

Monte Carlo studies of quantum and classical annealing on a double well

Lorenzo Stella,¹ Giuseppe E. Santoro,^{1,2} and Erio Tosatti^{1,2}

¹*International School for Advanced Studies (SISSA), and INFN Democritos National Simulation Center, Via Beirut 2-4, I-34014 Trieste, Italy*

²*International Center for Theoretical Physics (ICTP), P.O. Box 586, I-34014 Trieste, Italy*

(Received 2 December 2005; published 10 April 2006)

We present results for a variety of Monte Carlo annealing approaches, both classical and quantum, benchmarked against one another for the textbook optimization exercise of a simple one-dimensional double well. In classical (thermal) annealing, the dependence upon the move chosen in a Metropolis scheme is studied and correlated with the spectrum of the associated Markov transition matrix. In quantum annealing, the path integral Monte Carlo approach is found to yield nontrivial sampling difficulties associated with the tunneling between the two wells. The choice of fictitious quantum kinetic energy is also addressed. We find that a “relativistic” kinetic energy form, leading to a higher probability of long real-space jumps, can be considerably more effective than the standard nonrelativistic one.

DOI: [10.1103/PhysRevB.73.144302](https://doi.org/10.1103/PhysRevB.73.144302)

PACS number(s): 02.70.Uu, 02.70.Ss, 07.05.Tp, 75.10.Nr

I. INTRODUCTION

Solving difficult classical combinatorial problems by adiabatically controlling a quantum Hamiltonian is a fascinating route to quantum computation,^{1,2} which is an alternative to the main-stream approach³ and recently argued to be equivalent to it.⁴ The quantum annealing (QA) strategy^{2,5,6} is precisely based on the adiabatic switching off, as a function of time, of strong fictitious quantum fluctuations, suitably introduced by an extra kinetic term in the otherwise classical Hamiltonian under consideration.

A number of authors have by now applied QA to attack a variety of optimization problems. Some amount of success was obtained in several cases, notably in the folding of off-lattice polymer models,^{7,8} in the ground-state search of small Ising problems⁹ as well as of larger instances of the two-dimensional (2D) random Ising model^{10,11} and of the random-field Ising model,¹² in the optimization of Lennard-Jones clusters,^{13,14} in the traveling salesman problem,¹⁵ and in kinetically constrained models.¹⁶

Despite these successes—mostly obtained by path-integral Monte Carlo (PIMC-) based implementations of QA—there is no general theory addressing and even less predicting the performance of a QA algorithm. Different optimization problems are characterized by different “energy landscapes” of barriers and valleys about which little is known but which clearly influence the annealing process.¹⁷ The success of QA crucially depends on the efficiency with which the chosen kinetic energy and associated fictitious quantum fluctuations explores and influences these effective energy landscape. This is a uncomfortable situation, in view of the fact that it is *a priori* not obvious or guaranteed that a QA approach should do any better than, for instance, classical simulated annealing (CA). Indeed, for the interesting case of Boolean satisfiability (SAT) problems¹⁸—more precisely, a prototypical nondeterministic-polynomial- (NP-) complete problem such as 3-SAT—a recent study has shown that PIMC-QA performs definitely *worse* than simple CA.¹⁹

In a previous paper,²⁰ henceforth referred to as I, we compared classical and quantum annealing approaches in their

performance to optimize the simplest potential landscape $V(x)$: namely, the one-dimensional potential double well. Classical annealing was implemented by means of a Fokker-Planck (FP) equation with a time-dependent (decreasing) temperature $T(t)$. Quantum annealing was performed through Schrödinger’s equation, both in real and in imaginary time, describing a particle with a time-dependent (decreasing) inverse mass $m^{-1}(t)$ in the double-well potential $V(x)$. Given the textbook simplicity of the problem, it was possible to correlate in detail the features of the landscape (position and shape of valleys, barrier height) with the outcome of the annealing process. One saw, for instance, that Fokker-Planck CA is sensitive only to the ratio Δ_V/B of the energy splitting Δ_V between the two wells, or valleys, and the barrier height B , whereas on the contrary Schrödinger QA is sensitive to the Landau-Zener tunneling gap and, hence, among other things, to the barrier width.

Unfortunately, such a direct approach is applicable, for practical purposes, only to insignificantly small-sized optimization problems.^{9,21} To appreciate the difficulty, it is enough to consider that the number of possible configurations, and hence the Hilbert space size, of a small 32×32 square lattice Ising model is $2^{1024} \sim 10^{308}$, an astronomically large number which forbids any direct approach based on deterministic state evolution. Hard instances of typical optimization problems involve an even larger number of variables. Because of that, alternative strategies, based on Monte Carlo sampling, are mandatory. That notwithstanding, the various deterministic types of dynamics, Fokker-Planck, Schrödinger, or Monte Carlo dynamics, classical or quantum, are not at all equivalent. There is in particular no relationship between the (physical) time appearing in the Fokker-Planck or in the Schrödinger equation and the corresponding classical or quantum Monte Carlo time step. Furthermore, *many* different types of Monte Carlo dynamics are possible. While sampling the same equilibrium probability distribution and hence providing the same equilibrium averages,²² they will generally perform very differently in an out-of-equilibrium situation. In order to appreciate that, given the large variability of the

possible MC outcomes, it is once again wise to concentrate our attention on landscapes which are well under control.

In this paper we implement and investigate different Monte Carlo annealing strategies, both classical and quantum, to the very same optimization problem—the one-dimensional double well—which was studied in I by deterministic (Fokker-Planck and Schrödinger) annealing approaches. The model itself, the double well, finds innumerable applications in various physical contexts, including tunneling in amorphous metals,²³ tunneling of hydrogens in potassium dihydrogen phosphate (KDP) systems,²⁴ etc. The aspects we will concentrate on, however, have to do with the different possible realizations of classical and quantum *Monte Carlo dynamics*, as opposed to physical dynamics, thus making direct experimental implications of our findings not straightforward. We will first show (Sec. II) that the choice of the proposed move in a Metropolis MC scheme strongly influences the annealing performance. Such a strong influence of the move is seen to be correlated with the instantaneous spectrum of the Markov transition matrix associated to the MC scheme, as shown in detail in Sec. II A. Next, we will turn to a quantum MC approach, specifically to the path-integral Monte Carlo method, and show how the required tunneling dynamics is highly nontrivial. The outcome of an allegedly state-of-the-art PIMC annealing of a simple double well can even be very unsatisfactory, due to sampling difficulties of instanton events (see Secs. III, III A, and III B). Finally, we will analyze the choice of kinetic energy in the QA Hamiltonian by testing the effectiveness of an alternative *relativistic* kind of dispersion, which proves much more effective than the usual nonrelativistic choice (Sec. III C). Section IV will finally contain a summary of the main results and some discussion. Technical details of the spectral analysis of the classical Markov process, and of the generalization of the bisection algorithm of interest for the PIMC study of Sec. III C, are relegated to the two Appendices.

II. CLASSICAL MONTE CARLO ANNEALING OF A DOUBLE WELL

Let us start with the double-well one-dimensional potential on which we will test the different classical and quantum Monte Carlo annealing schemes presented in this paper. We assume $V(x)$ to describe two generally asymmetric wells of the form

$$V(x) = \begin{cases} V_0 \frac{(x^2 - a_+^2)^2}{a_+^4} + \alpha x & \text{for } x \geq 0, \\ V_0 \frac{(x^2 - a_-^2)^2}{a_-^4} + \alpha x & \text{for } x < 0, \end{cases} \quad (1)$$

where $a_+ \geq a_-$ (both positive) and the linear term αx , with $\alpha \geq 0$, splits the degeneracy of the two minima. (The discontinuity in the second derivative at the origin is of no consequence in our discussion.) The optimal value of the potential is obviously $E_{opt} = V(x_-)$. To linear order in the small parameter α , the two minima are located at $x_{\pm} = \pm a_{\pm} - \alpha a_{\pm}^2 / (8V_0)$ and the splitting between the two minima is $\Delta_V = V(x_+)$

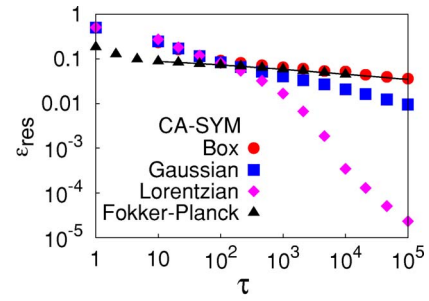


FIG. 1. (Color online) Classical annealing results. Comparison between Fokker-Planck (deterministic) annealing (triangles) and several different Metropolis Monte Carlo annealings which differ only by the choice of the proposed move: box, Gaussian, or Lorentzian (see text). The potential is a symmetric double well of Eq. (1) with $a_- = a_+ = 1$, $V_0 = 1$, and a linear term $\alpha = 0.1$. The initial condition was $T_0 = V_0 = 1$. The solid line through the points is a fit with a power law of the leading exponent $\tau^{-\Delta_V/B}$ times logarithmic corrections; see I. Clearly, the Fokker-Planck and Monte Carlo annealing “times” are not related (see text).

$-V(x_-) = \alpha(a_+ + a_-)$, while the second derivative of the potential at the two minima, to lowest order in α , is given by $V''(x = x_{\pm}) = 8V_0/a_{\pm}^2$. A “symmetric” double-well potential is recovered for $a_+ = a_-$.

Starting with some initial state or distribution, the key quantity under inspection will be the average residual energy $\epsilon_{res}(\tau) = \bar{E}_{pot}(\tau) - E_{opt}$ after annealing for a certain time τ , where $\bar{E}_{pot}(\tau)$ is the average final potential energy. In the classical MC case, annealing is performed by reducing the temperature $T(t) = T_0(1 - t/\tau)$. As shown in I (Ref. 20), the results of a Fokker-Planck annealing are roughly independent of the asymmetry of the potential and only controlled by the ratio Δ_V/B of the splitting Δ_V between the two minima and the barrier height $B = V_0 - V(x_+)$. In Fig. 1 we present the results of a standard Metropolis MC classical annealing for a symmetric ($a_- = a_+ = 1$, $V_0 = 1$) double-well potential, where a linear term αx with $\alpha = 0.1$ provides a splitting between the two minima of order $\Delta_V \approx 0.2$. Entirely similar results (not shown) are obtained for asymmetric double wells. The solid triangles represent the results of exact integration of the FP equation, via a fourth-order Runge-Kutta method, with a linear annealing schedule for the temperature $T(t) = T_0(1 - t/\tau)$, as reported in I. The circles, squares, and diamonds represent the results of three different classical Metropolis MC annealings—clearly showing very different behaviors—differing uniquely in the choice of the proposed move. To clarify this point, we recall that the Markov process associated with a Metropolis MC is fully specified by a transition probability $W(x'|x)$, to move from x to x' in the time step Δt . Here W is the product of two factors:

$$W(x'|x) = A(x'|x)P(x'|x) \quad \text{for } x' \neq x, \quad (2)$$

where $P(x'|x)$ is the probability of *proposing* a move from x to x' , while

$$A(x'|x) = \text{Min} \left\{ 1, \frac{P(x|x')}{P(x'|x)} e^{-[E(x') - E(x)]/k_B T} \right\} \quad (3)$$

is the probability of *accepting* the proposed move. The various results shown in Fig. 1 differ in the choice of $P(x'|x)$ as follows.

(i) (Box, circles) The new proposed position x' is uniformly distributed inside a box of length 2σ centered around x —i.e.,

$$P(x'|x) = P_{\text{Box}}(x'|x) = \frac{1}{2\sigma} \theta(\sigma - |x' - x|). \quad (4)$$

(ii) (Gaussian, squares) The new position x' is distributed according to a Gaussian of width σ centered around x —i.e.,

$$P(x'|x) = P_{\text{Gau}}(x'|x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x' - x)^2/(2\sigma^2)}. \quad (5)$$

(iii) (Lorentzian, diamonds) The new position x' is distributed according to a Lorentzian of width σ centered around x —i.e.,

$$P(x'|x) = P_{\text{Lor}}(x'|x) = \frac{1}{\pi} \frac{\sigma}{(x' - x)^2 + \sigma^2}. \quad (6)$$

In all three cases, the value of σ , which controls the move range, was decreased with temperature according to $\sigma = \sigma_0 \sqrt{T/T_0}$, with $\sigma_0=2.0$, for the box and Gaussian cases and $\sigma_0=2.9$ for the Lorentzian one. This choice was made empirically by monitoring the average acceptance and root-mean-square displacement $\sqrt{\langle (x_{i+1} - x_i)^2 \rangle}$ of the MC process.

As evident from Fig. 1, the results of the various choices of proposal moves differ in a substantial way. While the box choice leads to a residual energy very close to the deterministic Fokker-Planck result, the Gaussian choice and, even more so, the Lorentzian one lead to residual energies which decrease faster with the annealing time τ .

The reason for the nonuniqueness of the result of a classical MC annealing is not difficult to grasp. The Metropolis algorithm guarantees that the expectation values of a system at equilibrium are correctly reproduced by a MC time average, after a suitable equilibration time τ_{eq} . The equilibration time τ_{eq} depends on the quantity one is interested and, more crucially for our scope, on the type of proposed moves—i.e., on the $P(x'|x)$. For equilibrium MC simulations, this issue is of little concern: As long as one checks that the stochastic process has reached a stationary condition, the results are guaranteed to be correct. In an annealing simulations, however (or, in that respect, in any out-of-equilibrium evolution), the choice of the $P(x'|x)$, determining the instantaneous equilibration time of the process, strongly influences the overall performance of the algorithm.

Evidently, both the Gaussian and Lorentzian trial moves allow the attainment of residual energies $\epsilon_{res}(\tau)$ much below that of the box move (and of the FP dynamics). The Lorentzian case is particularly impressive. If we fit the asymptotic part of the residual energy data by $\epsilon_{res}(\tau) \propto \tau^{-\Omega_{CA}}$, the annealing exponent Ω_{CA} turns out to be ~ 0.31 for the Gaussian case and ~ 1.0 for the Lorentzian one—i.e., definitely much larger than what the Huse-Fisher theory²⁵ predicts for the Fokker-Planck case: namely, $\Omega_{CA} = \Delta_V/B \sim 0.2$.

The long Lorentzian tails clearly imply an abundance of

moves operating *long jumps*, which are beneficial to the annealing even at small temperatures T . References 26 and 27 indeed have reported a genuine improvement of the sampling by making use of the Lorentzian distribution, instead of the Gaussian, for the Thompson problem²⁷ (finding the equilibrium distribution of N charged particles on a sphere) and in optimizing the structure of a cluster made of N Ni atoms.²⁷ In order to analyze these issues in more quantitative detail, we carried out a spectral analysis of the transition matrices $W(x'|x)$ corresponding to the different proposal move schemes. In the following section we briefly recall the crucial steps behind this approach.

A. Spectral analysis of the Markov process in Monte Carlo simulations

Every discrete-time Markov process can be defined by means of an equation of the form

$$P(i, t + \Delta t) = \sum_j W(i|j) P(j, t), \quad (7)$$

where, for simplicity of notation, we have assumed that the possible states of the system, labeled i and j , are also discrete (we will indeed reduce our continuous problem to a discrete one in the following numerical calculations; see below). $P(i, t)$ is here the probability of finding the system in state i at time t , and $W(i|j)$ is the so-called *transition matrix*—i.e., the transition probability for going from state j to state i during the time step Δt . The transition matrix $W(i|j)$ (assumed to be time independent) is a stochastic matrix,²⁸ in the sense that $W(i|j) \geq 0$ and $\sum_i W(i|j) = 1$. In our particular case, the stochastic process behind the Metropolis sampling is given by Eq. (2), so that $W(i|j)$ also fulfills the detailed balance condition.²² The idea underlying the spectral analysis of $W(i|j)$ is very simple. For a fixed temperature T , the Markov process admits a spectral analysis in much the same spirit in which one analyzes a time-dependent Schrödinger equation via the spectrum of the associated Hamiltonian. Indeed, in continuous time—i.e., for $\Delta t \rightarrow 0$ —the result is quite simple to state. A similar analysis can be carried out for the original discrete-time Markov chain; see Ref. 28. As explained there, from $W(i|j)$ one can define a closely related matrix $\bar{W}(i|j)$ appearing in the short-time expansion, $W(i|j) = \delta_{i,j} + \Delta t \bar{W}(i|j)$, of the exponential map: $W = e^{\Delta t \bar{W}}$; in turn, $-\bar{W}$ possesses a set of right eigenvectors $\{P_k(i)\}$ with $P_0(i) = P^{(eq)}(i)$ (the equilibrium distribution), and the corresponding eigenvalues $\{\bar{\lambda}_k\}$ are such that $\bar{\lambda}_0 = 0$, $\bar{\lambda}_k > 0 \forall k > 0$. In terms of the spectrum of \bar{W} , one can formally write the general solution of the time-continuum ($\Delta t \rightarrow 0$) limit of Eq. (7) as²⁸

$$P(i, t) = P^{(eq)}(i) + \sum_{k>0} a_k e^{-\bar{\lambda}_k t} P_k(i), \quad (8)$$

where a_k are real coefficients depending on the initial ($t=0$) condition. Equation (8) shows that the approach to the equilibrium distribution is governed by a *relaxation time* $t_{rel} \propto \bar{\lambda}_1^{-1}$, with $\bar{\lambda}_1$ being the smallest nonzero eigenvalue of

$-\bar{W}$ (recall that $\bar{\lambda}_0=0$). Therefore, the larger the *spectral gap* $\Delta=\bar{\lambda}_1-\bar{\lambda}_0=\bar{\lambda}_1$, the faster the system reaches equilibrium. However, since in the Metropolis algorithm the time step is actually discrete (see Appendix A), it is better to consider the set of the W eigenvalues, given by $\lambda_i=e^{-\Delta t \bar{\lambda}_i}$, with $\lambda_0=1$. Therefore, in the rest of the paper we shall refer to the spectral gap of the original transition matrix W , which is $\Delta=1-\lambda_1$, instead of the corresponding time-continuum limit introduced above. As we shall show in a while, this choice will not change the conclusion of our spectral analysis.

How can we use the spectral gap concept in an annealing context? It is natural to think that, for a given schedule $T(t)$, it would be good to *maximize* the spectral gap for each given temperature $T(t)$. This is indeed what we will explore in the following. Generally speaking, finding explicitly the spectral gap Δ for a given \bar{W} is an impossible task; in the present one-dimensional double-well context, however, it turns out that the spectrum of $W(x'|x)$ is quite simple to obtain by, for instance, a straightforward discretization of the real axis.

We present now the results obtained by applying this spectral analysis to the proposal moves we defined in Sec. II. We recall that they are the box move [see Eq. (4)], the Gaussian move [see Eq. (5)], and the Lorentzian move [see Eq. (6)]. The main quantity under inspection is the *spectral gap* Δ of the transition matrix $W(x'|x)$ which is a function of both the temperature T and the proposal move range σ , $\Delta(T, \sigma)$. We obtain this gap (as well as the higher eigenvalues) by diagonalizing an appropriate symmetrized (see Appendix A) and discretized version of the transition operator $W(x'|x)$.²⁹ In Fig. 2 we plot the spectral gap Δ versus σ at two fixed temperatures, $T/V_0=0.1$ (top) and $T/V_0=0.001$ (bottom). As a general remark, we note how the behavior of $\Delta(\sigma)$ is remarkably different for different trial moves. Interestingly, for every proposal move and every T , there is a value of σ which maximizes the gap Δ . Recalling that the relaxation time is proportional to Δ^{-1} we anticipate that, at every temperature T , there is a $\sigma(T)$ which provides the fastest relaxation. Clearly this $\sigma(T)$ is *a priori* a better choice than taking just $\sigma(T)=\sigma_0(T/T_0)^{1/2}$, as previously implemented in Sec. II. At low T , however, the dependence of Δ on σ is no longer smooth. In particular, for $T/V_0=0.001$, the box trial move shows a sharp transition from vanishingly small gap values to finite ones, while both Gaussian and Lorentzian moves show a cusp maximum. At low temperatures, the maximum gap $\Delta_{max}^{(Lor)}$ of the Lorentzian move is larger than $\Delta_{max}^{(Gau)}$, which is in turn larger than $\Delta_{max}^{(box)}$; correspondingly, for small σ , we have $\Delta^{(Lor)}(\sigma) > \Delta^{(Gau)}(\sigma) > \Delta^{(box)}(\sigma)$. This is, essentially, the reason for the faster relaxation dynamics of the Lorentzian move during annealing. Notice also that the maximum value Δ_{max} of Δ decreases with temperature in all cases, as one can see by comparing the two panels in Fig. 2. Moreover, the value of σ that maximizes the Lorentzian and Gaussian gap decreases with temperature, while it seems to converge to a finite value (around $\sigma \approx 1.75$) for the box case.

The box case is particularly intriguing and deserves a closer inspection. In Fig. 3 we plot Δ versus inverse temperature $1/T$ for the case of the box move and for several values

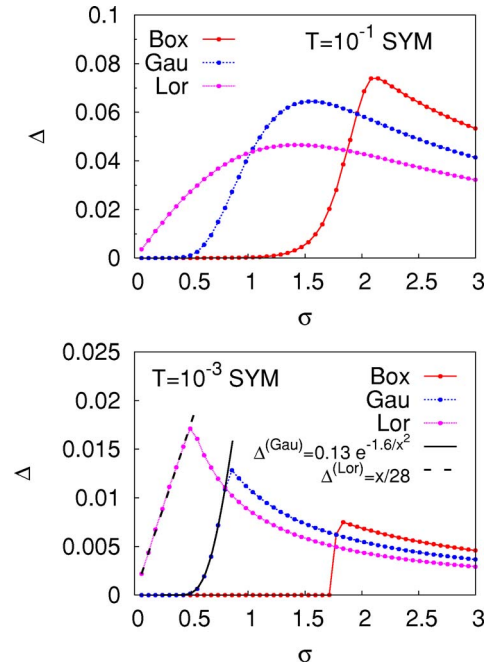


FIG. 2. (Color online) Exact diagonalization of the transition operator W for the symmetric potential case. We plot the spectral gap Δ versus the proposal move range σ . The temperature is $T/V_0=0.1$ (top) or $T/V_0=0.001$ (bottom) in the same units in which the barrier parameter is $V_0=1$. The solid and dashed black lines represent the fit provided by Eq. (A8) (Gaussian case) and Eq. (A9) (Lorentzian case). The functional form of those fits is recalled in the legend.

of σ . For a whole range of σ , $\sigma < \sigma_{cr} \approx 1.75$, we see a low-temperature behavior typical of an activated Arrhenius process $\Delta(T) \propto e^{-B_{eff}(\sigma)/T}$, where $B_{eff}(\sigma)$ is an *effective barrier*, clearly σ dependent, which the system experiences. Above a certain σ_{cr} , the behavior of Δ changes drastically, from Arrhenius to what appears, at first sight, just a constant. A closer inspection on an extended temperature range (see Fig. 3, top right inset), shows that at very small temperatures an (avoided) crossing between eigenvalues of W has taken place and Δ starts to decrease again toward zero as $\Delta \propto T^{1/2}$ (Ref. 30). The effective barrier $B_{eff}(\sigma)$ extracted from an Arrhenius fit, $\Delta \propto e^{-B_{eff}(\sigma)/T}$, goes to zero for $\sigma \rightarrow \sigma_{cr}$. There is a simple geometrical explanation for this critical behavior. First of all, we note that the value of σ_{cr} is close to the distance between the two minima. This suggests that the transition is related to the availability of a *direct jump* from the bottom of the metastable well to the other one. If we call x_{\pm} , respectively, the minimum in the right and left wells, there is a nontrivial solution of the equation $V(x_1)=V(x_+)$ with x_1 lying between the two minima, $x_- < x_1 < x_+$ (see Fig. 4): It turns out that $\sigma_{cr}=x_+-x_1$.

Indeed, for $\sigma > \sigma_{cr}$ there is a possible proposed move that brings the system from $x=x_+$ (the metastable minimum) to the point $x'=x_1$ on the other side of the barrier: Such a “non-local” move pays no energy ($\Delta E=0$) and is therefore certainly accepted by the Metropolis algorithm. In some sense, for $\sigma > \sigma_{cr}$ the barrier is not seen (or, at least, there are allowed and accepted moves that do not see it) and B_{eff} is zero.

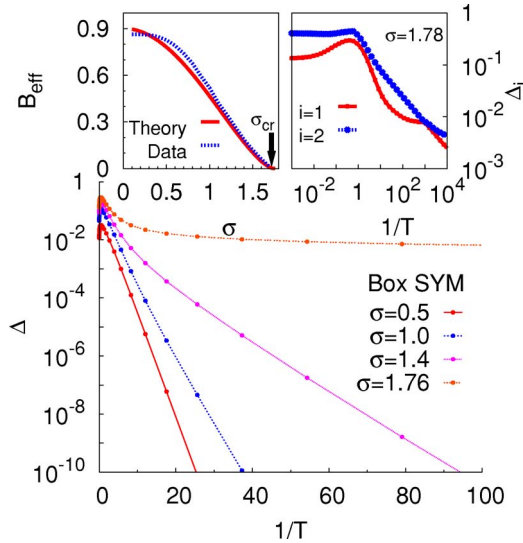


FIG. 3. (Color online) Exact diagonalization of the transition operator W for the symmetric potential with the box proposal move. We plot the spectral gap Δ versus the inverse temperature $1/T$ for several fixed values of the proposal move range σ . Top left inset: the value of the effective barrier B_{eff} seen by the system when a box proposal move with range σ is employed. The value σ_{cr} —which corresponds to a vanishing effective barrier—is indicated by a vertical arrow. Top right inset: the two first gaps $\Delta_1 = 1 - \lambda_1$ and $\Delta_2 = 1 - \lambda_2$ for a box proposal move versus the inverse temperature $1/T$ at fixed move range $\sigma = 1.78$.

Consider now the case $\sigma < \sigma_{cr}$. For every value $V = V(x_+) + \Delta V$, with $\Delta V > 0$, there are two equipotential solutions $x_1(\Delta V)$ and $x_2(\Delta V)$, such that $V(x_1) = V(x_2) = V$, lying between the two minima, but on opposite sides of the barrier—i.e., with $x_- < x_1(\Delta V) < x_2(\Delta V) < x_+$. Denote now by $d(\Delta V) = x_2(\Delta V) - x_1(\Delta V)$ the distance between such two equipoten-

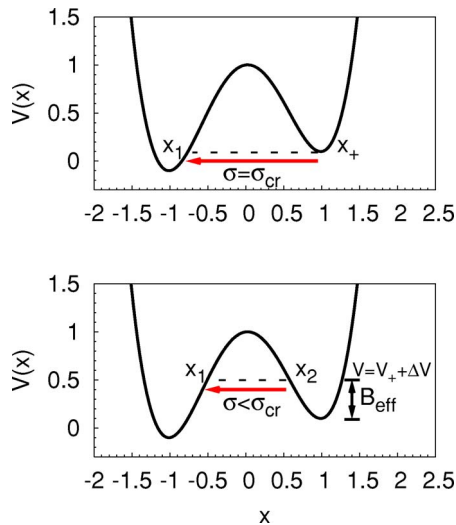


FIG. 4. (Color online) Illustrative sketch of a box move in a double-well potential. In the upper panel we consider the case $\sigma = \sigma_{cr}$, while in the lower one we illustrate the general case $\sigma < \sigma_{cr}$. Here $V_+ = V(x_+)$ is the potential at the bottom of the metastable minimum. The meaning of the other symbols is explained in the text.

tial points $[d(\Delta V)]$ is a monotonically decreasing function of ΔV . If the box width is σ , we can always find ΔV such that $\sigma = d(\Delta V)$ —i.e., such that there is a proposed move connecting $x = x_2$ to $x' = x_1$ which, once again, bypasses the top of the barrier. The effective barrier seen by the system close to the metastable minimum, for a given value of σ , is therefore just (see Fig. 4)

$$B_{\text{eff}}(\sigma) = \Delta V = d^{-1}(\sigma); \quad (9)$$

i.e., in essence, a piece of length σ is cut from the top of the barrier and is effectively not seen by the system. In other words, the effective barrier B_{eff} is just the potential energy drop ΔV which the system must overcome before a long jump $|x' - x| \sim \sigma$ is made available at no energy cost [$V(x') \sim V(x)$]. The value of $B_{\text{eff}}(\sigma)$ obtained through such a simple geometric construction is shown by a solid line in Fig. 3 (top left inset): The agreement with the numerical data—extracted from the Arrhenius fit—is remarkably good, with small deviations for very small values of σ , which are likely due to the effect of the finite grid employed in diagonalizing the transition operator W (see Sec. II A) and to finite- T effects.³¹

We would like stress that this geometrical picture is strictly true only at zero temperature and does not apply to the Gaussian or Lorentzian case, whose tails provide a small nonvanishing chance of making a “long jump” for any value of σ . The treatment of the Gaussian and Lorentzian cases needs, therefore, a more specific and technical discussion, which we sketch in Appendix A.

Summarizing, the box trial move shows a *sharp* (first-order-like) transition at a value of $\sigma_{cr} \approx 1.75$, where the effective Arrhenius barrier $B_{\text{eff}}(\sigma)$ vanishes and the gap starts to decrease as a power law, $\Delta \propto T^{1/2}$, for very small values of T (see Fig. 3).

One might suspect that the cusps shown in Fig. 2 for the Gaussian and Lorentzian cases signal, similarly to the box case, some kind of transition. A closer inspection, however, shows that this is not so: It is just a level crossing phenomenon. In Fig. 5 we plot Δ versus the inverse temperature $1/T$ for several values of σ for the Gaussian (top panel) and Lorentzian (bottom panel) proposal moves. After an initial Arrhenius-like behavior, particularly visible for the Gaussian small- σ cases, the system always, and smoothly, changes to a low- T behavior which is entirely similar to the large- σ box case: An apparent saturation of Δ followed by an avoided crossing between the first two excited states (not shown) and a final $\Delta \propto T^{1/2}$ (Ref. 30). No real transition exist in the Gaussian and Lorentzian cases: The cusp in Fig. 2 moves down toward smaller and smaller values of σ for decreasing T .

B. Classical annealing with optimal $\sigma(T)$

In the previous section we argued about the possibility of an optimal choice for the proposal move range $\sigma(T)$. This choice should guarantee the fastest *instantaneous* relaxation toward the instantaneous equilibrium distribution $P_{eq}(x)$ at any given temperature. The classical annealing performance is expected to be greatly improved by such a choice of $\sigma(T)$.

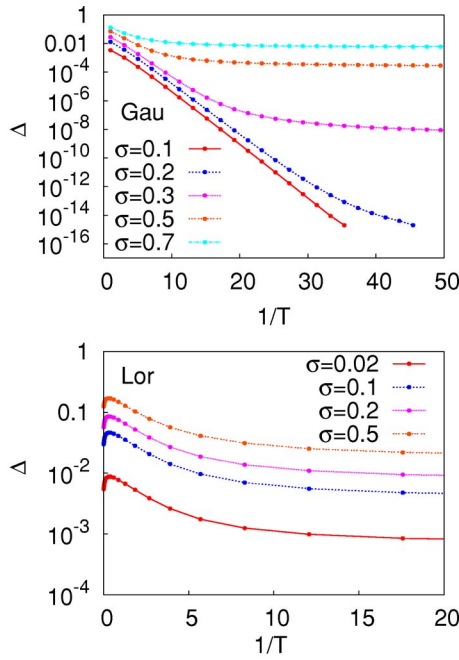


FIG. 5. (Color online) Exact diagonalization of the transition operator for the symmetric potential with Gaussian (top) or Lorentzian (bottom) move. We plot the gap Δ versus the inverse temperature $1/T$ for several fixed values of the proposal move range σ .

In practice, for each choice of proposal move (box, Gaussian, Lorentzian), we numerically determined the value of σ that maximizes the spectral gap Δ for each fixed temperature T . We will refer here to such an optimal σ as $\sigma_{opt}(T)$. In Fig. 6 we plot such an optimal choice for all the types of proposal moves introduced in Sec. II. For comparison, we also show the $\sigma(T)$ we employed in Sec. II, which is definitely smaller than σ_{opt} , for given value of T .

Having done this, we have then performed classical MC annealing runs, similarly to those reported in Sec. II, with the usual linear annealing schedule for the temperature $T(t) = T_0(1 - t/\tau)$, where τ is the annealing time. In Fig. 7 we show the MC annealing results. We stress again that the difference with respect to the runs illustrated in Fig. 1, where we took $\sigma = \sigma_0(T/T_0)^{1/2}$, is that, here, $\sigma_{opt}(T)$ is the optimal choice of σ obtained from the maximum instantaneous spectral gap of the transition matrix. We notice that the box and

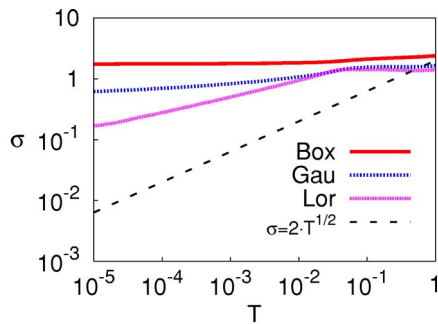


FIG. 6. (Color online) Plot of the optimal $\sigma_{opt}(T)$ for the box, Gaussian, and Lorentzian proposal moves. We also show, for comparison, the schedule $\sigma(T) \propto T^{1/2}$ employed in Sec. II.

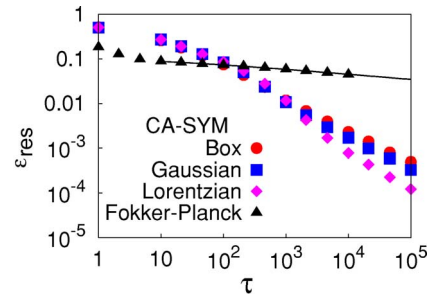


FIG. 7. (Color online) Plot of Monte Carlo classical annealings for the symmetric potential. τ is the annealing time and ϵ_{res} the residual energy (see text). We report the results of the exact integration (Fokker-Planck), together with the actual MC data for several proposal moves (box, Gaussian, Lorentzian). The MC simulations are performed with an optimal choice for $\sigma(T)$ obtained from the maximum instantaneous gap (see text).

Gaussian results are now very different from the previous ones (see Fig. 1). Moreover, interestingly, box and Gaussian data fall almost on the same curve, which is in turn very close to the Lorentzian case, the latter showing once again the best annealing results.³²

III. QUANTUM ANNEALING: PATH-INTEGRAL MONTE CARLO ANNEALING OF THE DOUBLE WELL

In I (Ref. 20) we studied, by direct numerical integration, the real- and imaginary-time Schrödinger dynamics provided by the *time-dependent* Hamiltonian

$$H = -\Gamma(t) \frac{\partial^2}{\partial x^2} + V(x), \quad (10)$$

where $\Gamma(t) = \hbar^2/[2m(t)] = \Gamma_0(1 - t/\tau)$ is the inverse mass parameter appearing in the usual nonrelativistic kinetic energy, providing the strength of the quantum fluctuations. This is a literal implementation of the quantum annealing strategy. However, as pointed out in the Introduction, this Schrödinger dynamics is suitable only for toy problems with a very limited Hilbert space and quite inapplicable for actual optimization. In order to be a viable strategy for realistic optimization, QA must resort to *stochastic*—i.e., quantum Monte Carlo (QMC)—approaches, mostly appropriate for an *imaginary-time* framework (we remind the reader that, as shown in I, working in imaginary time is actually beneficial for QA).

There are several possible QMC techniques on which a QA strategy can be build. By far, the simplest of these QMC techniques is the *path-integral Monte Carlo* method, which has already been used with some success in the QA context.^{10,11,15,19} The method does not addresses the imaginary-time Schrödinger time-dependent dynamics, but simulates an *equilibrium quantum system*, held at a small finite temperature T , where the relevant quantum parameter $\Gamma(t)$ is externally switched off in the course of a QMC simulation: t is not treated, therefore, as a proper physical time and is only replaced by a *Monte Carlo time*.³³

In the present section we will explore the potential of a PIMC-based QA strategy on the same simple one-

dimensional double-well potential. The reason for investing so much effort on this simple problem is that the simplicity of the problem landscape will allow us to have a perfect control of all the ingredients of the method. Indeed, we will learn a lot about PIMC-QA, particularly about the limitations of the method, from investigating this toy problem.

The equilibrium properties of our quantum system, for any fixed value of $\Gamma = \Gamma(t)$, are all encoded in the quantum partition function

$$Z(\beta) = \text{Tr}[e^{-\beta(\hat{T}+\hat{V})}] = \int dx \langle x | e^{-\beta(\hat{T}+\hat{V})} | x \rangle, \quad (11)$$

where

$$\hat{T} = -\Gamma \frac{\partial^2}{\partial x^2}, \quad (12)$$

from which all the thermodynamics follows. Here, as usual, $\beta = (k_B T)^{-1}$. We will drop the Boltzmann constant k_B from now on. As we see from Eq. (11), Z involves an integral over a positive distribution, $\langle x | e^{-\beta(\hat{T}+\hat{V})} | x \rangle$, which involves, however, the very difficult task of calculating the diagonal matrix element of the *exponential* of the Hamiltonian H .

The standard approach is to rewrite Eq. (11) as a Feynman *path integral*.³⁴ The main mathematical tool employed in such an approach is the so-called *Trotter theorem*, which reads

$$e^{-\beta(\hat{T}+\hat{V})} = \lim_{P \rightarrow \infty} (e^{-(\beta/P)\hat{T}} e^{-(\beta/P)\hat{V}})^P, \quad (13)$$

where P is the number of *Trotter's slices (or replicas)*, in which the imaginary time interval $[0, \hbar\beta]$ has been partitioned, each slice being of length $\Delta t = (\hbar\beta)/P$. By means of Eq. (13), and inserting $P-1$ identities in the form of $1_i = \int dx_i |x_i\rangle \langle x_i|$, one can rewrite the partition function (see, for instance, Ref. 35 for this simple derivation) as

$$Z(\beta) = \left(\frac{1}{4\pi\Gamma\Delta t} \right)^{P/2} \int \prod_{i=0}^{P-1} dx_i e^{-S_{PA}[x]} + O(\Delta t^2\beta), \quad (14)$$

where $S_{PA}[x]$ is the so-called (Euclidean) *primitive action*:

$$S_{PA}[x] = \Delta t \sum_{i=0}^{P-1} \left\{ \frac{m}{2} \left(\frac{x_{i+1} - x_i}{\Delta t} \right)^2 + V(x_i) \right\}, \quad (15)$$

with periodic boundary conditions $x_P = x_0$, as a consequence of the trace present in Eq. (11).

The similarity of Eq. (14) with the classical partition function of a *closed polymer* with P beads is evident: The polymer beads x_i can be seen as the imaginary-time snapshots $x_i = x(i\Delta t)$, $i=0, \dots, P-1$, of a fluctuating closed path $x(t)$ in the enlarged configuration space (x, t) , where t is the imaginary time. Two neighboring beads x_i and x_{i+1} interact with harmonic interactions of spring constant $K^\perp = mP^2/(\hbar\beta)^2$, originating from the propagator of the quantum kinetic term,

$$\langle x_i | e^{-(\beta/P)\hat{T}} | x_{i+1} \rangle \propto e^{-(\beta/P)(K^\perp/2)(x_{i+1} - x_i)^2}. \quad (16)$$

The strength of the harmonic interactions in imaginary time controls the amount of quantum fluctuations: A large mass m

(classical regime) results in a strong K^\perp and hence in a very “rigid” polymer, while a small mass m (quantum regime) results in a very soft and fluctuating polymer. All beads are also subject to the classical potential $V(x)$. Once we have reduced our problem to the path-integral form in Eq. (14), the standard classical Metropolis MC techniques can be used and the resulting algorithm is what is called a *path-integral Monte Carlo* algorithm. The most obvious MC moves to be used are just *single-bead* moves, exactly as in a classical MC simulation.

These are the bare bones of a PIMC approach. For the problem we are dealing with, a particle in a potential (or, more generally, for systems of quantum particles on the continuum), one can improve on the method just sketched in two possible directions: (i) by improving the quality of the approximation in Eq. (14), so as to get a smaller Trotter discretization error³⁶ for a given number of Trotter slices P used,³⁷ and (ii) by introducing MC moves which are more sophisticated than just moving a single bead at a time. Regarding (i), we will adopt a *fourth-order* approximation to the action,^{38,39} which improves the Trotter truncation error of the partition function from $O(\Delta t^2\beta)$ to $O(\Delta t^4\beta)$. More precisely, we used the so-called *Takahashi-Imada* approximation,^{38,39} which is especially suitable for continuous systems. The partition function is still given by an expression of the form of Eq. (14) with the primitive action S_{PA} replaced by the following Takahashi-Imada action:

$$S_{TIA}[x] = \Delta t \sum_{i=0}^{P-1} \left\{ \frac{m}{2} \left(\frac{x_{i+1} - x_i}{\Delta t} \right)^2 + V_{\text{eff}}(x_i) \right\}, \quad (17)$$

where the only difference with respect to S_{PA} is in the potential energy, which now reads

$$V_{\text{eff}}(x) = V(x) + \frac{1}{12} \Gamma(\Delta t)^2 \left(\frac{\partial V(x)}{\partial x} \right)^2. \quad (18)$$

As for the MC move choice (ii), we adopted smarter moves, known in the context of classical polymer simulations, which reconstruct entire pieces of the polymer, instead of a single bead at a time, through the *bisection algorithm*.^{35,40} This choice guarantees a fast relaxation to the instantaneous equilibrium distribution. Moreover, we also used global classical moves, in the form of rigid motions of the center of mass of the polymer,³⁵ so as to keep a good sampling in the final part of the annealing, where the mass m is large (Γ is small) and quantum moves are suppressed. In the next section, we will present the annealing results obtained with such an allegedly *state-of-the-art* PIMC method.

A. PIMC QA implementation I: IV-order action and bisection moves

Quantum annealing is performed by decreasing the inverse-mass parameter, $\Gamma(t) = \hbar^2/[2m(t)]$, appearing in Eq. (10), linearly to zero in a time τ , $\Gamma(t) = \Gamma_0(1-t/\tau)$ (it is understood here that both t and the total annealing time τ are MC times—i.e., measured in units of MC steps, each MC step consisting of one bisection move plus one global move). The initial condition is set to $\Gamma_0 = 0.5$ (as in I). For every

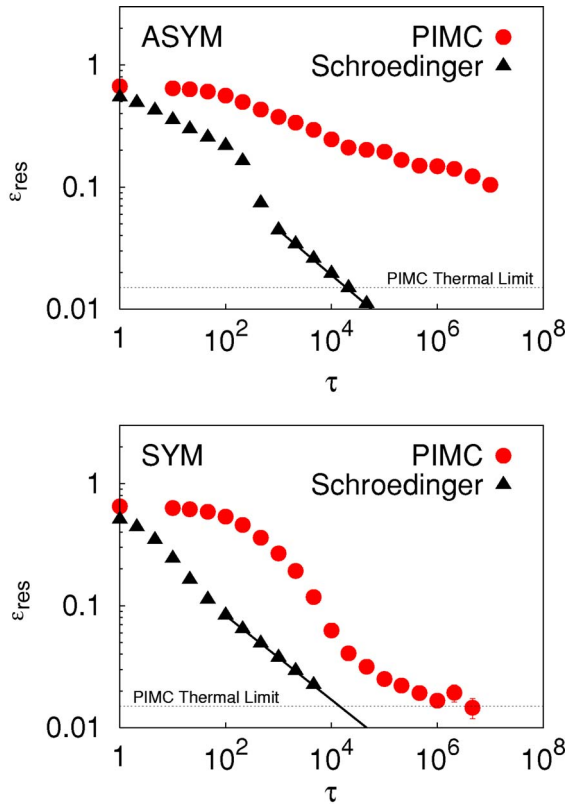


FIG. 8. (Color online) Top: PIMC-QA residual energy for the asymmetric potential using a fourth-order action and the bisection algorithm. The dashed line indicates the thermal equipartition limit $T/2$ for $T/V_0=0.03$. As a reference, the results obtained in Ref. 20 by exact integration of the imaginary-time Schrödinger equation are reported. Bottom: same as above for a symmetric potential.

value of the annealing time τ we calculated the relevant averages by repeating several times the same annealing experiment, starting from different randomly distributed initial conditions.⁴¹

In Fig. 8 we plot the final PIMC-QA residual energy obtained for both potential choices, asymmetric (top) and symmetric (bottom) (with the same parameters used in I and in Sec. II), for a fixed temperature $T/V_0=0.03$. The statistical errors are evaluated by 10^3 repetitions of each annealing run for $\tau < 10^5$ and 10^2 repetitions for $\tau > 10^5$. We employed $P=160$ with an $l=5$ bisection level (i.e., moving polymer pieces containing $2^5-1=31$ beads³⁵ for $\tau \leq 10^5$ and $P=20$ with $l=2$ (moving $2^2-1=3$ beads) for $\tau > 10^5$. We stress that the two series of data match perfectly, but obviously the use of a smaller P allows us to achieve larger annealing times τ , which would be otherwise too computationally heavy. (An analysis of the convergence behavior of the algorithm as a function of both Γ and P can be found in Ref. 42.) In the asymmetric potential case (top panel of Fig. 8)—which, we recall, presents a level crossing and a Landau-Zener (LZ) transition (see I)—the residual energy $\epsilon_{res}(\tau)$ as a function of the annealing time τ gets stuck around $\epsilon_{res} \approx 0.2$. Since this energy is comparable to the energy of the metastable state, $V(x_+)$, and definitely larger than the thermal limit $T/2 = 0.015V_0$, we see that the algorithm failed to follow adiabatically the ground state. For comparison, we also report in

Fig. 8 the residual energy data obtained by the exact imaginary-time Schrödinger annealing (reported in I), even if the time scales of the two algorithms are definitely different and unrelated. Considering, on the other hand, the symmetric potential case (bottom panel of Fig. 8), we notice that the algorithm succeeds in reaching the thermal limit $T/2$ in a reasonable amount of MC steps.

From this first comparison between PIMC-QA and exact Schrödinger QA, we appreciate that the LZ transition leads to a *severe slowing down* of the Monte Carlo algorithm, the actual tunneling event being essentially missed by the PIMC algorithm. This difficulty is also present in *static* simulations performed in the neighborhood of the LZ crossing (occurring at $\Gamma \sim 0.038$), where we find (results not shown) a dangerous loss of adiabaticity which calls for an improved sampling of the action: We will show how this improved sampling is achieved by the introduction of *instanton moves*.

B. PIMC implementation II: Adding the instanton move

In the previous section we tested an allegedly state-of-the-art PIMC-QA algorithm for the very simple problem of a particle in a double-well potential, with disappointing results. The problem is a *sampling crisis*: Our action is accurate, but its sampling misses the tunneling events, which is catastrophic when a Landau-Zener crossing occurs (asymmetric potential case). The well-known cure for this kind of sampling problem, for the case of a *perfectly symmetric* double well $V(x)=V_0(x^2-a^2)^2$, is the introduction of the so-called *instanton move* (see Ref. 43). In a nutshell, an instanton is defined as a solution of the classical equation of motion in the inverted potential, which goes from one minimum to the other. Moreover, the time-reversed path (anti-instanton) is also a solution. Instanton solutions are given by

$$x_{cl}(\tau) = \pm \tanh\left(\frac{\omega_{ins}(\tau - \tau_0)}{2}\right), \quad (19)$$

where the \pm signs denote instanton and anti-instanton, respectively, $\omega_{ins}=(8\Gamma V_0)/a^2$ is the instanton frequency, and τ_0 a configuration parameter (the instanton center). Strictly speaking, the whole classical trajectory takes an infinite (real) time, while overcoming the barrier is a very fast process (whence the name instanton). In our implementation, we made use of the imaginary-time version of an instanton–anti-instanton pair as a proposal MC move.^{42,43} In practice, the instanton move proposes to a subset of the Trotter’s slices an excursion from one minimum to the other. The move will be accepted according to the usual Metropolis criterion, with an energy competition between the possible gain in potential energy compensated by the increase in kinetic energy due to the spring stretchings (at the positions of the instanton and anti-instanton). Obviously, the instanton move will be not effective in the classical limit (small Γ , final part of the annealing), since it describes an inherently quantum effect: the tunneling process between the two wells. In particular, the instanton frequency ω_{ins} [defined in Eq. (19)] decreases with Γ , and when $1/\omega_{ins} \gtrsim \beta/2$, the tunneling time will exceed the thermal cutoff. As a consequence, the instanton–anti-

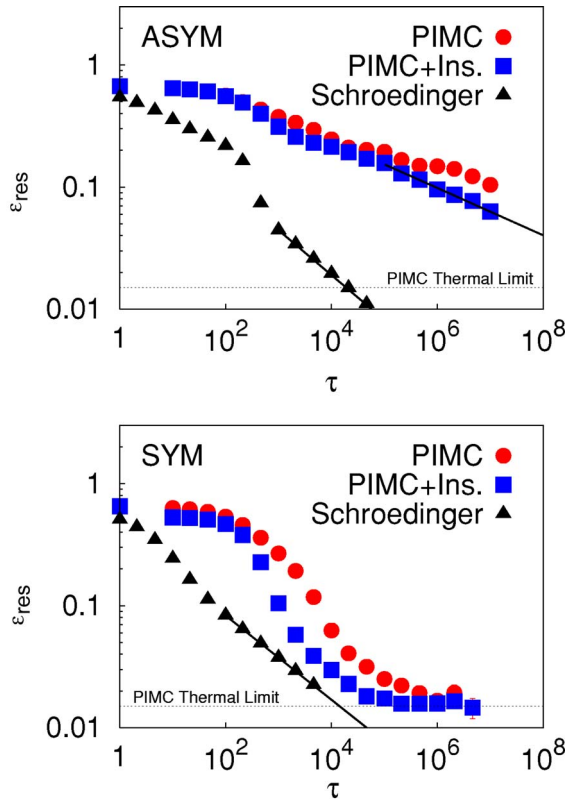


FIG. 9. (Color online) Top and bottom: same as in Fig. 8, with the instanton move allowed. As a reference the results obtained without instanton move are still reported. The data fit for the asymmetric potential case is discussed in the text.

instanton pair will reduce to a useless flat configuration over the interval $[0, \hbar\beta]$.

Before showing the results, we need to stress an important point: The instanton move is available only for potentials which are small deformations of a perfectly symmetric double-well potential—as our symmetric and asymmetric cases are (see Sec. II)—and is not the general key for solving sampling problems (ergodicity breaking) for a generic potential, whose landscape is generally poorly known. In essence, we are playing here, for demonstration purposes, an unfair game, using a *detailed information* on the potential landscape in order to correctly implement the tunneling dynamics in our PIMC.

Figure 9 shows the PIMC-QA residual energy results when the instanton move is introduced, for the case of asymmetric (top) and symmetric potential (bottom). As in the previous section, we employed $P=160$ Trotter’s slices with $l=5$ bisection level for $\tau \leq 10^5$ and $P=20$ with $l=2$ bisection level for $\tau > 10^5$. Statistical errors are evaluated with 10^3 repetitions of every annealing run.

We note how the introduction of the instanton move causes a visible improvement of the residual energy slope for the asymmetric potential case: The asymptotic power-law behavior is now quite evident, $\epsilon_{res}(\tau) \propto \tau^{-\Omega_{QA}}$, although with an exponent which is only $\Omega_{QA}=0.19$ (smaller than in the Schrödinger integration case, where $\Omega_{QA}=1/3$; see I). For the symmetric potential case, the instanton move leads to a

faster convergence to the thermal limit, but does not really give an overwhelming qualitative change. In particular we note that the partial acceptance ratio of the instanton moves is very low—around 1%—despite the fact that this proposal move was “tailored” to the potential under investigation. Nevertheless, the instanton move quantitatively changes the PIMC-QA performance. This is perhaps a generic feature of systems with barriers: namely, the crucial importance of rare events—tunneling in this case.

Unfortunately, as said, the instanton “recipe,” natural as it is for a double-well potential, cannot be generalized to generic landscapes, including more complicated potentials, not to mention actual combinatorial optimization problem. Nevertheless, the results obtained are instructive in many respects: They show clearly an important limitation of the PIMC approach, while demonstrating, once more, that a smart choice of the proposal move—even in the case of PIMC-QA—leads to immediate improvement of the annealing performance.

C. PIMC implementation III: Choosing the right kinetic energy—the Lorentzian move

It is by now a sort of *leitmotiv* of the paper that the choice of the MC move strongly influences the dynamics and, hence, the annealing behavior. We recall, in particular (see Sec. II), that in the classical annealing case the winning choice was given by a Lorentzian-distributed proposal move, which sometimes provides very long displacements.

In the present PIMC context, however, the nonrelativistic form of the kinetic energy part of the Hamiltonian, Eq. (10), *forces*, in a sense, the choice of Gaussian-distributed moves, because the free propagator of the nonrelativistic kinetic energy is a Gaussian [see Eq. (16)]. The whole bisection algorithm makes strong use³⁵ of the Gaussian nature of the free propagator.

It is natural to ask what would be the QA behavior if, instead of the standard nonrelativistic kinetic energy used so far, we employ, for example, a *relativistic* Hamiltonian of the form

$$H(t) = \Gamma(t)|p| + V(x). \tag{20}$$

One immediate consequence of this choice is that the free Gaussian propagator appearing in Eq. (16) is now replaced by a Lorentzian:

$$\langle x_i | e^{-\Delta t \Gamma |p|} | x_{i+1} \rangle = \frac{1}{\pi} \frac{\Delta t \Gamma}{(\Delta t \Gamma)^2 + (x_i - x_{i+1})^2}. \tag{21}$$

With a certain effort, we have succeeded in generalizing the bisection algorithm to implement in a very effective way the dynamics provided by this Lorentzian propagator (see Appendix B and Ref. 42 for more details).

We present here the results of a PIMC-QA approach which implements the relativistic kinetic energy through a bisection algorithm with such a “Lorentzian move.” In Fig. 10 we report the residual energy results for the case of the asymmetric (top) and symmetric (bottom) potentials. The results are obtained using $P=40$ Trotter’s slices and $l=2$ bisection steps. Averages and statistical errors are calculated, as

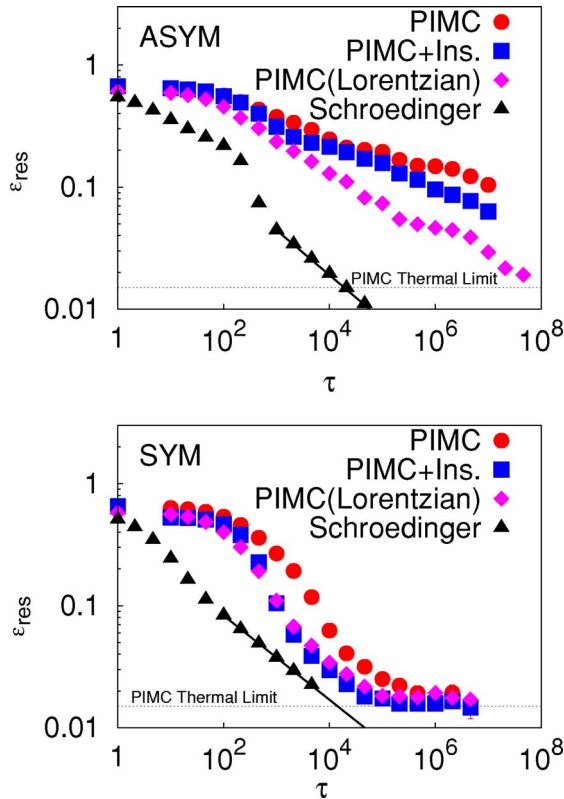


FIG. 10. (Color online) Same as in Fig. 8, but for a QA based on the relativistic kinetic energy of Eq. (20), implemented via a bisection algorithm adapted to Lorentzian moves. As a reference, the results obtained by exact integration of the imaginary-time Schrödinger and by the previous Gaussian-based PIMC-QA with and without instanton move are still reported.

usual, with 10^3 repetitions of every annealing run. It is clearly seen that, as in the classical case, the Lorentzian move—*alias*, in the present context, the *relativistic kinetic energy*—greatly accelerates the QA behavior for the difficult asymmetric potential case. As usual, the “simpler” symmetric potential case does not show a qualitative change.

IV. SUMMARY AND DISCUSSION

Summarizing, we would like to briefly stress some of the major conclusions reached in this paper.

(i) *Role of move choice in classical annealing.* The choice of Monte Carlo moves strongly influences the spectral properties of the MC Markov transition matrix, thus modifying in a strong way the annealing properties. In some sense, the *effective landscape* of the problem is really determined not only by the actual potential landscape, but also by the chosen Markov transition matrix, the latter determining the *neighborhood* of a given configuration.⁴⁴ In our double-well example a Metropolis MC simulation with Lorentzian-distributed proposed moves ended up showing the best annealing performances, outperforming not only the other classical MC choices, but also any path-integral Monte Carlo algorithm which we were able to implement.

(ii) *Tunneling and sampling difficulties of PIMC.* A tun-

neling event, and its associated Landau-Zener crossing, can cause severe difficulties in a “state-of-the-art” PIMC-QA algorithm. The difficulty is associated with a poor sampling of the action which misses the rare, but all-important, tunneling events. This problem can be cured by the *ad hoc* introduction of “instanton moves,” but this cure, while instructive, uses detailed information on the potential landscape that is generally not available. The generalization of these “instanton moves” to more complicated potentials, let alone to the generic hard optimization problem, seems basically impossible, for it would require, among other things, knowledge of the location of the minima where tunneling occurs.

(iii) *Other limitations of PIMC.* PIMC-QA suffers from a certain number of other limitations. First of all, it works with a *finite temperature* T , and this sets up a thermal energy lower bound below which we cannot possibly go (this limit was particularly clear in our double-well example). Second, a large number of Trotter slices P can cause additional sampling problems in that an effective uncorrelation of the configurations becomes harder and harder, even if a multistep bisection algorithm is employed. Moreover, the *Trotter breakup* itself [see Eq. (13)] can cause difficulties whenever, for a generic kinetic energy T , the form of the free propagator, generalizing the simple Gaussian result of Eq. (16), is not known analytically. This was indeed the difficulty met in a PIMC-QA study of the traveling salesman problem.^{6,15}

(iv) *Role of kinetic energy in QA.* The choice of the kinetic energy is clearly all important in QA: Section III C, illustrating the improvements in PIMC-QA upon using a relativistic kinetic energy, is particularly instructive. In our one-dimensional example, this choice corresponds to a free propagator of Lorentzian form, which we implemented by an appropriately modified bisection scheme (see Appendix B). The convenience of such a scheme in more general instances is not *a priori* obvious and further tests on higher-dimensional problems would be needed.

In view of the previous points, it is fair to say that exploring alternative quantum Monte Carlo schemes for performing stochastic implementations of quantum annealing remains an important and timely issue for future studies in this field. One such scheme, which in principle does not suffer from many of the limitations of the PIMC method, is the Green’s function Monte Carlo scheme. Test applications of a GFMC-QA strategy to the Ising spin glass are currently under way and will be reported shortly.⁴⁵

ACKNOWLEDGMENTS

This project was sponsored by MIUR through Grant Nos. FIRB RBAU017S8R004, FIRB RBAU01LX5H, PRIN/COFIN2003, and PRIN/COFIN2004 and by INFN (“Iniziativa trasversale calcolo parallelo”). We acknowledge illuminating discussions with Demian Battaglia. One of the authors (L.S.) is grateful to Llorenç Brulla for his generous help in discussing details about the PIMC algorithm.

APPENDIX A: A MODEL FOR THE LOW-LYING SPECTRUM OF THE METROPOLIS TRANSITION OPERATOR

In this appendix we shall use analytical tools to explain some of the most relevant features of the Markov transition

matrix spectra presented in Sec. II A. The question we want to address has to do with the low-lying spectrum of a master equation of the form

$$P_{n+1}(x') = \sum_x W(x'|x)P_n(x), \quad (\text{A1})$$

where $W(x'|x)$ is the transition operator associated with a certain choice of trial move in a Metropolis MC scheme.

As we have seen previously (see Sec. II A), the box-proposed-move case admits a simple geometrical interpretation. For the general case, which is much harder to treat, we shall sketch here the main steps of a more systematic procedure used to extract information on the gap of W , referring the reader to Ref. 42 for the many technical details.

As explained in Sec. II A, it is possible to find a complete set of eigenvectors and eigenvalues of the matrix $\bar{W}(i|j)$, defined through the short-time expansion of this equation: $W = e^{\Delta t \bar{W}}$ (Ref. 46). In particular, one can find a linear transformation from the (generally) nonsymmetric operator $-\bar{W}$ to a symmetric one \bar{H} , which preserves its spectral properties²⁸ and is easier to diagonalize in a standard way. Moreover, one can apply the same linear transformation to the original transition matrix W ; this leads to the expansion of the master equation solution reported in Eq. (8), having taken the appropriate time-continuum limit. However, in order to find the W eigenvectors and eigenvalues (and in particular its spectral gap Δ defined in Sec. II A), it is better to keep Δt finite and consider its symmetrized version, which turns out to be the exponential operator $e^{-\Delta t \bar{H}}$, where the appropriate Δt is uniquely determined by the move range σ . We notice that the spectrum of $-\bar{W}$, $\{\bar{\lambda}_i\}$, introduced in Sec. II A, is simply related to the spectrum of the original transition matrix W , $\{\lambda_i\}$, by the equation $\lambda_i = e^{-\Delta t \bar{\lambda}_i}$. In the following, we shall refer only to the gap of W .

The potential $V(x)$ will be always assumed to have the usual double-well form. At very small temperatures, the equilibrium distribution $P_{eq}(x)$ is concentrated around the local minima at $x = x_{\pm}$ (Ref. 20). Therefore, there is definitely a regime of parameters where a two-level system approximation for W must hold. By exploiting this approximation, one can show (see Ref. 42) that the spectral gap $\Delta = 1 - \lambda_1$ of the transition operator W can be expressed, at small temperature T , in terms of a well-to-well propagator of the evolution operator of \bar{H} as follows:

$$\Delta \approx \langle x_- | e^{-\Delta t \bar{H}} | x_+ \rangle e^{-(V_- - V_+)/2T}, \quad (\text{A2})$$

where $V_{\pm} = V(x_{\pm})$. The crucial quantity we need, therefore, is the well-to-well propagator $\langle x_- | e^{-\Delta t \bar{H}} | x_+ \rangle$. An expression for this propagator can be obtained by performing a Trotter decomposition and writing a Feynman path integral.³⁴ The final result can be cast—by an appropriate saddle-point approximation—in the form

$$\begin{aligned} \langle x_- | e^{-\Delta t \bar{H}} | x_+ \rangle &\propto \int_{V_+}^{V_0} dV \frac{\partial d}{\partial V} K_1(d(V - V_+); \Delta t) \\ &\times e^{-(1/T)[V - (V_+ + V_-)/2]} P(d(V - V_+)), \quad (\text{A3}) \end{aligned}$$

where $d(V - V_+)$ is the function introduced in Sec. II A (see Fig. 4) and K_1 accounts for the Gaussian fluctuations around the saddle-point classical solution of the Feynman integral, which do not modify the important exponentially activated behavior provided by the term $e^{-(1/T)[V - (V_+ + V_-)/2]}$. $P(d(V - V_+))$, finally, is just the proposal move distribution: $P(d(V - V_+)) = P(|x_1 - x_2|) = P(x_1 | x_2)$ (see Fig. 4).

This rather simple equation for the well-to-well propagator can be further treated by another standard saddle-point approximation of the relevant one-dimensional integral. We shall discuss here the cases we are interested in: box, Gaussian, and Lorentzian.

Box move. P was defined in Eq. (4). We recall that the function $d(V - V_+)$, introduced in Sec. II A, is a monotonically decreasing function of its argument, which exhibits an infinite first-order derivative for $V = V_0$. There is no true saddle point for such a move, since the functions involved in the exponential part of Eq. (A3) are all monotonic. On the other hand—in the case of the box move—the extremes of integration in Eq. (A3) can be taken as $[V_+, d^{-1}(\sigma) + V_+]$ instead of $[V_+, V_0]$, since its transition matrix $P(d(V - V_+))$ is zero outside the former interval. Therefore, the largest contribution to the whole integral is due either to $V = V_0$ or $V = d^{-1}(\sigma) + V_+$, which are the two integration extremes. It turns out that the latter choice guarantees the largest exponential and is therefore the right one. As a consequence, we can write down that

$$\langle x_- | e^{-\Delta t \bar{H}} | x_+ \rangle \propto e^{-(1/T)[d^{-1}(\sigma) + (V_+ - V_-)/2]}. \quad (\text{A4})$$

This transition amplitude is Arrhenius like and so is the gap function [see Eq. (A2)], which reads

$$\Delta(T, \sigma) \propto e^{-B_{\text{eff}}(\sigma)/T}, \quad (\text{A5})$$

where $B_{\text{eff}}(\sigma) = d^{-1}(\sigma)$, as anticipated in Sec. II A from geometrical considerations. This is indeed the behavior seen in Fig. 3.

Gaussian move. P was defined in Eq. (5). In this case, a true saddle point is possible if the following equation has a solution:

$$\frac{1}{T} - \frac{d(V - V_+)}{\sigma^2} \frac{\partial}{\partial V} d(V - V_+) = 0. \quad (\text{A6})$$

Since $\partial d / \partial V < 0$ and $d > 0$, then, in the limit $T \rightarrow 0$, we have that either $V = V_0$ or $V = V_+$ is a solution of Eq. (A6). The second choice guarantees the largest contribution to the integral in Eq. (A3) and must be therefore selected. The resulting expression for the transition amplitude is

$$\langle x_- | e^{-\Delta t \bar{H}} | x_+ \rangle \propto e^{-(V_+ - V_-)/2T} e^{-d(0)^2/2\sigma^2}, \quad (\text{A7})$$

and the corresponding gap value is

$$\Delta(T, \sigma) \propto e^{-d(0)^2/2\sigma^2}. \quad (\text{A8})$$

Figure 2 shows that such a dependence perfectly fits the simulation data. We emphasize that even the coefficient within the exponential part of the fitting function matches the theory [we recall that $d(0) = \sigma_{cr} \approx 1.75$]: Only an overall prefactor is used as fitting parameter.

As observed in Sec. II A, the previous equation will be no longer valid for very small values of the temperature T , where an (avoided) crossing between the first eigenvalue $\lambda_1 < 1$ and the second eigenvalue λ_2 will occur. On the other hand, for large values of σ we see from Fig. 2 that the data deviate from the behavior predicted by Eq. (A8). This is the consequence of another level crossing.

Lorentzian move. P was defined in Eq. (6). To cut a long story short (calculations proceed similarly to the box and Gaussian case), we write down the result. The gap value is

$$\Delta(T, \sigma) \propto \sigma, \quad (\text{A9})$$

and also in this case the model agrees with the data (see Fig. 2) within a multiplicative fitting coefficient.

APPENDIX B: THE BISECTION ALGORITHM WITH A LORENTZIAN MOVE

In this appendix we want to show how to approach, within a path-integral Monte Carlo annealing, a Hamiltonian with *relativistic* kinetic energy:

$$H = \Gamma(t)|p| + V(x). \quad (\text{B1})$$

The kinetic operator $\Gamma(t)|p|$ is singular in a real-space representation, but the corresponding density matrix operator can be written in a simple closed form

$$\rho_K(x', x; \Delta t) = \frac{1}{\pi} \frac{\Gamma \Delta t}{\Gamma^2 \Delta t^2 + (x' - x)^2}. \quad (\text{B2})$$

This is a *Lorentzian* (or *Cauchy*) *distribution*, while the kinetic density matrix operator considered in Sec. III, Eq. (16), was a Gaussian one.

It is possible to obtain a generalization of the Lévy construction³⁵ for this kind of Lorentzian move, observing that both Gaussian and Lorentzian distributions are stable under convolution (a property shared by all the so-called *Lévy distribution*⁴⁷). The main idea is to obtain a closed form of the constrained kinetic operator:

$$T_K^{(1)}(x'_i) = \frac{\rho_K(x_{i+1}, x'_i; \Delta t) \rho_K(x'_i, x_{i-1}; \Delta t)}{\rho_K(x_{i+1}, x_{i-1}; 2\Delta t)}, \quad (\text{B3})$$

which is the probability of moving the i th bead to x'_i , keeping fixed the $(i+1)$ th and $(i-1)$ th beads in x_{i+1} and x_{i-1} , respectively. We recall that this object is the building block of the bisection algorithm.³⁵ In the Lorentzian case, it reads

$$T_K^{(1)}(x'_i) = \frac{\Gamma \Delta t}{2\pi} \frac{\left(\frac{x_{i+1} - x_{i-1}}{\Gamma \Delta t}\right)^2 + 4}{\left[\left(\frac{x_{i+1} - x'_i}{\Gamma \Delta t}\right)^2 + 1\right] \left[\left(\frac{x'_i - x_{i-1}}{\Gamma \Delta t}\right)^2 + 1\right]}. \quad (\text{B4})$$

Unfortunately, this is no longer a Lorentzian distribution, and therefore it is not trivial to sample it (for instance, by making use of the usual transformation technique; see Ref. 48). However, after some algebra and a smart change of variables (taking due care of the Jacobian of such a transformation, which changes T onto \bar{T}), we obtain a simpler form

$$\bar{T}_K^{(1)}(y) = \frac{2}{\pi} \frac{a^2 + 1}{[(y+a)^2 + 1][(y-a)^2 + 1]}, \quad (\text{B5})$$

where $a = (x_{i+1} - x_{i-1}) / (2\Gamma \Delta t)$, $c = (x_{i+1} + x_{i-1}) / (2\Gamma \Delta t)$, and

$$y = \frac{x'_i}{\Gamma \Delta t} - c. \quad (\text{B6})$$

Finally, with a bit of more algebra, we find that

$$\bar{T}_K^{(1)}(y) = \frac{1}{\pi} \frac{y^2 + a^2 + 1}{[(y+a)^2 + 1][(y-a)^2 + 1]} - \frac{1}{\pi} \frac{y^2 - a^2 + 1}{[(y+a)^2 + 1][(y-a)^2 + 1]} \quad (\text{B7})$$

and also that $\bar{T}_K^{(1)}(y) < 2W_1(y)$. An efficient strategy to sample $\bar{T}_K^{(1)}(y)$ is, therefore, to sample $W_1(y)$ and then to make use of the usual rejection technique (see Ref. 48). The primitive of $W_1(y)$ can be easily computed. It reads

$$M_1(u) = \int_{-\infty}^u dy W_1(y) = \frac{1}{2\pi} \tan^{-1} \left(\frac{2u}{1 - u^2 + a^2} \right) + \frac{1}{2}. \quad (\text{B8})$$

Inverting $M_1(u)$ and making use of an equidistributed random number $u \in (0, 1]$, we find that

$$y = \begin{cases} -d - \sqrt{d^2 + (a^2 + 1)} & \text{for } u \in \left(0, \frac{1}{2}\right], \\ -d + \sqrt{d^2 + (a^2 + 1)} & \text{for } u \in \left(\frac{1}{2}, 1\right], \end{cases} \quad (\text{B9})$$

where

$$d = \cot \left[2\pi \left(u - \frac{1}{2} \right) \right]. \quad (\text{B10})$$

Provided $u' \in (0, 1]$ —another equidistributed random number—we shall accept the y obtained above according to the condition

$$2W_1(y)u' \leq W_1(y) - W_2(y). \quad (\text{B11})$$

It results from standard considerations (see Ref. 48) that the average acceptance of this method will be around 50%. [This means that in order to generate N random numbers distributed according to $T_K^{(1)}(x'_i)$, one has to try on average $2N$ times.] Finally, the original middle-point value x'_i appearing in $T_K^{(1)}(x'_i)$ is obtained by inverting the original change of variables in Eq. (B6),

$$x'_i = \Gamma \Delta t (y + c). \quad (\text{B12})$$

This sampling method is a bit cumbersome, but it can be easily implemented as a computer algorithm. Its generalization—namely, $T_K^{(l)}(x'_i)$, which represents the probability of moving the i th bead to x'_i having fixed the $(i + 2^{l-1})$ th and the $(i - 2^{l-1})$ th beads in $x_{i+2^{l-1}}$ and $x_{i-2^{l-1}}$ —can be easily derived (it is enough to change Δt with $2^{l-1}\Delta t$). This is what one needs in order to implement the bisection scheme.³⁵

In this way, a PIMC algorithm using such a bisection sampling scheme implements correctly a *relativistic* choice of Hamiltonian as in Eq. (B1). Moreover, since $[V, [|p|, V]] = 0$, for such a relativistic PIMC algorithm the simple primitive approximation is already exact to fourth order.³⁸ We finally report the virial-centroid estimators for the kinetic and potential energy for the relativistic-PIMC case:

$$K_{PA} = \frac{1}{\beta} + \frac{1}{P} \sum_{i=0}^{P-1} (x_i - \bar{x}) \frac{\partial V(x_i)}{\partial x_i},$$

$$V_{PA} = \frac{1}{P} \sum_{i=1}^P V(x_i),$$

where $\bar{x} = \frac{1}{P} \sum_{i=0}^{P-1} x_i$ is the centroid coordinate. The way to derive them is completely similar to the nonrelativistic case considered in Refs. 40 and 49.

¹W. V. Dam, M. Mosca, and U. Vazirani, in *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science* (IEEE, New York, 2001), p. 279.

²For a review, see *Quantum Annealing and Other Optimization Methods*, edited by Arnab Das and Bikas K. Chakrabarti, Lecture Notes in Physics (Springer, Berlin, 2005).

³M. Nielsen and I. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, England, 2000).

⁴D. Aharonov, W. van Dam, J. Kempe, Z. Landau, S. Lloyd, and O. Regev, quant-ph/0405098 (unpublished).

⁵A. Finnila, M. Gomez, C. Sebenik, C. Stenson, and J. Doll, Chem. Phys. Lett. **219**, 343 (1994).

⁶D. A. Battaglia, L. Stella, O. Zagordi, G. E. Santoro, and E. Tosatti, in *Quantum Annealing and Related Optimization Methods*, edited by A. Das and B. Chakrabarti (Springer-Verlag, Berlin, 2005), p. 171.

⁷Y. Lee and B. Berne, J. Phys. Chem. A **104**, 86 (2000).

⁸P. Liu and B. Berne, J. Chem. Phys. **118**, 2999 (2003).

⁹T. Kadowaki and H. Nishimori, Phys. Rev. E **58**, 5355 (1998).

¹⁰G. E. Santoro, R. Martoňák, E. Tosatti, and R. Car, Science **295**, 2427 (2002).

¹¹R. Martoňák, G. E. Santoro, and E. Tosatti, Phys. Rev. B **66**, 094203 (2002).

¹²M. Sarjala, V. Petäjä, and M. Alava, J. Stat. Mech.: Theory Exp. **2006**, P01008 (2006).

¹³Y. Lee and B. Berne, J. Phys. Chem. A **105**, 459 (2001).

¹⁴T. Gregor and R. Car, Chem. Phys. Lett. **412**, 125 (2005).

¹⁵R. Martoňák, G. E. Santoro, and E. Tosatti, Phys. Rev. E **70**, 057701 (2004).

¹⁶A. Das, B. K. Chakrabarti, and R. B. Stinchcombe, Phys. Rev. E **72**, 026701(R) (2005).

¹⁷P. Stadler and W. Schnabl, Phys. Lett. A **161**, 337 (1992).

¹⁸O. Dubois, R. Monasson, B. Selman, and R. Zecchina, (editors) (Special Issue) Theor. Comput. Sci. **265** 1–2, (2001).

¹⁹D. A. Battaglia, G. E. Santoro, and E. Tosatti, Phys. Rev. E **71**, 066707 (2005).

²⁰L. Stella, G. E. Santoro, and E. Tosatti, Phys. Rev. B **72**, 014303 (2005).

²¹T. Kadowaki, Ph.D. thesis, Tokyo Institute of Technology, 1998; quant-ph/0205020 (unpublished).

²²D. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge, England, 2000).

²³K. Vladár and A. Zawadowski, Phys. Rev. B **28**, 1596 (1983).

²⁴R. Blinc, J. Phys. Chem. Solids **13**, 204 (1960).

²⁵D. A. Huse and D. S. Fisher, Phys. Rev. Lett. **57**, 2203 (1986).

²⁶S. Szu and R. Hartley, Phys. Lett. A **122**, 157 (1987).

²⁷Y. Xiang and X. G. Gong, Phys. Rev. E **62**, 4473 (2000).

²⁸N. van Kampen, *Stochastic Processes in Physics and Chemistry*, revised and enlarged edition (North-Holland, Amsterdam, 1992).

²⁹We made use of a uniform grid of up to 10^3 points in the interval $[-3, +3]$ of the x space, just as in Ref. 20. This discretization appears to be suitable for the low-lying part of the spectrum we want to study.

³⁰The crossing occurs between the excitation due to the well-to-well transition and the first harmonic excitation in the single well. At very low temperature the latter is the smallest non-negative eigenvalue of W if the former is saturating (e.g., in the box case, when $\sigma \geq \sigma_{cr}$). For a theoretical derivation of the equation $\Delta_{\text{harm}} \propto T^{1/2}$, see Ref. 42.

³¹For $\sigma \rightarrow 0$, we have that $B_{\text{eff}} \rightarrow B = V_0 - V(x_+) \approx 0.9$, the geometrical barrier height.

³²Notice that, in the Lorentzian case, the annealing exponent with $\sigma_{\text{opt}}(T)$ is actually a bit smaller than that obtained with the naive choice $\sigma \propto T^{1/2}$ (compare Figs. 1 and 7). This small discrepancy is presumably the result of the MC implementation and of a slight ergodicity loss at small T . Indeed, we have noticed that, for $\tau > 10^4$, the system always ends up, at the end of the Lorentzian annealing, in the lower minimum, without any component

- on the metastable minimum. Since, as it turns out, the Lorentzian $\sigma_{opt}(T)$ is slightly larger than the σT used in Sec. II, this causes larger fluctuations of the results around the lower minimum.
- ³³This strategy parallels in a direct way the philosophy behind the classical simulated annealing technique, where the temperature $T(t)$ is also a parameter which is externally driven in the course of a standard equilibrium MC simulation.
- ³⁴R. Feynman, *Statistical Mechanics: A Set of Lectures* (Benjamin, New York, 1972).
- ³⁵D. Ceperley, *Rev. Mod. Phys.* **67**, 279 (1995).
- ³⁶T. Kashiwa, Y. Ohnuki, and M. Suzuki, *Path Integral Methods* (Clarendon, Oxford, 1997).
- ³⁷Since we want to use the smallest possible T (large β) and the computation cost of the PIMC algorithm scales with P , the use of a fourth-order approximation leads to a sensible speed-up of the QA simulations.
- ³⁸M. Takahashi and M. Imada, *J. Phys. Soc. Jpn.* **53**, 3765 (1984).
- ³⁹X.-P. Li and J. Broughton, *J. Chem. Phys.* **86**, 5094 (1987).
- ⁴⁰L. Brualla, K. Sakkos, J. Boronat, and J. Casulleras, *J. Chem. Phys.* **121**, 636 (2004).
- ⁴¹We made use of the virial-centroid estimators of the statistical functions (Refs. 40 and 49), in order to minimize the fluctuations in the averages.
- ⁴²L. Stella, Ph.D. thesis, SISSA, Trieste, 2005.
- ⁴³J. Negele and H. Orland, *Quantum Many-Particle Systems* (Addison-Wesley, Reading, MA, 1988).
- ⁴⁴P. Stadler, in *Complex Systems and Binary Networks*, edited by R. Lopez-Pena, R. Capovilla, R. Garca-Pelayo, H. Waelbroeck, and F. Zertuche (Springer-Verlag, Berlin, 1995), p. 77.
- ⁴⁵L. Stella and G. E. Santoro (unpublished).
- ⁴⁶In the definition of W , given in Sec. II, we employed σ instead of Δt since the first one is the relevant quantity that one can easily tune in a practical implementation of the Metropolis algorithm. Nevertheless, it is clear that there is a one-to-one correspondence between the move range σ and the time step Δt defined in the above exponential map.
- ⁴⁷R. Metzler and J. Klafter, *Phys. Rep.* **339**, 1 (2000).
- ⁴⁸M. Kalos and P. Whitlock, *Monte Carlo Methods* (Wiley, New York, 1986), Vol. 1.
- ⁴⁹M. Herman, E. Bruskin, and B. Berne, *J. Chem. Phys.* **76**, 5150 (1982).