PHYSICAL REVIEW A **106**, 042607 (2022)

Quantum optimization of complex systems with a quantum annealer

Steve Abel[*]
Institute for Particle Physics Phenomenology, Durham University, Durham DH1 3LE, United Kingdom
and Department of Mathematical Sciences, Durham University, Durham DH1 3LE, United Kingdom

Andrew Blance[†]
Institute for Particle Physics Phenomenology, Durham University, Durham DH1 3LE, United Kingdom;
Department of Physics, Durham University, Durham DH1 3LE, United Kingdom;
and Institute for Data Science, Durham University, Durham, DH1 3LE, United Kingdom

Michael Spannowsky[‡]
Institute for Particle Physics Phenomenology, Durham University, Durham DH1 3LE, United Kingdom
and Department of Physics, Durham University, Durham DH1 3LE, United Kingdom

(Received 1 August 2021; accepted 12 July 2022; published 10 October 2022)

We perform an in-depth comparison of quantum annealing with several classical optimization techniques, namely, thermal annealing, Nelder-Mead, and gradient descent. The focus of our study is large quasicontinuous potentials that must be encoded using a domain wall encoding. To do this, it is important to first understand the properties of a system that is discretely encoded onto an annealer, in terms of its quantum phases, and the importance of thermal versus quantum effects. We therefore begin with a direct study of the 2D Ising model on a quantum annealer, and compare its properties directly with those of the thermal 2D Ising model. These properties include an Ising-like phase transition that can be induced by either a change in "quantumness" of the theory (by way of the transverse field component on the annealer), or by scaling the Ising couplings up or down. This behavior is in accord with what is expected from the physical understanding of the quantum system. We then go on to demonstrate the efficacy of the quantum annealer at minimizing several increasingly hard two-dimensional potentials. For all potentials, we find the general behavior that Nelder-Mead and gradient descent methods are very susceptible to becoming trapped in false minima, while the thermal anneal method is somewhat better at discovering the true minimum. However, and despite current limitations on its size, the quantum annealer performs a minimization very markedly better than any of these classical techniques. A quantum anneal can be designed so the system almost never gets trapped in a false minimum, and rapidly and successfully minimizes the potentials.

DOI: 10.1103/PhysRevA.106.042607

I. INTRODUCTION

Finding the ground state of a complex system is a task of utmost importance in many research fields. In biology, for example, protein folding, namely, the dynamical process whereby a protein chain folds into its characteristic three-dimensional structure, is a well-known phenomenon. This unique structure corresponds to the energetic ground state of its configuration space [1]. In chemistry and drug design, bindings between molecules result in an energy potential that leads the compounds to self-assemble in a specific configuration into their energy ground states [2–4]. In finance, quantitative optimization of portfolios can yield increased profitability [5–7] and, in physics, the scope of applications ranges from complex systems in classical mechanics [8] through quantum mechanics [9–12] and condensed-matter systems [13–17] all the way to models in string theory [18–22] and high-energy physics [23–27]. Further, broad classes of mathematical problems, for example, finding the solution to a differential equation [28–30], can by variational methods be rephrased as an optimization task that ultimately corresponds to finding the extremum of a complicated system.

Due to the importance and prevalence of optimization problems, many methods have been devised to find the extremum of a system. Such methods include various sampling algorithms for discretized or latticized systems [31], optimization algorithms for continuous systems [32–34] and, in recent years, machine learning algorithms [35], i.e., self-adaptive neural networks. See Ref. [27] for a review of several popular classical optimization algorithms.

However, the performance of any such optimization algorithm is characterized by the speed at which it can reliably locate the global extremum of the problem. Often, even *a posteriori*, it is impossible to assess whether an optimization algorithm has actually managed to find the global optimum

---

[*] steve.abel@durham.ac.uk

[†] andrew.t.blance@durham.ac.uk

[‡] michael.spannowsky@durham.ac.uk

or whether it only settled in a local extremum. Surveying the entire configuration space of a problem can quickly become prohibitive and, for NP-hard problems, the difficulty typically increases exponentially with dimensionality. Some confidence can sometimes be regained by repeating the search from different starting conditions, but even then, classical optimization algorithms are often overwhelmed by the local structure of the configuration space, such that it is hard to prevent them falling into the domain of attraction of a nearby local minimum.

With the advent of powerful near-term quantum devices and quantum algorithms, it is natural to ask if these devices could provide a qualitatively different solution to the problem of finding the ground state of a complex system, specifically if they provide a novel avenue to find the global minimum of a system reliably and quickly. In fact, finding the ground state of a complex system was one of the first highly anticipated applications of quantum computers [36]. While several different quantum computing paradigms have been proposed, one type of quantum computer, even though non-universal, has been designed with the very task in mind of minimizing a potential: namely a quantum annealer (QA) [37–48].

The purpose of this paper is to demonstrate the qualitative difference between classical and quantum optimization algorithms. Our focus is on QAs, but we should remark that our discussion remains valid for other quantum computing paradigms, such as gate quantum computers. Indeed, as quantum gate computers are universal quantum computers, they can perform the same quantum calculations as a QA. Specifically, qubits can be connected so as to construct a domain wall encoded potential and the Hamiltonian of an Ising model can be encoded on a quantum gate device (see, for example, Ref. [48] for one of many such implementations). However, at the time of writing accessible QAs, for example, the devices provided by D-Wave Systems [49], have a significant advantage over quantum gate computers in that they offer systems with several thousand qubits [50].

The study we present is relatively straightforward, but to provide arguments in support of the procedures we follow, our discussion will be relatively methodical. Thus our study begins by comparing QAs to their close thermal annealing (TA) cousins in Sec. II. We discus how both kinds of annealers sample the configuration space, and we outline the different ways that these systems find the energetic ground state. We then apply both methods to solve the 2D Ising model in Sec. III. We choose the 2D Ising model as the simplest nontrivial example of a latticized system that has a vast configuration space, scaling as 2^N with N spins, and a highly degenerate energy landscape. The Ising model is of course of interest in and of itself, and we provide an analysis of the physics of its quantum incarnation.

To go beyond the basic Ising model, we use the domain wall encoding (DWE). Using DWE, one can encode a lattice approximation of a continuous function, as well as perform operations on such functions by employing the finite difference method. This extends the applicability of QAs to continuous systems. In Sec. IV, we first recap DWE, which we use for the quantum and thermal annealers. We then study three example functions to make a quantitative comparison between optimization using the gradient descent method (Appendix A 1), Nelder-Mead (NM) method (Appendix A 2), thermal annealing (Appendix A 3), and quantum annealing algorithms. We find a clear qualitative and quantitative difference in the performance of quantum versus classical algorithms in solving such optimization problems. We offer a summary and conclusions in Sec. V.

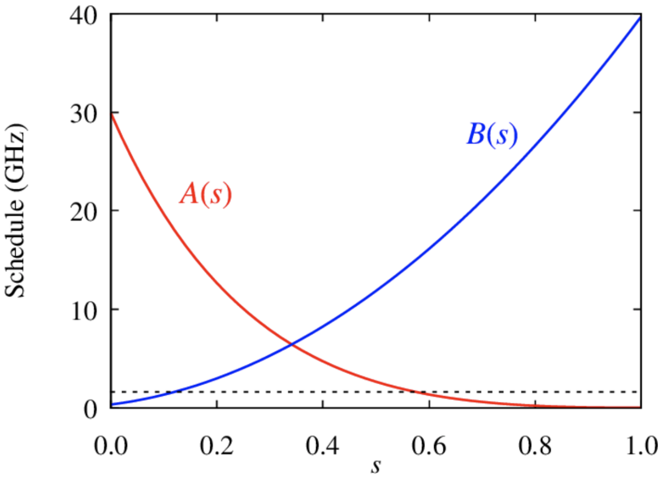


FIG. 1. Anneal schedule parameters. The thermal contribution is shown as a dotted line, while A and B are the coefficients scaling the transverse field and classical Ising contributions, respectively.

II. QUANTUM ANNEALING AND ISING MODEL ENCODING

Quantum versus classical annealers

The device that we will be studying here is the QA [51,52], a system of linked qubits with adjustable couplings. Such devices made available by D-Wave [49] have been able to successfully simulate condensed-matter systems, sometimes showing advantages over classical counterparts [38,53–58].

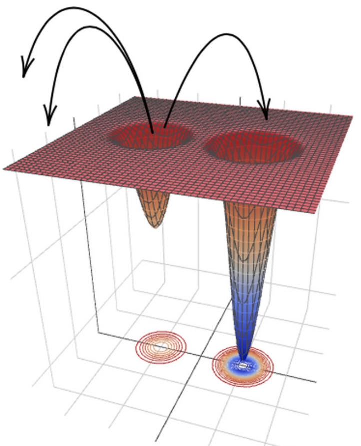
QAs operate in a dissipative rather than fully coherent way. This means that they are not useful for studying, for example, interference properties that depend on coherence, but can in principle be very effective for studying long timescale quantum processes, for example, those in which quantum tunneling allows barrier penetration that would classically be forbidden [59]. This makes them ideal for finding the energetic minimum in systems with complicated energetic landscapes and many barriers and minima.

The basic characteristics of the QA are embodied in its Hamiltonian, which takes the form of a generalized Ising model,

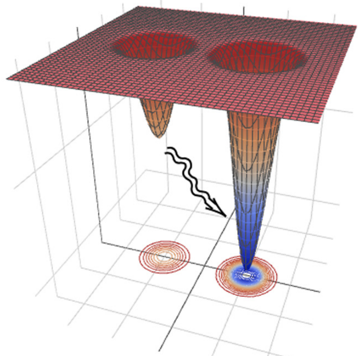
$$\mathcal{H} =, B(s)\left(\sum_{ij} J_{ij}\sigma_i^z\sigma_j^z + \sum_i h_i\sigma_i^z\right) + A(s)\sum_i \sigma_i^x, \quad (1)$$

where i, j label the qubits, σ_i^z are the $z-$spin Pauli matrices, and σ_i^x are the transverse field components, while the couplings h_i and J_{ij} between the qubits are set and kept constant.

The parameter $s(t)$ (with t being time) is a user-defined control parameter that can be adjusted, while $A(s)$ and $B(s)$ describe the consequent change in the quantum characteristics of the annealer. As shown in Fig. 1, smaller $s \in [0, 1]$ means



Thermal excitations over top of barrier at large T



Quantum excitations through barrier at small s

FIG. 2. Thermal versus quantum annealing.

larger transverse field parameter A compared to B, which induces more hopping of σ^z spins, which overall means a system that is more characteristically quantum.

To perform the task of finding a global optimization, the first objective is to encode the problem to be solved into the classical Ising model Hamiltonian represented by the B terms, such that the energetic minimum would correspond to the desired solution. One then adjusts s to alter the relative sizes of the parameters A, B to perform a so-called anneal in the hope that the systems ends up in the global minimum.

Indeed, although we will be comparing quantum optimization with several classical optimization methods, which are summarized in Appendix A, it is really TA which is the closest classical equivalent. For example, both classical and quantum annealers are typically based on the kind of Ising spin model displayed in Eq. (1). In the two panels of Fig. 2, we show schematically the different optimization strategies in a system with two competing minima that has been encoded on such an Ising model. The strategy for TA with the Metropolis algorithm, see Appendix A 3, is based on Boltzmann-weighted excitation and is shown in the first panel of Fig. 2. The generic approach is to perform many trials that begin with a raised temperature T which is gradually reduced, with the probability of energetically unfavorable transitions being suppressed by a factor $e^{-\Delta E/kT}$, where ΔE is the would-be gain in energy. Finessing the temperature as a function of time, i.e., the anneal schedule, is a crucial factor in the success or otherwise of the search strategy. Clearly, the initial temperature should be high enough that the system can clear the barriers. It must then be cooled slowly enough so as to avoid the system freezing into a suboptimal state. We can also appreciate from the figure what can go wrong with a thermal anneal. Aside from the possibility of getting stuck in a local minimum, a very tall thin barrier surrounding the global minimum is clearly hard for this technique to overcome. Moreover, high dimensionality is also likely to be a detrimental factor, as the excited state will have a large phase space to explore and may never return to the domain of attraction of the global minimum.

The second panel of Fig. 2 illustrates the contrasting behavior and search strategy for a QA. In quantum annealing, it is the parameter $s(t)$ that is adjusted to perform the anneal. The anneal schedule typically begins by taking the system to a highly quantum state, i.e., low s. This allows the system to tunnel to the global minimum but its wave function will be very wide, so precision will at that point be low. (Note that here we are thinking of the wave function in the potential encoded in the Ising σ^z spins, so the wave function is wide in the Z direction: It is really the quantum hopping induced by the transverse σ^x term that is dialling the quantumness in the Ising model up or down.)

Then, by increasing s in the anneal schedule, the system will remain trapped in the deeper minimum with the wave function becoming narrower as the system becomes more classical, and the precision increases. As in the thermal case, the solution is found by performing many trials. In the case of a QA, the tunneling is very rapid when the barriers are thin and the wells are deep but becomes less efficient when the barrier is wide compared to its depth. However, it is also somewhat less affected by the dimensionality of the problem. Barrier penetration goes more or less in a straight line between the two minima shown (although the rate does have some dimension dependence). In addition, it is less easy for the system to get lost in the phase space, for example, if the flat directions surrounded the minima in Fig. 2 were to become large.

We can appreciate this different behavior using a semiclassical approximation to the tunneling rate. Indeed, it is well established in field theory that the action of the tunneling instanton bounce solution can still be approximated by the WKB approximation even in multidimensional tunneling systems, as discussed at length in the review of Ref. [60]. From this, we can conclude that the rate at which quantum tunneling takes place is roughly proportional to the instanton tunneling factor $e^{-\omega\sqrt{2m\Delta E}/\hbar}$, where here ω is a measure of the width of the barrier, ΔE is its height, and the parameters m and $\hbar$ in this rate are implicit in the Schrödinger equation and are difficult to determine in the present context. The exponent in this approximation isgiven by the instanton action along a path that goes directly between the two minima. Note the different expected response to tall, thin barriers, namely, the exponent increases as only by the square root of the height, whereas the wide barriers are in principle more of an obstruction. (Although as we shall see, not much in practice). A particular aspect of this rate that will be important is that the *height* of the

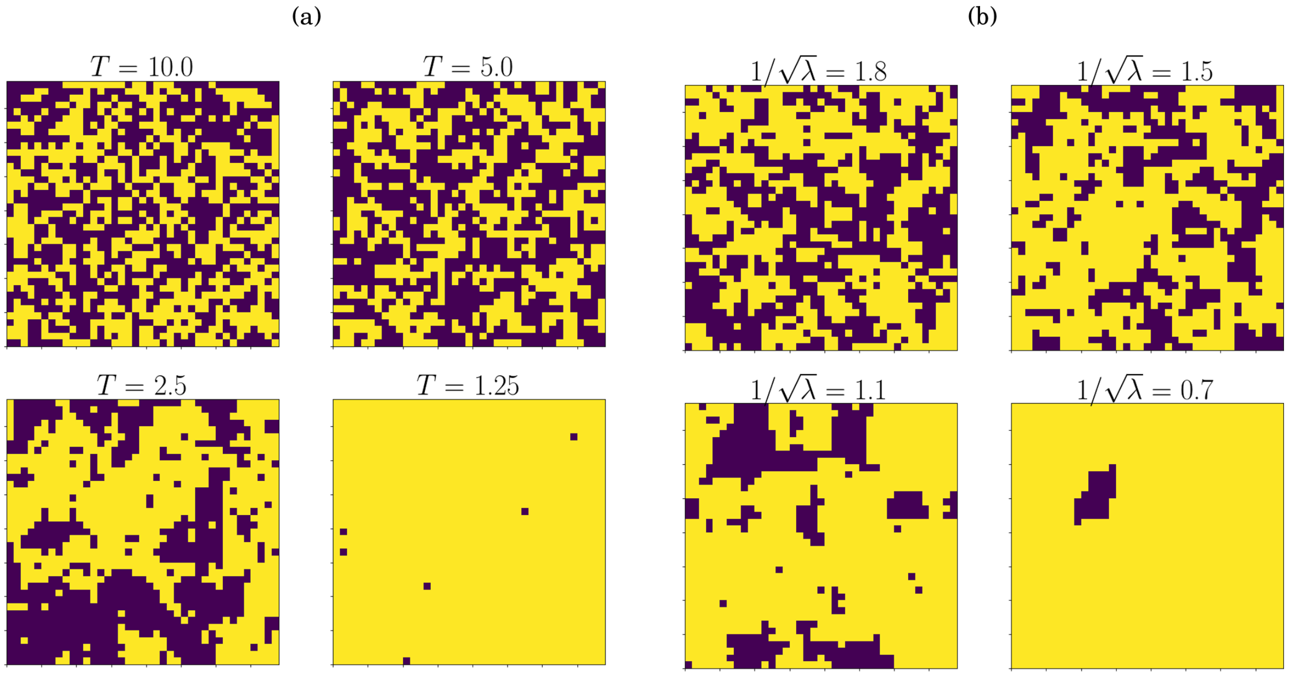


FIG. 3. Comparison of quantum (a) versus classical (b) annealers. Here we take $s = 0.2$ for all the quantum Ising cases, with an anneal time of 100μs. To produce these plots, the coupling is kept constant but the anneal schedule is set to turn on the $s = 0.2$ quantumness and then ramp it down as rapidly as possible (i.e., in 0.8μs) to take a snapshot of the system.

barriers, ΔE, clearly scales with the couplings in $\mathcal{H}$. Thus the parameter that corresponds most closely to the temperature T in the Metropolis algorithm is actually not s itself but rather the square root of the typical inverse J_{ij} coupling: It will therefore be useful to introduce an overall scaling of $\mathcal{H}$ by a factor λ, and then to compare the behavior of the thermal system at different T with the behavior of the quantum system at different λ but at constant s, with the expectation that in that case,

$$T \sim 1/\sqrt{\lambda}. \tag{2}$$

Note that by Eq. (1), this overall coupling factor is really just another way of changing the relative values of A and B, but it has the distinct advantage that the relation to the barrier heights is better understood, and also that the barriers and the quantumness of the system are not being simultaneously adjusted, so there is more control. In this sense, the coupling and s are not really separate parameters because $B(s)$ is itself changing as we vary s. For the quantum system, there are then two distinct kinds of strategies: studying the behavior of the system at constant s for different values of λ provides the most direct comparison of the physical properties of the quantum and the thermal Ising models, whereas for the more practical purpose of performing an optimization, the typical strategy would be to set λ to a constant value and implement a suitable anneal schedule $s(t)$. In the next section, we shall consider both possibilities: Indeed, as we shall see, the former provides a guide for how best to perform the latter.

III. SOLVING THE 2D ISING MODEL

Given the similar Ising encodings used on quantum and thermal annealers, the first interesting analysis that can be made on these two particular systems is simply to probe the phase diagram of the basic 2D Ising model in each case. This can be thought of as a way to understand and calibrate the behaviors of the two systems to guide the design of the anneal schedule for more complicated problems. However the 2D Ising model is, of course, a well-studied statistical mechanical system in its own right, and in the thermal case it famously displays a characteristic phase transition (which partly explains its ability to solve Ising-encoded problems). There has been surprisingly little equivalent analysis of the 2D Ising model on the QA [61]. In this section, we perform such a study of the behavior of the quantum Ising model to compare and contrast it with the thermal one.

Let us begin by defining the standard Ising 2D model. This model is essentially the Hamiltonian in Eq. (1) with only adjacent couplings, which are degenerate and negative (i.e., ferromagnetic). We can implement this model by defining a 2D $N \times N$ grid, with grid positions $(\hat{i}, \hat{j}) \in (1 \dots N,\ 1 \dots N)$ corresponding to N^2 qubits labeled $i \in (1 \dots N^2)$ as follows:

$$i = \hat{i}N + \hat{j}.$$

The negative coupling between adjacent qubits in the grid then corresponds to

$$\begin{aligned} J_{\hat{i}N+\hat{j},\hat{k}N+\hat{\ell}} = &-\lambda(\delta_{\hat{i},\hat{k}-1}\delta_{\hat{j},\hat{\ell}} + \delta_{\hat{i},\hat{k}+1}\delta_{\hat{j},\hat{\ell}} \\ &+ \delta_{\hat{i},\hat{k}}\delta_{\hat{j},\hat{\ell}-1} + \delta_{\hat{i},\hat{k}}\delta_{\hat{j},\hat{\ell}+1}), \end{aligned} \tag{3}$$

where $\delta_{\hat{i}\hat{j}}$ is the Kronecker-δ. (As each coupling provides a negative contribution if opposite spins are adjacent, it can also be thought of as an encoding that provides a solution to the following discrete and somewhat trivial problem: Given two different possible colors, how can one fill an $N \times N$ grid

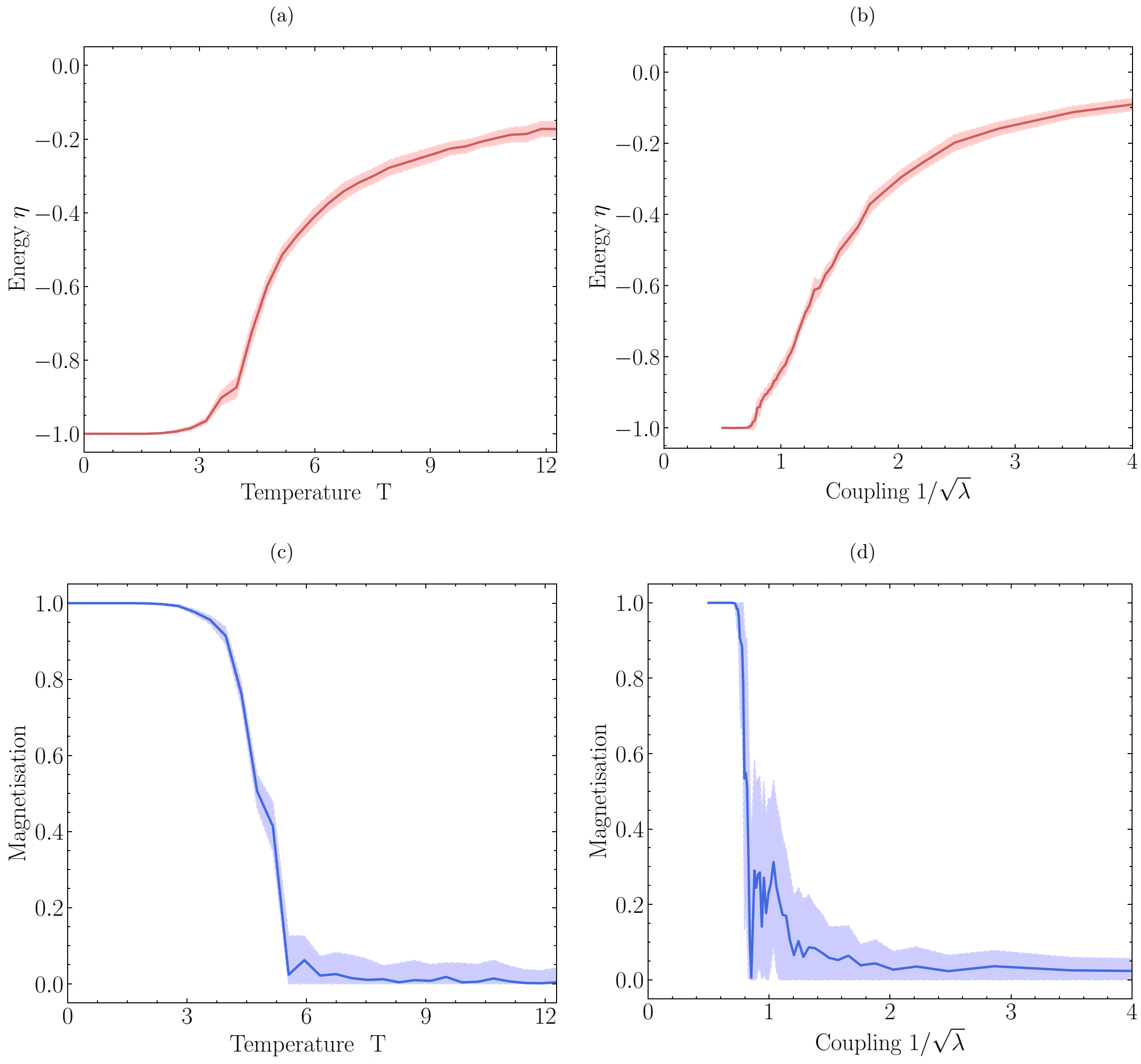


FIG. 4. Comparison of the calculation of the energy and the magnetization using either thermal [(a), (c)] or quantum annealing [(b), (d)]. (a) shows the energy η and magnetization $\mathcal{M}$ for the 2D Ising model at different temperatures and with $N = 40$. (b) shows the result for quantum annealing ($N = 40$, $s = 0.3$, $t = 100\mu$s) when, similarly for thermal and quantum annealing, the uncertainty bands are obtained by running the same setup 100 times.

so there are as few differently colored adjacent squares as possible?)

We begin by simply observing how the two systems behave in Fig. 3. As explained in the Introduction, the parameter in the QA that corresponds to temperature T of the thermal annealer is $1/\sqrt{\lambda}$. One can see that the behaviors of the two systems on varying these parameters are remarkably similar.

We, of course, wish to quantify this by, as usual, studying and comparing how the energy η and magnetization $\mathcal{M}$ of the Ising models varies with T and $1/\sqrt{\lambda}$. As λ is our variable for the quantum system, we must for comparison remove it from the Hamiltonian by defining these quantities as follows:

$$\begin{aligned}\eta &= \frac{1}{\lambda N^2}\sum_{i,j}^{N^2} J_{ij}\sigma_i^z\sigma_j^z,\\ \mathcal{M} &= N^{-2}\left\langle\left|\sum_{i=1}^{N^2}\sigma_i^z\right|\right\rangle,\end{aligned} \tag{4}$$

where they are defined as a density normalised per qubit (cf. volume). Note that with these normalizations, the energy is $0 \leqslant \eta \leqslant -1$, with -1 corresponding to the perfect solution (i.e., with all spins aligned), while the magnetization is, correspondingly, $0 \leqslant \mathcal{M} \leqslant 1$, with $\mathcal{M} = 1$ for perfectly aligned spins.

Beginning with the thermal system, the Ising model was placed on a grid with $N = 40$. For each temperature, we allowed the TA process to iterate 100 times to find the equilibrium before iterating another 100 times to find the energy of the system. From this latter 100 iterations, the mean energy was found and standard deviation error bars drawn. The same initial starting point was used in all runs, initially chosen by randomly assigning the spins. Figures 4(a) and 4(c) show how the energy and magnetization of the system then depends on temperature for a suitable choice of parameters and shows the expected Ising phase transition behavior.

Figures 4(b) and 4(d) show the equivalent set of plots for the QA, with the temperature replaced by $1/\sqrt{\lambda}$ and where

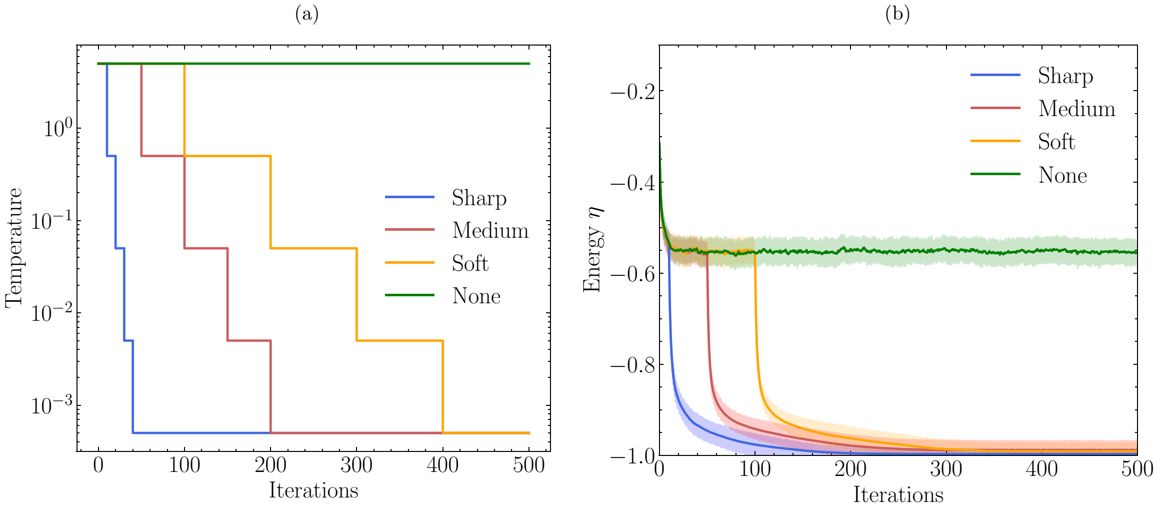


FIG. 5. A comparison of four thermal annealing schedules. In (a), we show our choice of different schedules, ranging from a very sharp drop in temperature to a slow drop. In (b), the results of performing thermal annealing with these schedules are shown. We choose a grid of $N = 40$ and run each setup 100 times to generate the error bars

we have again set the number of qubits in the system to be 40×40. Here and throughout, we perform *reverse* annealing, that is, the anneal schedule starts with the annealer at $s = 1$, the classical scenario, then it decreases s to a set value where it is held for a tunnelling period before then returning to $s = 1$. Unless otherwise specified, when we refer to a given value of s in a run, we are referring to the value during this tunnelling period. The ramp-up and ramp-down times in the schedule are limited by the physical constraints of the annealer, such that the minimum ramp-up or -down time is $(1 - s)\mu$s.

For the plots in Figs. 4(b) and 4(d), we took $s = 0.3$ and a relatively short quantum tunneling period of 100 μs. This optimizes the number of reads we can perform (i.e., 100 in practice) for each value of λ to find the mean and error bars. Despite the short time period for the runs, it was sufficiently long for the system to reach its equilibrium for $s = 0.3$, as we shall discuss shortly. Of course, to freeze the results and read them, the annealer must be brought back to the classical state $s = 1$ as quickly as possible, namely, in 0.7 μs for the choice of parameters. From these figures, it is clear that the quantum 2D Ising system exhibits a remarkably similar kind of phase transition to the thermal system.

We can now probe this characteristic behavior of the respective Ising models to devise a set of annealing procedures for both systems to carry forward to the study of optimization of more generic problems. For thermal systems, this will consist of a suitable temperature or time profile. For the quantum system, the procedure consists of a suitable choice of overall typical coupling λ, together with a time profile for the parameter $s(t)$. We are also interested in the dependence on lattice size N, as clearly this will limit the achievable precision.

Beginning with the thermal system, the equilibration is mainly determined by the anneal schedule, which is a profile of decreasing temperature. For comparison, we set four schedules: a constant schedule, one where the temperature falls off dramatically, a gentle decay, and one where there is a slow

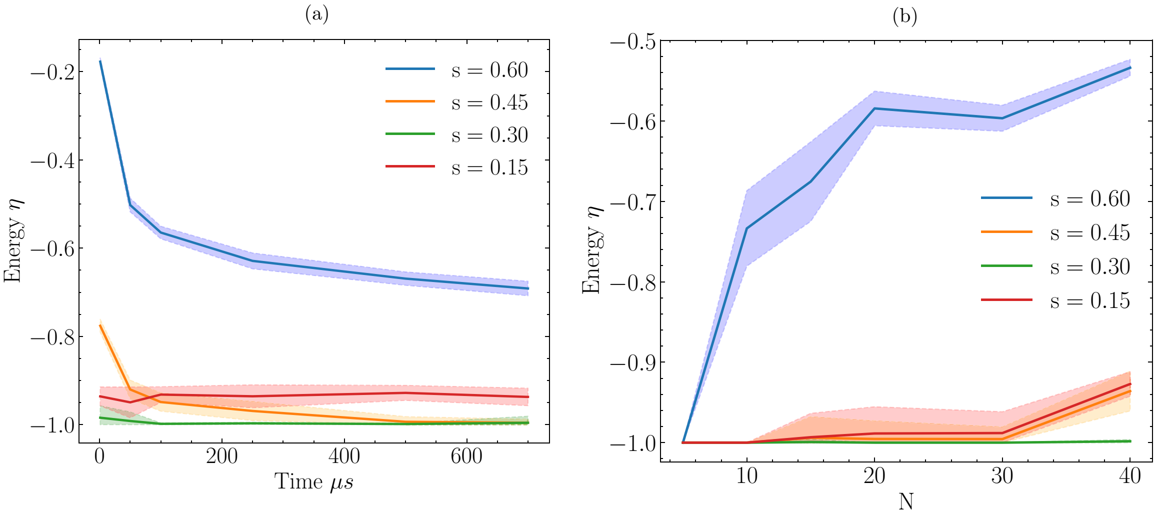


FIG. 6. Effect of changing the anneal time [in (a)] and N [in (b)] on the average energy, for several values of s.

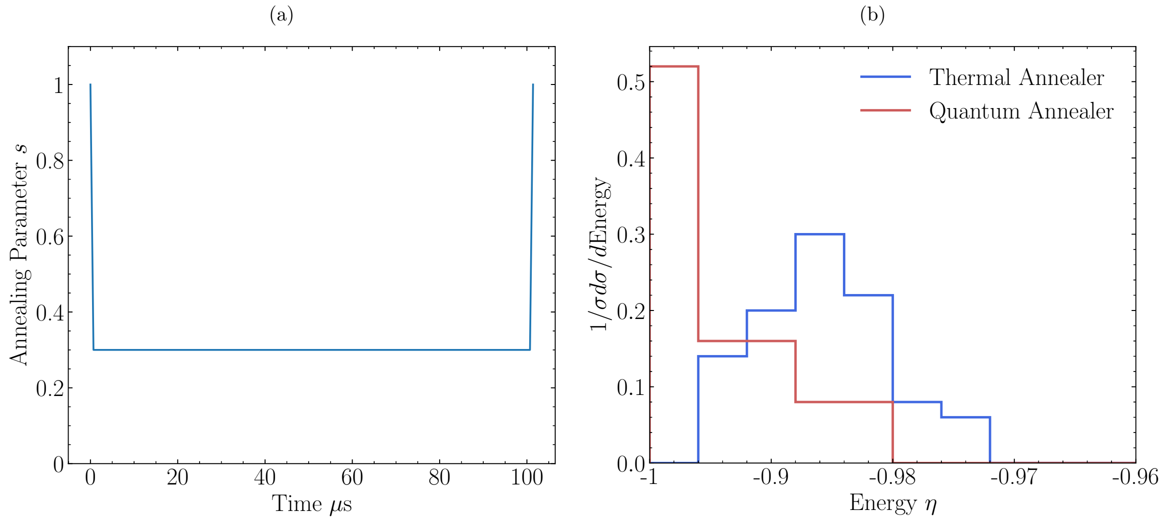


FIG. 7. A comparison of the distribution of minimum energies of the Ising model found using thermal and quantum annealers is shown in (b), starting from randomized initial states. The thermal annealer uses the medium schedule from (a). For the quantum annealer, we chose $s = 0.3$ with a quantum anneal period of 100 μs and the very simple reverse annealing schedule (which is shown for completeness) in (a). Both setups are applied to grids of $N = 40$. We use 50 and 25 randomized initial states for the thermal and quantum anneal, respectively.

decay. The schedules, and the corresponding results (which are displayed in terms of energy) are shown in Fig. 5. To draw the standard deviation bands, we ran each model 100 times. Each schedule scenario begins from its own randomized initial set of spins and is repeated 100 times.

Turning to the QA, the important parameter to gain control over is s. Figure 6 shows the effect of changing this value. We see the QA's ability to optimize is heavily dependent on our choice. However, with a value of $s = 0.3$ we consistently optimize the system even for very short times. To determine the mean and standard error bars for these plots, a run for each choice of s was performed 25 times. We note that the performance of the QA to find the ground state of the Ising model is not monotonic in s. Instead, we find that for $s = 0.3$ the QA finds the global minimum with a tunneling time of $t \simeq 100$ μs very efficiently, whereas smaller s tends to lead to randomized fluctuations that simply prevent the system from ever settling in the correct minimum. Indeed, we can appreciate from Fig. 3(b) that the optimal choice of s (and also λ) is in this sense (much like the maximum temperature in the thermal system) a compromise between focusing on deeper minima and enabling quantum tunneling. Note that it would be interesting in a future in-depth comparison of just these two methods to optimize the parameters using reinforcement learning techniques, as in Ref. [62].

In Fig. 6(b), we show how changing lattice size N affects the performance. Here, each model was again run for 25 times. Unsurprisingly, for larger values of N it becomes more difficult to sample the configuration space within the tunneling time of $t \simeq 100$ μs. However, while the performance of the QA for $s \gg 0.3$ or $s \ll 0.3$ clearly deteriorates with growing N, for $s = 0.3$ the QA finds the global minimum very reliably even for the largest N that can physically be placed on the annealer. It is also worth mentioning at this point that, of course, the physical limitations of the D-Wave annealer itself limits N. In particular, the annealer does not provide the most general Ising model as in Eq. (1), but only provides a subset of the J_{ij} couplings. Therefore, in practice, any desired model has to be embedded onto the annealer using locked spin chains. Thus, due to the physical constraints of the annealer (in particular, the fact that the maximum number of available qubits is ~5000), large N embeddings begin to become artificially frustrated for $N > 40$.

The results are shown in Fig. 7(b). The QA clearly shows a higher probability of finding the correct ground state and with a smaller variance. It is also notable that the evaluation time is vastly longer for TA: Although this is obviously a function of the available architecture, a TA run takes approximately $O(10^6)$ times longer than a run on the QA, which takes $O(100\,\mu\mathrm{s})$. We conclude that the QA has a clear edge in speed and reliability over simulated TA in calculating the ground state of the 2D Ising model, a discrete highly complex system.

We now wish to compare the thermal and quantum methods with *each other* to see how efficient each method is at finding solutions and, conversely, how dependent the results are on the choice of initial conditions. To do this, we ran both the quantum and thermal annealers for a number of randomized initial states with an optimal annealing configuration in each case.

For the QA, we selected a value of $s = 0.3$ and a schedule with a quantum period of 100 μs with ramp-up and -down times of $(1 - s)\,\mu$s as shown in Fig. 7(a). We then ran 25 initial states, 25 times each. For the TA, we chose the medium schedule from Fig. 5(a). For each initial state, we ran the TA process 50 times to find an average energy for 50 initial states. Both setups use $N = 40$ and $\lambda = -2.0$.

IV. FUNCTION OPTIMIZATION

We now wish to extend our study to consider the solution of more practical problems that are encoded onto nonminimal Ising models. The problem we shall discuss is the very generic one of finding the minimum of a continuous function of two

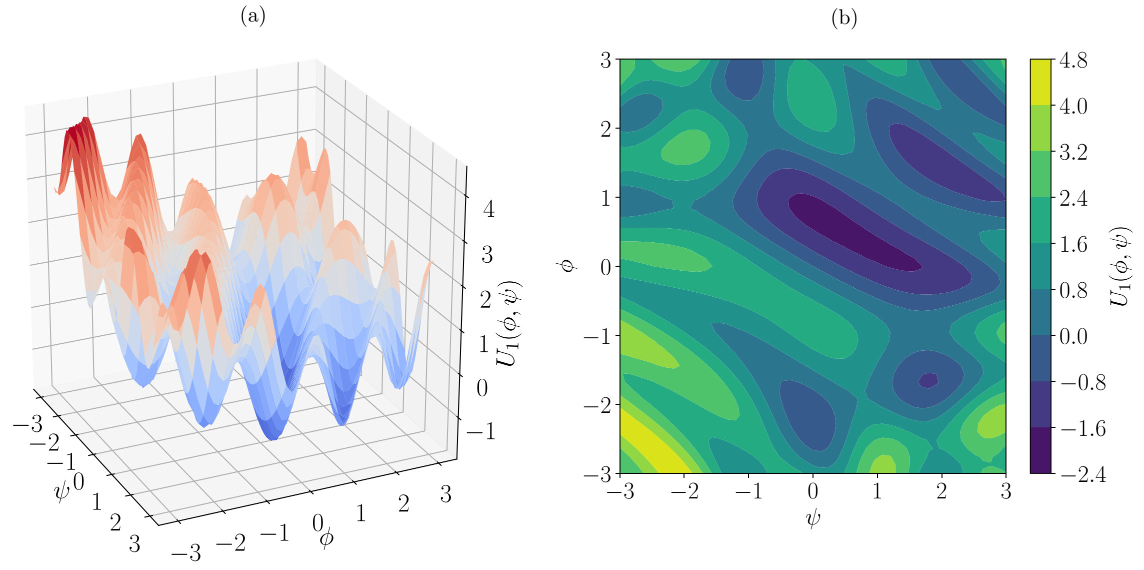


FIG. 8. The potential U_1 of Eq. (5) rendered in 3D in (a) and as a contour plot in (b).

variables. The useful aspect of such a generic study is that the analysis can now be broadened to bring in for comparison other optimization techniques that do not use Ising model encodings at all. In this paper, as mentioned in the Introduction, we also consider the gradient descent method and the NM method, described in Appendices A 1 and A 2, respectively. However, as a first step, let us describe the Ising encoding of a continuous function which will be central to this part of the paper. This is the subject of the next subsection.

A. Domain wall encoding of a continuous function

To have a specific problem in mind, we will consider minimizing the following function of two variables ϕ, ψ shown in Fig. 8:

$$\begin{aligned} U_1(\phi, \psi) = &- \lambda[\phi(1-\phi) + \psi(1-\psi) \\ &+ 12\cos(\phi\psi)\sin(\psi + 2\phi)]. \end{aligned} \tag{5}$$

This corrugated function has, of course, been chosen because it is particularly unpleasant to minimize due to its many local minima. Indeed, even in the domain shown in Fig. 8, namely, $\phi, \psi \in [-3, 3]$, we can identify six minima.

Naturally, the minimization of such a function will be implemented by adopting it directly as the Hamiltonian (i.e., we regard U_1 as a potential). Therefore, per the discussion in the Introduction and the previous section, we have introduced an overall scaling of the couplings λ that we can use to tune the efficiency of the annealers. Here Fig. 8 shows the potential for $\lambda = 0.5$.

The scalar values of the two variables ϕ, ψ can be represented using the DWE introduced in Ref. [63]. The basic idea of the DWE is to discretize the continuous ϕ and ψ domains into N elements of sizes ξ and ζ, respectively. Then the value of each of the scalars is represented by the position of a frustrated spin on a spin chain where the spin flips signs from negative to positive, the so-called domain wall. Thus, the total number of qubits in a chain is $2N$, with the first N qubits encoding ϕ and the second N qubits encoding ψ. (For convenience of presentation, we will use the same number of qubits for ϕ and ψ but, of course, this is not mandatory.) Clearly, to approximate the continuous function, N should be as large as possible (we will usually take either $N = 40$ or $N = 30$ depending on the difficulty of the problem we are considering). Two scalar values are then faithfully described when there is a single frustrated position encoding ϕ in the lower block at $i = r_1 \in [1, N]$, where the spin flips from negative for all lower indices to positive for higher indices up to N, and a second frustrated position encoding ψ in the upper block at $i = r_2 \in [N + 1, 2N]$. In other words, a faithful encoding of two scalar field values is defined to have the following spin structure:

$$\begin{array}{l} \begin{array}{|c|}\hline i \\ \hline \sigma_i^z \\ \hline \end{array} \overbrace{\begin{array}{|c|c|c|c|c|}\hline 1 & \dots & r_1 & \dots & \mathrm{N} \\ \hline - & - - - & + & + + + & + \\ \hline \end{array}}^{\phi} \\ \qquad\qquad\quad \underbrace{\begin{array}{|c|c|c|c|c|}\hline N+1 & \dots & r_2 & \dots & 2\mathrm{N} \\ \hline - & - - - & + & + + + & + \\ \hline \end{array}}_{\psi} \end{array} \tag{6}$$

The scalar values are extracted from such a configuration by counting the negative spins, as follows:

$$\begin{aligned} \phi &= \phi_0 + \xi(r_1 - 1) = \phi_0 + \frac{\xi}{2}\sum_{i=1}^{N}\left(1 - \sigma_i^z\right), \\ \psi &= \psi_0 + \zeta(r_2 - 1) = \psi_0 + \frac{\zeta}{2}\sum_{i=N+1}^{2N}\left(1 - \sigma_i^z\right), \end{aligned} \tag{7}$$

where ϕ_0, ψ_0 are fiducial minimum values.

This is the way that we wish the scalar values to be represented, but how do we enforce only such faithful representations of ϕ and ψ, and not, for example, meaningless configurations in which the spins flips signs at more positions or at no positions at all? This is done by adding two crucial components to the Hamiltonian, a domain-wall enforcing component in the linear h_i term of the Ising model in Eq. (1), and a ferromagnetic coupling in the J_{ij} terms that encourages spin alignment between neighboring sites:

$$\begin{aligned} h_i^{(\text{chain})} &= \Lambda' \, (\delta_{i,1} - \delta_{i,N} + \delta_{i,N+1} - \delta_{i,2N}), \\ J^{(\text{chain},\phi)}_{i<N,j<N} &= -\frac{\Lambda}{2}\, (\delta_{i,j+1} + \delta_{i+1,j}), \\ J^{(\text{chain},\psi)}_{i\geqslant N,j\geqslant N} &= -\frac{\Lambda}{2}\, (\delta_{i,j+1} + \delta_{i+1,j}), \end{aligned} \tag{8}$$

where Λ, Λ' are parameters that are somewhat larger than the largest energy scale in the problem. (For the best performance, they should not be very much larger.)

The couplings $h_i^{(\text{chain})}$ force the system to have negative spin $\sigma_1^z = -1$ at the bottom of each block, i.e., for $i = 1$ and $N + 1$, and positive spin $\sigma_i^z = +1$ at the top of each block, i.e., for $i = N$ and $2N$. This is in accord with the desired spin structure in Eq. (6) and it forces an odd number of spin flips between the two ends of a block. Meanwhile $J_{ij}^{(\text{chain},\phi)}$ gives a (relative) penalty of 2Λ for each domain wall in the lower block, and separately $J_{ij}^{(\text{chain},\psi)}$ gives a (relative) penalty of 2Λ for each domain wall in the upper block. This favors just a single wall in each block as required.

Having enforced only faithful representations of ϕ and ψ, it is then straightforward to see how one can now further encode the potential terms in $U_1(\phi, \psi)$. To describe this, let us split the potential as

$$U_1(\phi, \psi) = U_a(\phi) + U_b(\psi) + U_c(\psi, \phi), \tag{9}$$

where $U_{a,b}$ are the terms that involve only one of the two variables, ϕ and ψ respectively, while U_c is the mixed term, in this particular case:

$$U_c = -12\lambda \cos(\phi\psi) \sin(\psi + 2\phi).$$

Note that one has to be a little circumspect making this designation, as we shall see.

Consider first the simpler U_a term. This piece can be encoded into either h_i or J_{ij}. The encoding into h_i can be performed by noting that the desired potential can be written

$$U_a(\phi) = \frac{1}{2}\sum_{j=1}^{N} U_a(\phi_0 + \xi j)\big(\sigma^z_{j+1} - \sigma^z_j\big), \tag{10}$$

since there is a contribution only where σ_j^z and σ_{j+1}^z have different signs, namely, at the position of a domain wall where $j = r_1 - 1$. Assuming that U_a is differentiable, and $\xi \ll 1$, this is equivalent to a linear coupling of the form

$$h_j^{(U_a)} = -\frac{\xi}{2}\partial_\phi U_a(\phi_0 + \xi j). \tag{11}$$

Alternatively, the same coupling can be encoded into J_{ij} by instead adding the couplings

$$J_{ij}^{(U_a)} = \tfrac{1}{2} U_a(\phi_0 + \xi j)(\delta_{ij} - \delta_{(i-1)j}). \tag{12}$$

It is easy to see that again this picks out a contribution from the location of the domain wall similarly to h_i above. Indeed, clearly there is no contribution from Eq. (12) unless i, j are equal or adjacent. If we consider then summing i past a given j, the above term reduces to

$$\tfrac{1}{2} U_a(\phi_0 + \xi j)\big(1 - \sigma^z_{j+1}\sigma^z_j\big), \tag{13}$$

which is then to be summed over j. Upon performing the j sum, there again remains a contribution only at the domain wall, where $j = r_1 - 1$. The equivalent encoding for $U_b(\psi)$ is trivially performed by summing instead over the upper block of indices $i \in [N + 1, 2N]$, swapping ξ for ζ, and obviously swapping U_a for U_b.

In the present context, it does not matter if we decide to encode U_a and U_b into h_i or J_{ij}. Where it does, however, make a difference is if one wishes to alter the potential at some point in the anneal to study some particular physical process. This was the case, for example, in Ref. [59] where the system was first allowed to settle into a minimum before the potential was changed by scaling h_i (using the h-gain parameter) to develop a lower "true vacuum" into which the system could then tunnel by barrier penetration. This possibility will not be used here.

By contrast, the $U_c(\phi, \psi)$ terms in the potential that involve both variables can only be encoded in J_{ij}. The logic for such terms follows that of the encoding of a single variable term into h_i in Eq. (11). That is, we add

$$\begin{aligned} U_c(\phi, \psi) \equiv& \sum_{i,j=1}^{N} U_c(\phi_0 + i\xi, \psi_0 + j\zeta) \\ &\times \frac{1}{4}\big(\sigma^z_{i+1} - \sigma^z_i\big)\big(\sigma^z_{N+j+1} - \sigma^z_{N+j}\big) \\ =& \frac{\xi\zeta}{4}\sum_{i,j=1}^{N} \partial_\phi\partial_\psi U_c(\phi_0 + i\xi, \psi_0 + j\zeta)\, \sigma^z_i \sigma^z_{N+j}, \end{aligned} \tag{14}$$

which upon symmetrising implies that

$$J^{U_c}_{i,j+N} = J^{U_c}_{N+j,i} =, \frac{\mu}{8}\xi\zeta\, \partial_\phi\partial_\psi U_c(\phi_0 + i\xi, \psi_0 + j\zeta). \tag{15}$$

We now return to our somewhat cryptic comment above about care being needed to decide which encoding is appropriate, which is something of a technicality but we discuss it for completeness: The issue is that Eq, (14) may not be sufficient to encode the whole of U_c because it is a double derivative. Suppose, for example, that the coupling U_c contained a contribution $\phi\psi^2$. Then, written in discretized form, we see that there is a single-field term effectively given by the fiducial field values:

$$\phi\psi^2 \supset \phi_0\psi^2 + \psi_0^2\phi + \dots . \tag{16}$$

Such terms could not be incorporated by the double derivatives implicit in J_{ij} and Eq. (15), and instead h_i would have to be augmented to cancel spurious single field terms: In this particular case, the required additional terms would be

$$h^{(U_c)}_{j=1\dots N} = -\frac{\xi}{4}\mu\psi_0^2, \tag{17}$$

$$h^{(U_c)}_{N+j} = -\frac{\zeta}{2}\mu\bar{\phi}\,(\psi_0 + j\zeta), \tag{18}$$

where $\bar{\phi}$ is the average ϕ in the interval, i.e., $\bar{\phi} = \phi_0 + \frac{N\xi}{2}$. In principle, a similar consideration by trigonometric identities applies for the actual U_c terms in the U_1 of Eq. (5). In practice, we find that this issue is resolved by multiplying the potential by a factor that forces a constant value for U_1 at the boundaries, but is unity everywhere inside the domain away from the boundaries. For the functions we will consider, we find that this issue does not make a significant difference but it is something to be aware of for other studies.

B. Results for the corrugated potential U_1

To provide a baseline to compare the QA against, we will use three classical optimization methods. As well as TA, we will consider NM and conjugate gradient descent (GD). The NM method uses a simplex to traverse the function space to find a minimum. The GD method chosen is an extension to classical GD, which includes an adaptive step size. Both these methods, alongside TA, are discussed further in Appendix A.

While the NM and GD methods use the continuous form of the potential, both the QA and TA methods will use the discretized Ising encoding of the potential described in the previous subsection. Here, for the TA we chose to use $N = 50$. Due to limitations imposed by the QA device, we discretize the potential such that $N = 20$ for most of the QA runs. This decision is kept for most of the paper. However, later in Sec. IV C 2, we will discuss how the choice of N might effect our findings, and will in fact conclude that it does not. In the near future, when larger devices becomes available, we believe all the conclusions we reach here will carry forward.

Our choice of overall λ is also influenced by the method. For the classical methods, larger λ generally hinders the ability to find the minimum as it would be harder to escape local minima—although, of course, given the Boltzmann weighting, for the TA any change in λ can be absorbed in a redefinition of T. This is not the case with quantum annealing, where as explained in Secs. III and IV C 1, scaling the potential will make the global minimum more visible. Therefore, a slightly larger value for λ is taken for the QA: here we take $\lambda = 0.7$ instead of $\lambda = 0.5$ for the classical runs.

Guided by the studies in Sec. III, the thermal anneal schedule is set such that it runs for 4000 iterations, starting at a temperature of $T = 1.1$. After 500 iterations, the temperature is halved, then reduced to 5% of its value every 500 iterations afterward. The QA schedule is set such that the ramp-up period is 15 μs. It then tunnels for 100μs during which time s is set to 0.1, before ramping down for 250 μs. Note there is a somewhat different approach now: To find a global minimum we are free to choose a slow ramping down to ensure we do not excite the system.

Note that we refine the values of the QA and TA parameters to optimize the performance on the annealer, for example, to reduce the number of chain breaks occurring in the minor embedding which appears to be sensitive to these parameters: hence the parameter values differ somewhat for the three potentials, although this is not crucial for the conclusions which are derived from the successful runs. Indeed, the same schedule, Λ and s for the different examples, would result in a quite suboptimal optimization for some potentials, but essentially the same diagrams.

For each of the four optimization methods, we start the runs for a selection of initial starting points. This leads to Fig. 9. Here we show the distribution of results from the methods. For the classical methods, we use 2500 initial conditions while for the QA, fewer initial conditions are used, around 100 due to the limited time on the device. As is usual when performing quantum annealing, each initial position is run multiple times. Here, we choose the number of reads to be 100, with Figs. 9(g)

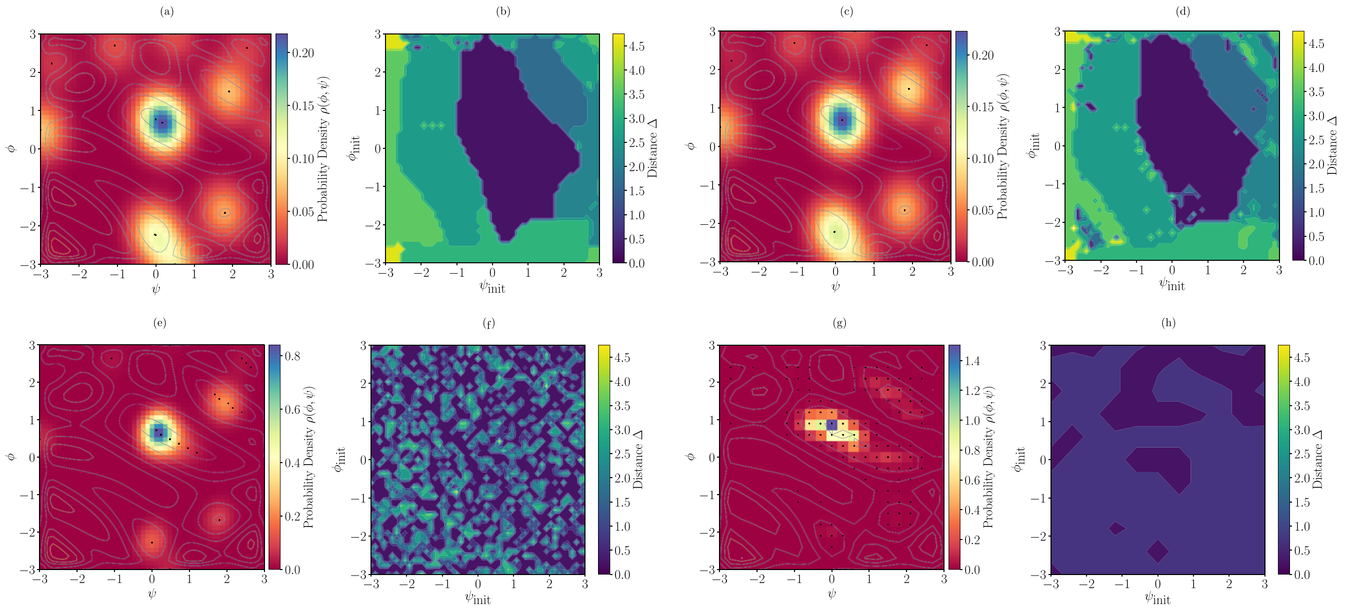


FIG. 9. For a range of initial ϕ and ψ values, the distance between the predicted minimum and true minimum for corrugated potential U_1 is given. This is done for Nelder-Mead in (a), (b); gradient descent in (c), (d); thermal annealing in (e), (f); and quantum annealing in (g), (h). For (a)–(f), a grid size of $N = 50$ and $\lambda = 0.5$ is chosen with the models being initialized from 2500 different points. For the quantum annealing in (g) and (h), we take $N = 20$ and $\lambda = 0.7$ and the model is initialized from ~100 points. Each point is used to produce 100 reads, with the peak of the probability distribution being identified as the predicted minimum for that setup.

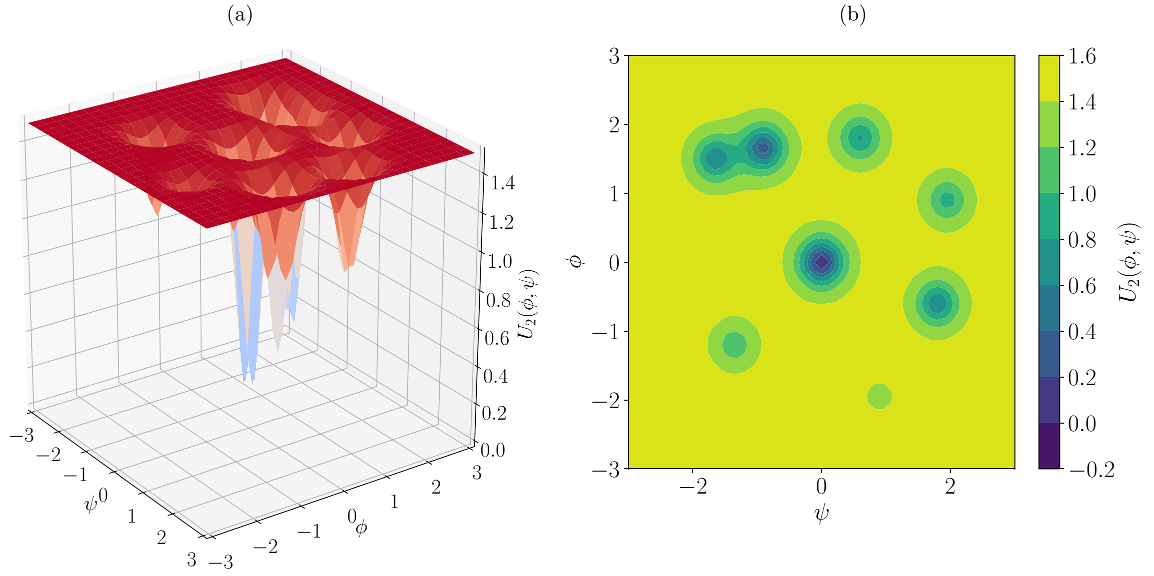


FIG. 10. The potential U_2 of Eq. (19) rendered in 3D in (a) and as a contour plot in (b).

and 9(h) showing every successful point. (An unsuccessful point would be one, for example, that had ended up with more than two domain walls so it does not faithfully describe the two variables). We see that the three classical methods are more liable to fall into and get trapped in false minima than the QA.

This effect is surprisingly marked, but]to quantify it, it is instructive to plot the mean distance from the true minimum that the system ends up, for each given starting point. That is, for each run [i.e., for each particular $(\phi_{\text{init}}, \psi_{\text{init}})$], we take the result and calculate the distance of the result to the correct true minimum, Δ. For the three classical methods, this is straightforwardly run by run. However, for each set of initial conditions, the QA has *already* been run multiple times when the results are collected. To select the predicted value from the set of 100 annealer runs corresponding to a given $(\phi_{\text{init}}, \psi_{\text{init}})$, we therefore take the peak of a fitted probability distribution. We justify this (namely, that taking the peak of the distribution of 100 reads is not unfairly favoring the QA) using the results presented in Appendix B, where we show the distribution of Δ for a selection of initial starting points. We see that for the QA runs, the distribution is already very focused around $\Delta = 0$ for each of the 100 reads for a choice of ϕ_{init} and ψ_{init}. (For example, taking instead the mean of the 100 reads would have produced very similar results.)

The right panels of Fig. 9 show the resulting Δ's. Clearly NM and GD perform very badly in such a potential, and we can identify basins of attraction around the local minima from which they never escape. The TA performs noticeably better in the sense that it is not susceptible to just getting trapped in the closest minimum to its starting point, but on the other hand there is a remaining randomness to the minimum that it eventually *does* end up in. However, we see again that the QA gets consistently much closer to the true minimum of the potential from virtually any starting point. Moreover, this does not appear to depend on the size of the array (i.e., N), as we shall shortly demonstrate.

C. Multiwell potential U_2

To show the versatility of our method, we now perform the same analysis on another potential that is even more difficult. This potential has the form of a flat plateau with multiple holes of varying depths, and is given by

$$U_2(\vec{\phi} = (\phi, \psi)) = \lambda\,\Big(p_0 \tanh^2(|\vec{\phi}|/\omega) - \sum_a p_a \operatorname{sech}^2(|\vec{\phi} - \vec{v}_a|/\omega)\Big), \tag{19}$$

where (p_0, p_a) is a choice of minimum depths with $p_0 > p_a$, where $\vec{v}_a$ is a choice of positions for each minimum, ω is the width of all the dips, and where λ is again an overall scaling parameter. We select ω to be 0.3 and λ to be 0.5 and 10 for the classical and quantum methods, respectively. The depths of the minima we took to be

$$(p_0, p_a) = (3, 0.9, 0.3, 1.2, 1.8, 1.5, 1.8, 2.4), \tag{20}$$

while the positions are given by

$$\begin{aligned}\vec{v}_a = {} & [(-1.2, -1.35), (-1.95, 0.9),\\ & \times (0.9, 1.95), (1.5, -1.65),\\ & \times (1.8, 0.6), (-0.6, 1.8), (1.65, -0.9)]_a.\end{aligned} \tag{21}$$

The resulting potential U_2 is shown in Fig. 10.

We now carry out the same analysis as in Sec. IV A. During this analysis, the thermal annealer is run on the same schedule as before. The QA is scheduled to ramp up for 15 μs before tunneling for 50 μs and ramping down for 50 μs. During the tunnel phase of the anneal, $s = 0.15$.

Running these optimizers, alongside the NM and CG methods, results in Fig. 11. As was the case with the first potential, the QA performs more consistently than the classical methods. Instead of falling into one of the many local minima. it is able to find the true minimum on almost every run. So far, we have focused on the consistency of the QA. Another benefit is its speed. The QA's runtime is decided by its schedule. Here,

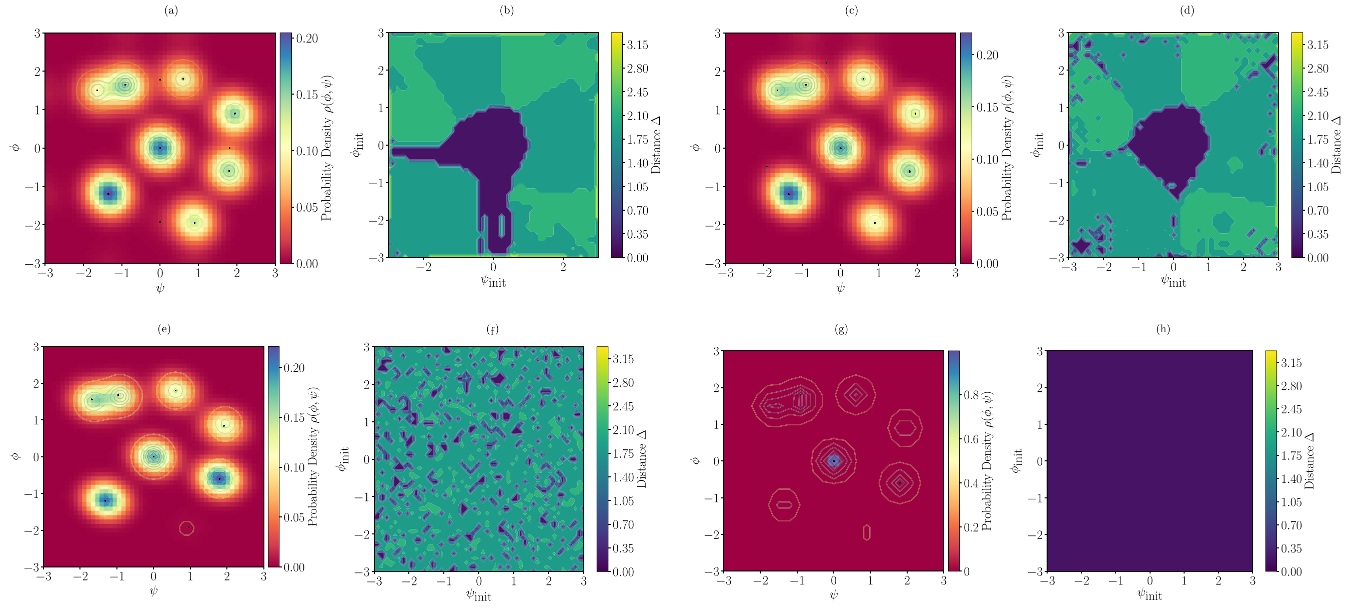


FIG. 11. For a range of initial ϕ and ψ values, the distance between the predicted minimum and true minimum for multiwell potential U_2 is given. This is done for Nelder-Mead in Figs. 9(a) and 9(b), gradient descent in Figs. 9(c) and 9(d), thermal annealing in Figs. 9(e) and 9(f), and quantum annealing in Figs. 9(g) and 9(h). For Figs. 9(a)–9(f), a grid size of $N = 50$ and $\lambda = 0.5$ is chosen with the models being initialized from 2500 different points. For the quantum annealing in Figs. 9(g) and 9(h), we take $N = 20$ and $\lambda = 0.7$ and the model is initialized from ∼100 points. Each point is used to produce 100 reads, with the peak of the probability distribution being identified as the predicted minimum for that setup.

that comes to 115], μs. The classical NM and GD methods are very dependent on the number of iterations each takes. By timing how long each optimization takes that is used to create Fig. 11, we can find an average run time for each method. Table I shows the results. For the NM, an average optimization takes 0.0049 s; for the GD method it takes 0.0029 s. TA, by far, takes the longest. One thermal anneal run takes around half a second. For fair comparison, the thermal annealer time is calculated for an $N = 20$ scenario. As the NM and GD runs largely depend on the step size of the optimizer, changing N does not effect the timing. However, for TA a change of N has a large impact on the time required to optimize, with a larger N resulting in a longer time. (These tests were performed on a 1.2 GHz Intel Core m3.)

1. Scaling the potential up or down

All three potentials chosen can be scaled up or shrunk down by adjusting the factor λ. Such a uniform operation does not change the position of the global minimum in the (ϕ, ψ) plane. Intuitively, it is clear what effect modifying this has on one of the classical methods. By increasing the depth of both local and global minima by the same amount, the probability for the classical algorithms to settle in a local minimum rather than the global remains unchanged to a good approximation. Classical algorithms that do not sample the entire configuration space are unaware of the depth of the extremum and whether they located the global one. However, the QA is instead aware of the depth of the potential well and by increasing the depth the quantum annealing process becomes more successful.

We demonstrate this using the the multiwell potential example of Sec. IV C. Figure 12 shows the results of running the QA for a variety of values of λ. Starting from a low value of λ, the QA has difficulty finding the minima. As we increase the depth of the potential's minima from $\lambda = 0.5$ to $\lambda = 10$ (the choice we made in Sec. IV C) the annealer's ability greatly increases.

TABLE I. Timings for the four optimization methods solving potential 2. For NM and CG, $N = 50$ while a value of $N = 20$ is used to calculate both annealing methods.

Method	Time/run (μs)
Nelder-Mead	4900
Gradient descent	2900
Thermal annealing	5×10^5
Quantum annealing	115

2. Effect of varying the grid size

It is worth briefly discussing our choice of N in all the examples we have discussed. In particular, an obvious remaining question is whether the different choices of N and the lower possible N values for the QA may somehow have favored it. A direct comparison shows that this possibility can be dismissed. In Fig. 13(a), we perform a thermal anneal taking $N = 20$. This shows us that the TA processes are not greatly influenced by the fineness of the discretization of the potential. We can see it is still unable consistently to find the true minimum. Likewise for the quantum annealer, the result is also independent of N. In Fig. 13(b), we present an $N = 30$ run, showing that the same good results obtain and that the superior performance of the quantum annealer

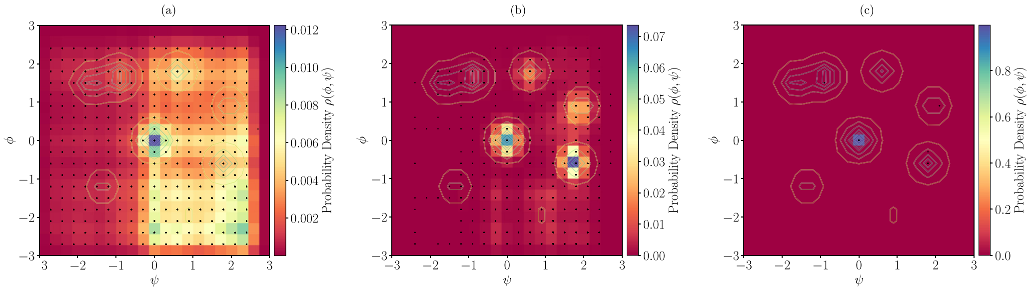


FIG. 12. The quantum annealer optimizing the multiwell potential, U_2, for three values of the scaling parameter $\lambda = 0.5,\ 5,\ 10$, in (a)–(c), respectively. All runs use $N = 20$, a tunneling and ramp-down period that are both 50 μs, and $s = 0.15$.

persists even when it has a finer lattice. The grid size does not effect the qualitative features of the annealer's ability to find the minimum. Indeed, the performance of both kinds of annealers is relatively independent of N as would be expected on physical grounds. As mentioned, our choice of $N = 20$ for most of the QA studies shown here is mostly a technical limitation, not a conceptual one.

D. Volcano crater potential U_3

Finally we consider a third potential that takes the following form:

$$U_3(\phi,\psi) = \lambda\,(2\,e^{-(\phi^4+\psi^4)/2} - 10\,e^{-20(\phi^2+\psi^2)}\cos^2(\psi)\cos^2(\phi)), \tag{22}$$

where λ is again a scaling factor, which is here chosen to be 1.7 for all the cases. This particular volcano crater potential is actually adapted directly from Ref. [27]. As can be seen from Fig. 14, it is perhaps the single most difficult potential of all to minimize as it has a very narrow global minimum, surrounded by a repulsive potential that in principle rejects all attempts to try and find it. We perform the same analysis as before, but this time the thermal annealer schedule is modified, allowing it to run for 1000 iterations before the temperature decay begins. This ensures that the result is not dependent on the initial starting condition. The quantum schedule ramps up for 15 μs, then the annealer tunnels for 100 μs at $s = 0.01$ before ramping down for 250 μs.

Figure 15 shows the distribution of results from the optimization processes. It is clear that the NM and GD methods are very poorly equipped to deal with this potential. From Figs. 15(a)–15(f), it appears that most of the results from these runs are no longer contained within the plotted grid or, rather, they run away to the boundary. Thermal and quantum annealing both perform much better. Once again, the quantum annealer in Figs. 15(g) and 15(h) is able to perform consistently.

V. CONCLUSIONS

optimization tasks are at the heart of many quantitative techniques across all sciences. Thus, various numerical optimization algorithms have been proposed and implemented on classically operating systems. As quantum computers, in particular, QAs, perform calculations in a fundamentally different way compared to classical computers, they are anticipated to have a qualitatively different and possibly transformative impact on the way optimization tasks are realized.

In this paper, we have performed a detailed comparison between quantum annealing and three of the most popular

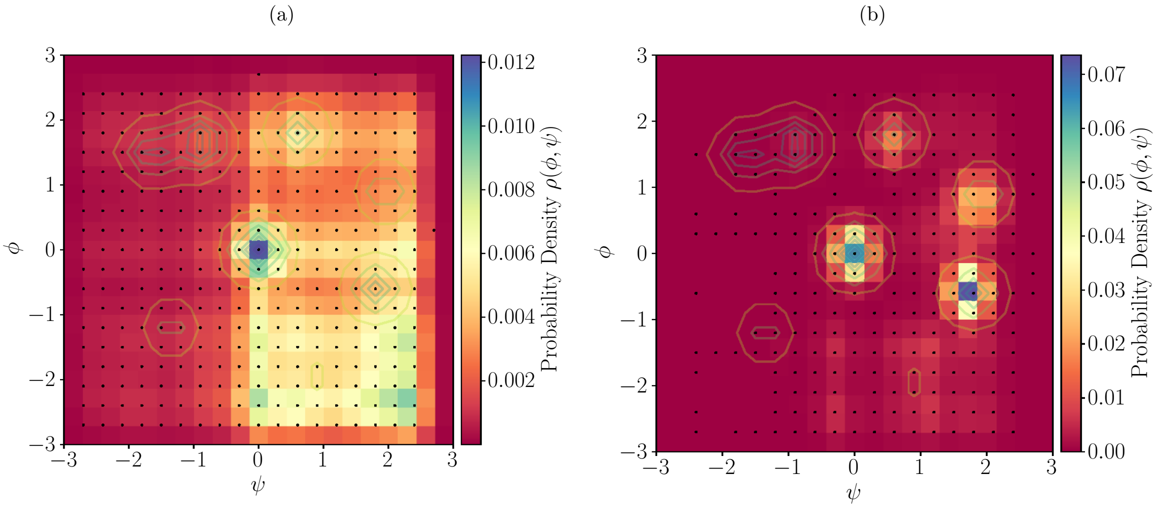


FIG. 13. The results of thermal annealing on an $N = 20$ grid in (a) and quantum annealing on an $N = 30$ grid in (b) for the multiwell potential.

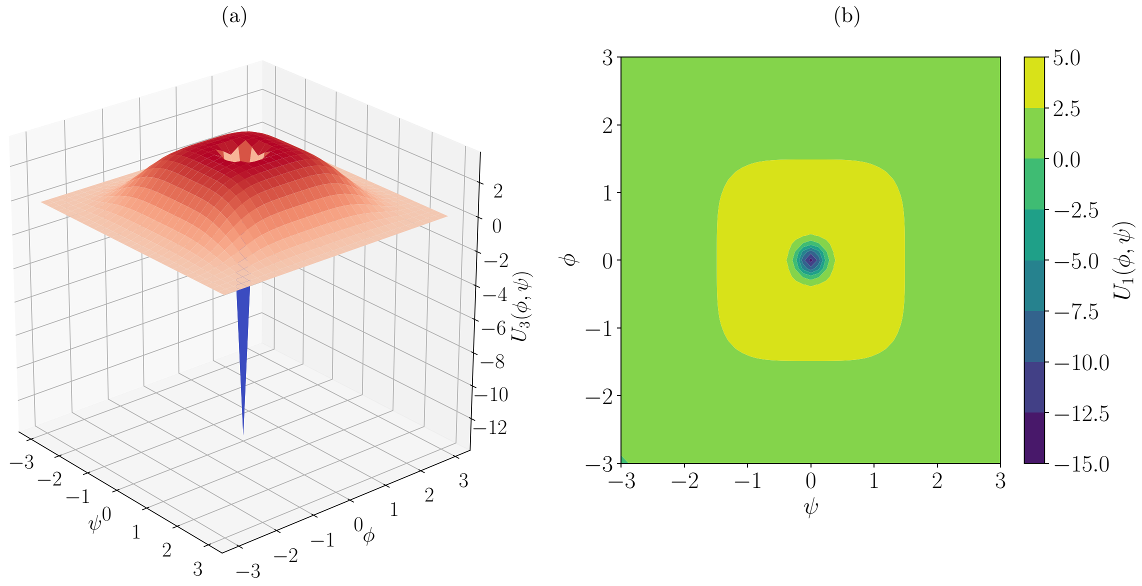


FIG. 14. The volcano potential U_3 of Eq. (22) rendered in 3D in (a) and as a contour plot in (b).

classical optimization algorithms, namely, TA, NM, and GD. First, we have shown that quantum annealing can quickly and reliably solve the 2D Ising model, a highly complex latticized system. Indeed, in many respects it displays a similar phase transition with respect to a change in overall scale of the potential, λ. For the QA, this scaling plays a similar role to the temperature in a thermal anneal.

Then, using DWE, we were able to show that a QA can find the ground state of various continuous two-dimensional functions much more reliably than classical methods. That is, it is far less dependent on the initial conditions of the optimization algorithm and on the topographical profile of the function, as well as being much faster than classical algorithms. It is notable that scaling the function $\lambda U_i(\phi, \psi)$ up by increasing λ improves the performance of the QA. This indicates that the QA samples the configuration space by tunneling underneath potential barriers, for which a deeper well increases the tunneling rate, rather than by sampling the configuration space by following the topographic structure of $U_i(\phi, \psi)$ in the lateral direction.

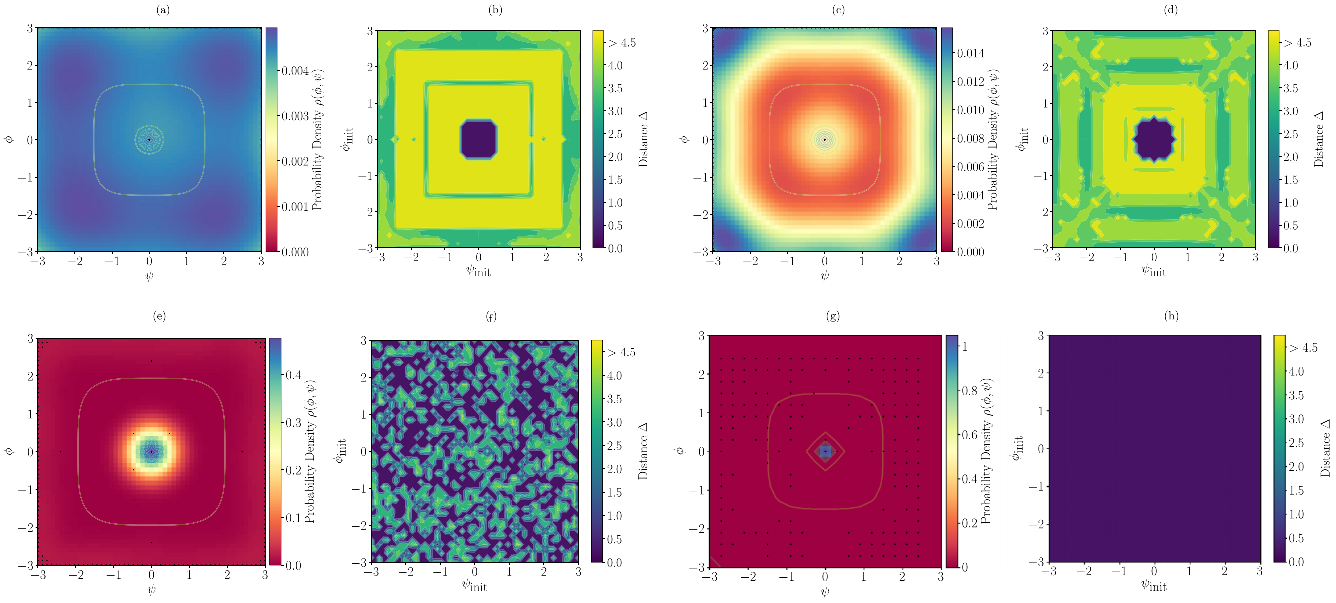


FIG. 15. For a range of initial ϕ and ψ values, the distance between the predicted minimum and true minimum for volcano potential U_3 is given. This is done for Nelder-Mead in (a), (b); gradient descent in (c), (d); thermal annealing in (e), (f); and quantum annealing in (g), (h). For (a)–(f), a grid size of $N = 50$ and $\lambda = 0.5$ is chosen with the models being initialized from 2500 different points. For the quantum annealing in (g) and (h), we take $N = 20$ and $\lambda = 0.7$ and the model is initialized from ∼100 points. Each point is used to produce 100 reads, with the peak of the probability distribution being identified as the predicted minimum for that setup.

For many problems it can be very hard, even *a posteriori*, to know if a classical optimization algorithm has settled into a local or a global minimum. Thus, finding the global minimum reliably is one of the biggest challenges for any optimization task, which is often only heuristically addressed by rerunning the algorithms with different initial conditions multiple times, or by performing other partial scanning manoeuvres. However, the depth awareness of a quantum algorithm seems to offer an entirely different approach to locating a global extrema and may have implications for optimization tasks in many areas.

ACKNOWLEDGMENT

We would like to thank Nicholas Chancellor for helpful discussions. S.A. and M.S. are supported by the STFC under Grant No. ST/P001246/1.

APPENDIX A: CLASSICAL OPTIMIZATION ALGORITHMS

1. Gradient descent method

Simple GD optimization techniques can face difficulties converging to the correct solution. An example of this is when they are getting close to the bottom of the valley they are descending into. By following the steepest gradient, depending on the step size, it is possible to step entirely over the solution. This results in the algorithm oscillating over the correct answer but never finding it.

Conjugate GD aims to resolve this by accounting for previous gradients when calculating its next step [33,34]. GD methods, generally, follow the form

$$x_{i+1} = x_i + \alpha d_i, \tag{A1}$$

where x_i and x_{i+1} is your current location and next location, respectively. The form also depends on α, a step size, and d_i, a vector describing the direction in which you will move. For a standard GD method, d_i will typically be $-\nabla f(x_i)$, the negative gradient of the function you are finding the minimum of.

In the conjugate gradient method, d_i will be a combination of gradients. Here, it will depend on $-\nabla f(x_i)$ and $-\beta_i \nabla f(x_{i-1})$. The term β_i is designed such that the new directional vector is conjugate to the previous. By accounting for the previous gradient when finding the next step, and using an adaptive step size, the likelihood of taking a step in the direction you have previously come from is reduced. This results in the optimization performance increasing.

2. Nelder-Mead method

NM is an optimization technique that uses a geometric structure known as a simplex [32]. The simplex, through the optimization procedure, traverses the function space to find a minimum. For an N-dimensional feature space, a simplex with $(N+1)$ vertices can be initialized. The vertices x_i will each have a corresponding value found from evaluating the function $f(x_i)$. At each iteration, these vertices are shifted according to a set of rules. The process repeats until either a maximum number of iterations is reached or the standard deviation of the points decreases under a threshold.

Specifically, we can describe the original NM optimization method with three repeating steps. In step (i), $f(x)$ is evaluated for each simplex vertex. The vertices are then put in ascending order (x_0, x_1,..., x_{N+1}) based on this evaluation. A centroid, x_c, is then calculated in step (ii). This is simply the mean of all the points evaluated in step (i), excluding x_{N+1}. At step (iii), the algorithm will shift the simplex. We will consider a few important vertices of the simplex, namely, the best point x_0, the second-worst point x_N, the worst point x_{N+1}, and the centroid. The updates are done using one of four rules:

1. Reflection: A new point, x_R, is found from $x_R = x_c + \alpha(x_c - x_{N+1})$, where $\alpha > 0$. If x_R is better than x_N but worse than x_0, then x_{N+1} (the worst point in the simplex) is replaced with x_R.
2. Expansion: If x_R was less than x_0 and therefore the new best point, another point $x_E = x_c + \gamma(x_R - x_c)$, where $\gamma > 1$, can be found. Here, we aim to check if another, even better, point can be found. If x_E is the new best point, it can replace x_{N+1}, otherwise x_{N+1} will be replaced by x_R.
3. Contraction: This is the case where $x_R \geqslant x_N$. A new point $x_{\text{cont}} = x_c + \beta(x_{N+1} - x_c)$, where $0 < \beta \leqslant 0.5$, is found. If this new value x_{cont} is better than the worst point, it can replace x_{N+1}.
4. Shrink: In the case the previous three checks fail, and x_{cont} is worse than x_{N+1}, the shrink procedure can be performed. Here, all points are replaced, except x_0. A point x_i is shifted such that $x_i = x_0 + \sigma(x_i - x_0)$, where σ is typically 0.5.

For a low-dimensional example, the simplex manipulations are shown in Fig. 16. After the appropriate update is performed, the iteration is complete and the process begins again. After step (i), the algorithm has the chance to be terminated. This depends on the standard deviation of the values found from evaluating the vertex values. If this is under a certain threshold or a maximum number of iterations is reached, then the optimization ends. The smallest value found from evaluating the vertices is returned as the minimum of the function.

3. Thermal annealing with Metropolis algorithm

We can consider our potential as being a system that contains various states. These states, each representing different points in the potential, will have an associated energy. The minimum of the potential will be the state with the lowest energy. TA aims to optimize a system by finding this state [31]. At each iteration of the algorithm, the energy of the system can be measured and compared against the energy of a neighboring state. Depending on the energy of the neighbor, there is a probability it will be accepted. If accepted, this new configuration replaces the current state and the next iteration can begin.

By beginning in a state A, an associated energy E_A can be found. This state can be changed to create B, a potential new state with an energy E_B. If $E_B < E_A$, then the new state has lower energy and can be accepted as our new state. If $E_B > E_A$, then the candidate state is not energetically preferable. However, there is still a chance to accept it with a probability $e^{-(E_B - E_A)/T}$. This probability, and hence the likelihood of the algorithm accepting a new configuration, is guided by a tem-

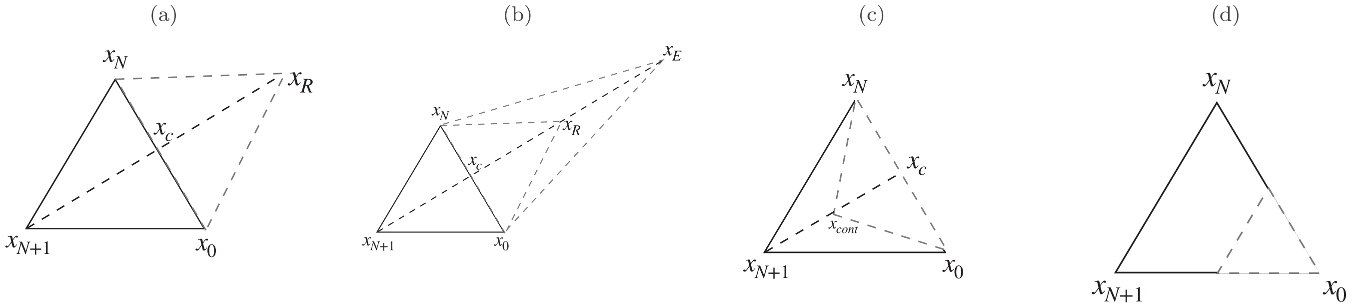


FIG. 16. The four variations of how the simplex vertices can be updated during a Nelder-Mead optimization iteration: Reflection in (a), expansion in (b), contraction in (c), and shrink in (d).

perature T. Being allowed to take steps away from a minimum reduces the likelihood of falling into local minima.

The choice of temperature is an important factor in TA. A higher temperature will result in more configurations being accepted. It is typical then to perform the annealing process with a schedule. By starting at a high temperature, space can be quickly explored. After a select number of iterations, the temperature can be reduced until the algorithm settles into the global minimum.

APPENDIX B: PERFORMANCE COMPARISON BETWEEN OPTIMIZERS WITH DISTINCT INITIAL CONDITIONS

To visualize the performance differences of the four optimizers, we choose four distinct initial conditions of $(\phi_{\text{init}}, \psi_{\text{init}})$, run each of them multiple times and plot the distribution of distances Δ in Figs. 17–19. That is, similarly to what was done previously for each of the potentials in Figs. 9, 11, 15. We again see that the classical methods are

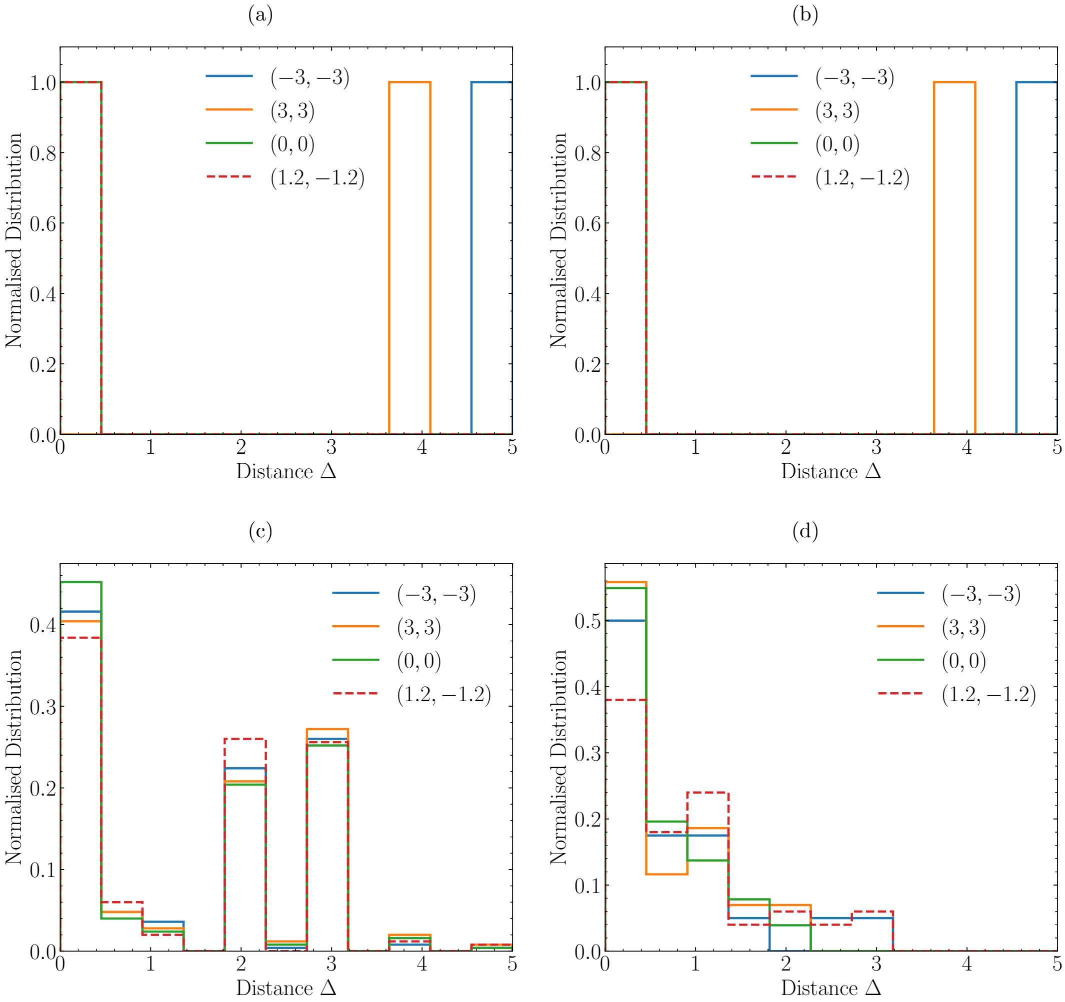


FIG. 17. A comparison of the distance a predicted point is away from the true minimum, for the four initial starting points in the legend, using potential U_1. Four optimization techniques are shown: Nelder-Mead in (a), gradient descent in (b), thermal annealing in (c), and quantum annealing in (d). Each initial condition was run 100 times.

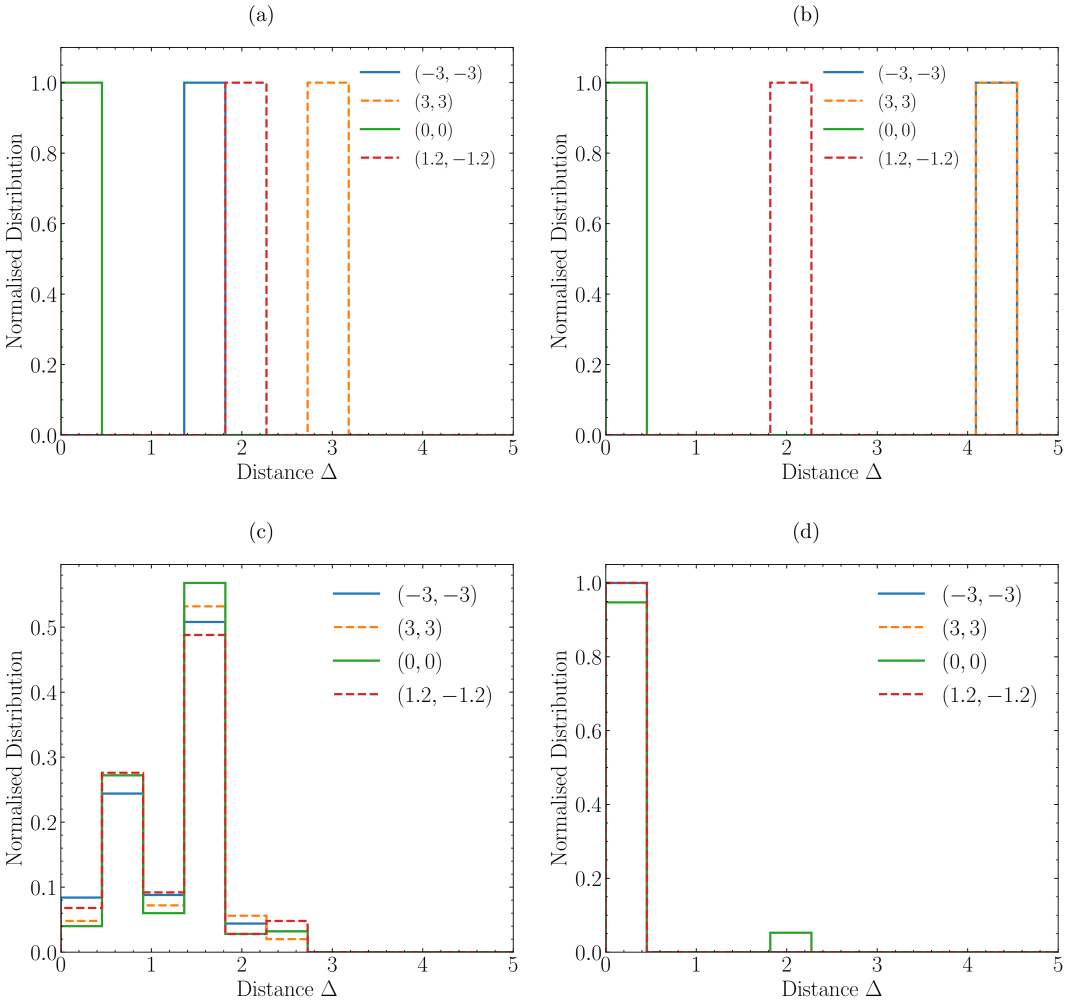


FIG. 18. A comparison of the distance a predicted point is away from the true minimum, for the four initial starting points in the legend, using the multiwell potential U_2. Four optimization techniques are shown: Nelder-Mead in (a), gradient descent in (b), thermal annealing in i(c), and quantum annealing in (d). Each initial condition was run 100 times.

more dependent on their initial starting point, and the behavior is in accord with what was concluded before: namely, NA and GD *always* get stuck in the wrong minimum for some of the points, and only (but always) find the true minimum if they happen to start in its basin of attraction. On the other hand, TA is reasonably good at jumping into the deepest well but gets stuck some of the time, while the QA almost always finds the deepest well but, at least for the limited N system available to us, retains a small spread. We would anticipate that larger N would allow us to eliminate this final spread.

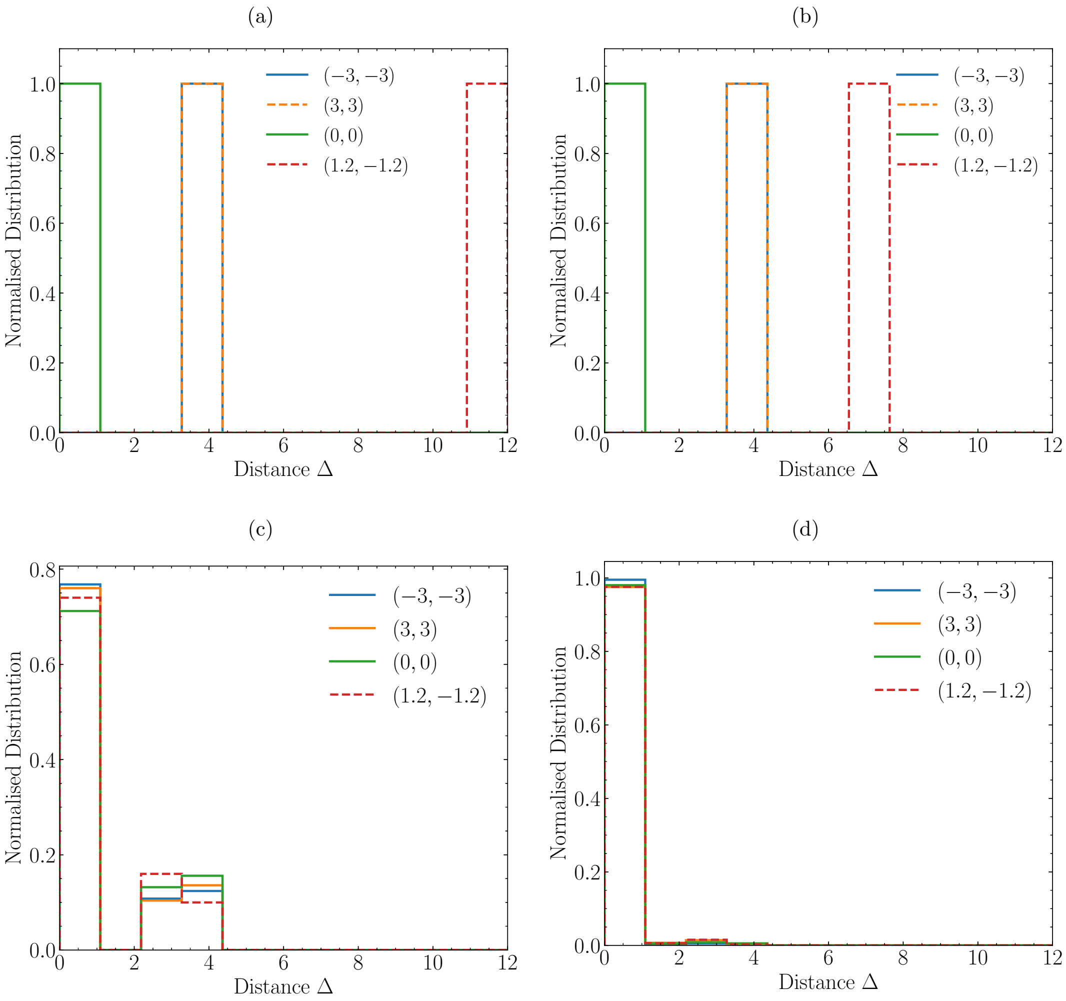


FIG. 19. A comparison of the distance a predicted point is away from the true minimum, for the four initial starting points in the legend, using the volcano potential U_3. Four optimization techniques are shown: Nelder-Mead in (a), gradient descent in (b), thermal annealing in (c), and quantum annealing in (d). Each initial condition was run 100 times.

[1] C. Levinthal, How to fold graciously, in *Mössbauer Spectroscopy in Biological Systems, Proceedings of a Meeting Held at Allerton House*, edited by J. T. P. DeBrunner and E. Munck (University of Illinois Press, Monticello, Illinois, 1969), Vol. 67, pp. 22–24.

[2] A. Aspuru-Guzik, A. Dutoi, P. Love, and M. Head-Gordon, Simulated quantum computation of molecular energies, Science **309**, 1704 (2005).

[3] B. Lanyon, J. Whitfield, G. Gillett, M. Goggin, M. Almeida, I. Kassal, J. Biamonte, M. Mohseni, B. Powell, M. Barbieri, A. Aspuru-Guzik, and A. White, Towards quantum chemistry on a quantum computer, Nat. Chem. **2**, 106 (2010).

[4] V. E. Elfving, B. W. Broer, M. Webber, J. Gavartin, M. D. Halls, K. P. Lorton, and A. Bochevarov, How will quantum computers provide an industrially relevant computational advantage in quantum chemistry?, arXiv:2009.12472.

[5] J. Moody, L. Wu, Y. Liao, and M. Saffell, Performance functions and reinforcement learning for trading systems and portfolios, J. Forecast. **17**, 441 (1998).

[6] A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Management: Concepts, Techniques and Tools*, Revised edition (Princeton University Press, 2015), pp. 1–699.

[7] A. Tsantekidis, N. Passalis, A. Tefas, J. Kanniainen, M. Gabbouj, and A. Iosifidis, *Using deep learning to detect price change indications in financial markets*, in *Proceedings of the 25th European Signal Processing Conference (EUSIPCO)* (IEEE, Piscataway, NJ, 2017), pp. 2511–2515.

[8] B. van Brunt, The calculus of variations,.

[9] V. Fock, Näherungsmethode zur lösung des quantenmechanischen mehrkörperproblems, Z. Phys. **61**, 126 (1930).

[10] J. MacDonald, Successive approximations by the Rayleigh-Ritz variation method, Phys. Rev. **43**, 830 (1933).

[11] S. R. White, Density matrix formulation for quantum renormalization groups, Phys. Rev. Lett. **69**, 2863 (1992).

[12] G. Drake and Z.-C. Van, Variational eigenvalues for the s states of helium, Chem. Phys. Lett. **229**, 486 (1994).

[13] A. Bogdanov and A. Hubert, Thermodynamically stable magnetic vortex states in magnetic crystals, J. Magn. Magn. Mater. **138**, 255 (1994).

[14] U. Rößler, A. Bogdanov, and C. Pfleiderer, Spontaneous skyrmion ground states in magnetic metals, Nature (London) **442**, 797 (2006).

[15] M. Segall, P. Lindan, M. Probert, C. Pickard, P. Hasnip, S. Clark, and M. Payne, First-principles simulation: Ideas, illustrations and the CASTEP code, J. Phys.: Condens. Matter **14**, 2717 (2002).

[16] S. Buhrandt and L. Fritz, Skyrmion lattice phase in three-dimensional chiral magnets from Monte Carlo simulations, Phys. Rev. B **88**, 195137 (2013).

[17] S. Schenk and M. Spannowsky, Exploring instantons with spin-lattice systems, Phys. Rev. B **103**, 144436 (2021).

[18] K. Becker, M. Becker, M. Haack, and J. Louis, Supersymmetry breaking and α'-corrections to flux induced potentials, J. High Energy Phys. 06 (2002) 060.

[19] S. Kachru, R. Kallosh, A. D. Linde, and S. P. Trivedi, De Sitter vacua in string theory, Phys. Rev. D **68**, 046005 (2003).

[20] V. Balasubramanian and P. Berglund, Stringy corrections to Kähler potentials, SUSY breaking, and the cosmological constant problem, J. High Energy Phys. 11 (2004) 085.

[21] V. Balasubramanian, P. Berglund, J. P. Conlon, and F. Quevedo, Systematics of moduli stabilisation in Calabi-Yau flux compactifications, J. High Energy Phys. 03 (2005) 007.

[22] S. AbdusSalam, S. Abel, M. Cicoli, F. Quevedo, and P. Shukla, A systematic approach to Kähler moduli stabilisation, J. High Energy Phys. 08 (2020) 047.

[23] D. Bardin, M. Bilenky, A. Chizhov, O. Fedorenko, S. Ganguli, A. Gurtu, M. Lokajicek, G. Mitselmakher, A. Olshevsky, J. Ridky, S. Riemann, T. Riemann, M. Sachwitz, A. Sazonov, A. D. Schaile, Yu. Sedykh, I. Sheer, and L. Vertogradov, ZFITTER: An Analytical program for fermion pair production in $e^+ e^-$ annihilation, arXiv:hep-ph/9412201.

[24] R. Lafaye, T. Plehn, and D. Zerwas, SFITTER: SUSY parameter analysis at LHC and LC, arXiv:hep-ph/0404282.

[25] H. Flächer, M. Goebel, J. Haller, A. Hocker, K. Monig, and J. Stelzer, Revisiting the global electroweak fit of the standard model and beyond with Gfitter, Eur. Phys. J. C **60**, 543 (2009); **71**, 1718(E) (2011).

[26] J. C. Criado, V. V. Khoze, and M. Spannowsky, The emergence of electroweak skyrmions through Higgs bosons, J. High Energy Phys. 03 (2021) 162.

[27] DarkMachines High Dimensional Sampling Group, C. Balázs *et al.*, A comparison of optimisation algorithms for high-dimensional particle and astrophysics applications, J. High Energy Phys. 05 (2021) 108.

[28] M. Raissi, P. Perdikaris, and G. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, J. Comput. Phys. **378**, 686 (2018).

[29] M. L. Piscopo, M. Spannowsky, and P. Waite, Solving differential equations with neural networks: Applications to the calculation of cosmological phase transitions, Phys. Rev. D **100**, 016002 (2019).

[30] J. Y. Araz, J. C. Criado, and M. Spannowsky, Elvet—a neural network-based differential equation and variational problem solver, arXiv:2103.14575.

[31] N. Metropolis, A. W. Rosenbluth, M. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. **21**, 1087 (1953).

[32] J. A. Nelder and R. Mead, A simplex method for function minimization, Comput. J. **7**, 308 (1965).

[33] M. R. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems, J. Res. Natl. Bur. Stand. **49**, 409 (1952).

[34] J. R. Shewchuk, An introduction to the conjugate gradient method without the agonizing pain (1994), http://www.cs.cmu.edu/∼quake-papers/painless-conjugate-gradient.pdf.

[35] D. Rumelhart, G. Hinton, and R. Williams, Learning representations by back propagating errors, Nature (London) **323**, 533 (1986).

[36] R. Feynman, Simulating physics with computers. int j theor phys 21(6, 7):467-488, Int. J. Theor. Phys. **21**, 467 (1982).

[37] A. B. Finilla, M. A. Gomez, C. Sebenik, and J. D. Doll, Quantum annealing: A new method for minimizing multidimensional functions, Chem. Phys. Lett. **219**, 343 (1994).

[38] T. Kadowaki and H. Nishimori, Quantum annealing in the transverse Ising model, Phys. Rev. E **58**, 5355 (1998).

[39] J. Brooke, D. Bitko, T. F. Rosenbaum, and G. Aeppli, Quantum annealing of a disordered magnet, Science **284**, 779 (1999).

[40] N. G. Dickson, M. W. Johnson, M. H. Amin, R. Harris, F. Altomare, A. J. Berkley, P. Bunyk, J. Cai, E. M. Chapple, P. Chavez, F. Cioata, T. Cirip, P. deBuen, M. Drew-Brook, C. Enderud, S. Gildert, F. Hamze, J. P. Hilton, E. Hoskinson, K. Karimi *et al.*, Thermally assisted quantum annealing of a 16-qubit problem, Nat. Commun. **4**, 1903 (2013).

[41] T. Lanting, A. J. Przybysz, A. Y. Smirnov, F. M. Spedalieri, M. H. Amin, A. J. Berkley, R. Harris, F. Altomare, S. Boixo, P. Bunyk, N. Dickson, C. Enderud, J. P. Hilton, E. Hoskinson, M. W. Johnson, E. Ladizinsky, N. Ladizinsky, R. Neufeld, T. Oh, I. Perminov *et al.*, Entanglement in a quantum annealing processor, Phys. Rev. X **4**, 021041 (2014).

[42] T. Albash, W. Vinci, A. Mishra, P. A. Warburton, and D. A. Lidar, Consistency tests of classical and quantum models for a quantum annealer, Phys. Rev. A **91**, 042314 (2015).

[43] T. Albash and D. A. Lidar, Adiabatic quantum computing, Rev. Mod. Phys. **90**, 015002 (2018).

[44] S. Boixo, V. N. Smelyanskiy, A. Shabani, S. V. Isakov, M. Dykman, V. S. Denchev, M. H. Amin, A. Y. Smirnov, M. Mohseni, and H. Neven, Computational multiqubit tunnelling in programmable quantum annealers, Nat. Commun. **7**, 10327 (2016).

[45] N. Chancellor, S. Szoke, W. Vinci, G. Aeppli, and P. A. Warburton, Maximum–entropy inference with a programmable annealer, Sci. Rep. **6**, 22318 (2016).

[46] M. Benedetti, J. Realpe-Gómez, R. Biswas, and A. Perdomo-Ortiz, Estimation of effective temperatures in quantum annealers for sampling applications: A case study with possible applications in deep learning, Phys. Rev. A **94**, 022308 (2016).

[47] S. Muthukrishnan, T. Albash, and D. A. Lidar, Tunneling and speedup in quantum optimization for permutation-symmetric problems, Phys. Rev. X **6**, 031010 (2016).

[48] A. Cervera-Lierta, Exact Ising model simulation on a quantum computer, Quantum **2**, 114 (2018).

[49] T. Lanting, The d-wave 2000q processor, presented at Adiabatic Quantum Computing Conference 2017, June 26-29, 2017, Tokyo, Japan.

[50] N. Dattani, S. Szalay, and N. Chancellor, Pegasus: The second connectivity graph for large-scale quantum annealing hardware, arXiv:1901.07636.

[51] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser, Quantum computation by adiabatic evolution, arXiv:quant-ph/0001106.

[52] H. Neven, V. S. Denchev, M. Drew-Brook, J. Zhang, W. G. Macready, and G. Rose, Nips 2009 demonstration: Binary classification using hardware implementation of quantum annealing, Quantum 4, 1 (2009).

[53] B. Heim, T. F. Rønnow, S. Isakov, and M. Troyer, Quantum versus classical annealing of Ising spin glasses, Science **348**, 215 (2014).

[54] R. Harris, Y. Sato, A. J. Berkley, M. Reis, F. Altomare, M. H. Amin, K. Boothby, P. Bunyk, C. Deng, C. Enderud, S. Huang, E. Hoskinson, M. W. Johnson, E. Ladizinsky, N. Ladizinsky, T. Lanting, R. Li, T. Medina, R. Molavi, R. Neufeld *et al.*, Phase transitions in a programmable quantum spin glass simulator, Science **361**, 162 (2018).

[55] A. D. King, J. Carrasquilla, J. Raymond, I. Ozfidan, E. Andriyash, A. Berkley, M. Reis, T. Lanting, R. Harris, F. Altomare, K. Boothby, P. I. Bunyk, C. Enderud, A. Fréchette, E. Hoskinson, N. Ladizinsky, T. Oh, G. Poulin-Lamarre, C. Rich, Y. Sato *et al.*, Observation of topological phenomena in a programmable lattice of 1,800 qubits, Nature (London) **560**, 456 (2018).

[56] A. D. King, J. Raymond, T. Lanting, S. V. Isakov, M. Mohseni, G. Poulin-Lamarre, S. Ejtemaee, W. Bernoudy, I. Ozfidan, A. Y. Smirnov, M. Reis, F. Altomare, M. Babcock, C. Baron, A. J. Berkley, K. Boothby, P. I. Bunyk, H. Christiani, C. Enderud, B. Evert *et al.*, Scaling advantage in quantum simulation of geometrically frustrated magnets, arXiv:1911.03446.

[57] T. Albash and J. Marshall, Comparing relaxation mechanisms in quantum and classical transverse-field annealing, Phys. Rev. Appl. **15**, 014029 (2021).

[58] H. Oshiyama, S. Suzuki, and N. Shibata, Classical Simulation and Theory of Quantum Annealing in a Thermal Environment, Phys. Rev. Lett. **128**, 170502 (2022).

[59] S. Abel and M. Spannowsky, Observing the fate of the false vacuum with a quantum laboratory, PRX Quantum **2**, 010349 (2021).

[60] A. Andreassen, D. Farhi, W. Frost, and M. D. Schwartz, Precision decay rate calculations in quantum field theory, Phys. Rev. D **95**, 085011 (2017).

[61] M. Schmitt, M. M. Rams, J. Dziarmaga, M. Heyl, and W. H. Zurek, Quantum phase transition dynamics in the two-dimensional transverse-field Ising model, arXiv:2106.09046.

[62] R. Ayanzadeh, M. Halem, and T. Finin, Reinforcement quantum annealing: A quantum-assisted learning automata approach, arXiv:2001.00234.

[63] N. Chancellor, Domain wall encoding of discrete variables for quantum annealing and QAOA, Quantum Sci. Technol. **4**, 045004 (2019).