Optimization using quantum mechanics: quantum annealing through adiabatic evolution

2006 J. Phys. A: Math. Gen. 39 R393

(http://iopscience.iop.org/0305-4470/39/36/R01)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **39** (2006) R393–R431 doi:10.1088/0305-4470/39/36/R01

TOPICAL REVIEW

Optimization using quantum mechanics: quantum annealing through adiabatic evolution

Giuseppe E Santoro[1,2] **and Erio Tosatti**[1,2]

[1] International School for Advanced Studies (SISSA) and CNR-INFM Democritos National Simulation Center, Via Beirut 2-4, I-34014 Trieste, Italy
[2] International Center for Theoretical Physics (ICTP), Trieste, Italy

E-mail: santoro@sissa.it and tosatti@sissa.it

Received 22 May 2006, in final form 21 July 2006
Published 18 August 2006
Online at stacks.iop.org/JPhysA/39/R393

Abstract

We review here some recent work in the field of quantum annealing, *alias* adiabatic quantum computation. The idea of quantum annealing is to perform optimization by a quantum adiabatic evolution which tracks the ground state of a suitable time-dependent Hamiltonian, where '$\hbar$' is slowly switched off. We illustrate several applications of quantum annealing strategies, starting from textbook toy-models—double-well potentials and other one-dimensional examples, with and without disorder. These examples display in a clear way the crucial differences between classical and quantum annealing. We then discuss applications of quantum annealing to challenging hard optimization problems, such as the random Ising model, the travelling salesman problem and Boolean satisfiability problems. The techniques used to implement quantum annealing are either deterministic Schrödinger's evolutions, for the toy models, or path-integral Monte Carlo and Green's function Monte Carlo approaches, for the hard optimization problems. The crucial role played by disorder and the associated non-trivial Landau–Zener tunnelling phenomena is discussed and emphasized.

PACS numbers: 03.67.Lx, 75.10.Nr, 03.65.Xp, 02.70.Ss, 05.10.Ln, 07.05.Tp

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Optimization has to do with finding the minimum of some given cost function; it is a very common task in science and a self-standing field of research. Quite often, the cost (or energy) function one is seeking to optimize depends on a large number N of variables (either discrete or continuous), and because of multiple constraints—which may lead to *frustration*—it will possess many local minima, a feature that makes the search for the optimal solution generically

'hard'. As non-trivial examples of hard energy functions, one can mention (i) finding the ground-state energy configuration of an assembly of N Ising spins with frustrating interactions (a *spin glass*) [1]; (ii) finding the minimal length itinerary of a travelling salesman problem (TSP) visiting N cities [2]; (iii) finding the satisfying bit assignment for an N-bit Boolean formula with many logical clauses each involving k bits (k-SAT) [3]. These examples, taken from the fields of physics (i) and computer science (ii, iii), are just a small repertoire from the vast world of disordered/frustrated physical systems, and of combinatorial optimization problems [3]; we will use them, in the present work, as our favourite illustration and benchmark problems. Common to all of them is the fact that the dimension of the *configuration space* of the N input variables increases exponentially with N—it is 2^N in the Ising and k-SAT cases, $(N-1)!$ in the TSP case—and that no obvious rule is available to find the optimal solution in such a space. On the other hand, the space is simply too large for an exhaustive enumeration to be feasible beyond a relatively small value of N ($N \sim 30$).

Theoretical computer science has classified the complexity of a given problem according to the time (and space) resources needed to solve it, in its *worst-case* instance, on a classical computer [4]. Without entering the details of such a classification (that will not be relevant in our work), we just mention that a problem belongs to the **P** or to the **NP** complexity classes if a polynomial (in N) algorithm is or is not known, respectively, which is able to solve worst-case instances. The examples quoted above all belong to the class of the **NP**-complete problems (the hardest of all the **NP** problems [3]), except for the Ising case on some two-dimensional lattices [5] or the 2-SAT problem, which are both polynomial.

Returning to physics, classical statistical mechanics has contributed a crucial input of concepts and methods to the field of optimization, starting from the pioneering idea of *thermal simulated annealing* by Kirkpatrick *et al* [6], down to the recent advances in the applications of replica methods and cavity techniques to general combinatorial optimization problems [7]. In simulated annealing, the optimization problem is tackled by the introduction of a (real or fictitious) temperature variable, which is slowly lowered (annealed) in the course of a Monte Carlo or molecular dynamics simulation [6]. This device allows exploration of the large configuration space of the problem at hand, effectively avoiding trapping at unfavourable local minima through thermal hopping above energy barriers. It makes for a very simple, robust and effective minimization tool [8, chapter 10], often much more effective than standard, gradient-based, minimization methods.

The idea of quantum annealing (QA) is an elegant and fascinating alternative to classical thermal simulated annealing (CA); it consists in helping the system escape the local minima using *quantum mechanics*—by tunnelling through the barriers rather than thermally overcoming them—with an artificial and appropriate source of quantum fluctuations (the counterpart of the temperature) initially present and slowly (adiabatically) switched off. To the best of our knowledge, this idea was first explicitly formulated, and tested in numerical simulations, in the early 1990s [9–11]. In the Ising spin glass context—more precisely, for the infinite range Sherrington–Kirkpatrick model [12]—the idea that the addition of a transverse field quantum term $-\Gamma \sum_i \sigma_i^x$ (σ_i^x being Pauli matrices at site i) might help the system in tunnelling through the infinitely high classical barriers separating the infinitely many metastable states was indeed put forward even earlier, in [13, 14]. More recently, experimental evidence in the disordered Ising ferromagnet $LiHo_{0.44}Y_{0.56}F_4$ in a transverse magnetic field showed that the QA strategy is not only feasible but presumably winning in certain cases [15, 16].

A parallel advance, again during the 1990s and with quantum mechanics as a protagonist, has been the blooming idea of quantum computation (QC) [17]. A remarkable result in this field, due to Schor [18], was the proof that the problem of factorizing a large integer of N

bits—believed to be classically **NP** (although not **NP**-complete)—can be solved with a number of quantum operations which are polynomial in N ($O(N^2)$), in the framework of the standard approach to QC (the so-called circuit theory) [17].

The remainder of this Introduction will be devoted to a brief list of basic concepts and issues.

1.1. The basic idea of quantum annealing

Let us consider QA, and concentrate discussion on classical optimization problems, where the energy function to be optimized is represented by a classical Hamiltonian, denoted by $H_{\rm cl}$. In QA, one supplements the classical Hamiltonian $H_{\rm cl}$ with a suitable, in principle arbitrary, *time-dependent* quantum kinetic term, $H_{\rm kin}(t)$, which is initially very large, for $t \leqslant 0$, and then gradually reduced to zero in a certain annealing time τ. The total Hamiltonian one aims at studying is therefore a time-dependent one given by $H(t) = H_{\rm cl} + H_{\rm kin}(t)$, where $H_{\rm kin}(0)$ dominates initially—that is, $H(0) \approx H_{\rm kin}(0)$—while, at the end of the annealing, $H_{\rm kin}(\tau) = 0$ so that we simply recover $H(\tau) = H_{\rm cl}$. The kinetic energy $H_{\rm kin}(t)$ can be chosen quite freely, with the only requirement of being represented by a local Hamiltonian. Using the Ising glass case as an illustrative example, we could take

$$H(t) = H_{\rm cl} + H_{\rm kin}(t) = -\sum_{\langle ij\rangle} J_{ij}\sigma_i^z\sigma_j^z - \Gamma(t)\sum_i \sigma_i^x, \tag{1}$$

where $H_{\rm cl} = -\sum_{\langle ij\rangle} J_{ij}\sigma_i^z\sigma_j^z$ is an Edward–Anderson disordered Ising model—with σ_i^z spin-1/2 Ising variables (z-component Pauli matrices) at each site i of a lattice, and J_{ij} random couplings—and represents the classical problem for which we are trying to find the ground state. Due to the fact that the nearest-neighbour spin–spin couplings J_{ij} are frustrating (i.e., both positive and negative) and random, finding the ground state of $H_{\rm cl}$ is generally hard. The choice of the kinetic term as a simple time-dependent transverse field $H_{\rm kin}(t) = -\Gamma(t)\sum_i \sigma_i^x$ (with σ_i^x the x-component Pauli matrix at site i) is in this case very natural, as is also directly realized in the experiment [15]. Indeed this form of $H_{\rm kin}$ is quite natural for all those problems that can be classically formulated in terms of spin-1/2 Ising variables (or bits, as in the case of k-SAT). The condition that $H_{\rm kin}(0)$ dominates initially is realized by $\Gamma(0) \gg {\rm Max}\{|J_{ij}|\}$; for $\Gamma(0) \to \infty$, the ground-state $|\Psi_0\rangle$ of $H(0) = H_{\rm cl} + H_{\rm kin}(0)$ approaches the simple product state $2^{-N/2}\prod_i(|\uparrow\rangle + |\downarrow\rangle)$, with all spins aligned in the x-direction; at the end of the annealing $\Gamma(\tau) = 0$, and the lowest energy of $H(\tau)$ coincides with the classical ground state of $H_{\rm cl}$ we are looking for. At zero temperature, the quantum state of the system $|\Psi(t)\rangle$, initially prepared in the fully quantum ground state $|\Psi_0\rangle$ of $H(t=0)$, $|\Psi(t=0)\rangle = |\Psi_0\rangle$, evolves according to the time-dependent Scrödinger equation

$$\mathrm{i}\hbar\frac{\mathrm{d}}{\mathrm{d}t}|\Psi(t)\rangle = H(t)|\Psi(t)\rangle, \tag{2}$$

to reach a final state $|\Psi(t=\tau)\rangle$. The key question is then how the residual energy $E_{\rm res}(\tau) = E_{\rm fin}(\tau) - E_{\rm opt}$ decreases for increasing annealing time τ, where $E_{\rm opt}$ is the absolute minimum of $H_{\rm cl}$, and $E_{\rm fin}(\tau)$ is the average energy attained by the quantum system after evolving for a time τ, i.e.,

$$E_{\rm res}(\tau) = E_{\rm fin}(\tau) - E_{\rm opt} = \frac{\langle\Psi(\tau)|(H_{\rm cl} - E_{\rm opt})|\Psi(\tau)\rangle}{\langle\Psi(\tau)|\Psi(\tau)\rangle}. \tag{3}$$

Generally speaking, the question posed is obviously related to the *adiabaticity* of the quantum evolution, i.e., whether the system is able, for sufficiently slow annealing (i.e., sufficiently long τ), to follow the instantaneous ground state of $H(t)$. Quite generally, also, the result

for $E_{\rm res}(\tau)$ will depend on the specific instance of the problem taken (given N, the specific choice of the couplings $\{J_{ij}\}$, in the Ising case), so that we have a *distribution* $P[E_{\rm res}(\tau)]$ of residual energies, and we can meaningfully talk about *average* (av) and *typical* (typ) residual energies $[E_{\rm res}(\tau)]_{\rm av/typ}$ (which might differ if the distribution is wide and non-Gaussian [19]). A crude, but useful, classification of the 'annealing complexity' of a problem might then be the following: A problem is on average/typically *easy* if, asymptotically for large τ,

$$\left[\frac{E_{\rm res}(\tau)}{N}\right]_{\rm av/typ} \propto \tau^{-\alpha}, \tag{4}$$

where the power-law exponent α does not vanish for $N \to \infty$; if, in contrast, $[E_{\rm res}(\tau)]_{\rm av/typ}$ decreases *slower* than a power law, for instance as

$$\left[\frac{E_{\rm res}(\tau)}{N}\right]_{\rm av/typ} \propto (\log \tau)^{-\zeta}, \tag{5}$$

then we say that the problem is on average/typically *hard*. In essence, if a problem as a power-law decrease of $[E_{\rm res}(\tau)]_{\rm av/typ}$ then we can increase the accuracy of our annealing calculation, reducing the residual energy by a given factor, say 2, by increasing the annealing time from τ to $2^{1/\alpha}\tau$, while, for instance, a logarithmically slow decrease would require, for the same improvement, an annealing time $\tau^{2^{1/\zeta}}$.

1.2. Quantum annealing and adiabatic quantum computation

As previously mentioned, the idea of annealing, in a quantum mechanics framework, is inherently related to the idea of adiabaticity. Computing by adiabatic evolution of a quantum system has become quite popular in the quantum computing (QC) community, where it is commonly known as *adiabatic quantum computation* (AQC) [20]. The standard reference for the QC community is [20], where the idea is formulated in the following alternative way: take $H_{\rm kin}$ to be time independent, and write $H(t) = (t/\tau)H_{\rm cl} + (1 - t/\tau)H_{\rm kin}$, so that $H(0) = H_{\rm kin}$ and, once again, $H(\tau) = H_{\rm cl}$. We stress, therefore, that quantum annealing and adiabatic quantum computation are two names given by two different scientific communities to the very same idea; we will therefore use QA and AQC interchangeably, or even together, QA–AQC. We mention, finally, that a generalized form of AQC has been recently shown to be polynomially equivalent [21] to the standard paradigm of QC, i.e., circuit theory, where computations are performed by the application of a sequence of universal gates (unitary operators) to a system of qubits [17]. In essence, if a problem can be efficiently solved by standard QC, then one can construct $H_{\rm init}$ and $H_{\rm final}$ (both local, the latter in general non-classical) such that a QA–AQC with $H(t) = (1 - t/\tau)H_{\rm init} + (t/\tau)H_{\rm final}$ leads to an equivalent result, at the price of an extra computational effort that scales polynomially with N. Therefore, the computational power of standard QC and that of such a generalized QA–AQC is the same [21]. The important point, however, is that $H_{\rm final}$ is not necessarily a classical Hamiltonian. Therefore, it is still not obvious whether the computational power of QA–AQC for classical optimization problems, where one insists in taking $H_{\rm final} = H_{\rm cl}$, is the same as that of standard QC. Indeed, we will suggest that even *polynomial* classical optimization problems can be *hard* for QA–AQC.

1.3. Adiabaticity and Landau–Zener tunnelling

Returning to the problem of adiabaticity, it is quite common that for a *given instance* of a problem of input size N, the Hamiltonian $H(t)$ will have an instantaneous spectrum of eigenvalues $E_n(t)$ with a strictly positive minimum gap $\Delta_{\rm min} = {\rm Min}_{0 \leqslant t < \tau}[E_1(t) - E_0(t)] > 0$ above the ground state. For instance, in the Ising case—more generally, in all cases in which

the matrix elements of $H_{\rm kin}(t)$ are non-positive, and any two configurations can be connected by the application of a suitable number of $H_{\rm kin}$—due to the Perron–Frobenius theorem [22, p 54ff], one can immediately conclude that the ground state of (1) for any non-vanishing transverse field $\Gamma(t) > 0$ has an everywhere-positive wavefunction and is *unique*, hence $\Delta_{\rm min} > 0$. *Ideally*, therefore, the adiabatic theorem of quantum mechanics [23] guarantees that the system will adiabatically follow the ground state if the annealing time τ is larger than a certain characteristic time, $\tau_{\rm LZ} \propto \Delta_{\rm min}^{-2}$, beyond which $E_{\rm res}(\tau)$ decreases very fast (see, for instance, [24], where a behaviour $E_{\rm res}(\tau) \propto \tau^{-2}$ is argued). (The earliest realization of a such a phenomenon is just the textbook 2×2 Landau–Zener problem [25, 26], see section 3.) In order for this argument to be of any real use, however, one should make sure that $\Delta_{\rm min}$ does not get very small over the ensemble of instances one is considering, and, more generally, when $N \to \infty$. In practice, unfortunately, due to disorder and the associated Landau–Zener tunnelling between broad barriers, such a minimum gap $\Delta_{\rm min}$ can be so tiny (even in systems with a relatively small value of N) that the annealing time needed to adiabatically follow the ground state is astronomically large. Later on, we will present a simple case, based on a disordered Anderson model, where this happens (see section 3). As a general remark, therefore, beware of QA–AQC results for cases where N is so small that, typically, the system will have a relatively large $\Delta_{\rm min}$: they are not really representative of the actual computational power of QA–AQC on that problem.

1.4. Real-time versus imaginary-time Schrödinger evolution

In practical implementations of QA on an ordinary classical computer, the task of following directly the time-dependent Schrödinger evolution in (2) is feasible only for toy models with a sufficiently manageable Hilbert space [11, 27, 28]; for spin-1/2 problems, this restricts N to the range $N < 20 - 30$. Actual optimization problems of practical interest usually involve astronomically large Hilbert spaces, a fact that calls for alternative quantum Monte Carlo (QMC) approaches. These QMC techniques, in turn, are usually based, in some way or another, on an *imaginary-time* (IT) framework. For instance, it is much easier to implement, stochastically, an imaginary-time Schrödinger evolution,

$$-\hbar \frac{\mathrm{d}}{\mathrm{d}t} |\Psi(t)\rangle = H(t)|\Psi(t)\rangle, \tag{6}$$

rather than the standard real-time (RT) evolution in (2). Obviously, such an evolution is non-unitary; hence the norm of $|\Psi(t)\rangle$ is not conserved. It is still perfectly legitimate to ask if, at the end of the annealing, the residual energy $E_{\rm res}^{\rm IT}(\tau)$, which is formally given by the same expression (3), is better or worse than the corresponding one obtained in a standard real-time evolution. We have recently addressed such a question in the context of simplified toy problems [28], finding that, as far as annealing is concerned, IT is essentially equivalent to RT, and, as a matter of fact, it can be quantitatively better [28]; we conjecture that, quite generally, $E_{\rm res}^{\rm IT}(\tau) < E_{\rm res}(\tau)$ for a given instance of the problem. This aspect will be directly addressed in section 2.3.

1.5. Quantum Monte Carlo studies

A number of recent studies have applied strategies related to path-integral Monte Carlo (PIMC) [29] or to diffusion quantum Monte Carlo (DMC) [30], to perform QA. PIMC is based on an imaginary-time framework, but now an equilibrium one, rather than a time-dependent one. Instead of solving (6), one simulates the equilibrium properties of the quantum partition function $Z = \mathrm{Tr}\, \mathrm{e}^{-\beta H(\Gamma)}$ for a given fixed value of the quantum coupling $\Gamma(t)$ appearing in

$H_{\rm kin}$ and sufficiently small temperature $T = 1/\beta$, reducing the value of $\Gamma(t)$ as a function of the Monte Carlo 'time', in much the same way as temperature is reduced in standard Metropolis Monte Carlo-simulated annealing. A certain success has been obtained in a number of optimization problems, such as the folding of off-lattice protein models [31, 32], the ground-state structure of classical Lennard–Jones clusters [9, 10, 33, 34], the random Ising model [35, 36] and the random field Ising model ground-state problems [37] (see section 4.2.1), the travelling salesman problem [38] (see section 4.2.2). On the other hand, for the interesting case of Boolean satisfiability (SAT)—more precisely, a prototypical **NP**-complete problem such as 3-SAT—a recent study of our group shows that PIMC annealing performs *worse* than simple CA [39] (see section 4.2.3). The same conclusion might apply to the one-dimensional (1D) Ising ANNNI model case [40].

In view of these results, it is fair to stress that it is *a priori* not obvious or guaranteed that a QA approach should do better than, for instance, CA, on a given problem. Evidently, the comparative performance of QA and CA depends in detail on the energy landscape of the problem at hand, in particular on the nature and type of barriers separating the different local minima, a problem about which very little is known in many practical interesting cases [41]. That in turn depends crucially on the type and effectiveness of the kinetic energy chosen. Unfortunately, there is still no reliable theory predicting the performance of a QA algorithm, in particular correlating it with the energy landscape of the given optimization problem.

In order to gain understanding on these problems, we have moved, more recently, one step back and concentrated attention on the simplest textbook problems where the energy landscape is well under control: essentially, one-dimensional potentials, starting from a double-well potential, the simplest form of barrier. On these well-controlled landscapes we carried out a detailed and exhaustive comparison between quantum adiabatic Schrödinger evolution, both in real and in imaginary time, and its classical deterministic counterpart, i.e., Fokker–Planck evolution [28]. This work will be illustrated in section 2.2. On the same double well potential, we also studied [42] the performance of different stochastic approaches, both classical Monte Carlo and path-integral Monte Carlo. Some of this work, which turns out to be quite instructive, is briefly presented in section 4.3.

The scope of this review is to give an overview of some of the recent work in the field of QA, from the authors' perspective. No systematic attempt has been made in being exhaustive in quoting references in the field, and we apologize for omissions: the reader can find a more extensive source of references in the recent book on the subject by Das and Chakrabarti [43]. A review of the experimental work on the disordered Ising ferromagnet $LiHo_{0.44}Y_{0.56}F_4$ is given in [16].

The rest of this review is organized as follows: section 2 illustrates the deterministic annealing approaches applied to toy problems, essentially the minimization of a function of a continuous coordinate. Section 3 discusses the crucial role played by disorder and the issue of Landau–Zener tunnelling in QA. Section 4 introduces the path-integral Monte Carlo techniques, and illustrates some of the recent applications, notably on the random Ising model, on the travelling salesman problem, and on Boolean satisfiability problems. Section 5 introduces the Green's function Monte Carlo technique, and illustrates its recent application to the same random Ising model instance previously studied by PIMC–QA as well as CA. Section 6, finally, contains a brief summary of the main points, and some concluding remarks.

2. Deterministic approaches on toy problems

The deterministic annealing approaches (both classical and quantum) described in the present section are directly related to the work of Kadowaki and Nishimori on Ising models on

small clusters ($N \leqslant 8$) [11]. Further references along the same lines include [27], where a Schrödinger evolution QA–AQC is applied to small ($N \leqslant 20$) random instances of a Boolean satisfiability problem (exact cover), and recent work on the disordered Anderson model (with $N \leqslant 20$ sites) and on the $\pm J$ Ising model on a 3×3 lattice [24].

Conceptually, one of the simplest optimization problems to illustrate is that of finding the global minimum of an ordinary function of several continuum variables with many minima. Suppose the classical Hamiltonian $H_{\rm cl}$ mentioned in the introduction is just a potential energy $V(\mathbf{x})$, (with $\mathbf{x}$ being a Cartesian vector of arbitrary dimension), of which we need to determine the absolute minimum ($\mathbf{x}_{\rm opt}, E_{\rm opt} = V(\mathbf{x}_{\rm opt})$). Assume, generally, a situation in which a steepest descent approach, i.e., the strategy of following the gradient of V, would lead to trapping into one of the many local minima of V, and would thus not work. Classically, as an obvious generalization of a steepest descent approach, one could imagine performing a stochastic (Markov) dynamics in x-space according to a Langevin equation

$$\dot{\mathbf{x}} = -\frac{1}{\eta(T)}\vec{\nabla}V(\mathbf{x}) + \xi(t), \tag{7}$$

where the strength of the noise term ξ is controlled by the squared noise correlations $\overline{\xi_i(t)\xi_j(t')} = 2D(T)\delta_{ij}\delta(t-t')$, with $\bar{\xi} = 0$. Both $D(T)$ and $\eta(T)$—with dimensions of a diffusion constant and of a friction coefficient and related, respectively, to fluctuations and dissipation in the system—are temperature-dependent quantities which can be chosen, for the present optimization purpose, with a certain freedom. The only obvious constraint is in fact that the correct thermodynamical averages will be recovered from the Langevin dynamics only if $\eta(T)D(T) = k_B T$, an equality known as Einstein's relation [44]. Physically, $D(T)$ should be an increasing function of T, so as to lead to increasing random forces as T increases, with $D(T=0) = 0$, since noise is turned off at $T = 0$. Classical annealing can in principle be performed through this Langevin dynamics, by slowly decreasing the temperature $T(t)$ as a function of time, from some initially large value T_0 down to zero. Instead of working with Langevin equation—a stochastic differential equation—one might equivalently address the problem by studying the probability density $P(\mathbf{x}, t)$ of finding a particle at position $\mathbf{x}$ at time t. The probability density is well known to obey a *deterministic* time-evolution equation given by the *Fokker–Planck* (FP) equation [44]

$$\frac{\partial}{\partial t}P(\mathbf{x}, t) = \frac{1}{\eta(T)}\,\mathrm{div}(P\vec{\nabla}V) + D(T)\nabla^2 P. \tag{8}$$

Here, the second term on the right-hand side represents the well-known *diffusion term*, proportional to the diffusion coefficient $D(T)$, whereas the first term represents the effect of the *drift* force $-\vec{\nabla}V$, inversely proportional to the friction coefficient $\eta(T) = k_B T/D(T)$ [44]. Annealing can now be performed by keeping the system for a long enough equilibration time at a large temperature T_0, and then gradually decreasing T to zero as a function of time, $T(t)$, in a given annealing time τ. We can model this by assuming $T(t) = T_0 f(t/\tau)$, where $f(y)$ is some assigned monotonically decreasing function for $y \in [0, 1]$, with $f(y \leqslant 0) = 1$ and $f(1) = 0$. In this manner the diffusion constant D in equation (8) becomes a time-dependent quantity, $D_t = D(T(t))$. The FP equation should then be solved with an initial condition given by the equilibrium Boltzmann distribution at temperature $T(t=0) = T_0$, i.e., $P(\mathbf{x}, t=0) = \mathrm{e}^{-V(\mathbf{x})/k_B T_0}$. The final average potential energy after annealing, in excess of the true minimum value, will then be simply given by

$$\epsilon_{\rm res}(\tau) = \int \mathrm{d}\mathbf{x}\, V(\mathbf{x})P(\mathbf{x}, t=\tau) - E_{\rm opt} \geqslant 0, \tag{9}$$

where $E_{\rm opt}$ is the actual absolute minimum of the potential V.

In a completely analogous manner, we can conceive using Schrödinger's equation to perform a deterministic quantum annealing (QA) evolution of the system, by introducing quantum fluctuations through a standard non-relativistic kinetic term $H_{\rm kin}(t) = -(\hbar^2/2m_t)\nabla^2$, with a fictitious time-dependent mass m_t. We are therefore led to studying the time-dependent Schrödinger problem

$$\xi \frac{\partial}{\partial t}\psi(\mathbf{x}, t) = [-\Gamma(t)\nabla^2 + V(\mathbf{x})]\psi(\mathbf{x}, t), \tag{10}$$

where $\xi = \mathrm{i}\hbar$ for a *real-time* (RT) evolution, while $\xi = -\hbar$ for an *imaginary-time* (IT) evolution. Here $\Gamma(t) = \hbar^2/2m_t$ will be our annealing parameter, playing the role that the temperature $T(t)$ had in classical annealing. Once again we may take $\Gamma(t)$ varying from some large value Γ_0 at $t \leqslant 0$—corresponding to a small mass of the particle, hence to large quantum fluctuations—down to $\Gamma(t = \tau) = 0$, corresponding to a particle of infinite mass, hence without quantum fluctuations. Again, we can model this with $\Gamma(t) = \Gamma_0 f(t/\tau)$, where f is a preassigned monotonically decreasing function. The initial condition here will be $\psi(\mathbf{x}, t = 0) = \psi_0(\mathbf{x})$, where $\psi_0(\mathbf{x})$ is the *ground state* of the system at $t \leqslant 0$, corresponding to the large value $\Gamma(t) = \Gamma_0$ and hence to large quantum fluctuations. The residual energy after annealing will be similarly given by equation (9), where now, however, the probability $P(\mathbf{x}, t = \tau)$ should be interpreted, quantum mechanically, as

$$P(\mathbf{x}, t) = \frac{|\psi(\mathbf{x}, t)|^2}{\int \mathrm{d}\mathbf{x}'|\psi(\mathbf{x}', t)|^2}.$$

In general, the residual energy will be different for RT or IT Schrödinger evolutions. We will comment further on RT versus IT Schrödinger evolution later on.

In the remaining part of this section, we will present some of the results obtained along the previous lines on simple one-dimensional potentials [28].

2.1. The harmonic potential: a warm-up exercise

Preliminary to any further treatment of a potential with barriers, and as a warm-up exercise which will be useful later on, we start here with the simple case of a parabolic potential in one dimension, $V(x) = kx^2/2$, which has a trivial minimum in $x = 0$, with $E_{\rm opt} = 0$, and no barriers whatsoever.

Let us consider classical FP annealing first. As detailed in [28], it is a matter of simple algebra to show that, for the harmonic potential, one can write a simple closed linear differential equation [45] for the average potential energy $\epsilon_{\rm pot}(t)$, which has the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\epsilon_{\rm pot}(t) = kD_t\left[1 - \frac{2}{k_B T(t)}\epsilon_{\rm pot}(t)\right], \tag{11}$$

the initial condition being simply given by the equipartition value $\epsilon_{\rm pot}(t = 0) = k_B T_0/2$. As with every one-dimensional linear differential equation, equation (11) can be solved by quadrature for any choice of $T(t)$ and $D_t = D(T(t))$. Assuming the annealing schedule to be parameterized by an exponent $\alpha_T > 0$, $T(t) = T_0(1 - t/\tau)^{\alpha_T}$, τ being the annealing time, and the diffusion constant $D(T)$ to behave as a power law of temperature, $D(T) = D_0(T/T_0)^{\alpha_D}$ with $\alpha_D \geqslant 0$, we can easily extract from the analytical solution for $\epsilon_{\rm pot}(t)$ the large-τ asymptotic behaviour of the final residual energy $\epsilon_{\rm res}(\tau) = \epsilon_{\rm pot}(t = \tau)$. That turns out to be

$$\epsilon_{\rm res}(\tau) \approx \tau^{-\Omega_{\rm CA}} \qquad \text{with} \qquad \Omega_{\rm CA} = \frac{\alpha_T}{\alpha_T(\alpha_D - 1) + 1}. \tag{12}$$

Trivial as it is, annealing proceeds here extremely fast, with a power-law exponent $\Omega_{\rm CA}$ that can increase without bounds (for instance if $\alpha_D = 1$) upon increasing the exponent α_T of the annealing schedule.

Consider now the Schrödinger evolution problem for this potential,

$$\xi \frac{\partial}{\partial t}\psi(x,t) = \left[-\Gamma(t)\nabla^2 + \frac{k}{2}x^2\right]\psi(x,t) \qquad \psi(x,t=0) = \psi_0(x), \tag{13}$$

where $\psi_0(x) \propto \exp(-B_0 x^2/2)$ is the ground-state Gaussian wavefunction corresponding to the initial value of the Laplacian coefficient $\Gamma(t=0) = \Gamma_0$, and $\xi = i\hbar$ or $\xi = -\hbar$ for RT or IT evolution, respectively. As shown in [28], a Gaussian *ansatz* for $\psi(x,t)$, of the form $\psi(x,t) \propto \exp(-B_t x^2/2)$ with $\mathrm{Real}(B_t) > 0$, satisfies the time-dependent Schrödinger equation as long as the inverse variance B_t of the Gaussian satisfies the following ordinary nonlinear first-order differential equation:

$$-\xi \dot{B}_t = k - 2\Gamma(t)B_t^2 \qquad B_{t=0} = B_0 = \sqrt{\frac{k}{2\Gamma_0}}. \tag{14}$$

Contrary to the classical case, there is no simple way of recasting the annealing problem in terms of a closed linear differential equation for the average potential energy $\epsilon_{\mathrm{pot}}(t)$. The final residual energy $\epsilon_{\mathrm{res}}(\tau) = \epsilon_{\mathrm{pot}}(t=\tau)$ is still expressed in terms of B_τ (or better, of its real part $\Re(B_\tau)$),

$$\epsilon_{\mathrm{res}}(\tau) = \frac{\int \mathrm{d}x\, V(x)|\psi(x,t=\tau)|^2}{\int \mathrm{d}x|\psi(x,t=\tau)|^2} = \frac{k}{4\Re(B_\tau)}, \tag{15}$$

but the behaviour of B_t must be extracted from the study of the nonlinear equation (14). The properties of the solutions of equation (14) are studied in detail in [28], where it is shown that

(i) $\epsilon_{\mathrm{res}}(\tau)$ cannot decrease faster than $1/\tau$, for large τ, i.e., a power-law exponent $\epsilon_{\mathrm{res}}(\tau) \approx \tau^{-\Omega_{\mathrm{QA}}}$ is bounded by $\Omega_{\mathrm{QA}} \leqslant 1$.

(ii) Adopting a power-law annealing schedule $\Gamma(t) = \Gamma_0(1 - t/\tau)^{\alpha_\Gamma}$, the exponent Ω_{QA} for the IT case is

$$\Omega_{\mathrm{QA}} = \frac{\alpha_\Gamma}{\alpha_\Gamma + 2}, \tag{16}$$

increasing towards the upper bound 1 as α_Γ is increased towards ∞. In particular, a linear annealing schedule, $\alpha_\Gamma = 1$, leads to $\Omega_{\mathrm{QA}} = 1/3$.

(iii) RT quantum annealing proceeds with exactly the same exponent Ω_{QA} as IT quantum annealing—although $\epsilon_{\mathrm{res}}^{\mathrm{RT}}(\tau) \geqslant \epsilon_{\mathrm{res}}^{\mathrm{IT}}(\tau)$ in general—except that the limit $\alpha_\Gamma = \infty$ (abrupt switch-off of the Laplacian coefficient) is singular in the RT case.

Note that, in the present continuum context, the residual energy cannot decrease faster $1/\tau$, contrary to the $1/\tau^2$ decrease observed in small-sized discrete problems (disordered Anderson and random Ising models) [24].

Summarizing, we have learned that, for a single parabolic valley in configuration space, both CA and QA proceed with power laws, but CA can be much more efficient than QA, with an arbitrarily larger power-law exponent. We underline however that this is merely an academic matter at this point, steepest descent being much more efficient than both CA and QA in such a simple case. The power of annealing, in particular of QA, shows up only when potentials with barriers are considered.

2.2. The simplest barrier: a double-well potential

Consider, as a potential $V(\mathbf{x})$ to be optimized, a slightly generalized double-well potential in one dimension

$$V_{\rm asym}(x) = \begin{cases} V_0 \dfrac{\left(x^2 - a_+^2\right)^2}{a_+^4} + \delta x & \text{for} \quad x \geqslant 0 \\ V_0 \dfrac{(x^2 - a_-^2)^2}{a_-^4} + \delta x & \text{for} \quad x < 0, \end{cases} \tag{17}$$

with, in general, $a_+ \neq a_-$, both positive, V_0, and δ real constants. (The discontinuity in the second derivative at the origin is of no consequence in our discussion.) In the absence of the linear term ($\delta = 0$), the potential has two degenerate minima located at $x_- = -a_-$ and $x_+ = a_+$, separated by a barrier of height V_0. When a small linear term $\delta > 0$ is introduced, with $\delta a_\pm \ll V_0$, the two degenerate minima are split by a quantity $\Delta_V \approx \delta(a_+ + a_-)$, the minimum at $x \approx -a_-$ becoming slightly favoured. For reasons that will be clear in a moment, it is useful to consider the situation, which we will refer to as 'asymmetric double well', in which the two wells possess definitely distinct *curvatures* at the minimum (i.e., their widths differ too), realized by taking $a_+ \neq a_-$. (To lowest order in δ, we have $V''(x = x_\pm) = 8V_0/a_\pm^2$.) In particular, we shall examine the case in which the metastable 'valley' at x_+ is 'wider' than the absolute minimum at x_-, a situation realized by choosing $a_+ > a_-$. This curvature asymmetry will have a rather important effect on the quantum evolution, since, as we shall see, for small and intermediate values of the mass of the particle, the wavefunction of the system will be predominantly located on the metastable minimum. Obviously, if we set $a_+ = a_- = a$, and $\delta = 0$ we recover the standard double-well potential. QA work on the latter problem, via an approximate solution of the imaginary-time Schrödinger problem with an Gaussian ansatz for the wavefunction, is discussed in [9].

We now present the results obtained by the annealing schemes introduced in section (2). The Fokker–Planck and the Schrödinger equation (both in RT and in IT) can be integrated numerically using a fourth-order adaptive Runge–Kutta method [8], after discretizing the x variable in a sufficiently fine real-space grid [28]. For the FP classical annealing, the results shown are obtained with a linear temperature schedule, $T(t) = T_0(1 - t/\tau)$, and a diffusion coefficient simply proportional to $T(t)$, $D_t = D_0(1 - t/\tau)$. (Consequently, the friction coefficient is kept constant in t, $\eta_t = k_B T(t)/D_t = k_B T_0/D_0$.) Similarly, for the Schrödinger quantum annealing we show results obtained with a coefficient of the Laplacian $\Gamma(t)$ vanishing linearly in a time τ, $\Gamma(t) = \Gamma_0(1 - t/\tau)$.

Figure 1 shows the results obtained for the final annealed probability distribution $P(x, t = \tau)$ at different values of τ, for both the Fokker–Planck (FP-CA, panel (a)) and the Schrödinger imaginary-time case (IT-QA, panel (b)), for an 'asymmetric' double-well potential $V_{\rm asym}(x)$, with $V_0 = 1$ (our unit of energy), $a_+ = 1.25, a_- = 0.75$, $\delta = 0.1$. Figure 1(*c*) summarizes the results obtained for the residual energy $\epsilon_{\rm res}(\tau)$ in equation (9).

We note immediately that QA wins over FP-CA, giving a better annealing for large enough value of τ. The RT-QA, which behaves as its IT counterpart for a symmetric double-well ($a_+ = a_-$, see [28]), shows a slightly different behaviour from IT-QA in the asymmetric case (see below for comments). We discuss first the FP-CA data (panel (*a*) and (*c*) of figure 1). Starting from an initially broad Boltzmann distribution at a high $T = T_0 = V_0$, $P(x, t = 0)$ (solid line), the system quickly sharpens the distribution $P(x, t)$ into two well-defined and quite narrow peaks located around the two minima $x_\pm$ of the potential. This agrees very well with what CA does for a harmonic potential [28]. If we denote by $p_\pm$ the integral of each of the two narrow peaks, with $p_- + p_+ = 1$, it is clear that the problem has effectively been reduced

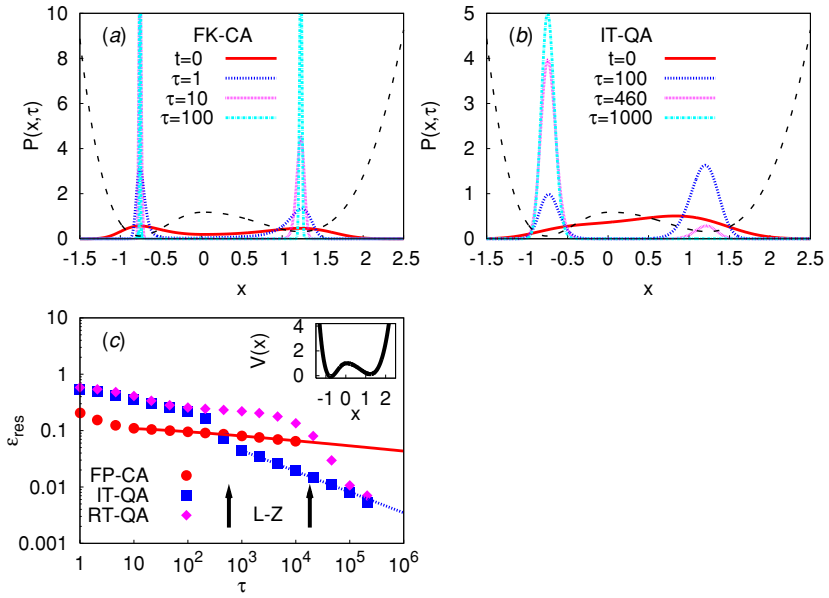

Figure 1. (a, b): the annealed final probability distribution $P(x, t = \tau)$ at different values of the annealing time τ, for both the Fokker–Planck classical annealing (FP-CA, panel (a)), and the imaginary-time Schrödinger quantum annealing (IT-QA, panel (b)). (c): final residual energy $\epsilon_{\rm res}(\tau)$ versus annealing time τ for quantum annealing in real-time (RT-QA) and imaginary-time (IT-QA), compared to the Fokker–Planck classical annealing (FP-CA). The solid line in (c) is a fit of the FP-CA data. Here the double-well potential (dashed line in (a, b), inset of (c)) is given by equation (17) with $a_+ = 1.25$, $a_- = 0.75$, $V_0 = 1$, $\delta = 0.1$.

to a *discrete* two-level problem. The time evolution of $p_\pm$, therefore, obeys a discrete master equation which involves the thermal promotion of particles over the barrier, V_0, of the form presented and discussed by Huse and Fisher in [46]. They show that, apart from logarithmic corrections, the leading behaviour of the residual energy is of the form $\epsilon_{\rm res} \sim \tau^{-\Delta_V/B}$, with the power-law exponent controlled by the ratio Δ_V/B between the energy splitting of the two minima Δ_V and the barrier $B = V_0 - V(x_+)$. As shown in figure 1(c) (solid line through solid circle), the power-law asymptotic behaviour anticipated by Huse and Fisher fits nicely our FP-CA residual energy data (solid circles), as long as the logarithmic corrections are accounted for in the fitting procedure [28]. Obviously, we can make the exponent as small as we wish by reducing the linear term coefficient δ, and hence the ratio Δ_V/B, leading to an exceedingly slow classical annealing.

The behaviour of the QA evolution is remarkably different. Observe, as a first point, that the final annealed wavefunctions narrow only slowly around the minimum of the potential, although the residual energy asymptotics of QA are clearly winning over FP-CA. The asymptotic behaviour of the QA residual energy is $\epsilon_{\rm res}(\tau) \propto \tau^{-1/3}$, indicated by the dashed line in figure 1(c). This rather strange exponent was anticipated by equation (16) for the Schrödinger annealing with a linear schedule $\Gamma(t)$ within a harmonic potential (the lower minimum valley). Going back to figure 1(b), the initial wavefunction squared $|\psi(x, t = 0)|^2$ corresponds to a quite small mass (a large $\Gamma_0 = 0.5$), and is broad and delocalized over both minima (solid line). As we start annealing, and if the annealing time τ is relatively short—that is, if $\tau < \tau_c$, with a characteristic time τ_c which depends on which kind of annealing,

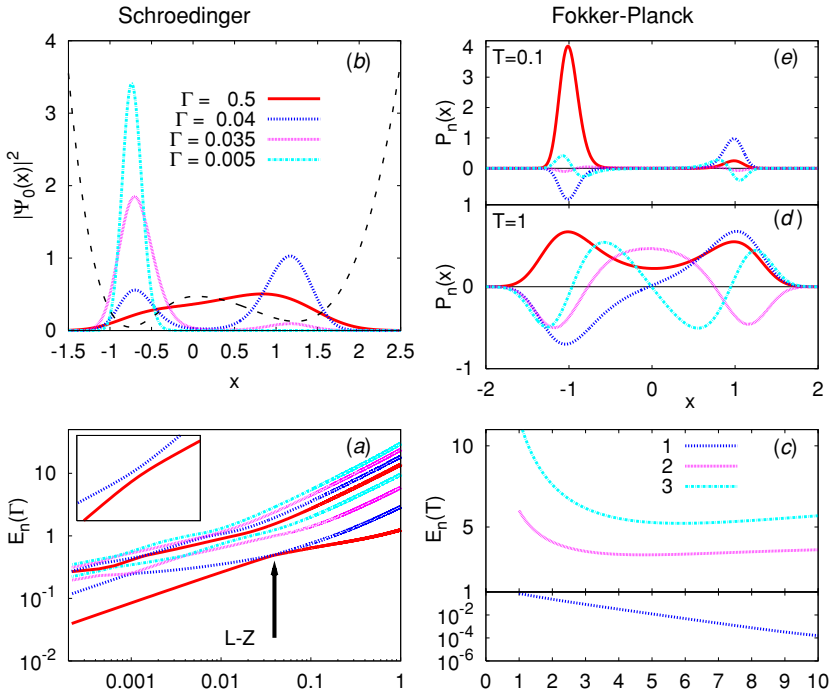

Figure 2. (a, b): instantaneous eigenvalues (a) and ground-state wavefunctions (b) of the Schrödinger problem $H\psi = E\psi$ for different values of Γ, for the potential in equation (17) with $a_+ = 1.25, a_- = 0.75, V + 0 = 1, \delta = 0.1$. Note the clear Landau–Zener avoided crossing in (a), indicated by the arrow and magnified in the inset. (c)–(e): instantaneous Fokker–Planck eigenvalues (c) as a function of temperature T, and the corresponding eigenstates for $T = 1$ (d), and $T = 0.1$ (e), for the same potential used in (a, b).

RT or IT, we perform—the final wavefunction becomes mostly concentrated on the *wrong minimum*, roughly corresponding to the ground state with a still relatively large $\Gamma_1 < \Gamma_0$ (see also figure 2 and accompanying discussion). The larger width of the wrong valley is crucial in causing that, giving a smaller quantum kinetic energy contribution, so that tunnelling to the other (deeper) minimum does not immediately occur, and the system is trapped in the wrong but broader valley. By increasing τ, there is a crossover: the system finally recognizes the presence of the other minimum, and effectively tunnels into it, with a residual energy that, as previously mentioned, decays asymptotically as $\epsilon_{\rm res}(\tau) \propto \tau^{-1/3}$ (dashed line in figure 1(c)). There is a characteristic annealing time τ_c—different in the two Scrödinger cases, RT and IT—above which tunnelling occurs, and this shows up as the clear crossover in the residual energy behaviour of both IT and RT, shown in figure 1(c).

These findings can be rationalized by looking at the *instantaneous* (adiabatic) eigenvalues and eigenstates of the associated time-independent Schrödinger problem, which we show in figures 2(a) and (b). Looking at the instantaneous eigenvalues shown in figure 2(a) we note a clear avoided crossing occurring at $\Gamma = \Gamma_{\rm LZ} \approx 0.038$, corresponding to a resonance condition between the states in the two different valleys of the potential. For $\Gamma > \Gamma_{\rm LZ}$ the ground-state wavefunction is predominantly concentrated in the wider but metastable valley, while for $\Gamma < \Gamma_{\rm LZ}$ it is mostly concentrated on the deeper and narrower global minimum valley. In the full time-dependent RT evolution, transfer to the lower valley is a Landau–Zener problem

[25, 26]: the characteristic time τ_c for the tunnelling event is given by $\tau_{\rm LZ} = \hbar\alpha\Gamma_0/2\pi\Delta^2$, where α is the relative slope of the two crossing branches as a function of Γ, 2Δ is the gap at the avoided crossing point and Γ_0 is the initial value of the annealing parameter. (For the case shown in figure 2, we have $2\Delta = 0.0062, \alpha = 2.3$, hence $\tau_{\rm LZ} \approx 18\,980$, see rightmost arrow in figure 1(*c*).) The Landau–Zener probability of jumping, during the evolution, from the ground state onto the 'wrong' (excited) state upon fast approaching of the avoided level crossing is $P_{\rm ex} = {\rm e}^{-\tau/\tau_{\rm LZ}}$, so that adiabaticity applies only if the annealing is slow enough, $\tau > \tau_{\rm LZ}$. Note that the gap 2Δ, and hence the probability of following adiabatically the ground state, can be made arbitrarily small by increasing the asymmetry of the two well, i.e., by making $a_+ \gg a_-$. The IT characteristic time is smaller, in the present case, than the RT one. This point is discussed in some detail in [28]. In a nutshell, the reason for this is the following. After the system has jumped into the excited state, which occurs with a probability $P_{\rm ex} = {\rm e}^{-\tau/\tau_{\rm LZ}}$, the residual IT evolution will filter out the excited state; this relaxation towards the ground state is controlled by the annealing rate as well as by the average gap seen during the residual evolution. Numerically, the characteristic time τ_c seen during the IT evolution is of the order of $\hbar/(2\Delta)$, see leftmost arrow in figure 1(*c*), rather than being proportional to $1/\Delta^2$ as $\tau_{\rm LZ}$ would imply.

Obviously, instantaneous eigenvalues/eigenvectors can be studied for the Fokker–Planck equation as well; their properties, however, are remarkably different from the Landau–Zener scenario just described for the Schrödinger case. Figure 2(*c*) shows the first four low-lying eigenvalues of the FP equation as a function of T, while figures 2(*d*) and (*e*) show the corresponding eigenstates for two values of the temperature, $T/V_0 = 1$ and $T/V_0 = 0.1$, for the same asymmetric potential used in figures (1) and (2). The lowest eigenvalue of the FP operator is identically 0, and the corresponding eigenvector [44] is the Boltzmann distribution ${\rm e}^{-V(x)/k_BT}$, with roughly symmetric maxima on the two valleys. The first excited state corresponds to a function peaked on the two valleys but with a node at the origin, and is separated from the ground state by an exponentially small Arrhenius-like gap ${\rm e}^{-B/k_BT}$. Higher excited states are separated by a very large gap, so that, effectively, only the two lowest lying states dominate the dynamics at small temperature. The reduction of a continuum double-well FP classical dynamics onto a discrete effective two-level system, previously noted, is quite evident from this form of the spectrum. In contrast, the true quantum case does not allow a discrete two-level system description to hold for small enough Γ. Indeed, when $\Gamma < \Gamma_{\rm LZ}$ the tower of oscillator states within the valley at x_- is always very close and collapsing in energy onto the actual ground state, and the quantum annealing evolution reduces effectively to a particle in a single harmonic well. This explains the rather large width of the final distributions $P(x, \tau)$ observed in the quantum case.

Summarizing, we have found that QA and CA proceed in a remarkably different way. CA is sensitive to the *height* of the barrier, more precisely to the ratio Δ_V/B between the energy offset Δ_V of the two minima, and the barrier height B. In contrast, QA crucially depends on the tunnelling probability between the two valleys, which is reflected in a Landau–Zener (avoided crossing) gap: a *wide* tunnelling barrier is obviously bad for QA. Finally, we noted that RT and IT proceed with different characteristic times: we discuss this issue a bit more in the following section.

2.3. Real- versus imaginary-time Schrödinger evolution

A Schrödinger dynamics in imaginary time (IT) is clearly much more convenient than that in real time (RT) for QA simulations on current classical computers: but does it make a

difference in the final results? The answer to this question is, we believe: no, it does not make a difference, in essence, although IT does give quantitatively better results.

To qualify this statement, let us denote by $|\Psi^{(\xi)}(t)\rangle$ the solution of the Scrödinger equation

$$\xi \frac{\mathrm{d}}{\mathrm{d}t}|\Psi^{(\xi)}(t)\rangle = [H_{\mathrm{cl}} + H_{\mathrm{kin}}(t)]|\Psi^{(\xi)}(t)\rangle \qquad |\Psi^{(\xi)}(t_0)\rangle = |\Psi_0\rangle, \tag{18}$$

where we assume that $|\Psi_0\rangle$ is the ground state of the initial Hamiltonian at time t_0, $H_{\mathrm{cl}} + H_{\mathrm{kin}}(t_0)$, while $\xi = \mathrm{i}\hbar$ for RT or $\xi = -\hbar$ for IT. By definition, the final residual energy after annealing up to time $t_f = t_0 + \tau$, where the kinetic energy is finally turned off, is given by

$$\epsilon_{\mathrm{res}}^{(\xi)}(\tau) = \frac{\langle\Psi^{(\xi)}(t_0+\tau)|H_{\mathrm{cl}}|\Psi^{(\xi)}(t_0+\tau)\rangle}{\langle\Psi^{(\xi)}(t_0+\tau)|\Psi^{(\xi)}(t_0+\tau)\rangle} - E_{\mathrm{opt}}. \tag{19}$$

We conjecture that the residual energies for the two alternative ways of doing a Schrödinger evolution verify the following: (i) the IT residual energy is *not larger* than the RT one, that is

$$\epsilon_{\mathrm{res}}^{(\mathrm{IT})}(\tau) \leqslant \epsilon_{\mathrm{res}}^{(\mathrm{RT})}(\tau), \tag{20}$$

and (ii) in many problems, the leading asymptotic behaviour, for $\tau \to \infty$, might be identical for $\epsilon_{\mathrm{res}}^{(\mathrm{IT})}(\tau)$ and $\epsilon_{\mathrm{res}}^{(\mathrm{RT})}(\tau)$.

Expectation (i) seems very reasonable, and is inspired by the time-independent case, where it is well known that the IT Schrödinger dynamics tends to 'filter the ground state' out of the initial trial wavefunction, as long as the gap between the ground state and the first excited state is non-zero. However, we have here a time-dependent situation, and the result is *a priori* not guaranteed. We do not have a proof of this statement, but we have verified it in all the cases where an explicit integration of the Schrödinger equation was possible (see, for instance, the results of the previous section). Needless to say, we have no proof of (ii) either, but, again, it never failed in all our tests.

The simplest time-dependent problem where one can test our conjectures is the discrete two-level system (TLS) problem. Here, in terms of Pauli matrices, $H_{\mathrm{cl}} = \Delta\sigma^z$, while $H_{\mathrm{kin}}(t) = -\Gamma(t)\sigma^x$, with $\Gamma(t) = -vt$. The full $H(t)$ is therefore

$$H(t) = \Delta\sigma^z - \Gamma(t)\sigma^x. \tag{21}$$

The annealing interpretation is very simple: the classical optimal state is $|\downarrow\rangle$, with energy $E_{\mathrm{opt}} = -\Delta$, separated from the excited state $|\uparrow\rangle$ by a gap 2Δ. The kinetic term induces transitions between the two classical states. Starting from the ground state of $H(t_0)$ at time $t_0 = -\tau$ we let the system evolve up to time $t_f = t_0 + \tau = 0$, at which point the Hamiltonian is entirely classical, $H(t_f = 0) = H_{\mathrm{cl}} = \Delta\sigma^z$. The probability of missing the instantaneous final ground-state $|\downarrow\rangle$, ending up instead with the excited-state $|\uparrow\rangle$, is $P_{\mathrm{ex}}(0) = |\langle\uparrow|\Psi^{(\xi)}(0)\rangle|^2/\langle\Psi^{(\xi)}(0)|\Psi^{(\xi)}(0)\rangle$. In principle, P_{ex} depends, for given Δ, both on the initial $\Gamma(t_0) = \Gamma_0 = v\tau$ and on the annealing time τ. The really important parameter, however, turns out to be the ratio v between these two quantities, which determines the 'velocity of annealing': taking $\tau \to \infty$ (i.e., $t_0 \to -\infty$), and $\Gamma_0 \to \infty$ with $\Gamma(t) = -vt$ for every t, the problem can be solved analytically (in terms of parabolic cylinder functions, see for instance [11] for the RT case) for both RT and IT. The probability $P_{\mathrm{ex}}(0)$ of ending into the excited state can be expressed in terms of the variable $\gamma = \Delta^2/4v$. The explicit expressions, in terms of Gamma functions, are

$$P_{\mathrm{ex}}(0) = \frac{|R+1|^2}{2(1+|R|^2)} \qquad \text{with} \qquad R = \mathrm{e}^{\mathrm{i}\phi_0}\frac{1}{\sqrt{\gamma}}\frac{\boldsymbol{\Gamma}(1+z)}{\boldsymbol{\Gamma}(1/2+z)}, \tag{22}$$

where $\phi_0 = 3\pi/4$ and $z = \mathrm{i}\gamma$ for RT, while $\phi_0 = \pi$ and $z = \gamma$ for IT. A plot of P_{ex} for both RT and IT is shown in figure 3(*a*) as a function of $\gamma = \Delta^2/4v$. Note that (i) the IT result

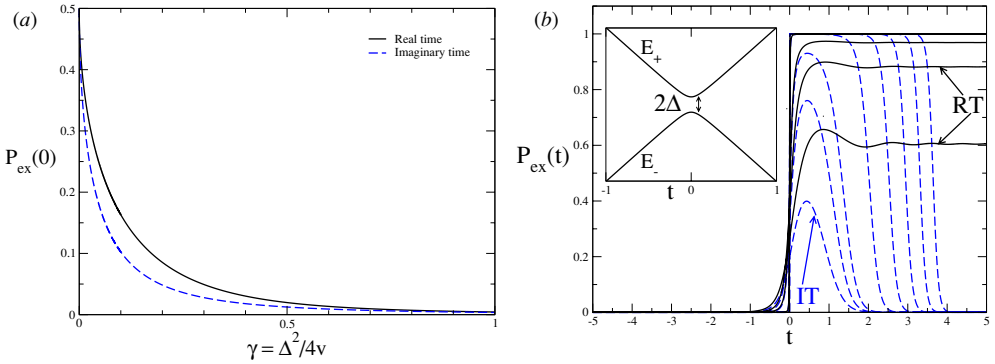

Figure 3. (*a*) The probability $P_{\rm ex}(0)$ of ending up into the excited state, given by equation (22), for the discrete two-level system problem in equation (21), for both imaginary-time (IT, dashed line) and real-time (RT, solid line) Schrödinger annealing. The large γ behaviour of $P_{\rm ex}(0)$ is, in both cases, given by $P_{\rm ex}(0) \approx 1/(256\gamma^2)$. (*b*) Comparison between the RT (solid lines) and the IT (dashed lines) evolution of a Landau–Zener problem, equation (23), for several values of the tunnelling gap 2Δ (the values of Δ shown are $\Delta = 0.4, 0.2, 0.1, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}$, while $v = 1$). The inset shows the two instantaneous eigenvalues of the problem, $E_{\pm}(t)$, as a function of t.

for $P_{\rm ex}$ (dashed line) is always below the RT result, (ii) the difference between the two curves is only quantitative: one can verify analytically that the leading behaviour for large γ is the same in both cases, i.e., $P_{\rm ex} \approx 1/(256\gamma^2)$. Similar results are obtained by direct numerical integration of the Schrödinger equation for finite Γ_0 and τ, and with other forms of $\Gamma(t)$.

With the same toy model, we can illustrate another point raised in the previous section: what happens to the IT evolution *after* a Landau–Zener avoided crossing gap is encountered? The Hamiltonian we consider is essentially that of equation (21), simply rotated in spin space,

$$H(t) = -vt\sigma^z - \Delta\sigma^x. \tag{23}$$

In the absence of the tunnelling amplitude Δ, the two energy levels would cross at $t = 0$, while for $\Delta > 0$ the two instantaneous eigenvalues are simply $E_{\pm}(t) = \pm\sqrt{(vt)^2 + \Delta^2}$ (see inset in figure 3(*b*)). Starting with the system in the ground state at $t = -\infty$, we can monitor the probability of jumping onto the excited state at any time t, which we plot in figure 3(*b*) for both the RT and the IT evolution and for several values of Δ (taking $v = 1$). The RT data provide an illustration of the well-known Landau–Zener result: after a (relatively short) tunnelling time, and possibly a few oscillations, the probability of getting onto the excited state saturates to a value given by $P_{\rm ex}(t = \infty) = {\rm e}^{-\pi\Delta^2/\hbar v}$. As for the IT data, the initial (tunnelling) part and the subsequent plateau of the curves are similar to the RT case: the plateau value attained, call it $P^*_{\rm ex}$, is indeed very close to the RT saturation value (in fact, asymptotically the same for $\Delta \to 0$); after that, the IT evolutions start to filter out the ground-state component—initially present in the state with a small amplitude $1 - P^*_{\rm ex}$—through the usual mechanism of suppression of excited states, leading to $P_{\rm ex}(t)$ which is nicely fit by the curve

$$P_{\rm ex}(t) = \frac{P^*_{\rm ex}\exp\left(-2\int_0^t {\rm d}t'[E_+(t') - E_-(t')]\right)}{(1 - P^*_{\rm ex}) + P^*_{\rm ex}\exp\left(-2\int_0^t {\rm d}t'[E_+(t') - E_-(t')]\right)}, \tag{24}$$

which asymptotically goes to zero as $t \to \infty$. This rather trivial effect of filtering, if on one hand explains the discrepancy between the IT and the RT evolution observed in the asymmetric double-well case of the previous section, is, on the other hand, of no harm at all: in contrast, it provides a quantitative improvement of IT over RT.

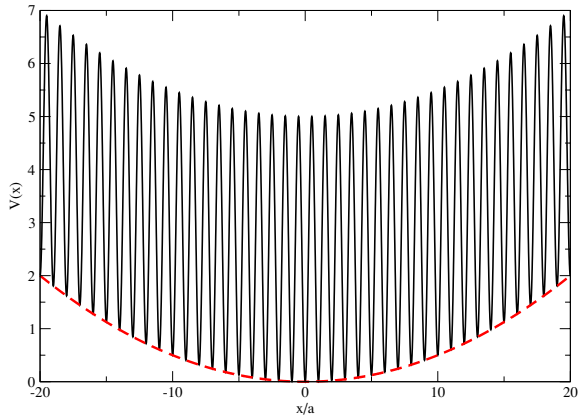

Figure 4. Parabolic washboard potential resulting in a logarithmically slow classical annealing. The minima are regularly located at positions $x_i = ia$, and the dashed line shows the parabolic envelope potential.

In summary, the essential equivalence of IT and RT Schrödinger annealing (with, moreover, a quantitative improvement of IT over RT) justifies practical implementations of quantum annealing based on imaginary-time quantum Monte Carlo schemes.

2.4. Other simple one-dimensional potentials with many minima

Moving on to multi-minima problems, we would like to mention one interesting one-dimensional potential (see [28] for details) which shows a remarkably different behaviour of CA and QA. The problem was proposed and solved, for CA, by Shinomoto and Kabashima in [45, 47], and consists in a parabolically shaped washboard potential, sketched in figure 4. This example will display a logarithmically slow classical annealing, showing that CA may run into trouble even in simple models with no complexity whatsoever, whereas quantum mechanics can do much better in this case. The problem consists in a wiggly one-dimensional potential with barriers of individual height $\approx B$ separating different local minima, regularly located a distance a apart one from each other, i.e., at positions $x_i = ai$, $i = 0, \pm 1, \pm 2, \ldots$. The ith local minimum is at energy $\epsilon_i = ka^2 i^2/2$, so that the resulting envelope is parabolic. By writing the appropriate master equation governing the probability $P_i(t)$ that the particle is found in the ith valley at time t, and taking the continuum limit $a \to 0$, Shinomoto and Kabashima [45, 47] showed that the equation governing the evolution of $P(x, t)$ turns out to be a Fokker–Planck (FP) equation, equation (8), with an effective diffusion constant of the form

$$D_{\text{eff}}(T) = \gamma a^2 \, \mathrm{e}^{-B/k_B T}, \tag{25}$$

$\eta(T) = k_B T / D_{\text{eff}}(T)$, and an effective drift potential $V(x) = kx^2/2$ given by the macroscopic parabolic envelope potential. This exponentially activated $D_{\text{eff}}(T)$ makes the annealing behaviour of the $P(x, t)$ exceedingly slow. In fact, the surprising result of this exercise [45, 47] is that the optimal annealing schedule $T(t)$ is *logarithmic* and the residual energy converges to 0 at best as $\epsilon_{\text{res}}(t) \sim \log(t)^{-1}$. The physical reason behind such a slow CA annealing is that the relaxation time $t_{\text{relax}} = k_B T/(2\gamma k a^2) \, \mathrm{e}^{B/k_B T}$ for the system to thermalize at any temperature T diverges exponentially at low T. As a result, the system will never be able to follow the decreasing T till the end of the annealing, by maintaining roughly the equilibrium

value $\epsilon_{\rm pot} = k_B T/2$. Indeed, if we assume for instance $T(t) = T_0(1 - t/\tau)$, the relaxation of the systems will cease to be effective—i.e., the system will fall *out of equilibrium*—at a time t^*, and temperature $T^* = T(t^*)$, at which $t_{\rm relax} \approx \tau$, i.e., when $k_B T^* \approx B/\log \gamma\tau$. The residual energy at this point cannot be smaller than the equipartition value $k_B T^*/2$, hence $\epsilon_{\rm res} \approx B/\log \gamma\tau$ as well. This freezing and falling out of equilibrium for classical systems with barriers seem to provide an ubiquitous source of logarithms in classical annealing [46].

The quantum mechanical approach to the same problem has been illustrated in [28]. In essence, starting from a tight-binding description in which the on-site energies ϵ_i are supplemented by a time-dependent nearest-neighbour hopping term which contains the inverse mass $\Gamma = \hbar^2/2m$ in the typical semi-classical (WKB) form $\sim e^{-\sqrt{V_h/\Gamma}}$ (V_h being an energy related to the details of the barrier), one can take, once again, the continuum limit $a \to 0$. The dynamics for the $\psi(x, t)$ reduces, in strict analogy with the classical case, to an effective Schrödinger equation for a particle moving in the parabolic envelop potential $V(x) = kx^2/2$, with an effective Laplacian coefficient $\Gamma_{\rm eff}(t) \propto e^{-\sqrt{V_h/\Gamma(t)}}$, which plays here the role that the effective diffusion constant in equation (25) played in the FP case. Contrary to the classical case, however, where an exponentially activated behaviour of the diffusion constant $D_{\rm eff}$ was *strongly detrimental* to the annealing (turning a power law into a logarithm), here the exponential WKB-like behaviour of $\Gamma_{\rm eff}$ *does no harm at all*: surprisingly, it improves the annealing. Indeed, as shown in [28], the power-law exponent $\Omega_{\rm QA}$ determining the decrease of the residual energy for a particle in a harmonic well, $\epsilon_{\rm res}(\tau) \propto \tau^{-\Omega_{\rm QA}}$, increases as one switches off the Laplacian coefficient more and more rapidly, tending to the value $\Omega_{\rm QA} \to 1$ for an infinitely fast switching off.

We believe that one of the important points that makes QA so different from CA in this case is that the spectrum of the instantaneous eigenvalues of the quantum problem does not show any dangerous Landau–Zener avoided crossing, and, correspondingly, the ground-state wavefunction is always more peaked in the central valley (the minimum at $x_i = 0$) than elsewhere. As in the two-level case, a disorder in the width of the different valleys would drastically change this result, and likely ruin the power-law behaviour of QA.

3. Role of disorder and Landau–Zener tunnelling

Despite their disarming simplicity, the cases illustrated above turn out to be extremely informative in qualifying the profound difference of QA from CA, and their surprising consequences. Of course the cases studied, although instructive, do not possess the real ingredient which makes annealing, especially QA, difficult, i.e., some form of disorder in the distribution of the minima.

As an explicit demonstration of the effect of disorder in a very simple framework, consider the quantum walk of a particle in a given lattice, say cubic in D dimensions, of N sites. If an on-site energy ϵ_i is associated with each site of the lattice, the Hamiltonian describing the quantum walk is just a disordered Anderson model

$$H(t) = \sum_i \epsilon_i \hat{n}_i - \Gamma(t) \sum_{\langle i,j \rangle} \left[c_j^\dagger c_i + \text{H.c.} \right], \tag{26}$$

where $H_{\rm cl} = \sum_i \epsilon_i \hat{n}_i$ represents the disordered on-site energy term, while the hopping term, proportional to $-\Gamma(t)$, provides the quantum kinetic term. The problem is to search for the lowest ϵ_i in the lattice—with the on-site energy ϵ_i randomly distributed in a given energy interval, say [0, 1]—by adiabatic evolution of a single-electron wavefunction from a very large $\Gamma_0 = \Gamma(t \leqslant 0)$ (when the wavefunction is extended over all sites) down to $\Gamma(\tau) = 0$. Recent work on this problem has been reported in [24], where small instances ($N = 20$ sites)

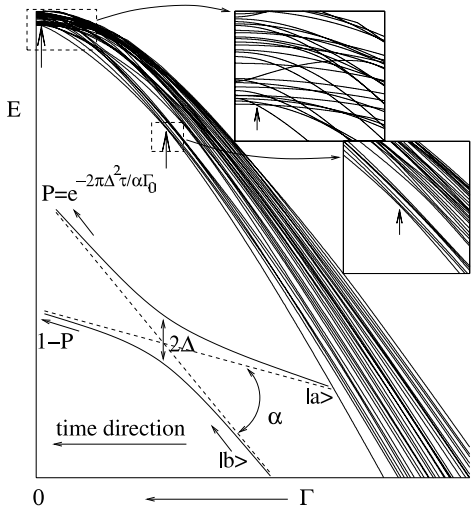

Figure 5. Instantaneous low-lying eigenvalues of a three-dimensional disordered Anderson model, on a lattice of size $10 \times 10 \times 10$, as a function of the hopping integral Γ. Note the Landau–Zener avoided crossings, particularly the one occurring at smaller Γ, which can have a tremendously small gap 2Δ. The inset shows a schematic of the Landau–Zener process.

of the problem are shown to be annealed rather fast by QA–AQC, with $E_{\rm res}(\tau) \sim 1/\tau^2$. This disordered Anderson model is clearly related to the problem of searching for an item in an unsorted database of N entries, Grover's problem [17, 48], one of the cases where QC is known to provide a speed-up, from the classical exhaustive search, $O(N)$, to $O(\sqrt{N})$. Indeed, the spatial search Grover's problem associated with a *structured database* can be formulated in terms of equation (26) with $\epsilon_{\bar{i}} = 0$ while $\epsilon_{j \neq \bar{i}} = 1$, where $\bar{i}$ is the desired entry of the lattice (database), and $\langle ij \rangle$ denote neighbouring lattice sites (database entries) connected by the hopping in the given lattice structure; in this case, for instance, a quadratic speed-up has been demonstrated for $D \geqslant 4$ [49]. However, the disorder associated with the on-site energies ϵ_i, in a more general circumstance, makes the spectrum of the Hamiltonian in (26) much more complicated than in the spatial search Grover's case. Obviously, the properties of the instantaneous spectrum of the Hamiltonian (26) depend on the dimensionality D of the lattice, and on the nearest-neighbour kinetic energy terms included (the so-called *adjacency matrix* of the quantum walker, in other words, on the topography of the landscape of the problem, specifying who is neighbour of whom). On quite general grounds, Anderson's localization [50, 51] would predict that instantaneous wavefunctions of $H(t)$ are *localized*, for a genuinely disordered potential and for large enough mass (i.e., small enough $\Gamma(t)$ and hence kinetic energy bandwidth) in any $D > 2$ (this localization occurs for all values of the mass in $D = 1, 2$). Therefore, quantum annealing should always, via a cascade of Landau–Zener events [35], end up into some localized state which has, *a priori*, nothing to do with the actual potential minimum, i.e., the smallest ϵ_i. Figure 5 shows the instantaneous eigenvalues of a disordered three-dimensional Anderson model, on a lattice of size $10 \times 10 \times 10$, as a function of the nearest-neighbour hopping integral Γ. In the process of reducing Γ, the ground state encounters several 'identity crisis' associated with tunnelling from one region of the lattice to another. Some of these tunnelling amplitudes can be tremendously small, so that one has to wait for an astronomically large Landau–Zener characteristic time $\tau_{\rm LZ}$ (see section (2.2))

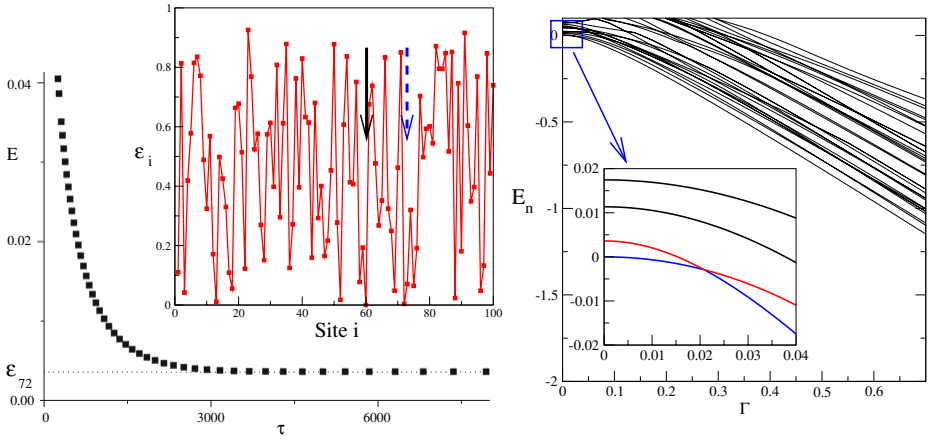

Figure 6. (Right) Instantaneous low-lying eigenvalues of a one-dimensional disordered Anderson model, on a lattice of size $N = 100$, as a function of the hopping integral Γ. Note the Landau–Zener avoided crossings, particularly the one occurring at smaller Γ, which has a very small gap 2Δ. (Left) The residual energy obtained by a Schrödinger imaginary-time evolution versus the annealing time τ. The inset shows the particular realization of the disordered on-site energies ϵ_i, with the two arrows signalling the lowest minimum, at $i = 60$ and the competing one at $i = 72$.

in order for the probability of 'adiabatically following the ground state' to be non-negligible. An even more concrete illustration of this problem is obtained for the $D = 1$ case shown in figure 6. Here, for the realization of the disorder shown in the left inset, there are two competing minima: the absolute minimum at $i = 60$, and a higher one, very close in energy, at $i = 72$. For reasons that have to do with the quantum kinetic energy—the valley around $i = 72$ being broader—the energy of the state localized around $i = 72$ is the lowest one, down to very small values of $\Gamma > \Gamma_{\rm LZ} \approx 0.02$. At such a value of Γ, a Landau–Zener crossing with an exceedingly small gap occurs, see right inset of figure 6. The system is, however, unable to follow the instantaneous ground state, for all practical purposes, and the final annealed energy of an IT–QA Schrödinger evolution remains trapped at $\epsilon_{i=72}$: no $E_{\rm res} \sim 1/\tau^2$ behaviour [24] is seen until astronomically large τ values are reached.

Moving to slightly more complicated models, let us consider the crucial role that disorder plays in the one-dimensional ($D = 1$) disordered Ising ferromagnet:

$$H = -\sum_i J_i \sigma_i^z \sigma_j^z - \Gamma \sum_i \sigma_i^x, \tag{27}$$

where $J_i \geqslant 0$ are non-negative random variables in the interval [0, 1], and Γ is the transverse field inducing quantum fluctuations. Obviously, the ground state of the classical model at $\Gamma = 0$ is the ferromagnetic state with all spins aligned up (or down). So, once again we are considering a *trivial minimization problem*, without any frustration at all (frustration requires competition between different couplings, and is therefore absent in $D = 1$ nearest-neighbour models). However, even at $\Gamma = 0$ there are important low-energy defects—*domain walls* between up and down ferromagnetic regions—which are typically pinned at the bonds where the couplings J_is are the smallest. (For a finite system with periodic boundary conditions, domain walls appear in pairs, and separate sections of the system with alternating $\uparrow$ and $\downarrow$ ferromagnetic ground states.) Given two domain walls, pinned at weak $J_i < \epsilon \ll 1$ bonds, typically at a distance $L \sim \epsilon^{-1} \gg 1$ apart, healing the system via single spin flip moves requires flipping L spins, which can be a formidable barrier to tunnel through (or to go over,

thermally). The situation is even worse at the critical point [19], $\ln \Gamma_c = \overline{\ln J} = \int_0^1 \mathrm{d}J \ln J$, where it is known that the typical gap Δ between the ground state and the first relevant excited state is exponentially small, in the lattice size N [52],

$$[\Delta]_{\mathrm{typ}} \sim \mathrm{e}^{-\mathrm{const}\sqrt{N}}. \tag{28}$$

The critical point Γ_c—separating the gapped large Γ quantum paramagnetic phase with all spins aligned along x, from the gapless low Γ classically disordered phase—is therefore a crucial source of small Landau–Zener gaps. We expect that, due to such exponentially small gaps, the typical characteristic Landau–Zener time $\tau_{\mathrm{LZ}} \propto \Delta_{\mathrm{min}}^{-2}$ beyond which the system will be able to adiabatically follow the ground state will be exponentially large in N, resulting in a residual energy per spin $\epsilon_{\mathrm{res}}(\tau) = E_{\mathrm{res}}(\tau)/N$ which, on average, will decrease only logarithmically as a function of the annealing time τ,

$$[\epsilon_{\mathrm{res}}(\tau)]_{\mathrm{av}} \propto (\log \tau)^{-\zeta}, \tag{29}$$

for $N \to \infty$. Work on this problem is currently in progress [53]. In conclusion, the system will have a very slow annealing while showing, at the same time, no complexity whatsoever: simple disorder is enough to yield slow annealing logarithmically.

4. Path-integral Monte Carlo quantum annealing

In order to move from toy problems with a manageable Hilbert space to real optimization problems, stochastic approaches are mandatory. As discussed in the introduction and in section 2.3, imaginary-time stochastic approaches are perfectly suitable to the goal: there is no gain in doing, on a classical computer, a Schrödinger evolution in real time [28].

A very simple quantum Monte Carlo approach, suitable to the proposed goal, is the PIMC approach. PIMC work in the QA context has been pioneered by Berne and collaborators, who have combined clever thermal and quantum annealing schemes within a path-integral framework, and applied these strategies to finding the minimal energy configurations of protein models on the continuum (with up to $N = 46$ monomers) [31, 32], and the minimal energy configurations of Lennard-Jones clusters of up to $N = 100$ atoms [33]. More recent PIMC-based QA work on the Lennard-Jones problem is reported in [34].

We start by briefly sketching the idea behind this approach with two introductory examples: the Ising case, as representative of discrete optimization problems, and the particle in a potential, as representative of continuum problems.

4.1. Path-integral Monte Carlo: introduction

The first, crucial, observation is that PIMC is intended to simulate the *equilibrium* behaviour of a system at *finite temperature* T. Both these features are potential limitations of the method. To clarify this point, consider, for instance, the Edward–Anderson Ising glass in a transverse field: $H = H_{\mathrm{cl}} + H_{\mathrm{kin}} = -\sum_{\langle ij \rangle} J_{ij}\sigma_i^z\sigma_j^z - \Gamma(t)\sum_i \sigma_i^x$. Strictly speaking, in the quantum annealing context, this is a *time-dependent* Hamiltonian, of which we would like to follow the low-lying states (ideally, the ground state) as a function of time, by turning off the transverse field $\Gamma(t)$. PIMC allows you to simulate the thermodynamics (at fixed strictly positive temperature T) for a fixed value of $\Gamma(t)$, by an approximate sampling of the quantum partition function

$$Z(T, \Gamma) = \mathrm{Tr}\, \mathrm{e}^{-\beta(H_{\mathrm{cl}}+H_{\mathrm{kin}})} = \sum_{s^1} \langle s^1 |\, \mathrm{e}^{-\beta(H_{\mathrm{cl}}+H_{\mathrm{kin}})} | s^1 \rangle, \tag{30}$$

where s^1 denotes a generic configuration of all the N spins. The idea behind the path integral is to reduce equation (30) to a classical partition function which is then sampled in the usual

way using, for instance, a Metropolis Monte Carlo. In order to do that, one needs to split the exponential of the Hamiltonian, appearing in equation (30), into products of exponentials. This is allowed by the Trotter theorem, stating that

$$\mathrm{e}^{-\beta(H_{\mathrm{cl}}+H_{\mathrm{kin}})} = \lim_{P\to\infty} \left(\mathrm{e}^{-\frac{\beta}{P}H_{\mathrm{cl}}}\,\mathrm{e}^{-\frac{\beta}{P}H_{\mathrm{kin}}}\right)^P. \tag{31}$$

Using this relationship, and inserting identities between the various exponentials, we get

$$Z(T,\Gamma) = \lim_{P\to\infty} \sum_{s^1\ldots s^P} \exp\left(-\frac{\beta}{P}\sum_{k=1}^{P} H_{\mathrm{cl}}(s^k)\right) \langle s^1|\,\mathrm{e}^{-\frac{\beta}{P}H_{\mathrm{kin}}}|s^2\rangle \cdots \langle s^P|\,\mathrm{e}^{-\frac{\beta}{P}H_{\mathrm{kin}}}|s^1\rangle. \tag{32}$$

The various configurations s^k ($k = 1\cdots P$) are often referred to as *Trotter replicas* of the original configuration s^1. The next thing one needs to do is to calculate explicitly the relevant exponential of the kinetic term, $\langle s^k|\,\mathrm{e}^{-H_{\mathrm{kin}}/PT}|s^{k+1}\rangle$, between two generic configurations. This is sometimes very easy to do—like in the Ising transverse field case, or in the Laplacian case (see below)—but can also be very difficult, for other choices of H_{kin} (like in the travelling salesman case, see section (4.2.2)). In the Ising case, the problem factorizes into N independent sites, each of which involves a simple Pauli matrix expectation value, yielding

$$\langle s^k|\,\mathrm{e}^{-\frac{\beta}{P}H_{\mathrm{kin}}}|s^{k+1}\rangle = C^N \exp\left(\frac{\beta}{P}J^{\perp}\sum_i s_i^k s_i^{k+1}\right), \tag{33}$$

where the transverse coupling $J^{\perp}$ is given by

$$J^{\perp} = -\frac{P}{2\beta}\ln(\tanh\beta\Gamma/P) > 0 \tag{34}$$

(the constant C is not relevant for our discussion). This kinetic term has a very transparent interpretation: $J^{\perp}$ gives a ferromagnetic Ising-like coupling between nearest-neighbour (k and $k+1$) Trotter replicas of the same spin. In order to implement the problems numerically, a finite number of Trotter replicas P is mandatory. This leads to an approximation, the error of which is proportional to the square of the Trotter break-up time, $O((\beta/P)^2)$ [54]. (Better Trotter break-ups, for given finite values of P, can lead to smaller errors, see section (4.3), but we will concentrate here on the basic form proposed above.) For the full partition function we thus finally get, in the Ising case

$$Z(T,\Gamma) \approx C^{\mathrm{NP}} \sum_{s^1}\cdots\sum_{s^P} \mathrm{e}^{-\frac{\beta}{P}S_{D+1}} \tag{35}$$

$$S_{D+1} = -\sum_{k=1}^{P}\left(\sum_{\langle ij\rangle} J_{ij}s_i^k s_j^k + J^{\perp}\sum_i s_i^k s_i^{k+1}\right), \tag{36}$$

which represents the partition function of a classical $(D+1)$-dimensional anisotropic Ising system at temperature $P/\beta = PT$. The system has couplings J_{ij} along the original D-dimensional lattice bonds (same for all Trotter replicas), and $J^{\perp}$ (same for all sites i) along the extra Trotter dimension where the system has a finite length P.

Similar expressions hold, for instance, for the problem of a particle in a potential $V(\mathbf{x})$, where $H_{\mathrm{cl}} = V(\mathbf{x})$, $H_{\mathrm{kin}} = -\Gamma\nabla^2$, and sums over configurations $\sum_{s^k}$ transform into integrals over the variables $\mathbf{x}^k$. Similarly, the kinetic term contribution

$$\langle \mathbf{x}^k|\,\mathrm{e}^{-H_{\mathrm{kin}}/PT}|\mathbf{x}^{k+1}\rangle = \left(\frac{K^{\perp}}{2\pi PT}\right)^{D/2} \mathrm{e}^{-\frac{1}{PT}\frac{K^{\perp}}{2}(\mathbf{x}^k-\mathbf{x}^{k+1})^2}, \tag{37}$$

where D is the dimension of the $\mathbf{x}$-space and $K^{\perp} = (PT)^2/(2\Gamma)$ admits a perfectly transparent

interpretation: the transverse coupling $K^{\perp}$ between different Trotter replicas has the form of a *spring* coupling neighbouring configurations $\mathbf{x}^k$ and $\mathbf{x}^{k+1}$.

In all cases, QA, in the present context, consists in externally controlling, during the PIMC dynamics, the value of the transverse field Γ—leaving T untouched—in much the same way as one externally controls T in classical simulated annealing [6]. This approach, which we will refer to as PIMC-QA, does not lead, therefore, to the simulation of a true quantum mechanical annealing dynamics, of the type implied by equation (2), but only to a MC annealing dynamics [42]. A clever mixing of thermal (reduction of T) and quantum annealing (reduction of Γ)—possibly combined with decimation techniques on the number of Trotter replicas P as Γ is reduced—has been successfully applied in [31–33] to various continuum optimization problems (protein folding models and Lennard-Jones clusters). In the following, however, we will stick to the simple PIMC-QA scheme outlined above, whereby T is untouched during a QA simulation and the number of Trotter replicas P is also kept fixed, in order to more clearly discriminate possibly genuine quantum effects from thermal ones. We now move on to describe some of the results obtained so far with this technique on discrete optimization problems.

4.2. PIMC-QA applied to combinatorial optimization problems

In a quite general way, one could define a combinatorial optimization problem as the algorithmic task of minimizing any given *cost function* which depends on the configuration of variables assuming discrete values [4]. Further classifications are of course possible, but they somehow hide the fundamental fact that it is straightforward to map such problems over the search for the ground state of some Hamiltonian depending on Potts (or Ising) spin degrees of freedom [7, 55]. This is the case, for instance, of the travelling salesman problem [56–59], Boolean satisfiability [60, 61], vertex covering [62], graph colouring [63] and many others.

Random problem instances are of particular interest, because they can be investigated resorting to powerful techniques developed in the context of disordered statistical mechanics systems [1]. The physical approach to combinatorial optimization has often allowed the derivation of *phase diagrams*, telling us in which range of some control parameters hard instances are expected to be found [7, 55]. These analyses provide insight about the typical-case complexity of problem solving, in contrast to the more rigorous but less informative worst-case complexity theory, which constitutes one of the corner stones of theoretical computer science [2]. The basic distinction between the **P** and **NP** complexity classes (that is, between problems for which a polynomial algorithm able to solve worst-case instances is or is not known) can sometimes be misleading. Easy instances of **NP**-complete problems (the hardest of all the **NP** problems, [3]) can easily be found (see e.g. [64]), while sometimes the optimization of instances of problems in **P** can take an exponential time using local search techniques (see e.g. [65]).

In the following, we shall briefly illustrate the results obtained on three classes of problems which have been recently addressed using PIMC-QA.

4.2.1. Random ising spin models. Determining the ground state of an Ising spin glass model can be an extraordinarily difficult task. To get the picture, it is enough to think that the number of possible configurations of a very small 32×32 square lattice Ising model is of order 10^{308}, while the number of electrons in the universe is 'just' of the order 10^{80}![3] It can be rigorously

[3] Other examples of this kind can be found in the appendices of MacKay (2003) *Information Theory, Inference and Learning Algorithms* (Cambridge, MA: Cambridge University Press).

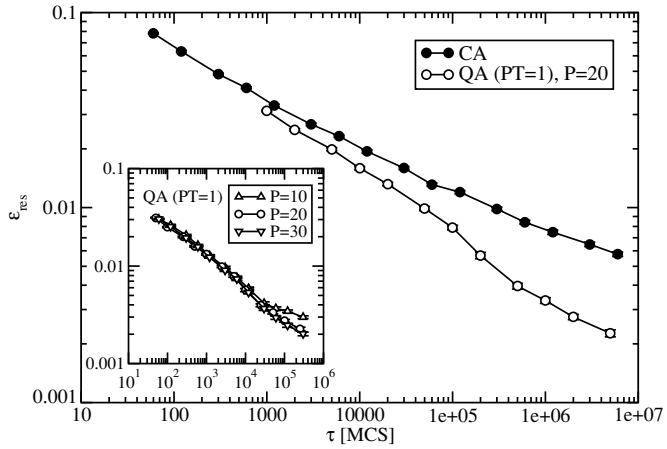

Figure 7. Residual energy per site for an 80×80 disordered 2D Ising model after CA and path-integral Monte Carlo (PIMC) QA. We show the PIMC-QA data for the optimal value of $PT = 1$, with $T = 0.05$ and $P = 20$ Trotter replicas. The actual inverse annealing rate τ used in QA has been rescaled (multiplied by P) for fair comparison with CA. Still, QA is faster than CA.

shown that, in the 3D lattice case, the ground-state determination belongs to the **NP**-complete complexity class [5], but here we shall report results on the simpler 2D lattice case, where E_{GS} can be calculated up to sufficiently large lattice sizes [66]. The Hamiltonian of an Ising spin glass has been already discussed in section 4.1 and is given by equation (1).

For a given 2D lattice size $L \times L$ (L up to 80), and for various quenched realizations of the random couplings J_{ij}, drawn from a flat distribution in the interval $(-2, 2)$, we carried out several repeated classical and quantum annealings (for more details, see [35, 36]). At the end of both QA and CA, the system remains generally trapped at energy $E_{final} = E_{GS} + \epsilon_{res}$ and the efficiency of each protocol is monitored by considering the average residual energy $\epsilon_{res}(\tau)$ as a function of τ.

The annealing parameters T (CA) or Γ (QA) were decreased linearly from the initial value of $T_0 = 3$ or $\Gamma_0 = 2.5$ down to zero, with a total of τ MC steps per spin. In QA we used fixed values of $T_q = PT = 1, 1.5, 2$ at several P values and prepared the initial state (same for all replicas) by a classical pre-annealing stage. The computational cost scales linearly with P, but increasing P beyond a certain characteristic length (see inset in figure 7) does not produce any further improvement. The choice of $P = 20$ was found to be optimal. The moves proposed in both CA and QA are single-spin flip moves, but QA also attempts slightly more 'global moves' by proposing a spin flip for the spins $s_i^k, k = 1, \ldots, P$ of all the Trotter replicas of a given site i. Figure 7 shows that QA is definitely superior to CA in the case of the Ising spin glass, in agreement with the experimental observation of significantly faster frequency-dependent relaxations during QA of the disordered magnet [15]. Theoretical arguments, however, can be given in favour of the fact that the residual energy will decrease, in both cases, logarithmically slow [35]

$$\epsilon_{res}(\tau) \propto (\log \tau)^{-\zeta},$$

albeit with a possibly different annealing exponent ζ, which might favour QA over SA, $\zeta_{QA} > \zeta_{CA}$.

Very recently, PIMC-QA and CA have been extensively applied [37] to the problem of finding the ground state of a *random field* Ising model (once again, a **P**-problem)

$$H = -J \sum_{\langle ij \rangle} \sigma_i^z \sigma_j^z - \sum_i h_i \sigma_i^z,$$

for 1D, 2D and 3D cubic lattices, via the usual strategy, for QA, of adding and annealing a transverse field term. The results of [37] seem to confirm that the log behaviour is present in both cases, but that QA is indeed quantitatively better, with a larger ζ.

Finally, let us mention that there is an Ising model case where PIMC-QA might be possibly worse than CA, according to [40]. This is the 1D Ising ANNNI model

$$H = -J_1 \sum_i \sigma_i^z \sigma_{i+1}^z + J_2 \sum_i \sigma_i^z \sigma_{i+2}^z,$$

with the point $J_2 = J_1/2$ representing a highly frustrated case. While CA slows down only close to the frustrated case $J_2 = J_1/2$, PIMC-QA seems to have a uniformly poor performance even for small values of J_2, deep in the ferromagnetic phase of the model [40].

4.2.2. Travelling salesman problem. Given N cities and their tabulated distances d_{ij}, TSP consists in finding the shortest path connecting them, visiting each city only once and returning to the starting point. An account of the vast literature about algorithms for TSP can be found e.g. in [67], while three classical papers analysing physics approaches to the problem are [57–59].

As a first step to a QA optimization we have to choose a representation for the classical potential energy $H_{\rm pot}$ of a given configuration (in our case, the length of a tour), and, most crucially, a suitable source of quantum fluctuations $H_{\rm kin}$. TSP can be mapped to a highly constrained Ising-like model—in a way similar to [56, 68]—in which each configuration of the system (a valid tour) is associated with an $N \times N$ 0/1-matrix $\hat{T}$. For every ordered sequence of cities, $\hat{T}_{i,j} = 1$ if the tour visits city i immediately after city j, and $\hat{T}_{i,j} = 0$ otherwise. For the *symmetric* TSP problem we wish to consider (a TSP with symmetric distance matrix $d_{ij} = d_{ji}$) the direct tour represented by a $\hat{T}$, and the reversed tour, represented by the transposed matrix $\hat{T}^t$, have exactly the same length. It is then convenient to introduce the symmetric matrix $\hat{U} = \hat{T} + \hat{T}^t$ as representative of *undirected* tours. The length of a tour can now be written as

$$H_{\rm pot}(\hat{U}) = \frac{1}{2} \sum_{(ij)} d_{ij} \hat{U}_{i,j} = \sum_{\langle ij \rangle} d_{ij} \hat{U}_{i,j}, \tag{38}$$

where $\langle ij \rangle$ signifies counting each link only once. $H_{\rm kin}$ should be chosen in order to induce fluctuations generating the important elementary 'moves' of the problem, changing a tour into another one. Deciding which configurations are to become direct neighbours of a given configuration is indeed a crucial step, because it determines the problem's effective landscape [69]. A very important move that is often used in heuristic TSP algorithms is the so-called *two-opt move*, which consists in eliminating two links in the current tour, $(c_1 \to c_2)$ and $(c_{1'} \to c_{2'})$, and rebuilding a new tour in which the connections are exchanged, $(c_1 \to c_{1'})$ and $(c_2 \to c_{2'})$ (see figure 8). Associating a spin variable +1 (−1) with each entry 1 (0), the whole two-opt move, when working with $\hat{U}$ matrices, can be represented by just four spin-flip operators

$$S^+_{\langle c_{1'}, c_1 \rangle} S^+_{\langle c_{2'}, c_2 \rangle} S^-_{\langle c_2, c_1 \rangle} S^-_{\langle c_{2'}, c_{1'} \rangle},$$

where, by definition, each $S^{\pm}_{\langle i,j \rangle}$ flips an Ising spin variable (defined as $S^z_{\langle i,j \rangle} = (2\hat{U}_{i,j} - 1) = \pm 1$) at position (i, j) and at the symmetric position (j, i), i.e., $S^{\pm}_{\langle i,j \rangle} = S^{\pm}_{i,j} S^{\pm}_{j,i}$. However, this

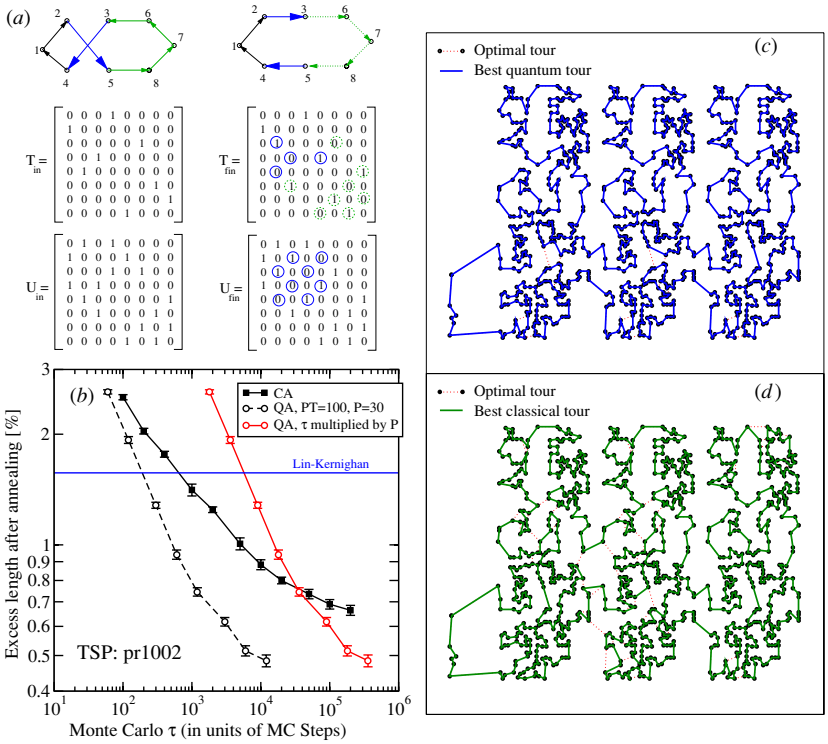

Figure 8. (*a*) (Left) Representation of an eight-city tour, with the corresponding matrices $\hat{T}_{\rm in}$ and $\hat{U}_{\rm in} = \hat{T}_{\rm in} + \hat{T}^t_{\rm in}$. (Right) The final tour obtained when a two-opt move is performed, with a whole section reversed (dotted line). The matrices $\hat{T}_{\rm fin}$ and $\hat{U}_{\rm fin}$ are shown, the circles indicating the entries that have been switched ($0 \leftrightarrow 1$) by the two-opt move. The dotted circles in $\hat{T}_{\rm fin}$ are entries related to the trivial reversal of a section of the tour. (*b*) Average residual excess length found after CA and PIMC-QA for a total time τ (in MC steps), for the $N = 1002$ instance `pr1002` of the TSPLIB. PIMC-QA is once again faster than CA. (*c*), (*d*) the `pr1002` instance considered, with its optimal tour and best tours obtained in the QA (*c*) and CA (*d*) cases.

kinetic Hamiltonian does not allow for an obvious Trotter discretization of the path integral (see discussion in section 4), and the PIMC scheme cannot deal with it (for this purpose, Green's function MC methods, that do not use a Trotter break-up, would be in principle more effective, see section 5). We then introduce a drastic simplification to our kinetic energy term, replacing it altogether with a standard transverse Ising form, arriving finally at the Hamiltonian

$$\tilde{H}_{\rm TSP} = \sum_{\langle ij \rangle} d_{ij} \frac{\left(S^z_{\langle i,j \rangle} + 1\right)}{2} - \Gamma(t) \sum_{\langle ij \rangle} \left[S^+_{\langle j,i \rangle} + \text{H.c.}\right]. \tag{39}$$

This simplified form of kinetic energy no longer fulfils the constraint to take a valid tour to another valid tour, but this problem is avoided by proposing exclusively two-opt moves in the MC algorithm [38].

We tested our QA algorithm against CA [38] on a standard benchmark TSP problem, namely the printed circuit board instance `pr1002` of the TSPLIB [70]. It is a structured TSP problem with $N = 1002$ cities whose optimal tour length $L_{\rm opt}$ is known exactly. For CA, we chose an optimal initial temperature T_0 by first performing several CA with various short cooling times τ and starting from sufficiently high temperatures. The point where the cooling

curves for different τs start to differ identifies an approximate 'dynamical temperature' $T_{\rm dyn}$. For pr1002, we obtained $T_{\rm dyn} \sim 100$. As expected [67], the optimal T_0 for CA approximately coincides with $T_{\rm dyn}$. Not surprisingly, for QA the same choice $PT \sim T_{\rm dyn}$ yields the optimal results, together with the choice $\Gamma_0 = 300$. Figure 8 shows the results obtained [38] for the average percentage best-tour excess length $\epsilon_{\rm exc}(\tau) = (\bar{L}_{\rm best}(\tau) - L_{\rm opt})/L_{\rm opt}$, both with CA (filled squares) and with QA (open circles). As a reference, the best out of 1000 runs of the Lin–Kernighan algorithm (one of the standard local-search algorithms for TSP [67]) is also plotted (dashed line in figure 8). The results show that, once again, QA anneals more efficiently, even accounting for the extra factor P in the total CPU time (rightmost open circles), reducing the error at a much steeper rate than CA.

4.2.3. Random Boolean satisfiability. In order to state the problem, consider a set of N Boolean variables $z_1, \ldots, z_N$, where $z_i = 1$ or 0 ('true' or 'false'). Denoting by ζ_i the variable z_i or its negation $\bar{z}_i$, one then considers the disjunction (logical OR) of three variables $C = (\zeta_i \vee \zeta_j \vee \zeta_k)$, which is called a three-*clause*. The random 3-SAT problem consists in deciding if the conjunction (logical AND) of M different clauses $C_1 \wedge C_2 \cdots \wedge C_M$—each clause being formed by three variables extracted at random among the N available, and appearing negated or direct with uniform probability—can be simultaneously satisfied by a truth value assignment $\{z_i\}$. (In the exact cover variant studied in [27], a clause C is satisfied if one of the three variables is 1, the other two being 0.) If we associate an Ising spin variable $S_i = (-1)^{z_i}$ with each Boolean variable z_i, we can assign to any clause C_a involving three variables z_i, z_j, z_k an energy E_a given by

$$E_a = \frac{(1 + J_{a,i}S_i)(1 + J_{a,j}S_j)(1 + J_{a,k}S_k)}{8}, \tag{40}$$

where the coupling $J_{a,i}$ assumes the value -1 if the variable z_i appears negated in clause a, $+1$ otherwise. Evidently, $E_a = 0$ if the corresponding clause is satisfied, $E_a = 1$ otherwise.

As in the case of TSP, archives of hard structured instances exist [71]. In addition, statistical mechanics techniques can be used to determine the *phase diagram* of the random 3-SAT problem [7, 60, 61]. The main parameter determining the hardness of a formula is the ratio $\alpha = M/N$ between the number, M, of clauses and the number, N, of variables. For $\alpha < \alpha_c \simeq 4.26$ it is typically possible to find satisfying assignments, but instances particularly hard to solve are expected to be found if $\alpha > \alpha_G \simeq 4.15$ [72]. It is expected that, due to the proliferation of an exponential number of metastable states acting as dynamical traps, local search gets trapped at an energy close to some finite threshold level, lower bounded by the so-called *Gardner energy* [73]. The trapping effect induced by the threshold states cannot be neglected when the instancesize is large ($N \geqslant 10\,000$) and large statistical fluctuations become sufficiently rare [72]. Smaller random formulae are, on the other hand, often easily solvable by classical simulated annealing and cannot be used as significant benchmarks.

We performed a first set of annealings over a single hard 3-SAT random instance with $N = 10^4$ and $\alpha = 4.24$ [39]. The kinetic term was given by a simple transverse field inducing single spin-flip fluctuations, like in the Ising case, since no clever sets of moves are known for 3-SAT, unlike the TSP case [74]. Using an efficient *ad hoc* algorithm (described in [72]), we verified that the chosen formula was actually satisfiable, as expected from theory for $\alpha < \alpha_c$. As in the case of the TSP optimization, we set both T_0 for CA and PT for QA equal to $T_{\rm dyn} = 0.3$. The optimal field-ramp range was found to be between $\Gamma_0 = 0.7$ and $\Gamma_f \simeq 10^{-3}$.

A comparison between the performance of the optimal CA and the optimal QA at $P = 50$, both with and without global (i.e., all $s_i^k, k = 1, \ldots, P$, are flipped) moves [39], is shown in figure 9. For each point, an average has been taken over 50 different realizations of the

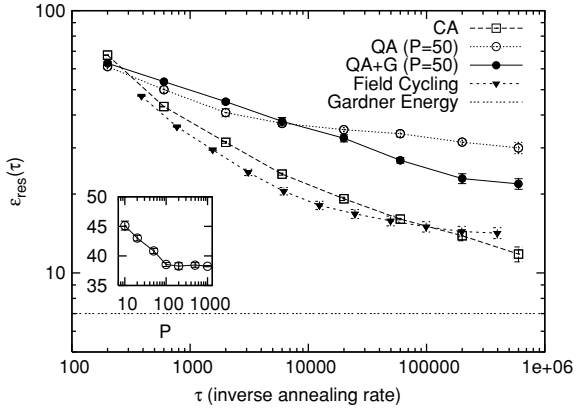

Figure 9. Comparison between optimal linear-schedule classical (CA) and path-integral Monte Carlo quantum annealing (PIMC-QA) for a 3-SAT problem with $N = 10^4$ and $\alpha = M/N = 4.24$. CA always performs better than PIMC-QA simulated with $P = 50$ Trotter replicas. The average performance of linear PIMC-QA is worse than that of CA, even if an improvement in the results can be obtained by introducing global moves (G) and by increasing P (in the inset the final average energy found by PIMC-QA after 2000 iterations for increasing P is plotted and compared with the average result of a CA of the same length, dashed line). The solid triangles are the data obtained by the field cycling PIMC-QA hybrid strategy described in [39].

same experiment; in the case of QA, a second average was performed among the energies of the P replicas, which are in general different. It can be seen that the linear-schedule CA always performs better than the linear-schedule QA. No further improvement can be obtained for $P \geqslant 100$, see inset of figure 9—a much larger value than in the case of the Ising spin glass and the TSP instance—but we chose $P = 50$ in order to extend the simulation time as much as possible. The asymptotic slope of the linear-schedule QA curves seems indeed to be definitely less steep than that of CA, independently of the number of replicas involved in the simulation and of the use of global moves.

The sobering message converged by this failure is that superiority of QA over CA is not universal, and is only achieved when we can use some understanding of the problem, especially when building the kinetic energy operator.

4.3. PIMC-QA of a double-well: lessons from a simple case

We would like to finish our discussion about path-integral Monte Carlo -based QA by mentioning recent results on a very simple case from which one can learn much about the limitations of the method [42]. Suppose we want to perform a QA optimization of the simple double-well potential which was investigated in section 2.2 using PIMC. One is then lead to simulate the behaviour of a closed polymer made up of P Trotter replicas $\{x^k\}(k = 1, \ldots, P)$ of the original particle, held at temperature β/P and moving in the potential V_{asym} with a nearest-neighbour spring coupling, as shown in equation (37). One can actually be more sophisticated than that, and perform a higher order Trotter break-up, correct to $O(\beta/P)^4$ instead of $O(\beta/P)^2$, using, for instance, the Takahashi–Imada approximation [75]. Moreover, instead of performing single-bead moves, i.e., moves involving a single x^k at a time, one can reconstruct, during the move, entire sections of the polymer, using the *bisection* method [29]. We have applied this rather sophisticated PIMC to our textbook double-well problem, working

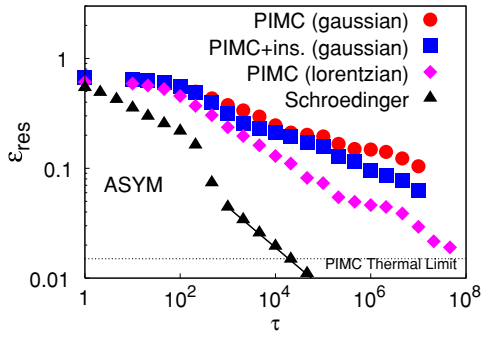

Figure 10. Comparison between the Schrödinger annealing data, solid triangles and different types of PIMC-QA, on the double-well potential of section 2.2.

with a temperature $T = 0.03\,V_0$, a number of Trotter slices up to $P = 160$, and a bisection of level up to 4, i.e., involving up to 2^4+1 replicas x^k at each move. The initial temperature value of $\Gamma = \hbar^2/2m$ was taken to be $\Gamma_0 = 0.5$, and its value was reduced linearly to 0 in a certain total number τ of Monte Carlo moves. The results, shown in figure 10 by solid circles, are altogether disappointing: the data barely start to cross below the level $\epsilon_{\rm res} \approx 2\delta = 0.2$, which corresponds to the metastable minimum of the potential, for the largest values of τ simulated. This means that the system had no occasion, up to these large values of τ, of realizing that there was another minimum available through tunnelling. Moreover, the overall slope of the data is definitely less steep than that of direct Schrödinger annealing, shown for comparison by solid triangles in figure 10. (The absolute values of τ are not comparable between the two sets of data, because they refer to different quantities: a real Schrödinger dynamics versus a Monte Carlo dynamics.) The situation improves substantially (see solid squares in figure 10) if we introduce, as candidate Monte Carlo moves, also the so-called *instanton moves*, i.e., basically classical trajectories that move from one minimum to the other (plus fluctuations) [76]. This is, however, not a fair game: we have substantially exploited a crucial information on the landscape which is generally *not available* for a complicated optimization problem!

One other lesson we can learn, in the present context, is the role of the kinetic energy operator $H_{\rm kin}$ on which the quantum fluctuations are based. Up to now, we were using as $H_{\rm kin}$ the usual non-relativistic kinetic energy $H_{\rm kin} = p^2/2m$, and annealed the system by increasing the mass m of the particle. (The propagator of this kinetic term is just the Gaussian, as shown in equation (37).) Imagine now we pretend that the particle has a relativistic 'photon-like' dispersion

$$H_{\rm kin} = \Gamma|p| \quad \rightarrow \quad \langle \mathbf{x}^k | \, \mathrm{e}^{-\frac{\beta}{P} H_{\rm kin}} | \mathbf{x}^{k+1} \rangle = \frac{1}{\pi} \frac{\hbar\Gamma\beta/P}{(x^k - x^{k+1})^2 + (\hbar\Gamma\beta/P)^2}, \qquad (41)$$

and that we anneal the system by reducing to 0 the velocity Γ of the dispersion. The bisection method can be generalized for this kind of kinetic energy, and the results obtained are shown by solid diamonds in figure 10. As one notes, the residual energy versus τ is now considerably lower than in the non-relativistic (Gaussian) case, and even levels-off, for large τ, to the thermal limit $k_B T/2 = 0.015$ set by our finite temperature T. This example shows the important role played by the choice of the kinetic energy $H_{\rm kin}$ (as well as the limitations imposed by the unavoidable finite temperature T).

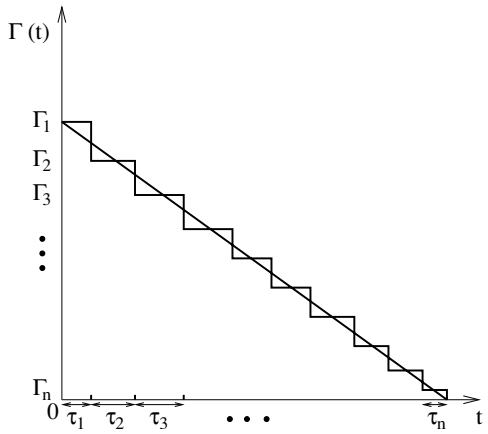

Figure 11. Sketch of the annealing of the transverse field Γ, linearly in a time $\tau = \tau_1 + \tau_2 + \cdots \tau_n$, or in a stepwise fashion.

Summarizing, PIMC-QA suffers evidently of a number of problems: (i) it is only a fake Monte Carlo annealing dynamics, in principle not fully representative of the true imaginary-time Schrödinger dynamics; (ii) the sampling of relevant 'action' might be highly inefficient (recall the instanton problem above) and the cure for that might not be obvious at all; (iii) the finite temperature T imposes a lower thermal limit for the residual energy $\epsilon_{\rm res}(\tau)$ below which we can never possibly go; (iv) the calculation of the propagator of $e^{-H_{\rm kin}/PT}$ might be very difficult for a kinetic term which we would like to implement (see section 2.2). These various problems clearly call for an alternative to PIMC in order to implement a stochastic QA approach. In principle, a useful alternative is given by a Green's function Monte Carlo (GFMC) method, which we now turn to discuss.

5. Green's function Monte Carlo quantum annealing

Green's function Monte Carlo is a very effective quantum Monte Carlo approach for correlated quantum Hamiltonians on a lattice, the method of choice, for instance, for studying zero temperature properties of Heisenberg spin models [77]. It is naturally a method that works at zero temperature, by projecting out the ground-state wavefunction from an arbitrary initial state. We now illustrate the general ideas behind a GFMC-based quantum annealing approach.

5.1. Green's function Monte Carlo quantum annealing: ideas

As outlined in the introduction, the ideal scope of a QA approach is to take some initial-state $|\Psi(0)\rangle$ and let it evolve according to the Schrödinger dynamics, see equation (2), associated with a time-dependent Hamiltonian $H(t)$ interpolating between an extreme quantum regime and the classical problem one is interested in. For the random Ising problem, for instance, a very natural choice is given by $H(t) = -\sum_{\langle i,j\rangle} J_{i,j}\sigma_i^z\sigma_j^z - \Gamma(t)\sum_i \sigma_i^x$, where $\Gamma(t)$ is initially very large, and slowly decreased towards zero in a given time τ, as sketched in figure 11. We will use the Ising case, throughout this section, to exemplify the ideas of the method, which can be adapted essentially to any other model in a straightforward way.

We already argued, see section 2.3, that an imaginary-time Schrödinger evolution is, for optimization purposes on a classical computer, equally good, and most likely even superior, to the standard real-time evolution. With this in mind, we can aim at solving the imaginary-time evolution in equation (6), instead of the real-time one in (2). If we also imagine the gradual decrease of Γ to be made stepwise, as sketched in figure 11, then the solution for $|\Psi(t)\rangle$ is obtained by repeated applications of an imaginary-time evolution operator

$$|\Psi(\tau)\rangle = \mathrm{e}^{-H(\Gamma_n)\tau_n} \cdots \mathrm{e}^{-H(\Gamma_1)\tau_1} |\Psi(0)\rangle, \tag{42}$$

where $\tau = \tau_1 + \cdots \tau_n$ is the total annealing time, $\Gamma_1 > \Gamma_2 > \cdots > \Gamma_n \sim 0$ is a decreasing sequence of transverse fields and $H(\Gamma_i)$ is a shorthand for $H(t)$ with a value Γ_i of the transverse field. Each application of the imaginary-time evolution operator $\mathrm{e}^{-H(\Gamma_i)\tau_i}$ effectively tends to *filter out* the corresponding ground state of $H(\Gamma_i)$ from the state to which it is applied. In turn, $\mathrm{e}^{-H(\Gamma_i)\tau_i}$ can be obtained by a repeated application of many infinitesimal propagators of the type $[1 - \Delta t H(\Gamma_i)]$, i.e.,

$$|\Psi(t + \tau_i)\rangle = \mathrm{e}^{-H(\Gamma_i)\tau_i} |\Psi(t)\rangle = \lim_{m\to\infty} [1 - \Delta t H(\Gamma_i)]^m |\Psi(t)\rangle, \tag{43}$$

where $\Delta t = \tau_i/m$. The Green's function Monte Carlo (GFMC) is just a stochastic technique which implements equation (43). More precisely, if we define, recursively,

$$\psi_{n+1}(x') = \sum_x G^{(\Gamma)}_{x',x} \psi_n(x), \tag{44}$$

where $\psi_n(x) = \langle x|\psi_n\rangle$, $|x\rangle$ is a shorthand for a generic spinconfiguration describing the Hilbert space of the problem, and the Green's function $G^{(\Gamma)}_{x',x}$ is given by

$$\begin{aligned} G^{(\Gamma)}_{x',x} &= \langle x'|G^{(\Gamma)}|x\rangle = \langle x'|1 - \Delta t[H(\Gamma) - E_T]|x\rangle \\ &= (1 + \Delta t E_T)\delta_{x',x} - \Delta t \langle x'|H(\Gamma)|x\rangle, \end{aligned} \tag{45}$$

one can show that—for large n—the iterated state ψ_n converges (apart from a normalization constant) to the ground state $\psi^{(\Gamma)}_{\mathrm{GS}}(x)$ of $H(\Gamma)$, if Δt is chosen to be suitably small [77]. (E_T is an estimate of the ground-state energy which allows us to reduce the statistical fluctuations, see also [78].)

The problem in equation (44) looks superficially similar to an ordinary Markovian master equation [44], with a few very crucial differences: (i) $\psi_n(x)$ are not *probabilities*, but *amplitudes*; (ii) the Green's function $G^{(\Gamma)}_{x',x}$ in equation (45), unlike the transition probability of a master equation, is not necessarily made of non-negative elements, and is, in general, not columnnormalized, $\sum_{x'} G^{(\Gamma)}_{x',x} \neq 1$, unlike a Markov transition probability [44]. In summary, the process underlying the iterated-power method is *not* a properly defined Markov chain, and, therefore, it cannot be immediately simulated, as it stands, with a Monte Carlo approach.

Problem (ii) above can be quite serious: if some of the matrix elements of $G^{(\Gamma)}_{x',x}$ are *negative*, no possible interpretation of it as a 'transition probability' is possible. This is at the heart of the so-called *sign problem* [78] in quantum Monte Carlo. In the following, we will assume that a choice of basis is possible in which no sign problem exists, i.e., all matrix elements of $G^{(\Gamma)}$ are non-negative, $G^{(\Gamma)}_{x',x} \geqslant 0$. This is certainly true for the Ising glass in a transverse field. More generally, since the choice of the kinetic energy to be used in QA is at our disposal, it is wise to choose the signs of H_{kin} in equation (1) such that no sign problem occurs. Still, we miss the correct column-normalization:

$$\sum_{x'} G^{(\Gamma)}_{x',x} = 1 + \Delta t E_T - \Delta t \sum_{x'} H_{x',x} \stackrel{\mathrm{def}}{=} b_x \neq 1. \tag{46}$$

A way out of this difficulty is to factorize $G^{(\Gamma)}$ in terms of a stochastic matrix $p_{x',x}$—by definition [44], a matrix with all positive elements $p_{x',x} \geqslant 0$, and with the normalization

condition $\sum_{x'} p_{x',x} = 1$ for all columns—times the scale factor b_x defined above. Indeed, with the previous definition (46) of b_x, the matrix

$$p_{x',x} = G^{(\Gamma)}_{x',x} / b_x \tag{47}$$

is trivially positive and column normalized and, therefore, it is a suitable transition matrix for a Markov chain in x-space.

The crucial idea is then to *extend* the configuration space where the Markov process is defined, adding to x a *non-negative* weight factor (hereafter, the weight) w. This extended configuration space will be labelled by (x, w). The pair (x, w) is often called a *walker*, because it will be the basic entity in the Markov chain 'random walk'. The weight part will take care of b_x, while x will be taken care of by $p_{x',x}$. More precisely, if (x_n, w_n) indicates a walker at iteration time n, in this extended configuration space, we set up the following Markov process:

$$\begin{aligned} &\text{(a) Generate } x_{n+1} = x' \text{ with probability } \ p_{x',x_n}. \\ &\text{(b) Update the weight with } w_{n+1} = w_n b_{x'}. \end{aligned} \tag{48}$$

In words, the walker performs a *random walk* in the Hilbert space (x) of the system and in the weight space (w); such a random walk is composed of a standard Markov chain in x-space, associated with $p_{x',x}$, and a multiplicative process for the weight $w_n \to w_{n+1} = w_n b_{x'}$. By moving in this way, the walkers visit every point in the (x, w)-space with a probability $P_n(x_n, w_n)$ whose first w-moment can be shown to be proportional to the wavefunction ψ_n, $\psi_n(x) \propto \int \mathrm{d}w_n w_n P_n(x, w_n)$ [79].

The algorithm in (48) is the basic version of a GFMC [77]. In this form, however, the algorithm simply *does not work* in practice. The reason for this failure is not difficult to grasp. While $x_n \to x_{n+1}$ is a plain Markov process, the weight update $w_n \to w_{n+1} = w_n b_{x_n}$ is a *multiplicative* process with random factors b_{x_n}, which is prone to very large fluctuations [77, 79]: w_n might grow large, or become negligibly small, in just a few iterations of the algorithm, and the whole method would go wild, because error bars in the calculations of the averages grow in an uncontrolled way. The cure to this disease goes through the introduction of *many walkers*, simulated in parallel, and through performing occasional 'reconfigurations' of their weights, via the so-called *branching* [78]. In practice, one propagates simultaneously a set of M walkers defined by weights w_i and configurations x_i, for $i = 1, \ldots, M$. Before the variance of the weights w_i becomes too large, one appropriately redefines the set of walkers—by reproducing some of them and deleting some others—in such a way as to drop those with excessively small weight, and to generate copies of the more important ones[4].

The analogy of such a many-walker GFMC with a genetic-like algorithm [80] is worth noting. Each walker (x, w) plays the role of an individual that propagates (mutates) increasing or decreasing its *fitness*—represented by the accumulated weight w, related to the wavefunction amplitude $\psi(x)$. A mutation is here simply a step of the algorithm, which attempts a single spin flip of a certain site in the configuration, this is what the off-diagonal matrix elements $\langle x'|H|x\rangle$ do for an Ising model in the transverse field. At certain times, branching occurs, which modifies the population of individuals by favouring the survival of those with highest fitness (largest w). The only genetic feature that is missing in the quantum mechanical case is the possibility of cross-breeding (mixing of genetic codes of two configurations, to give rise to new configurations); this would correspond to *non-local moves* which change the configurations in a global way.

[4] The branching is just a particular Markov process applied to the configurations (x_j, w_j), which leads to new walkers (x'_j, w'_j). Each new walker (x'_j, w'_j), with $j = 1, \ldots, M$, will have the same weight $w'_j = \bar{w} = \sum_j w_j / M$ and an arbitrary configuration x'_j picked up among the M possible old ones x_k, $k = 1, \ldots, M$, with a probability p_k proportional to the weight of that configuration, $p_k = w_k / \sum_j w_j$, see [79].

The final, important, ingredient that makes the algorithm work is the so-called *importance sampling* [30]. It can be seen, in the genetic analogy proposed before, as a way of proposing mutations (single spin flips) that instead of being equally probable, with a common matrix element Γ, are biased by a function which guides the system towards the most representative configurations. More precisely, suppose we have a reasonable guess of the ground-state $\psi_{\rm GS}(x)$ in the form of some nodeless wavefunction $\psi_T(x)$, known as *trial (or guiding) wavefunction*. It is then enough to substitute the Green's function $G^{(\Gamma)}$ with the so-called *importance-sampling* Green's function,

$$\bar{G}_{x',x} = \psi_T(x') G^{(\Gamma)}_{x',x} \psi_T^{-1}(x), \tag{49}$$

which just rescales by an extra factor $\psi_T(x')/\psi_T(x)$ a transition from x to x', thus favouring those transitions $x \to x'$ where $\psi_T(x')/\psi_T(x)$ is largest. In general, $\bar{G}_{x',x}$ is not symmetric, but one can still apply to it the same decomposition in equation (47), defining the corresponding Markov chain (48) with

$$\begin{aligned} &p_{x',x} = \bar{G}_{x',x}/\bar{b}_x, \\ &\bar{b}_x = \sum_{x'} \bar{G}_{x',x} = 1 + \Delta t\, E_T - \Delta t \frac{\sum_{x'} \psi_T(x') H_{x',x}}{\psi_T(x)} = 1 - \Delta t\,(E_L(x) - E_T), \end{aligned} \tag{50}$$

where the local energy $E_L(x)$ is defined as

$$E_L(x) = \frac{\langle \psi_T | H | x \rangle}{\langle \psi_T | x \rangle}. \tag{51}$$

Quite amusingly, if the guessed trial wavefunction ψ_T coincides with the ground-state wavefunction $\psi_T(x) = \psi_{\rm GS}(x)$, then $E_L(x) = E_{\rm GS}$ and $\bar{b}_x = 1 - \Delta t\,(E_{\rm GS} - E_T)$ are constant, and one can show that statistical fluctuations in the calculation vanish exactly. (This is the so-called *zero variance property* [78].) Therefore, by variationally improving the quality of the guiding wavefunction $\psi_T(x)$ one can substantially improve the quality of the calculation (e.g., reducing the error bars).

We will soon show that the use of a good importance-sampling guiding wavefunction is indeed a crucial ingredient of a GFMC–QA application; without importance sampling, GFMC–QA is simply worthless. Therefore, we start illustrating our application of GFMC on the random Ising model with a discussion of possible trial wavefunctions we have tested, and the difficulties encountered.

5.2. GFMC–QA on random ising model

Finding a good trial wavefunction for a random Ising model in a transverse field is a highly non-trivial task. The first idea that comes to mind is a kind of 'mean-field' wavefunction, made up of a product of single-site factors as

$$\left|\psi_T^{(\rm MF)}\right\rangle = \prod_{i=1}^{N} \left(\frac{{\rm e}^{+\frac{h_i}{2}} |\uparrow\rangle_i + {\rm e}^{-\frac{h_i}{2}} |\downarrow\rangle_i}{\sqrt{2\cosh(h_i)}} \right), \tag{52}$$

where $\{h_i\}$, the local fields on each site i, are variational parameters to be optimized for each given value of the transverse field Γ. The optimization of $\{h_i\}$ amounts to finding the minimum of the variational energy,

$$\begin{aligned} E_T^{(\rm MF)}(\{m_i\};\Gamma) &= \left\langle \psi_T^{(\rm MF)} \middle| H \middle| \psi_T^{(\rm MF)} \right\rangle \\ &= -\sum_{\langle i,j \rangle} J_{i,j} m_i m_j - \Gamma \sum_i \sqrt{1 - m_i^2}, \end{aligned} \tag{53}$$

where $m_i = \tanh(h_i)$ are the local magnetizations. The stationary conditions required by the minimization read

$$0 = \frac{\partial E_T^{(\mathrm{MF})}}{\partial h_i} = -\left(1 - m_i^2\right)\left(\sum_{j \in \mathcal{N}(i)} J_{i,j} m_j\right) + \Gamma m_i \sqrt{1 - m_i^2} \qquad \forall\, i, \tag{54}$$

where $\mathcal{N}(i)$ indicates the set of nearestneighbours of site i. As it turns out, finding solutions of equation (54) with optimal variational energies is simple only for large enough Γ, where the quantum paramagnetic solution $h_i = m_i = 0$ is found. Such a solution, representing the $\Gamma = +\infty$ ground state with all spins aligned along the $+\hat{x}$ direction, survives down to some value Γ_{cr} of the transverse field, below which non-trivial solutions of equation (54)—with non-vanishing local magnetizations $m_i \neq 0$—start to appear. However, the number of those solutions (local minima)—found for instance by a straightforward conjugate gradients algorithm [8]—is large [81]. In the low-Γ region, our minimization problem is just the quantum counterpart of the well-known Weiss mean-field approach for the classical random Ising model [82], which is known to run into difficulties in the classical glassy phase. In a sense, minimizing equation (53) in the low-Γ glassy phase is not much simpler than finding the classical ground state of the problem for $\Gamma = 0$. We transformed a minimization task in a discrete space of variables, $S_i = \pm 1$, into one where the variables are continuous, $m_i \in (-1, 1)$, but the task itself is of comparable difficulty. We refer the reader to [81] for more details on this aspect.

A second, quite natural, choice of trial wavefunction that comes to mind is a Boltzmann-like wavefunction of the form

$$\psi_T^{(\beta)}(\{S_i\}) = \mathcal{N}(\beta)\, \mathrm{e}^{-(\beta/2) H_{\mathrm{cl}}(\{S_i\})}, \tag{55}$$

where $1/\beta$ plays the role of an effective temperature, with β a variational parameter to be optimized, and $H_{\mathrm{cl}}(\{S_i\})$ (see equation (1)) is the classical energy of a given configuration $\{S_i\}$. $\mathcal{N}(\beta)$ is an appropriate normalization factor, which we will not need to calculate. Once again, for large Γ we expect to find $\beta = 0$ (the exact $\Gamma = +\infty$ solution), while, by decreasing Γ, larger and larger values of β will favour regions where the 'potential energy' $H_{\mathrm{cl}}(\{S_i\})$ has a local minimum, until we get, for $\Gamma = 0$, to the asymptotic limit $\beta \to \infty$ (ideally), required by a wavefunction which is perfectly localized in the global minimum (see below for a discussion of this point).

To calculate the expectation value of energy with the Boltzmann-like trial wavefunction in equation (55), as a function of the single parameter β, we used a standard variational Monte Carlo (VMC) algorithm [78]. Figure 12 shows (left panel) the optimal value β_{opt} of β which minimizes the variational energy $E_{\mathrm{tot}}^{(\mathrm{Boltz})} = \left\langle \psi_T^{(\beta)} \middle| H \middle| \psi_T^{(\beta)} \right\rangle$, for several values of the transverse field Γ. Note that β_{opt} saturates for small Γ to about $\beta_{\mathrm{opt}} \approx 2$, somewhat surprisingly at first sight, since, for $\Gamma \to 0$, one would expect $\beta_{\mathrm{opt}} \to +\infty$, in such a way that the classical ground state dominates (i.e., equation (55) becomes a delta-like function localized in the exact classical ground state). This is clearly an effect of a severe ergodicity loss of the VMC algorithm, which is not difficult to understand. For a given β, indeed, the VMC samples a probability distribution $\left|\psi_T^{(\beta)}(x)\right|^2 = \mathrm{e}^{-\beta H_{\mathrm{cl}}(x)}$ with single spin-flip moves; its efficiency in exploring the phase space, therefore, is exactly identical to that of a classical Metropolis Monte Carlo at the temperature $T = 1/\beta$. Finding an optimal β for a given Γ is therefore totally equivalent to asking what is the *effective temperature* of a classical Ising spin glass which provides the best approximation to the wavefunction of a quantum Ising glass at zero temperature and non-zero Γ. Now, from classical spin-glass physics [1, 83] we know that a threshold energy E_{th} exists below which the system has a finite complexity, i.e., it displays an

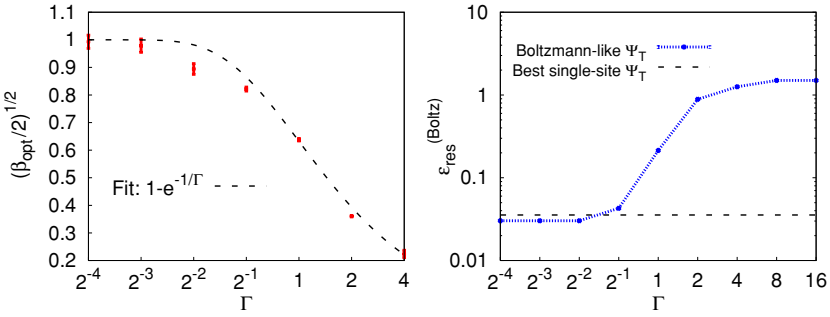

Figure 12. (Left) Plot of the optimal β, $\beta_{\rm opt}$, for the 'Boltzmann' trial wavefunction $|\psi_T^{(\beta)}\rangle$ defined in equation (55), for several values of Γ. The dashed line is a fit to the data. (Right) The variation residual diagonal energy $\epsilon_{\rm res}^{\rm (Boltz)} = \langle\psi_T^{(\beta_{\rm opt})}|H_{\rm cl}|\psi_T^{(\beta_{\rm opt})}\rangle/N - \epsilon_{\rm GS}$ corresponding to the $\beta_{\rm opt}$ is shown in the left panel, for several Γ. The dashed horizontal line represents the best residual energy ever achieved, for $\Gamma > 0.01$, by employing the mean-field trial wavefunction in [52].

exponentially ($\sim \exp N$) large number of metastable minima. Close to this threshold energy, the relaxation of any local algorithm towards equilibrium becomes exceedingly slow—the algorithm gets stuck for a long time in every minimum visited—and the average quantities measured are not representative of their true thermodynamical values. Evidently, for $\Gamma \to 0$, the VMC algorithm does not visit the regions near the true minima of the classical energy, but wanders in a high-energy band of metastable states, separated by moderate energy barriers. In such a case, a small and finite value of β allows us to (still) overcome such barriers, so as to find slightly more favourable local minima, while perfect localization ($\beta \to +\infty$) in a wrong excited state would lead to an average bigger residual energy.

The right panel in figure 12 shows the optimal variational residual energy $\epsilon_{\rm res}^{\rm (Boltz)} = \left\langle\psi_T^{(\beta_{\rm opt})}\middle|H_{\rm cl}\middle|\psi_T^{(\beta_{\rm opt})}\right\rangle/N - \epsilon_{\rm GS}$ corresponding to the optimal β shown in the left panel, for several values of the transverse field Γ. For large Γ values, the variational residual energy per spin is of order 1, as a random spin configuration would do. By decreasing Γ, we note that the variational residual energy saturates, for small Γ, to finite non-zero values, of order 0.03 per spin, in agreement with the previously noted saturation in the optimal $\beta_{\rm opt}$, due to ergodicity breaking. (A closer inspection shows that the variational residual energy is actually *non-monotonic* for $\Gamma < 0.25$, again an artefact of sampling difficulties.) Note, however, that this saturation value is definitely below the best results provided by the previously discussed $\psi_T^{\rm (MF)}$, which is of order 0.035 per spin (shown for comparison by a dashed horizontal line). Therefore, with all its pitfalls, the Boltzmann-like trial wavefunction in equation (55) provides, at low Γ, a marginally better approximation of the true GS, than that obtained by the mean-field *Ansatz* in equation (52). Moreover, $\psi_T^{(\beta)}$ is also much better behaved, and simpler to optimize, as far as the minimization problem is concerned. For these reasons, we decided to work out our GFMC results using the Boltzmann-like wavefunction only.

We finally present the results of a GFMC-based QA approach, where the transverse field Γ is decreased stepwise during the simulation, while, at the same time, the importance sampling Boltzmann-like wavefunction is changed according to the corresponding value of the variational parameter $\beta_{\rm opt}(\Gamma)$. As a benchmark, we will compare GFMC-QA outcomes with the path-integral Monte Carlo quantum annealing (PIMC-QA) and classical simulated annealing (CA) results described in [35, 36]. We reduce the coupling Γ in equation (1) at each Monte Carlo step (MCS) in a linear way. We start from an initial large enough value

of the transverse field, $\Gamma_0 = 2.5$, and then we set $\Gamma_n = \Gamma_0(1 - n/\tau)$ during the nth MCS ($0 \leqslant n < \tau$). τ is the total annealing time (see figure 11) measured as the total number of MCS performed by the algorithm. We used $M = 20$ walkers, and performed branching at each MCS, because the low Γ region is affected by severe weight instabilities which would otherwise make the algorithm unstable (for the initial, large Γ, part of the annealing one could consider branching less often, as weights are well under control). However, even with this very conservative choice, the weights go sometimes completely wild if a Γ of order 10^{-7} is reached (i.e., for long annealing times τ). As a consequence, we decided to cut off the Γ_n annealing schedule in such a way that the final Γ is $\approx 10^{-4}$ and not 0. This finally guarantees a good weight stability. (Whenever it was possible to perform annealings with smaller cut-offs on Γ, and approximately the same slope $1/\tau$, we checked that the results obtained are not very sensitive to the cut-off chosen.)

For each value of Γ_n the trial wavefunction used is the Boltzmann-like one, defined in equation (55), with a variational parameter $\beta_{\rm opt}(\Gamma_n)$, which corresponds to the instantaneous optimal value. (Practically, we used for $\beta(\Gamma)$ the fitting function shown in figure 12, left panel.) Figure 13 shows the best residual energy per spin ever reached during the annealing simulation, for several values of τ, averaged over ten independent repetitions of the whole annealing process (due to computer-time limitations, a single run is shown for the largest, $\tau > 10^8$, annealings). For comparison, the CA and PIMC–QA data obtained in [35] are also shown. Note first that the τ axes of the three calculations are completely unrelated. The GFMC τ is measured in units in which a MCS is just a *single spin flip*, while MCS for CA and PIMC–QA are intended as *sweeps of the entire lattice* of N spins (including all the $P = 20$ Trotter slices, for the PIMC case). For this reason, we also present the CA and PIMC–QA data in a shifted time axis where τ is multiplied by a factor $N = 80^2$ (rightmost curves). Although the GFMC–QA data are strictly below both CA and PIMC–QA data, on the same perspin timeunit (i.e., compared to the shifted CA and PIMC–QA data), it is clear that the GFMCslope is still *worse* than that of PIMC–QA, and indeed surprisingly similar to CA. Moreover, the CPUtime needed for a single spin flip in GFMC is much larger than the corresponding single-spin move in CA or PIMC–QA (each GFMC move costs of order N operations).

Let us pause to consider the similarity between the CA and the GFMC–QA slopes that figure 13 suggests. This similarity must be somehow related to the fact that we have used, as importance wavefunction for the GFMC, a Boltzmann-like wavefunction, $\psi_T(x) \propto \mathrm{e}^{-(\beta/2)H_{\rm cl}(x)}$. At the level of a plain variational Monte Carlo (VMC), we already pointed out that such a choice of wavefunction amounts to sampling $\left|\psi_T^{(\beta)}(x)\right|^2 = \mathrm{e}^{-\beta H_{\rm cl}(x)}$, and is therefore totally equivalent to a classical Metropolis MC at temperature $T = 1/\beta$. If, during the GFMC simulation, we neglect the weights associated with the walkers (as well as the associated branching process), we will be carrying over a completely classical simulation where the generated configurations are distributed (see equation (50)) according to

$$p_{x',x} = \frac{\bar{G}_{x',x}}{\bar{b}_x} \propto \Gamma \frac{\psi_T^{(\beta)}(x')}{\psi_T^{(\beta)}(x)} = \Gamma \exp(-(\beta_{\rm opt}(\Gamma)/2)[H_{\rm cl}(x') - H_{\rm cl}(x)]).$$

As a consequence, the Markov process in x-space will obey a classical detailed balance condition

$$p_{x',x}\, \mathrm{e}^{-\beta_{\rm opt} H_{\rm cl}(x)} = p_{x,x'}\, \mathrm{e}^{-\beta_{\rm opt} H_{\rm cl}(x')}. \tag{56}$$

In other words, a GFMC–QA without weights would be just a computationally heavy way of doing a CA with a peculiar form of annealing of the effective temperature $\beta_{\rm opt}(\Gamma)$ (note, in passing, that such an effective temperature never gets too low, since $\beta_{\rm opt}$ saturates to around

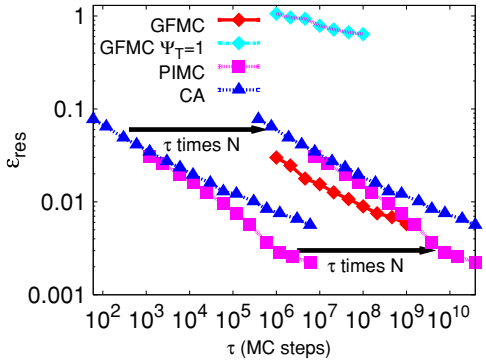

Figure 13. The average best residual energy obtained by Green's function Monte Carlo (GFMC) QA for the 80×80 instance of the random Ising model studied in section 4.2.1, versus the total annealing-time τ. Previous results obtained by classical simulated annealing (CA) and by path-integral Monte Carlo quantum annealing (PIMC-QA) with $P = 20$ Trotter slices [35, 36] are shown for comparison. The GFMC time-unit is a single spin flip, while CA and PIMC–QA Monte Carlo time units are sweeps of the entire lattice (see [36]). Importance sampling is performed by using the optimal trial wavefunction $\psi_T^{(\mathrm{Boltz})}$. The transverse field is linearly reduced down to 10^{-4} in a total annealing time τ, starting from $\Gamma_0 = 2.5$. We used here $M = 20$ walkers and performed branching at every Monte Carlo step (MCS) ($t_B = 1$).

$\beta_{\mathrm{opt}} \approx 2$ for low Γ). Quantum mechanics enters, therefore, only through the weights that the GFMC carries over (and the unavoidable branching process which makes the multiplicative process of weight updating numerically stable). Evidently, such a weight updating is, in the present disordered case, not sufficiently strong and effective as to make the resulting averages really different from the underlying classical physics governing the Markov chain in x-space, and the resulting GFMC–QA data are rather close to the CA ones (although they are much more computer-time demanding).

6. Summary and conclusions

We have illustrated several applications of quantum annealing strategies to a range of problems going from textbook toy models—displaying in a clear way the crucial differences between classical and quantum annealing—all the way to challenging optimization problems (random Ising models, TSP, 3-SAT). The techniques used to implement QA are either deterministic Schrödinger's evolutions, for the toy models, or path-integral Monte Carlo (PIMC), and Green's function Monte Carlo (GFMC) approaches, for the hard optimization problems.

Worth mentioning, here, are a few other approaches or applications which we have had no occasion to discuss so far. An alternative $T = 0$ projection QMC scheme (similar in spirit to the GFMC discussed above, but not using importance sampling) is presented in [84], and applied to an infinite range $\pm J$ spin glass model. Quantum annealing strategies have also been applied to a one-dimensional kinetically constrained model [84, 85], based on Ising spins σ_i^z, subject to a downward longitudinal field h, and experiencing a barrier for flipping, whenever the neighbouring spin σ_{i-1}^z is $\downarrow$; based on a semiclassical treatment of the problem, it has been shown [84, 85] that a quantum tunnelling of the barrier is more effective than the thermal barrier jump in annealing the system towards the ultimate ground state, with all spins $\downarrow$. Finally, let us also mention the application of PIMC–QA to problems of image reconstruction and information processing [86].

As a way of summary, we would like to mention here some of the major points touched upon.

Is quantum really better? Although, in the examples illustrated, QA often wins over CA, sometimes it does not (the negative result for the 3-SAT case is particularly instructive in this respect). The fact that QA wins over its classical counterpart is *a priori* not guaranteed, the outcome of the battle being strongly related to the landscape of the problem one deals with. The role of the disorder is particularly crucial in the quantum case. Both the two major algorithmic achievements of quantum computation, i.e., Grover's and Schor's algorithm, deal with problems without disorder, where the intrinsic parallelism of a quantum algorithm can be fully exploited. We suspect that even simple optimization problems, like for instance the random Ising model in one-dimension, can be quite hard, if tackled by an adiabatic quantum computation, precisely due to the all-important role played by disorder.

Role of kinetic energy. The choice of the kinetic energy is clearly all important in QA. Section 4.3, illustrating the improvements in annealing a double-well potential upon using a relativistic kinetic energy, is particularly instructive.

Concerning more technical points.

Limitations of PIMC. Path-integral Monte Carlo QA, although often successful over its classical counterpart, suffers from several limitations (finite temperature T, possible sampling problems for the action, difficulties with the Trotter break-up), which suggest investigating alternative quantum Monte Carlo approaches to QA.

Limitations of GFMC. Green's function Monte Carlo seems, in principle, a very good tool for QA. However, GFMC needs, as a crucial ingredient, a reasonably good guess of trial variational wavefunctions. Finding a good trial function $\psi_T(x)$ for the problem at hand is an essential part of a GFMC–QA application, and constitutes the delicate point of the whole algorithm.

In conclusion, it is quite plausible that quantum annealing (adiabatic quantum computation), although potentially useful and sometimes more convenient than classical annealing, is not capable, in general, of finding solutions to **NP**-complete problems in polynomial time. Nevertheless, understanding when and how quantum mechanics can quantitatively improve on the solution of hard optimization problems is still an open and timely issue. Moreover, there are fundamental physics issues behind the dynamics of quantum disordered systems which certainly deserve further study.

Acknowledgments

The work described has been sponsored, over the years, by MIUR through FIRB RBAU017S8R004, FIRB RBAU01LX5H, COFIN2003 and COFIN2004, and by INFM ('Iniziativa trasversale calcolo parallelo'). It is a real pleasure to thank Roman Martoňák, Roberto Car, Demian Battaglia, Lorenzo Stella and Osvaldo Zagordi for their collaboration on various stages with the work described in the present review. Finally, instructive discussions with Rosario Fazio, Riccardo Zecchina, Sandro Sorella and Silvio Franz are gratefully acknowledged.

References

[1] Parisi G, Mézard M and Virasoro M 1987 *Spin Glass Theory and Beyond* (Singapore: World Scientific)
[2] Papadimitriou C H and Steiglitz K 1998 *Combinatorial Optimization: Algorithms and Complexity* (New York: Dover)
[3] Garey M R and Johnson D S 1979 *Computers and Intractability* (San Francisco: Freeman)

[4] Papadimitriou C H 1993 *Computational Complexity* (Reading, MA: Addison-Wesley)
[5] Barahona F 1982 *J. Phys. A: Math. Gen.* **15** 3241
[6] Kirkpatrick S, Gelatt C D Jr and Vecchi M P 1983 *Science* **220** 671
[7] Dubois O, Monasson R, Selman B and Zecchina R (ed) 2001 Special Issue on NP-hardness and phase transitions *Theor. Comput. Sci.* **265** 1–307
[8] Press W H, Teukolsky S A, Vetterling W T and Flannery B P 1992 *Numerical Recipes in C: The Art of Scientific Computing* 2nd edn (Cambridge: Cambridge University Press)
[9] Amara P, Hsu D and Straub J E 1993 *J. Phys. Chem.* **97** 6715
[10] Finnila A B, Gomez M A, Sebenik C, Stenson C and Doll J D 1994 *Chem. Phys. Lett.* **219** 343
[11] Kadowaki T and Nishimori H 1998 *Phys. Rev.* E **58** 5355
[12] Sherrington D and Kirkpatrick S 1975 *Phys. Rev. Lett.* **35** 1792
[13] Ray P, Chakrabarti B K and Chakrabarti A 1989 *Phys. Rev.* B **39** 11828
[14] Thirumalai D, Li Q and Kirkpatrick T R 1989 *J. Phys. A: Math. Gen.* **22** 3339
[15] Brooke J, Bitko D, Rosenbaum T F and Aeppli G 1999 *Science* **284** 779
[16] Aeppli G and Rosenbaum T F 2005 *Quantum Annealing and Related Optimization Methods* ed A Das and B K Chakrabarti (Berlin: Springer) p 159
[17] Nielsen M and Chuang I L 2000 *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press)
[18] Shor P W 1997 *SIAM J. Comput.* **26** 1484
[19] Fisher D S 1995 *Phys. Rev.* B **51** 6411
[20] Farhi E, Goldstone J, Gutmann S and Sipser M 2000 Quantum computation by adiabatic evolution *Preprint* quant-ph/0001106
[21] Aharonov D, van Dam W, Kempe J, Landau Z, Lloyd S and Regev O 2004 Adiabatic quantum computation is equivalent to standard quantum computation *Preprint* quant-ph/0405098
[22] Gantmakher F R 1960 *The Theory of Matrices* (New York: Chelsea)
[23] Messiah A 1962 *Quantum Mechanics* vol 2 (Amsterdam: North-Holland)
[24] Suzuki S and Okada M 2005 *J. Phys. Soc. Japan* **74** 1649
[25] Landau L D 1932 *Phys. Z. Sowjetunion* **2** 46
[26] Zener C 1932 *Proc. R. Soc.* A **137** 696
[27] Farhi E, Goldstone J, Gutmann S, Lapan J, Lundgren A and Preda D 2001 *Science* **292** 472
[28] Stella L, Santoro G E and Tosatti E 2005 *Phys. Rev.* B **72** 014303
[29] Ceperley D M 1995 *Rev. Mod. Phys.* **67** 279
[30] Kalos M H and Whitlock P A 1986 *Monte Carlo Methods* vol 1 (New York: Wiley)
[31] Lee Y H and Berne B J 2000 *J. Phys. Chem.* A **104** 86
[32] Lee Y H and Berne B J 2001 *J. Phys. Chem.* A **105** 459
[33] Liu P and Berne B J 2003 *J. Chem. Phys.* **118** 2999
[34] Gregor T and Car R 2005 *Chem. Phys. Lett.* **412** 125
[35] Santoro G E, Martoňák R, Tosatti E and Car R 2002 *Science* **295** 2427
[36] Martoňák R, Santoro G E and Tosatti E 2002 *Phys. Rev.* B **66** 094203
[37] Sarjala M, Petäjä V and Alava M 2006 *J. Stat. Mech.* P01008
[38] Martoňák R, Santoro G E and Tosatti E 2004 *Phys. Rev.* E **70** 057701
[39] Battaglia D A, Santoro G E and Tosatti E 2005 *Phys. Rev.* E **71** 066707
[40] Sen P and Das P K 2005 *Quantum Annealing and Related Optimization Methods* ed A Das and B K Chakrabarti (Berlin: Springer) p 305
[41] Stadler P F and Schnabl W 1992 *Phys. Lett.* A **161** 337
[42] Stella L, Santoro G E and Tosatti E 2006 *Phys. Rev.* B **73** 144302 (*Preprint* cond-mat/0512064)
[43] Das A and Chakrabarti B K 2005 *Quantum Annealing and Related Optimization Methods (Lecture Notes in Physics)* (Berlin: Springer)
[44] van Kampen N G 1992 *Stochastic Processes in Physics and Chemistry ed* (Amsterdam: North-Holland) (Revised and enlarged)
[45] Shinomoto S and Kabashima Y 1991 *J. Phys. A: Math. Gen.* **24** L141
[46] Huse D A and Fisher D S 1986 *Phys. Rev. Lett.* **57** 2203
[47] Kabashima Y and Shinomoto S 1991 *J. Phys. Soc. Japan* **60** 3993
[48] Grover L K 1997 *Phys. Rev. Lett.* **79** 325
[49] Childs A M and Goldstone J 2003 Spatial search by quantum walk *Preprint* quant-ph/0306054
[50] Anderson P W 1958 *Phys. Rev.* **109** 1492
[51] Lee P A and Ramakrishnan T V 1985 *Rev. Mod. Phys.* **57** 287
[52] Fisher D S and Young A P 1998 *Phys. Rev.* B **58** 9131

[53] Caneva T, Fazio R and Santoro G E Quantum annealing of a random Ising chain in a transverse field (in preparation)
[54] Suzuki M 1976 *Prog. Theor. Phys.* **56** 1454
[55] Nishimori H 2001 *Statistical Physics of Spin Glasses and Information Processing* (Oxford: Oxford University Press)
[56] Hopfield J J and Tank D W 1986 *Science* **233** 625
[57] Mézard M and Parisi G 1986 *Europhys. Lett.* **2** 913
[58] Percus A G and Martin O C 1999 *J. Stat. Phys.* **94** 739
[59] Dean D S, Lancaster D and Majumdar S N 2005 *J. Stat. Mech.* L01001
[60] Mézard M, Parisi G and Zecchina R 2002 *Science* **297** 812
[61] Mézard M and Zecchina R 2002 *Phys. Rev.* E **66** 56126
[62] Weigt M and Hartmann A K 2000 *Phys. Rev. Lett.* **84** 6118
[63] Braunstein A, Mulet R, Pagnani A, Weigt M and Zecchina R 2003 *Phys. Rev.* E **68** 036702
[64] Mitchell D, Selman B and Levesque H 1992 *Proc. of the Tenth National Conf. on Artificial Intelligence* (Menlo Park, CA: AAAI Press) p 459
[65] Mézard M, Ricci-Tersenghi F and Zecchina R 2003 *J. Stat. Phys* **111** 505
[66] De Simone C, Diehl M, Jünger M, Mutzel P, Reinelt G and Rinaldi G 1995 *J. Stat. Phys.* **80** 487
[67] Johnson D S and McGeoch L A 2003 *Local Search in Combinatorial Optimization* (Princeton, NJ: Princeton University Press) p 215
[68] Kadowaki T 2002 Study of optimization problems by quantum annealing *Preprint* quant-ph/0205020
[69] Stadler P F 1995 *Complex Systems and Binary Networks* ed R Lopez-Pena, R Capovilla, R Garca-Pelayo, H Waelbroeck and F Zertuche (Berlin:Springer) p 77
[70] See http://www.iwr.uni-heidelberg.de/groups/comopt/software/TSPLIB95
[71] See http://www.intellektik.informatik.tu-darmstadt.de/SATLIB/
[72] Battaglia D A, Kolář M and Zecchina R 2004 *Phys. Rev.* E **70** 036107
[73] Montanari A, Parisi G and Ricci-Tersenghi F 2004 *J. Phys. A: Math. Gen.* **37** 2073
[74] Michalewicz Z 1996 *Genetic Algorithms + Data Structures = Evolution Programs* 3rd edn (Berlin: Springer)
[75] Takahashi M and Imada M 1984 *J. Phys. Soc. Japan* **53** 3765
[76] Negele J W and Orland H 1988 *Quantum Many-Particle Systems* (Reading, MA: Addison-Wesley)
[77] Trivedi N and Ceperley D M 1990 *Phys. Rev.* B **41** 4552
[78] Foulkes W M C, Mitas L, Needs R J and Rajagopal G 2001 *Rev. Mod. Phys.* **73** 33
[79] Calandra Buonaura M and Sorella S 1998 *Phys. Rev.* B **57** 11446
[80] Holland J H 1971 *Associative Information Processing* (New York: Elsevier)
[81] Stella L and Santoro G E Quantum annealing of an Ising spin glass by Green's function Monte Carlo (in preparation)
[82] Binder K and Young A P 1986 *Rev. Mod. Phys.* **58** 801
[83] Castellani T and Cavagna A 2005 *J. Stat. Mech.* P05012
[84] Das A and Chakrabarti B K 2005 *Quantum Annealing and Related Optimization Methods* ed A Das and B K Chakrabarti (Berlin: Springer) p 217
[85] Das A, Chakrabarti B and Stinchcombe R B 2005 *Phys. Rev.* E **72** 026701
[86] Inoue J-I 2005 *Quantum Annealing and Related Optimization Methods* ed A Das and B K Chakrabarti (Berlin:Springer) p 171

J. Phys. A: Math. Theor. **41** (2008) 209801 (2pp) doi:10.1088/1751-8113/41/20/209801

Corrigendum

Optimization using quantum mechanics: quantum annealing through adiabatic evolution

Giuseppe E Santoro and Erio Tosatti 2006 *J. Phys. A: Math. Gen.* **39** R393-R431

In our recent paper (Giuseppe E. Santoro and Erio Tosatti 2006 J. Phys. A: Math. Gen. **39**, R393-R431) we reviewed some of the recent work in the field of quantum annealing, *alias* adiabatic quantum computation. Here we point out two early references, due to Apolloni, de Falco and collaborators and dating back to 1988, where the idea of quantum annealing was first put forward and tested on hard combinatorial optimization problems.

The idea of quantum annealing is an elegant and fascinating alternative to classical thermal simulated annealing; it consists in helping the system escape the local minima using *quantum mechanics* — by tunneling through the barriers rather than thermally overcoming them, with an artificial and appropriate source of quantum fluctuations (the counterpart of the temperature) initially present and slowly (adiabatically) switched off.

In our recent review on this subject [1] we erroneously indicated what appeared to us as being the earliest references [2–4] in which similar ideas were first explicitly formulated, and tested in numerical simulations.

We have in the meantime learned that the idea of using stochastic processes based on quantum mechanics with the goal of minimizing classical complex functions was indeed formulated and tested much earlier, in 1988, by the group of Apolloni, de Falco and collaborators, who applied these ideas to combinatorial optimization problems like graph-partitioning [5, 6].

We also stress that the idea of quantum annealing is inherently related to the idea of adiabaticity. Computing by adiabatic evolution of a quantum system has become a quite popular idea in the Quantum Computing community, where it is commonly known as *adiabatic quantum computation*, and commonly traced back to Ref. [7]. Quantum annealing and adiabatic quantum computation are, however, two names given by two different scientific communities to the very same idea.

Acknowledgments

It is a pleasure to thank Professor de Falco for bringing to our attention the references presented here.

References

[1] Santoro G E and Tosatti E 2006 J. Phys. A: Math. Gen. **39**, R393-R431
[2] Amara P, Hsu D and Straub J E 1993 *J. Phys. Chem.* **97** 6715
[3] Finnila A B, Gomez M A, Sebenik C, Stenson C and Doll J D 1994 *Chem. Phys. Lett.* **219** 343
[4] Kadowaki T and Nishimori H 1998 *Phys. Rev. E* **58** 5355
[5] Apolloni A, Cesa-Bianchi N and de Falco D (1990) in *Stochastic Processes, Physics and Geometry, Proceedings*

of the Ascona-Locarno Conference, 4-9 July 1988, S. Albeverio *et al.* editors, pp. 97-111 (World Scientific, 1990)

[6] Apolloni A, Carvalho C and de Falco D (1989) Stoc. Proc. Appl. **33**, 233-244

[7] Farhi E, Goldstone J, Gutmann S and Sipser M 2000 Quantum computation by adiabatic evolution (*Preprint* quant-ph/0001106)