proc. Suppl.) 9 (1989) 276.

[8] M.F.L. Golterman and D.N. Petcher, A local interactive lattice model with supersymmetry, preprint UCD-88-08, and work in progress.

[9] F.F. Assaad, work in progress.

[10] Ph. de Forcrand and J. Potvin, work in progress.

[11] S. Sorella, E. Tosatti, S. Baroni, R. Car and M. Parrinello, Int. J. Mod. Phys. B1 (1988) 993.

[12] Tests on a one-dimensional model have been performed by Michael Creutz, private communication.

[13] E. Dagotto, A. Moreo, D. Scalapino, R. Sugar and D. Toussaint, Santa Barbara preprint.

[14] G. Bhanot, A. Gocksch and P. Rossi, Phys. Lett. B199 (1987) 101.

[15] A. Gocksch, Phys. Rev. D37 (1988) 1014.

[16] S.R. White and J.W. Wilkins, Phys. Rev. B37 (1988) 5024.

Figure captions

Fig. 1. Langevin history of x_2 versus x_1 for a one-dimensional harmonic potential. The system is attracted to $x_2 = -x_1$.

Fig. 2. Langevin history of W_M for first simulation in half-filled 2d Hubbard model. See text, eq.(22).

Fig. 3. Langevin history of W_M for second simulation in half-filled 2d Hubbard model.

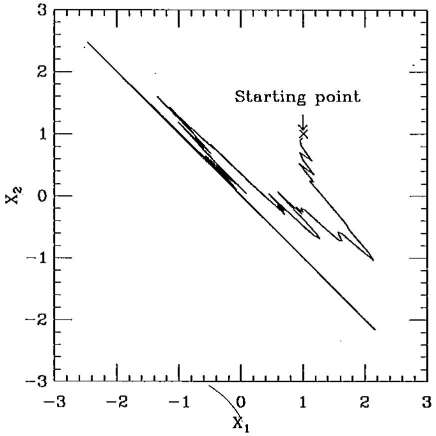
Fig. 4. Langevin history of W_M for combined simulations in half-filled 2d Hubbard model.

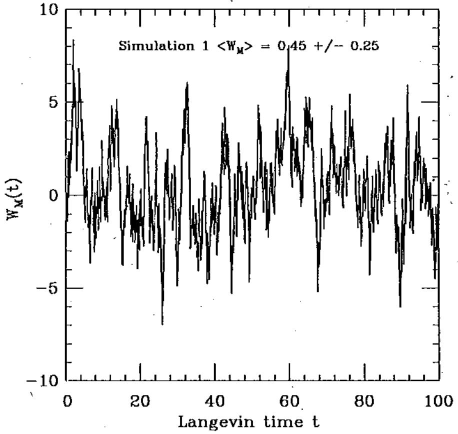
Fig. 5 to 7. Same as Fig. 2 to 4, for W_{SM}. See text, eq.(23). The variance is reduced by $\mathcal{O}(40)$ for the combined simulations.

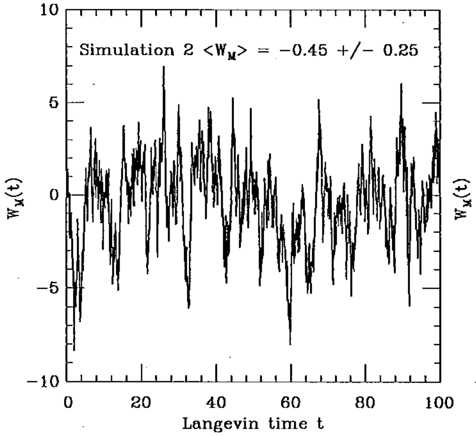
Fig. 8. Effective action S_{eff} for one-dimensional determinant model. See text, eq.(25).

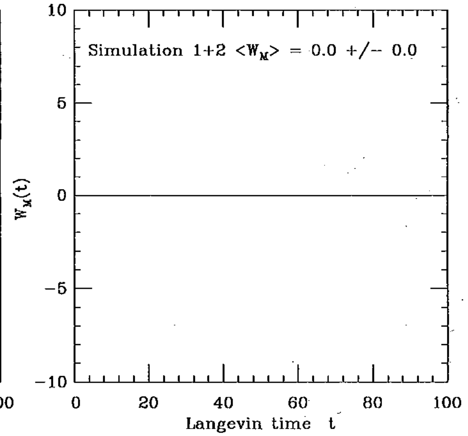
Fig. 9. ϕ_2 versus ϕ_1.

Fig. 10. Average sign for the single and combined simulations at successive Langevin times. Individual simulations give a wrong answer.









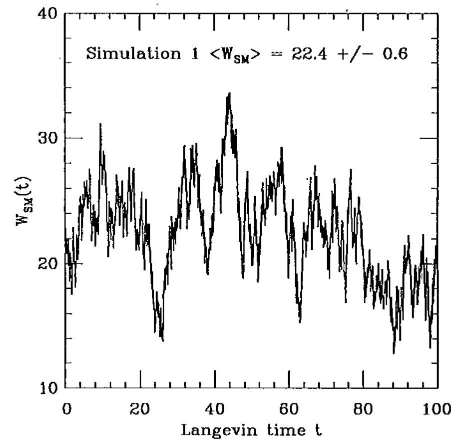
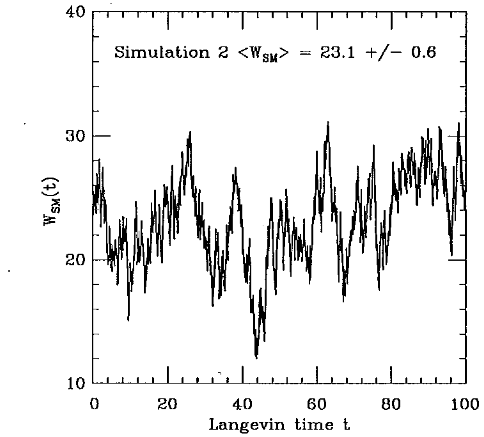
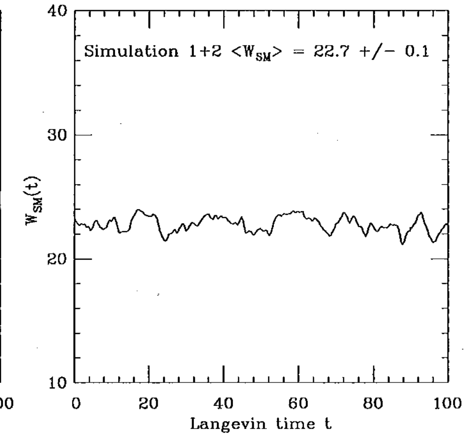


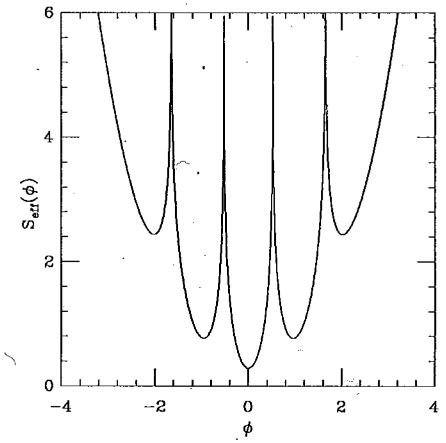
Fig. 1 Fig. 2

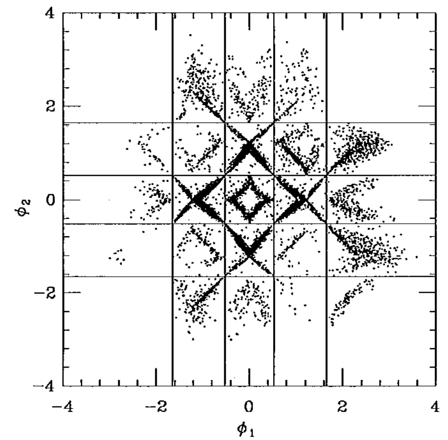
Fig. 3 Fig. 4

Fig. 5









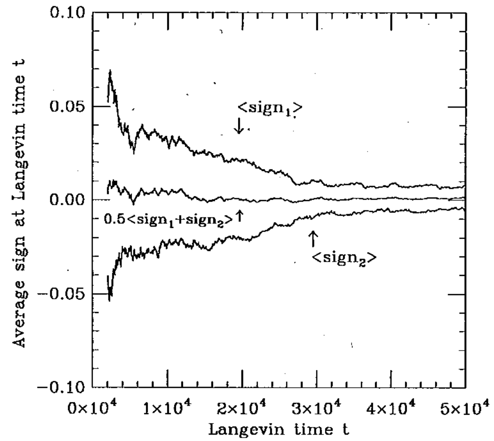


Fig. 6 Fig. 7

Fig. 8, Fig. 9

Fig. 10