

**Effectiveness of quantum annealing for continuous-variable optimization**Shunta Arai <sup>1,\*</sup>, Hiroki Oshiyama <sup>2</sup>, and Hidetoshi Nishimori <sup>1,2,3</sup><sup>1</sup>*International Research Frontiers Initiative, Tokyo Institute of Technology, Shibaura, Minato-ku, Tokyo 108-0023, Japan*<sup>2</sup>*Graduate School of Information Sciences, Tohoku University, Sendai 980-8579, Japan*<sup>3</sup>*RIKEN, Interdisciplinary Theoretical and Mathematical Sciences, Wako, Saitama 351-0198, Japan*

(Received 11 May 2023; accepted 28 August 2023; published 2 October 2023)

The application of quantum annealing to the optimization of continuous-variable functions is a relatively unexplored area of research. We test the performance of quantum annealing applied to a one-dimensional continuous-variable function with a rugged energy landscape. After domain-wall encoding to map a continuous variable to discrete Ising variables, we first benchmark the performance of the real hardware, the D-Wave 2000Q, against several state-of-the-art classical optimization algorithms designed for continuous-variable problems to find that the D-Wave 2000Q matches the classical algorithms in a limited domain of computation time. Beyond this domain, classical global optimization algorithms outperform the quantum device. Next we examine several optimization algorithms that are applicable to the Ising formulation of the problem, such as the time-evolving block decimation (TEBD) to simulate ideal coherent quantum annealing, simulated annealing, simulated quantum annealing, and spin-vector Monte Carlo. The data show that TEBD's coherent quantum annealing achieves far better results than the other approaches, demonstrating the effectiveness of coherent tunneling. From these two types of benchmarks, we conclude that the hardware realization of quantum annealing has the potential to significantly outperform the best classical algorithms if thermal noise and other imperfections are sufficiently suppressed and the device operates coherently, as demonstrated in recent short-time quantum simulations.

DOI: [10.1103/PhysRevA.108.042403](https://doi.org/10.1103/PhysRevA.108.042403)**I. INTRODUCTION**

Quantum annealing (QA) is a metaheuristic designed to solve combinatorial optimization problems by controlling quantum fluctuations [1–11]. The hardware implementation of QA is provided by D-Wave Quantum Inc. [12–16] with more than 5000 qubits in the latest version [17].

Combinatorial optimization problems with discrete variables are the main target of QA, but the basic idea of using quantum fluctuations to search for the lowest-energy state in the rugged energy landscape is also applicable to problems with continuous variables. An example is the early paper by Finnila *et al.* [18], although they used an imaginary-time version of the Schrödinger equation without quantum coherence. Stella *et al.* [19] studied a few one-dimensional problems with and without rugged energy landscapes and compared QA with its classical counterpart of simulated annealing (SA) by directly solving the Schrödinger equation for QA and the Fokker-Planck equation for SA.

Those studies found that the relative efficiency of QA and SA depends on the shape of the energy landscape. Simulated annealing is influenced by the height of the energy

barrier, whereas QA reflects the structure of the energy spectrum, which affects the probability of quantum tunneling. An apparent difference in performance between SA and QA was clearly seen in a parabolic washboardlike potential with a single global minimum and many local minima as proposed by Shinomoto and Kabashima [20]. These authors used a coarse-grained phenomenological version of the problem to show that the best possible performance of SA can be achieved when the rate of temperature decrease is proportional to the inverse of logarithm, which results in a very slow logarithmic decrease of the residual energy as a function of the annealing time. Stella *et al.* discussed also QA phenomenologically for the same washboardlike potential by considering a coarse-grained tight-binding Hamiltonian under a semiclassical approximation, which consists of a harmonic potential and a Laplacian term with a coefficient representing a semiclassical Wentzel–Kramers–Brillouin (WKB) tunneling effect. They found that the residual energy shows a power-law decrease, which is faster than the logarithmic decrease in SA as found by Shinomoto and Kabashima. Inack and Pilati studied the performance of QA by the projective quantum Monte Carlo method [21] and found that this imaginary-time formulation behaves better than SA in the long-time regime. Koh and Nishimori recently analyzed a problem with a Shinomoto-Kabashima-like potential with many local minima [22], which is known as the Rastrigin function in the field of nonconvex continuous-optimization problems [23]. They solved the time-dependent Schrödinger equation directly numerically for QA and the master equation for SA and confirmed that the residual energy under QA decreases in a power law with an

\*arai.s.ao@m.titech.ac.jp

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

arbitrarily large power achievable by adjusting the annealing schedule, whereas SA has a smaller power. This is a simple but nontrivial example of a continuous-optimization problem, in which QA has an advantage over SA.

The studies mentioned above were based on theoretical or numerical approaches. Several experimental investigations on real hardware have recently been reported using domain-wall encoding [24] to represent continuous variables in terms of Ising spins. Examples include the work of Chancellor and co-workers [25,26], who observed quantum tunneling behavior from a local minimum to the global minimum in a double-well potential. In another paper, Abel *et al.* [27] applied the same method to two-dimensional continuous-optimization problems with many local minima using reverse annealing (RA) [28–31], which is a variant of QA starting from a classical spin configuration.<sup>1</sup> They graphically compared the output of QA on the D-Wave hardware with data from classical continuous-variable-optimization algorithms and found that the former reached the global minimum much more reliably than the latter. This is a very interesting development, but detailed analysis is still needed to quantitatively demonstrate a clear advantage of QA over state-of-the-art classical algorithms tailored for continuous-variable optimization.

In the present paper we systematically investigate whether QA has an advantage over classical algorithms for optimization of the Rastrigin function in one dimension, adopting domain-wall encoding to map a continuous variable to Ising spins. The first part of our test compares the hardware D-Wave 2000Q with several well-established classical algorithms for continuous-variable optimization: Nelder-Mead [32], conjugate gradient descent [33], basin hopping [34], and differential evolution [35]. We use the D-Wave 2000Q, not the latest D-Wave Advantage, because the former produces better results than the latter, as will be described later, in particular in Appendix B 1. The second part also runs the hardware, but we additionally test a classical algorithm to faithfully simulate quantum annealing, time-evolving block decimation (TEBD) [36,37] based on a matrix-product representation of the wave function, in order to clarify to what extent the hardware runs coherently. We also study several classical algorithms for combinatorial optimization, SA [38], simulated quantum annealing (SQA) [39], and spin-vector Monte Carlo (SVMC) [40] to investigate how classical algorithms are affected by the barrier height compared to the quantum case.

This paper is organized as follows. In Sec. II we explain the Rastrigin function and how to implement continuous-variable optimization with domain-wall encoding. In Sec. III we perform benchmark tests for the Rastrigin function on the D-Wave 2000Q and several classical algorithms developed for continuous-variable optimization. In Sec. IV we compare D-Wave 2000Q, TEBD, SA, SQA, and SVMC, all of which are for discrete-variable problems, to see how thermal effects influence the performance. Section V discusses the results and summarizes the paper. A few technical details are described in the Appendixes.

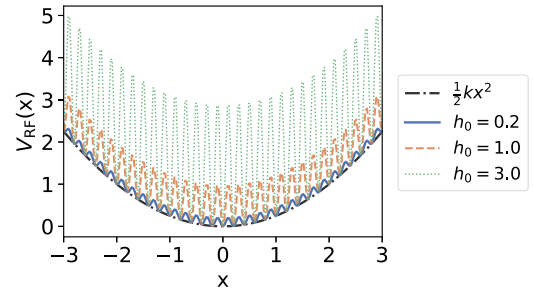


FIG. 1. Illustration of the Rastrigin function with  $k = 0.5$  and  $w_0 = 0.2$  for different values of  $h_0$ .

## II. PROBLEM AND METHODS

In this section we explain the mathematical definition of the Rastrigin function and how to optimize the continuous function with the domain-wall-encoding method on the D-Wave 2000Q.

### A. Rastrigin function

The Rastrigin function is one of the standard benchmark functions for nonconvex optimization problems [23]. We consider the one-dimensional Rastrigin function, which consists of a harmonic potential term with a unique global minimum at  $x = 0$  and a cosine term that gives many local minima as

$$V_{\text{RF}}(x) = \frac{1}{2}kx^2 + \frac{h_0}{2} \left[ 1 - \cos\left(\frac{2\pi x}{w_0}\right) \right], \quad (1)$$

where  $k$  is the spring constant and  $h_0$  and  $w_0$  are the control parameters of the height and width of each local minimum, respectively. The above mathematical formulation of the function was introduced by Koh and Nishimori [22] and they called it the Shinomoto-Kabashima-like potential. The original formulation of the Rastrigin function has no control parameters and corresponds to  $k = 2$ ,  $w_0 = 1$ , and  $h_0 = 20$  in Eq. (1). In this study we do not adopt the original formulation, but Koh and Nishimori's formulation to examine the dependence of algorithm performance on the shape of the energy landscape. We illustrate the one-dimensional Rastrigin function with  $k = 0.5$  and  $w_0 = 0.2$  for different values of  $h_0$  in Fig. 1. Due to the second cosine term, the simple gradient descent algorithm stops at a local minimum and fails to find the global minimum efficiently.

### B. Domain-wall encoding for continuous-function optimization

We briefly formulate QA and then describe the way to perform continuous-function optimization by quantum annealing with domain-wall encoding.

The generic Hamiltonian to be implemented by QA is

$$H(s) = \frac{B(s)}{2}H_0 - \frac{A(s)}{2}H_{\text{TF}}, \quad (2)$$

$$H_0 = \sum_{(i,j) \in E} J_{ij} \hat{\sigma}_i^z \hat{\sigma}_j^z + \sum_{i=1}^N h_i \hat{\sigma}_i^z, \quad (3)$$

$$H_{\text{TF}} = \sum_{i=1}^N \hat{\sigma}_i^x, \quad (4)$$

<sup>1</sup>In Ref. [27], RA is referred to as QA.

where  $\hat{\sigma}_i^z$  and  $\hat{\sigma}_i^x$  are the Pauli operators acting at site  $i$ ,  $J_{ij}$  is the coupling constant on the edge  $(i, j)$  in the graph  $E$  describing the connectivity between qubits,  $h_i$  is the local field at site  $i$ , and  $N$  represents the number of spins (qubits). The coefficients of the transverse field and the problem are denoted by  $A(s)$  and  $B(s)$ , respectively, controlled by the time-dependent annealing parameter  $s = t/t_a \in [0, 1]$ . The initial state is the ground state of Eq. (4) with the coefficient  $-A(0)/2$ , and the system evolves from  $t = 0$  to  $t = t_a$  following the Schrödinger equation with the coefficients evolving as  $A(0) > 0 \rightarrow A(1) = 0$  and  $B(0) \approx 0 \rightarrow B(1) > 0$ . The precise forms of the scheduling functions of  $A(s)$  and  $B(s)$  depend on the physical devices [41]. If the system evolves adiabatically, the final state of Eq. (2) approaches the ground state of Eq. (3).

Following Refs. [24–26], we use the domain-wall-encoding method to rewrite a continuous-variable-optimization problem to a corresponding problem with Ising spins. In this method, we map the value of a continuous variable to the position of a domain wall within a spin chain. Concretely, we discretize a continuous variable  $x \in [x_{\min}, x_{\max}]$  into  $N - 1$  values as

$$x_j = x_{\min} + (j - 1)\Delta x \quad (j = 1, \dots, N - 1), \quad (5)$$

with the step size

$$\Delta x = \frac{x_{\max} - x_{\min}}{N - 1}, \quad (6)$$

and map each value  $x_j$  to an  $N$ -spin state  $|x_j\rangle$ , which has a domain wall at the  $j$ th bond: All spins at sites 1 to  $j$  are down and the spins from  $j + 1$  to  $N$  are up. Using the fact that  $\langle x_j | (\hat{\sigma}_{i+1} - \hat{\sigma}_i)/2 | x_j \rangle = \delta_{ij}$  under this spin configuration, where  $\delta_{ij}$  is the Kronecker delta, we can construct an  $N$ -spin Hamiltonian  $H_{\text{RF}}$  such that its diagonal element in the basis  $|x_j\rangle$  corresponds to  $V_{\text{RF}}(x_j)$  as

$$H_{\text{RF}} = \frac{1}{2} \sum_{i=1}^{N-1} V_{\text{RF}}(x_i) (\hat{\sigma}_{i+1}^z - \hat{\sigma}_i^z). \quad (7)$$

To enforce that the ground state is a single-domain-wall state, we use a penalty Hamiltonian that punishes multikink and no-kink states as

$$H_{\text{DW}} = -J \sum_{i=1}^{N-1} \hat{\sigma}_i^z \hat{\sigma}_{i+1}^z + h (\hat{\sigma}_1^z - \hat{\sigma}_N^z), \quad (8)$$

where the first term with  $J > 0$  is the ferromagnetic coupling to align neighboring spins and the second term with  $h > J$  is the magnetic field at both ends of the chain to fix the spin at the first site in the down state and the spin at the  $N$ th site in the up state. The degenerate ground states of  $H_{\text{DW}}$  are  $|x_j\rangle$  for  $j = 1, \dots, N - 1$  and the penalty energy of each additional kink is  $2J$ .

Combining these two Hamiltonians, the domain-wall-encoding Hamiltonian for optimizing  $V_{\text{RF}}(x)$  is given as

$$H_0 = \lambda H_{\text{RF}} + H_{\text{DW}}, \quad (9)$$

where  $\lambda$  is a parameter that controls the relative strength of the penalty term  $H_{\text{DW}}$ . For sufficiently small  $\Delta x$  and  $\lambda$ , the ground state of  $H_0$  corresponds to the global minimum of  $V_{\text{RF}}(x)$ . If

$\lambda$  is too large, the ground state of  $H_0$  is no longer a single-domain-wall state.

Assuming that the potential function is differentiable and  $\Delta x \ll 1$ , Eq. (7) can be reduced to

$$\begin{aligned} H_{\text{RF}} &= -\frac{1}{2} \left( \Delta x \sum_{i=2}^{N-1} \frac{V_{\text{RF}}(x_i) - V_{\text{RF}}(x_{i-1})}{\Delta x} \hat{\sigma}_i^z \right. \\ &\quad \left. + V_{\text{RF}}(x_1) \hat{\sigma}_1^z - V_{\text{RF}}(x_{N-1}) \hat{\sigma}_{N-1}^z \right) \\ &\approx -\frac{1}{2} \left( \Delta x \sum_{i=2}^{N-1} \frac{dV_{\text{RF}}(x)}{dx} \Big|_{x=x_i} \hat{\sigma}_i^z \right. \\ &\quad \left. + V_{\text{RF}}(x_1) \hat{\sigma}_1^z - V_{\text{RF}}(x_{N-1}) \hat{\sigma}_{N-1}^z \right) \\ &= \sum_{i=1}^N h_i \hat{\sigma}_i^z, \quad (10) \\ h_i &= \begin{cases} -\frac{V_{\text{RF}}(x_1)}{2}, & i = 1 \\ -\frac{\Delta x}{2} \frac{dV_{\text{RF}}(x)}{dx} \Big|_{x=x_i}, & i = 2, \dots, N - 1 \\ \frac{V_{\text{RF}}(x_{N-1})}{2}, & i = N. \end{cases} \quad (11) \end{aligned}$$

The above construction of the Ising Hamiltonian for continuous-function optimization can be easily generalized to higher-dimensional problems [26].

### III. COMPARISON OF QA WITH CLASSICAL ALGORITHMS FOR CONTINUOUS OPTIMIZATION

In this section we show the results of benchmarks for QA and several classical algorithms designed for the optimization of continuous-variable functions to investigate how QA applied to continuous-variable optimization compares with dedicated classical algorithms. In the present section QA means the hardware implementation, i.e., the D-Wave 2000Q.

#### A. Optimization algorithms

We run QA and the four classical algorithms for the optimization of continuous-variable functions: Nelder-Mead (NM) [32], conjugate gradient descent (CGD) [33], basin hopping (BH) [34], and differential evolution (DE) [35], which are easily performed on the classical computer and are commonly used for the optimization of continuous functions with a complex potential landscape. While the NM and CGD methods are for local optimization, BH and DE are for global optimization. The NM method searches for a minimum by updating the polyhedron with the observed functional value. The CGD method utilizes the gradient of the potential function with the adjustable step size to find the minimum. Basin hopping performs two steps iteratively: local optimization around the proposed point and perturbation to search for the neighborhood domain of the current solution. Differential evolution is a kind of evolutionary computation and updates a proposed point by following the evolutionary procedure. Details of the classical algorithms are explained in Appendix A. Domain-wall encoding is applied only to QA on the D-Wave 2000Q, and classical algorithms use the continuous variable directly.

We apply these algorithms to Eq. (1) using the SCIPY package [42]. All numerical simulations for the classical optimization algorithms are performed on a laptop with Apple M1 Max chip with 10 CPU cores, which is one of the fastest chips.<sup>2</sup> For QA, we optimize Eq. (9) with the D-Wave 2000Q, more precisely the DW\_2000Q\_6 device. We also run the same experiments on two types of the more recent D-Wave Advantage. Since DW\_2000Q\_6 yields the best results, we utilize it in the main text. The comparison of performance among a few models of D-Wave devices is described in Appendix B 1.

### B. Measured quantities

We compute two observables. The first is the ground-state probability

$$P_{\text{GS}} = \frac{\mathcal{N}_{\text{GS}}}{n_{\text{init}}}, \quad (12)$$

where  $\mathcal{N}_{\text{GS}}$  is the number of cases where the global minimum was obtained by each algorithm and  $n_{\text{init}}$  stands for the number of trials (the number of different initial conditions). The second is the absolute error in the energy,

$$E_{\text{abs}} = |\langle V_{\text{RF}}(x) \rangle - V_0|, \quad (13)$$

where the angular brackets represent the average over the different initial conditions and  $V_0 = 0$  is the minimum value of Eq. (1).

### C. Parameter settings

For the classical optimization algorithms, we set  $n_{\text{init}} = 10^3$  and fix the search space as  $x \in [-3, 3]$ .<sup>3</sup> For a fair comparison with QA, we regard the output  $x$  obtained by the classical optimization algorithms as the solution to Eq. (1) if the condition  $|x| < \Delta x$  is satisfied. We control the execution time  $t_{\text{ex}}$  by changing the maximum number of iterations of classical algorithms as  $t_{\text{max}} = 2^i$  with the integer  $i$  running from 0 to 7. The execution time can be computed as  $t_{\text{ex}} = \sum_{k=1}^{t_{\text{max}}} \tau_k$ , where  $\tau_k$  represents the  $k$ th iteration time.

For the NM and CGD methods, the default parameter settings in SCIPY are used. For BH, we apply the CGD method in the local optimization process. For DE, the number of populations is set to 10 and 120. Each algorithm stops on the way if it satisfies the stopping criterion as  $|x^t - x^{t-1}| < 10^{-7}$  ( $t = 1, \dots, t_{\text{max}}$ ).

Our experimental settings are carefully designed for a fair comparison of QA and the classical algorithms in execution time and performance. Although there may be room for further improvements in execution time and performance, our studies are expected to lead to a general understanding of the

features of QA for continuous-function optimization against classical optimization algorithms.

For QA, we set the system size  $N = 211$  and fix the coefficients as  $J = 1$ ,  $h = 2$ , and  $\lambda = 1$ . The value  $N = 211$ , which corresponds to  $\Delta x = 0.0286$  with  $-x_{\text{min}} = x_{\text{max}} = 3$ , is large enough to represent the Rastrigin function with sufficient precision in its hills and valleys. The gradient of the potential in Eq. (1) is equal to 0 at the global minimum  $x = 0$  and yields the vanishing of the magnetic field  $h_{j'} = 0$ , where  $j'$  represents the place of the zero magnetic field. The vanishing of the magnetic field according to Eq. (11) gives two degenerate ground states of Eq. (9) as  $\hat{\sigma}_{\pm}^z = (-1, \dots, -1, \hat{\sigma}_{j'}^z = \pm 1, 1, \dots, 1)$ . The ground state  $\hat{\sigma}_{-}^z$  yields  $x_{j'} = 0$  with Eq. (5). On the other hand, the global minimum of Eq. (1) cannot be recovered from the ground state  $\hat{\sigma}_{+}^z$ .<sup>4</sup> Since the problem is negligible when  $N$  is large enough, we regard the two degenerate ground states of Eq. (9) as equally qualified solutions of Eq. (1).

We operate the execution time of QA by varying the annealing time  $1 \mu\text{s} \leq t_a \leq 500 \mu\text{s}$ . The total execution time of QA is defined as  $t_{\text{ex}} = t_p + t_s$ , where  $t_p$  is the programming time and  $t_s$  is the sampling time, which includes annealing time and readout time. The  $t_{\text{ex}}$  can be obtained from `qpu_access_time` in the D-Wave Ocean SDK [41]. To compute the observables, we repeat the annealing process 20 times independently and sample  $n_{\text{reads}} = 10^3$  spin configurations for each run. To reduce control errors, we apply flux bias calibration and gauge transformation [46,47]. The flux bias is estimated before each run. A gauge is generated from a uniform distribution and fixed in each independent run. The ground-state probability is calculated from  $n_{\text{reads}} = 10^3$  spin configurations. The absolute error in the energy is computed from samples that meet the single-domain-wall condition.

### D. Result for a high-energy barrier

We first consider the problem with a high-energy barrier  $h_0 = 1.0$ . We show the dependence of the ground-state probability and the absolute error in the energy on the execution time in Fig. 2. Data points labeled D-Wave (mean) are the averaged result for 20 independent runs. Data with the maximum ground-state probability among all runs are labeled D-Wave (best). The error bars show the standard error computed by bootstrapping. Execution time  $t_{\text{ex}}$  is always expressed in seconds.

Figure 2(a) shows that around the region  $0.7 \text{ s} \leq t_{\text{ex}} \leq 0.9 \text{ s}$ , the ground-state probability obtained by D-Wave (best) can be larger than the other algorithms except for DE (popsize = 120). The data for the NM and CGD methods are not shown in the left-side range beyond each leftmost point because they fail to find the solution in this short-time region. Also, the data for the NM and CGD methods as well as for

<sup>2</sup>Extensive CPU benchmarks are performed in Geekbench [43]. The single-core score of the Apple M1 Max chip in Geekbench version 6 is higher than those of the Intel Xeon E5-1650 used in an old benchmark [44] and the Intel Xeon E5-2695 version 4 in a recent benchmark [45].

<sup>3</sup>In the original formulation of the Rastrigin function, the search space is defined in  $x \in [-5.12, 5.12]$  [23]. Taking into account the computational precision of QA, we reduce the search space.

<sup>4</sup>By substituting  $x_j = 0$  into the right-hand side of Eq. (5), the location of the zero magnetic field can be computed as  $j' = 106$  with the present numerical values,  $N = 211$  and  $x_{\text{max}} = -x_{\text{min}} = 3$ . For  $\hat{\sigma}_{106}^z = -1$ , the ground state  $x = 0$  can be recovered as  $x_{106} = 0$ . For  $\hat{\sigma}_{106}^z = +1$ , the domain wall is located between  $j = 105$  and  $j = 106$  and  $x_{105} = -\Delta x (\neq 0)$  results.



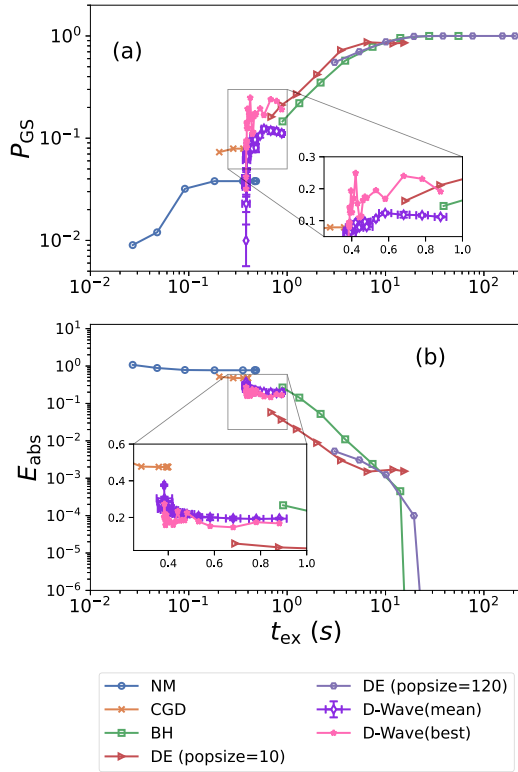


FIG. 2. Dependence of observables on the execution time for  $k = 0.5$ ,  $w_0 = 0.2$ , and  $h_0 = 1.0$ . The vertical axes denote the observables: (a) the probability of the ground state and (b) the absolute error in energy. The insets enlarge the region  $0.25 \text{ s} \leq t_{\text{ex}} \leq 1 \text{ s}$ .

QA are not shown to the right of their respective rightmost points since they show saturation and no improvements in these long-time regions.<sup>5</sup> On the other hand, BH and DE find the solution in the longer-time region  $10 \text{ s} \leq t_{\text{ex}}$  and their performance improves as the execution time increases. In the shorter-time region, the minimum execution time for BH and DE cannot be reduced further than shown in Fig. 2 with our computational resources.

Figure 2(b) shows that the absolute error in the energy obtained by QA is smaller than that of the NM and CGD methods and BH around the region  $0.4 \text{ s} \leq t_{\text{ex}} \leq 0.9 \text{ s}$ . Therefore, QA can find the local minima closer to the global minima than the NM and CGD methods and BH in this region. The outputs of BH and DE with large populations concentrate around the global minima as we increase the execution time, as we see in the right half of Fig. 2(b). If the number of populations is small, DE falls into local minima even in the long-execution-time region. Similarly to the ground-state probability, the absolute error in the energy of QA saturates at a finite value.

For the higher-energy barrier  $h_0 = 3.0$ , the qualitative behaviors of the classical algorithms are similar to those of

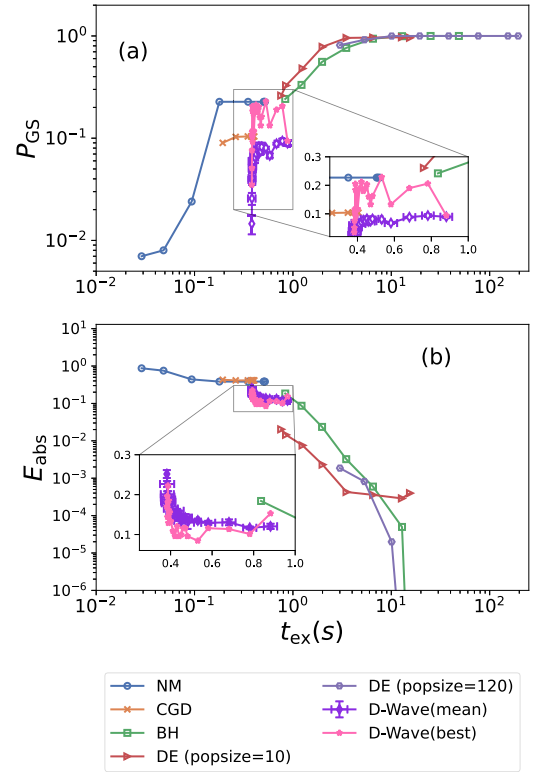


FIG. 3. Dependence of observables on the execution time for  $k = 0.5$  and  $w_0 = h_0 = 0.2$ . Both axes are the same as those in Fig. 2. The insets enlarge the region  $0.25 \text{ s} \leq t_{\text{ex}} \leq 1 \text{ s}$ .

$h_0 = 1.0$ . On the other hand, the performance of QA deteriorates significantly, as will be described in Sec. IV A.

### E. Result for a low-energy barrier

Next we consider the problem with a lower-energy barrier  $h_0 = 0.2$ . Figure 3(a) shows an overall improvement of classical algorithms compared to the case of a higher-energy barrier in Fig. 2(a). Nevertheless, the performance of QA deteriorates slightly. The best result from QA does not reach the level obtained by BH and DE, even with a minimum execution time of about 0.8 s. The saturation of the ground-state probability of QA for a longer execution time is observed beyond about 0.9 s as in Fig. 2(a) and is thus not drawn explicitly in Fig. 3(a).

Figure 3(b) demonstrates that the absolute error in the energy generally decreases compared to the case of a high-energy barrier in Fig. 2(b). The lowest value of the absolute error in the energy for the NM method is seen to be slightly lower than that of the CGD method, in contrast to the case of Fig. 2(b). This result implies that the one-dimensional structure with unique global minima fits the search strategy of the NM method. Except for this observation, the behaviors of the absolute error in the energy remain similar to those in Fig. 2(b).

In summary, the performance of QA is slightly better than those of local optimization algorithms and is comparable to those from global algorithms in a limited region of execution time for the potential function with a higher-energy barrier. On the other hand, if the energy barrier is low, we do not observe the clear benefits of QA over classical algorithms

<sup>5</sup>The NM and CGD methods get stuck in local minima before the number of iterations of the algorithms reaches the maximum value. Therefore, an increase in the maximum number of iterations does not affect the execution time and their rightmost points have the same values for the execution time.

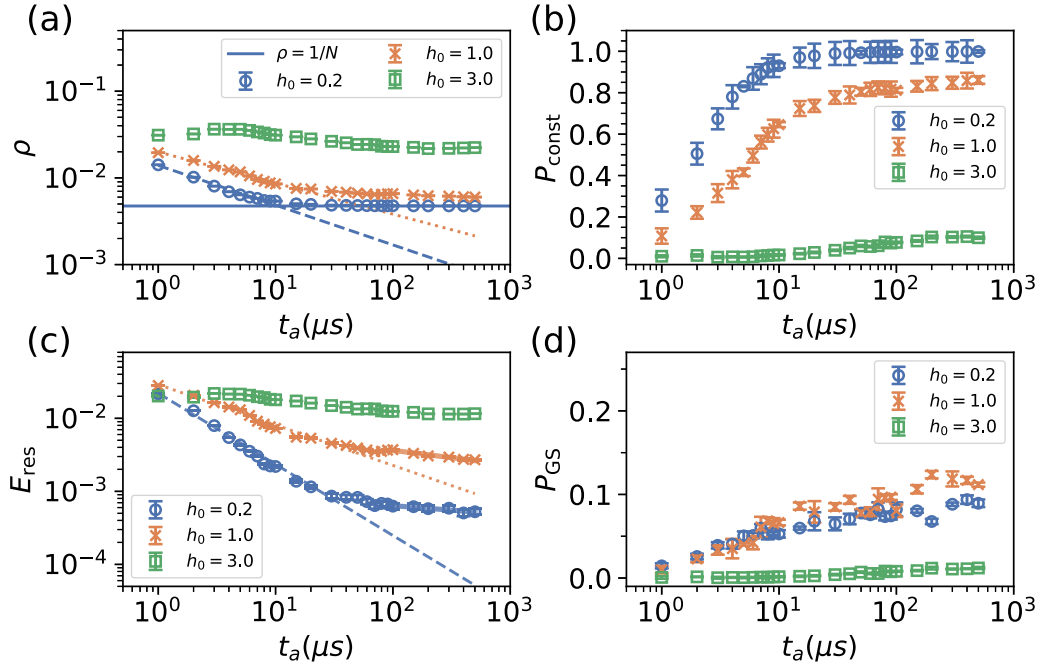


FIG. 4. Dependence of the observables obtained by the D-Wave on the annealing time for three values of  $h_0$ . The vertical axes are the observables: (a) the kink density, (b) the probability satisfying the single-domain-wall condition, (c) the residual energy, and (d) the ground-state probability. The horizontal axes represent the annealing time.

for the present problem of continuous-function optimization. Whereas increasing the execution time enhances the performance of the global optimization algorithms, QA shows saturation beyond intermediate timescales. In the next section we investigate whether this saturation originates in the problem in hardware implementation or for other reasons, including the mapping of the continuous variable to Ising spins by the domain-wall encoding.

#### IV. COMPARISON OF QA WITH ALGORITHMS FOR DISCRETE-VARIABLE OPTIMIZATION

In this section we compare the performance of the D-Wave 2000Q with several classical discrete-variable optimization algorithms, as well as with TEBD, a direct implementation of noise-free coherent QA on a classical computer, to examine how the D-Wave 2000Q compares with these other methods on the same footing of domain-wall coding. The results may help us estimate how the hardware would perform if the noise level were significantly reduced. The classical algorithms used in this section are SA, SQA, and SVMC. We carefully distinguish between coherent QA simulated by TEBD and hardware implementation of QA, the latter being represented by the product name. For the sake of simplicity, the D-Wave 2000Q is hereafter referred to as the D-Wave.

We compute four quantities. The first one is the kink density

$$\rho = \frac{1}{2N} \sum_{i=1}^{N-1} (1 - \langle \hat{\sigma}_i^z \hat{\sigma}_{i+1}^z \rangle). \quad (14)$$

The second is the probability  $P_{\text{const}}$  that the result satisfies the condition of a single domain wall and the third one is the

residual energy

$$E_{\text{res}} = \frac{\langle H_0 \rangle - E_0}{N}. \quad (15)$$

Here the angular brackets denote the sample average and  $E_0$  is the ground-state energy of Eq. (9). The last quantity is the ground-state probability  $P_{\text{GS}}$  that the ground state is successfully found. The kink density  $\rho$  represents the number of defects in spin chains. When a single domain wall exists,  $\rho = 1/N$  holds. Unlike the absolute error  $E_{\text{abs}}$  used in Figs. 2(b) and 3(b), the residual energy  $E_{\text{res}}$  is calculated from all samples, including the samples which do not satisfy the single-domain-wall condition. To understand the behavior of the entire system, we do not adopt  $E_{\text{abs}}$  but use  $E_{\text{res}}$  in this section.

Since the execution time of the D-Wave contains the measurement, initialization, and other auxiliary-processing times, we mainly use the pure annealing time as the  $x$  axis of each figure for fair comparison. The experimental setting of the D-Wave is the same as those in Figs. 2 and 3.

##### A. D-Wave

Figure 4 shows the dependence of the observables obtained from the D-Wave on the annealing time for  $h_0 = 0.2, 1.0,$  and  $3.0$ . The error bars represent the standard error obtained by bootstrapping and computed in  $2 \times 10^4$  samples for  $\rho$  and  $E_{\text{res}}$  and in 20 independent runs for  $P_{\text{const}}$  and  $P_{\text{GS}}$ . All error bars in the figures of this section are computed in this way. The data for  $h_0 = 0.2$  and  $1.0$  are the same as those used in Figs. 2 and 3.

TABLE I. Fitting exponents of the observables in the short-annealing-time region: (a) the kink density and (b) the residual energy.

$h_0$	D-Wave	Coherent QA	SA	SQA
		(a)		
0.2	$0.458 \pm 0.014$	$0.517 \pm 0.003$	$0.616 \pm 0.006$	$0.631 \pm 0.014$
1.0	$0.359 \pm 0.011$	$0.529 \pm 0.003$	$0.474 \pm 0.017$	$0.571 \pm 0.01$
3.0		$0.596 \pm 0.008$	$0.164 \pm 0.011$	$0.148 \pm 0.026$
		(b)		
0.2	$0.974 \pm 0.022$	$0.543 \pm 0.010$	$0.632 \pm 0.012$	$0.651 \pm 0.015$
1.0	$0.553 \pm 0.025$	$0.601 \pm 0.006$	$0.483 \pm 0.014$	$0.591 \pm 0.011$
3.0		$0.657 \pm 0.002$	$0.182 \pm 0.01$	$0.244 \pm 0.015$

### 1. Kink density

Let us first look at the kink density  $\rho$  in Fig. 4(a). We see that it needs some time for the data to reach  $\rho = 1/N$ , where the condition of a single domain wall is satisfied. In other words, this is the minimum amount of time it takes to find the correct solution on average. If the kink density behaves as  $\rho \propto t_a^{-b}$  below this threshold time, the annealing time required for the kink density to reach  $\rho = 1/N$  scales as  $t_a \propto N^{1/b}$ . Therefore, the value of the exponent  $b$  is important to understand how the performance of the domain-wall encoding depends on  $N$ , and a larger  $b$  is preferable. Except for the case of  $h_0 = 3.0$ , Fig. 4(a) demonstrates that the kink density gradually decreases as we increase the annealing time.

Each line is fitted with the polynomial function  $\rho \propto t_a^{-b}$  in the range  $1 \mu\text{s} \leq t_a \leq 10 \mu\text{s}$ . Extracted exponents are summarized in Table I. A smaller value of the exponent reflects the effect of the higher-energy barrier as seen in Table I(a). In addition, the higher-energy barrier gives a larger absolute value of  $\rho$ , not just a smaller exponent. For  $h_0 = 3.0$ , the kink density behaves nonmonotonically. Such behavior has also been observed in several experiments on the D-Wave devices and is believed to be caused by the open-system nature of the hardware [47,48].

Kink densities for  $h_0 = 1.0$  and  $3.0$  do not reach  $\rho = 1/N$  and saturate at finite values for long annealing time. Since the behaviors of all observables in Fig. 4 do not change even though we increase the annealing time beyond  $t_a = 500 \mu\text{s}$ , we only show the data in  $1 \mu\text{s} \leq t_a \leq 500 \mu\text{s}$ .

### 2. Single-domain-wall condition

Figure 4(b) illustrates that almost all samples satisfy the single-domain-wall condition for  $h_0 = 0.2$  beyond  $t_a = 10 \mu\text{s}$ . On the other hand, the probability of satisfying the single-domain-wall condition saturates around  $P_{\text{const}} \approx 0.8$  for  $h_0 = 1.0$  and around  $P_{\text{const}} \approx 0.1$  for  $h_0 = 3.0$ . For those cases with higher-energy barriers, the number of samples that satisfy the single-domain-wall condition deteriorates and an excessive number of domain walls appear.

### 3. Residual energy

Figure 4(c) indicates that the residual energy polynomially decreases as  $E_{\text{res}} \propto t_a^{-c}$  in the short-annealing-time region  $1 \mu\text{s} \leq t_a \leq 10 \mu\text{s}$  except for  $h_0 = 3.0$ . As with the kink density, the exponent of  $E_{\text{res}}$  becomes smaller for higher-energy barriers [see Table I(b)]. Higher-energy barriers yield larger

absolute values of the residual energy in addition to smaller exponents.

Next we investigate the behaviors of  $E_{\text{res}}$  in the long-annealing-time region  $t_a \geq 100 \mu\text{s}$ . The scaling analysis in this region gives us a hint at the performance of the D-Wave as an optimization solver. We omit the data for  $h_0 = 3.0$  because few samples satisfy the single-domain-wall condition. We fit the polynomial function and show the exponents in Table II. For the problem with  $h_0 = 1.0$ , the slope is slightly larger than in the case of  $h_0 = 0.2$ . This result indicates interestingly that the process is more efficient in the problem with the higher-energy barrier.

Although the residual energy has a polynomial decay as found by Koh and Nishimori [22], the exponent of  $E_{\text{res}}$  is not close their result  $E_{\text{res}} \sim t_a^{-2}$ . This discrepancy may come from the effects of noise on the hardware. In Fig. 4 we omit the data for  $t_a \geq 500 \mu\text{s}$  because the behavior of  $E_{\text{res}}$  hardly changes and the effect of noise increases in the longer annealing time. This result corresponds to the fact that in Figs. 2 and 3 the observables hardly show improvements if we increase the execution time beyond a certain value and the data are omitted over  $t_{\text{ex}} \geq 0.9$  s. Therefore, the execution time region  $t_{\text{ex}} \geq 0.9$  s corresponds to the annealing-time region  $t_a \geq 500 \mu\text{s}$ .

### 4. Ground-state probability

The ground-state probability for  $h_0 = 1.0$  is slightly larger than for  $h_0 = 0.2$  but remains small for  $h_0 = 3.0$ , as seen in Fig. 4(d). For  $h_0 = 0.2$  and  $1.0$ , the residual energy shows slightly decreasing behavior even in the long-annealing-time region  $t_a \geq 100 \mu\text{s}$ , but the ground-state probability saturates around  $P_{\text{GS}} \approx 0.1$ . This result indicates that the transition to the ground states hardly appears, whereas the transition to the neighborhoods of the ground states occurs in this timescale. The ground-state probability for  $h_0 = 3.0$  is suppressed due to the existence of many defects with few samples satisfying the single-domain-wall condition as shown in Figs. 4(a) and 4(b).

TABLE II. Fitting exponents of the residual energy in the long-annealing-time region.

$h_0$	D-Wave	SA	SQA	SVMC
0.2	$0.124 \pm 0.03$	$0.307 \pm 0.025$	$0.66 \pm 0.044$	$1.043 \pm 0.059$
1.0	$0.201 \pm 0.02$	$0.291 \pm 0.031$	$0.786 \pm 0.032$	$1.722 \pm 0.169$
3.0			$0.751 \pm 0.009$	

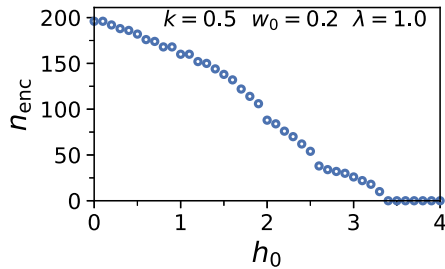


FIG. 5. Number of single-kink states encoded in the lowest-energy states of the classical Hamiltonian  $H_0$  as a function of the height of the energy barrier  $h_0$ . The parameters of the Rastrigin function are set to  $k = 0.5$  and  $w_0 = 0.2$ , and the weight of  $H_{\text{RF}}$  is set to  $\lambda = 1.0$ .

### 5. Kinks in low-energy states

To understand the reason for the significant performance deterioration observed for  $h_0 = 3.0$ , we examine the behavior of the low-energy states of the classical Hamiltonian  $H_0$  in Eq. (9). To evaluate how many single-kink states are properly encoded as low-energy states of  $H_0$ , we introduce  $n_{\text{enc}}$  defined as

$$n_{\text{enc}} = \min\{j \in \{0, 1, \dots\} \mid n_{\text{kink}}(\mathcal{S}_j) \neq 1\}, \quad (16)$$

where  $n_{\text{kink}}(\mathcal{S})$  denotes the number of kinks in the spin state  $\mathcal{S}$  and  $\mathcal{S}_j$  is the spin state corresponding to the  $j$ th energy level. The value of  $n_{\text{enc}}$  represents the index of the lowest-energy level for which the number of kinks in the corresponding spin state is not equal to 1. If  $n_{\text{enc}} > 0$ , the ground state is a single-kink state. If  $n_{\text{enc}} = 0$ , on the other hand, the ground state is a multikink state, indicating a failure of the encoding. The value of  $n_{\text{enc}}$  depends on the relative magnitude of  $\lambda H_{\text{RF}}$  and  $H_{\text{DW}}$ . If we set  $\lambda$  to be sufficiently small relative to  $J$ , we can make  $n_{\text{enc}} > 0$ .

Figure 5 shows the  $h_0$  dependence of  $n_{\text{enc}}$  obtained by dynamic programming [49]. As  $h_0$  increases, the values of the magnetic field  $h_i$  determined by Eq. (17) become larger, resulting in a smaller  $n_{\text{enc}}$ . For  $h_0 = 0.2$  and  $1.0$ , we have  $n_{\text{enc}} = 196$  and  $160$ , respectively, indicating that most of the single-kink states are encoded as low-energy states of  $H_0$ . For  $h_0 = 3.0$ , we have  $n_{\text{enc}} = 26$ , indicating that the lowest-energy states of  $H_0$  are contaminated with multikink states. The reason for the extremely low constraint satisfaction probability  $P_{\text{const}}$  for  $h_0 = 3.0$  in Fig. 4(b) is considered to be that the low-energy multikink states trap the annealing process. Reducing  $\lambda$  increases  $n_{\text{enc}}$ , which leads to an increase in  $P_{\text{const}}$ . However, reducing  $\lambda$  also increases the thermal effect and therefore the optimization performance may not necessarily improve. The dependence of the observables on  $\lambda$  is shown in Appendix B 2.

### 6. Summary

The values of observables obtained by the D-Wave depend on the height of the energy barrier. From the viewpoint as an optimization solver, achievement of the condition  $P_{\text{const}} = 1$  is important. For the problem with the higher-energy barrier, the probability of satisfying the single-domain-wall condition does not reach 1 in our experiments even if we increase the annealing time. The ground-state probability also saturates

at a finite value and does not improve any further. Since the decoherence effects by thermal fluctuations may be dominant in the long-annealing-time region, these effects hamper the transition to the ground states and degrade the performance. Although the residual energy  $E_{\text{res}}$  shows a polynomial decay in this region, the exponents are different from the direct simulation of the Schrödinger dynamics [22] due to the effects of thermal noise.

### B. TEBD for coherent QA

In order to understand the origin of the suboptimal performance of the D-Wave described in the preceding section, we now present the results of coherent QA based on TEBD to elucidate the differences between the D-Wave and the ideal QA that follows the Schrödinger dynamics without thermal noise. The TEBD algorithm is based on the Suzuki-Trotter decomposition [39] of the time-evolution operator and the matrix-product state (MPS) representation to approximately follow the time evolution of the wave function [50,51]. Time and energy are measured in units of  $J$  in Eq. (8) and we adopt the natural unit where  $\hbar = 1$ . In this unit, if we choose  $B(1)/2$  for the D-Wave as  $J$ , then  $t_a = 1$  corresponds to approximately 0.026 ns. The time-step size of the Suzuki-Trotter decomposition is set to  $\Delta t = 0.025/J$ . The maximum bond dimension of the MPS, which controls the accuracy of the approximation, is 32. We confirmed that the results converged well with respect to these parameters. The annealing schedules are the linear function, as  $A(s)/2 = 1 - s$  and  $B(s)/2 = s$ .

Figure 6 illustrates the results obtained by TEBD. A significant difference from the experimental results by the D-Wave in Fig. 4 is that the behavior of the kink density  $\rho$  exhibits only a very slight dependence on the height of the energy barrier  $h_0$ , which may originate in the weak dependence of coherent quantum tunneling effects on the barrier height. Moreover, in the case of coherent QA, for  $t_a \geq 10^3$ , the single-domain-wall condition is almost completely satisfied and the residual energy begins to decay quite rapidly, while in the case of the D-Wave in Fig. 4, the single-domain-wall condition is not satisfied even in the long-time regime and the residual energy does not decay. These results suggest that deviations from the ideal Schrödinger dynamics in the D-Wave lead to low performance, particularly for high  $h_0$ .

In Figs. 6(a) and 6(c) the kink density and the residual energy have the polynomial decay in the region  $t_a \leq 100$ ; the fitting exponents are summarized in Table I. In the limit where  $N$  is large and  $\lambda$  is small, the theory of the Kibble-Zurek mechanism [52–54] predicts that  $\rho \propto t_a^{-b}$  and  $E_{\text{res}} \propto t_a^{-c}$ , with  $b = c = 1/2$  for the one-dimensional transverse-field Ising model without longitudinal fields. In the simulation for Fig. 6,  $\lambda$  is set to 1 and therefore the magnitude of the longitudinal magnetic field is not negligible [see Eqs. (7) and (9)], but the exponents show a relatively mild dependence on  $h_0$ .

As shown in Figs. 6(b) and 6(d), for  $t_a > 100$ , the constraint satisfaction probability  $P_{\text{const}}$  starts to increase from 0 and the ground-state probability  $P_{\text{GS}}$  also begins to increase. In this region, the residual energy deviates from the polynomial behavior and decreases more rapidly.

In a longer-annealing-time region ( $t_a > 2 \times 10^3$ ), where the single-domain-wall condition is completely satisfied,  $P_{\text{GS}}$



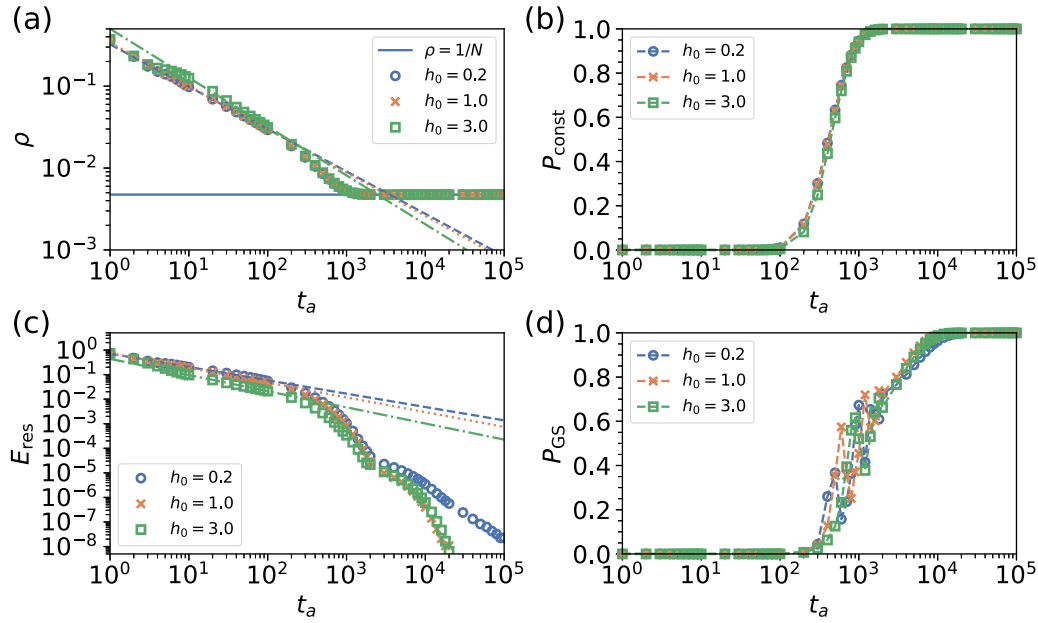


FIG. 6. Results of TEBD simulations for different values of  $h_0$  as functions of annealing time. The vertical axes are the same as those in Fig. 4. The annealing time  $t_a$  is in units of  $J$  and does not correspond to that of Fig. 4.

approaches 1. In this regime, the energy expectation value can be expressed as

$$\langle H_0 \rangle = E_0 P_{\text{GS}} + \mathcal{O}(1 - P_{\text{GS}}) \approx E_0 + \mathcal{O}(1 - P_{\text{GS}}) \quad (17)$$

and therefore the residual energy is proportional to  $1 - P_{\text{GS}}$  as

$$E_{\text{res}} = \frac{\langle H_0 \rangle - E_0}{N} = \mathcal{O}(1 - P_{\text{GS}}). \quad (18)$$

We can confirm this behavior by plotting the relation between  $E_{\text{res}}$  and  $1 - P_{\text{GS}}$  in Fig. 7(a), as the slope is approximately 1 in the region where  $1 - P_{\text{GS}} < 0.1$ . The annealing-time dependence of  $1 - P_{\text{GS}}$  is shown in Fig. 7(b), and in the region where  $1 - P_{\text{GS}} < 0.1$ , it can be seen that  $1 - P_{\text{GS}}$  follows an exponential decay as  $1 - P_{\text{GS}} \propto \exp(-Ct_a)$ , which is reminiscent of the Landau-Zener formula. Consequently, the annealing-time dependence of the residual energy can be expressed as  $E_{\text{res}} = C' \exp(-Ct_a)$ , where  $C'$  and  $C$  are

positive constants. The solid lines in Fig. 7(c) represent the result of fitting this expression to the data in the range of  $3 \times 10^3 \leq t_a \leq 10^4$ . In Ref. [22] an analysis of QA by direct numerical solution of the time-dependent Schrödinger equation for a single-particle system in a continuous space showed that for a linear annealing schedule, the residual energy decays as the inverse square of the annealing time beyond the Landau-Zener regime, as suggested in Ref. [55]. The deviation of the residual energy for  $h_0 = 0.2$  from the exponential behavior in the longer-time region  $t_a > 10^4$  may be an indication of a transition to such a power-law behavior, although the exponent is much larger than 2. To confirm the existence of this transition and to determine the power-law exponents, simulations for much longer annealing times are necessary, which is challenging due to the accumulation of numerical errors in TEBD. To understand the effect of the height of the energy barrier on the coherent QA, we analyze

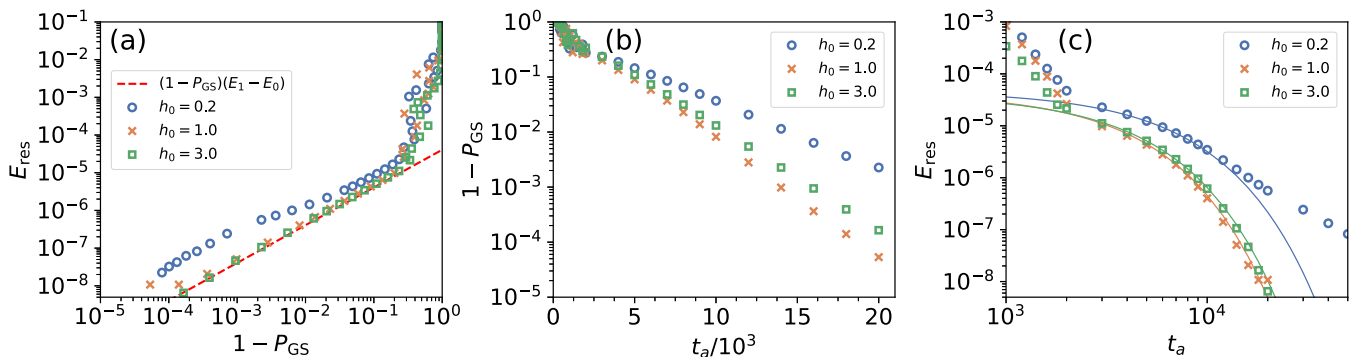


FIG. 7. Results of TEBD simulations. (a) Relation between the residual energy and the deviation of the ground-state probability from unity. The dashed line denotes  $(1 - P_{\text{GS}})(E_1 - E_0)$ , where  $E_1$  is the first excited energy of  $H_0$ . (b) Deviation of the ground-state probability from unity as a function of the annealing time. (c) Residual energy and its fitting curve as a function of the annealing time. The solid lines represent the curves of  $E_{\text{res}} = C' \exp(-Ct_a)$  fitted to the data in the range of  $3 \times 10^3 \leq t_a \leq 10^4$ .

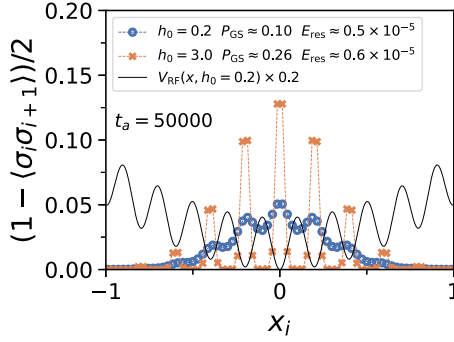


FIG. 8. Local number of kinks at  $t_a = 5 \times 10^4$  for  $\lambda = 0.01$  as a function of spatial coordinate  $x_i = x_{\min} + i\Delta x$ , where  $i$  is the site index. The solid line represents the potential function of Eq. (1) for  $h_0 = 0.2$ . The dashed lines are guides to the eye.

the spatial distribution of kinks in the wave function after the process of QA has terminated. Figure 8 shows the local number of kinks  $(1 - \langle \hat{\sigma}_i^z \hat{\sigma}_{i+1}^z \rangle) / 2$  at  $t_a = 5 \times 10^4$  as a function of spatial coordinate, as well as the values of the ground-state probability  $P_{\text{GS}}$  and residual energy  $E_{\text{res}}$ . For a lower-energy barrier  $h_0 = 0.2$ , kinks are distributed around the global minimum of the potential function at  $x = 0$  without significantly avoiding the maxima. This result suggests that for the low-energy barrier, the kink is more likely to move around, making it difficult to remain at the bottom of the potential minima. On the other hand, for a high-energy barrier  $h_0 = 3.0$ , kinks localize in the vicinity of the minima and hardly exist in the vicinity of the maxima due to the high-energy barrier. The increase in the population at the local (but not global) minima for  $h_0 = 3.0$  leads to larger residual energy than that of  $h_0 = 0.2$ . However, the probability of obtaining the ground-state probability  $P_{\text{GS}}$  is higher when  $h_0 = 3.0$ . This behavior suggests that a lower-energy barrier case is not necessarily easier to optimize, contrary to naive intuition. The characteristic of the kink spatial distribution discussed here becomes indistinct when  $\lambda$  is larger, but it can still be observed. Indeed, for  $\lambda = 1$ , the optimization performance does not decrease even in the case of a high-energy barrier, as shown in Fig. 6.

In summary, we have found that the optimization performance of coherent QA realized by TEBD is almost independent of the height of energy barrier  $h_0$  and the system converges to the ground state with high probability for sufficiently long annealing times. Therefore, the significant dependence of performance on  $h_0$  and the premature saturation of the ground-state probability in the data of the D-Wave can be attributed to imperfect realization of QA in the hardware, possibly mainly by thermal noise. As discussed in Appendix C, in order to fully understand the thermal effects on the hardware, it is necessary to analyze an open quantum system, not just an isolated system as done in this study. The performance of QA with a more coherent device would approach the results of TEBD simulations presented in this section and can be expected to be equivalent to or even better than that of global optimization algorithms such as BH and DE. Additional evidence will be provided in the following sections to show that the  $h_0$  dependence of data originates in thermal effects.

### C. Simulated annealing

Next we consider SA to investigate the effects of thermal fluctuations. In SA, we use a single-spin update with the Metropolis rule. The temperature is linearly decreased as

$$T(t) = T_0 + \frac{T_1 - T_0}{t_{\text{MCS}} - 1} t \quad (t = 0, \dots, t_{\text{MCS}} - 1), \quad (19)$$

with  $T_0 = 1$  and  $T_1 = 10^{-5}$ . The maximum number of Monte Carlo steps is denoted by  $t_{\text{MCS}}$ . The number of spin updates is  $t_{\text{MCS}}N$ . We repeat SA 20 times independently and sample  $10^3$  spin configurations for each independent run.

#### 1. Kink density

Figure 9 shows the results obtained by SA. As in Fig. 4(a), Fig. 9(a) indicates the power-law behaviors of the kink density in the short-annealing-time region  $2 \leq t_{\text{MCS}} \leq 100$ . Unlike the results obtained from the D-Wave in Fig. 4(a), the power-law behaviors can be seen even in the case of a high-energy barrier  $h_0 = 3.0$ . For  $h_0 = 0.2$ , the exponents are larger than those of the results obtained by the D-Wave and the coherent QA (see Table I). On the other hand, for larger  $h_0$  with a higher-energy barrier, the exponents are larger than those of the D-Wave and smaller than those of the coherent QA for the problem. In this time region, the values of the kink density become larger than those of the D-Wave and are similar to the case of coherent QA except for  $h_0 = 3.0$ . For  $h_0 = 3.0$ , the kink density does not reach  $\rho = 1/N$  even at  $t_{\text{MCS}} = 10^5$ . We need more annealing time to reduce the defects for the problem with the larger value of  $h_0$ .

#### 2. Single-domain-wall condition

Figure 9(b) illustrates that few samples satisfy the single-domain-wall condition in the short-time region where the kink density has the polynomial decay. While this result is similar to the result obtained by the coherent QA in the short-time region as seen in Fig. 6(b), a clear difference from the D-Wave in Fig. 4(b) is observed. Since the system is not sufficiently equilibrated in the short-annealing-time region, many defects remain. For the D-Wave result in Fig. 4(b),  $P_{\text{const}}$  has a finite value except for  $h_0 = 3.0$  in this region. Although this D-Wave result is similar to the behavior of SA in the intermediate-annealing-time region  $10^2 \leq t_{\text{MCS}} \leq 10^3$  in Fig. 9(b), it is impossible to reproduce all D-Wave data quantitatively by SA, for example, the residual energy, as shown below.

In the long-annealing-time region  $t_{\text{MCS}} \geq 10^4$ , almost all samples satisfy the single-domain-wall condition except for  $h_0 = 3.0$ . For  $h_0 = 3.0$ ,  $P_{\text{const}}$  is suppressed. Similarly to the D-Wave data in Fig. 4(b), SA is also affected by the height of the energy barrier.

#### 3. Residual energy

As with the kink density in Figs. 9(a) and 9(c) shows that the residual energy scales polynomially in the short-annealing-time region  $2 \leq t_{\text{MCS}} \leq 100$ . The exponents are smaller than the D-Wave results, as seen in Table I(b). The exponents are larger than those by the coherent QA for  $h_0 = 0.2$  and smaller for  $h_0 = 1.0$  and  $3.0$ . These results imply that

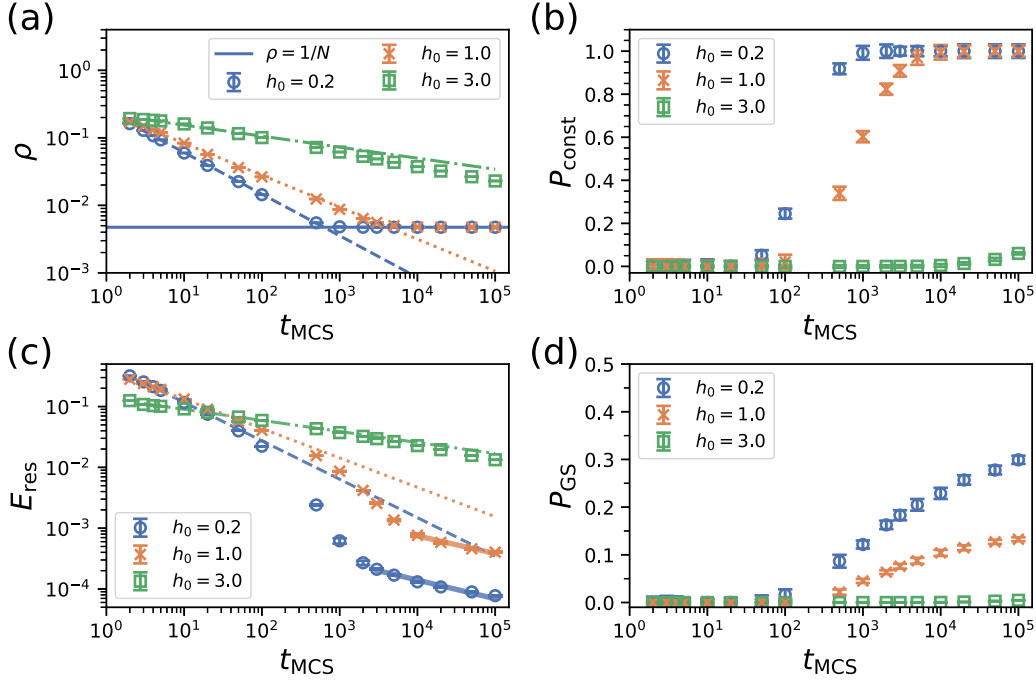


FIG. 9. Dependence of the observables obtained from SA on the annealing time for different values of  $h_0$ . The vertical axes are the same as those in Fig. 4. The horizontal axes represent the Monte Carlo step.

the dynamics based on thermal fluctuations is hampered by the higher-energy barrier, and quantum fluctuations are more efficient than thermal fluctuations for higher barriers.

The residual energy of SA decreases polynomially in the long-annealing-time regions  $2 \times 10^3 \leq t_{\text{MCS}} \leq 10^5$  for  $h_0 = 0.2$  and  $10^4 \leq t_{\text{MCS}} \leq 10^5$  for  $h_0 = 1.0$ , where  $P_{\text{const}} \approx 1$  holds. According to Table II, the slope of the fitting line becomes smaller for the problem with the higher-energy barrier. The higher-energy barrier impedes the transition to the ground states by thermal fluctuations. On the other hand, the exponents obtained by the D-Wave show the opposite trend. Note that  $P_{\text{const}} = 1$  does not hold for  $h_0 = 1.0$  in Fig. 4(b) for the D-Wave. Whether or not this is related to the quantumness of the D-Wave is not clear at this point. Although the exponents of SA are better than those of the D-Wave, the inverse square scaling of  $E_{\text{res}}$  is not observed.

#### 4. Ground-state probability

Figure 9(d) shows that the ground-state probability starts to increase monotonically from  $t_{\text{MCS}} = 10^2$ . The performance depends on the height of the energy barrier. The higher the energy barrier is, the smaller the success probability becomes. The behavior is different from that in Figs. 4(d) and 6(d). Since the ground-state probability is not close to 1 even in the long-annealing-time region, the exponential decay of  $E_{\text{res}}$  observed in the coherent QA in Fig. 7(c) does not appear in SA, as seen in Fig. 9(c).

#### 5. Summary

The observables of SA clearly depend on  $h_0$ , while the observables of the coherent QA hardly depend on  $h_0$ . This is a significant difference between SA and coherent QA. In coherent QA, quantum tunneling effects play an important role in

the state transition between local minima. On the other hand, SA shows deteriorated performance under higher-energy barriers. This behavior is consistent with the physical picture that SA goes over the energy barrier with thermal fluctuations. Also in the case of the D-Wave, dependence of the observables on  $h_0$  can be clearly seen. Premature saturation of  $P_{\text{const}}$  and  $P_{\text{GS}}$  at lower values appears in the long-annealing-time region for the D-Wave. Such saturation of these observables does not exist in SA. This suggests that the degraded performance of the D-Wave may not be due to thermal noise alone, but to other factors as well [56,57].

#### D. Simulated quantum annealing

Next we study SQA and see its difference from SA and other protocols in some detail. Simulated quantum annealing is the algorithm to simulate some aspects of QA on a classical computer. In SQA, we apply the Suzuki-Trotter decomposition [39] to the partition function of Eq. (2). The quantum system at a finite temperature is then mapped to the effective classical Ising model as

$$H_{\text{eff}} = \frac{B(s)}{2M} \sum_{k=1}^M H_0(k) + \frac{M}{2\beta} \ln \tanh \left( \frac{\beta A(s)}{2M} \right) \times \sum_{k=1}^M \sum_{i=1}^N \sigma_{ik} \sigma_{ik+1}, \quad (20)$$

$$H_0(k) = \lambda \sum_{i=1}^N h_i \sigma_{ik} - J \sum_{i=1}^{N-1} \sigma_{ik} \sigma_{i+1k} + h(\sigma_{1k} - \sigma_{Nk}), \quad (21)$$

where  $M$  is the Trotter number,  $\beta = 1/T$  is the inverse temperature, and  $\sigma_{ik} \in \{\pm 1\}$  ( $i = 1, \dots, N, k = 1, \dots, M$ ) is the

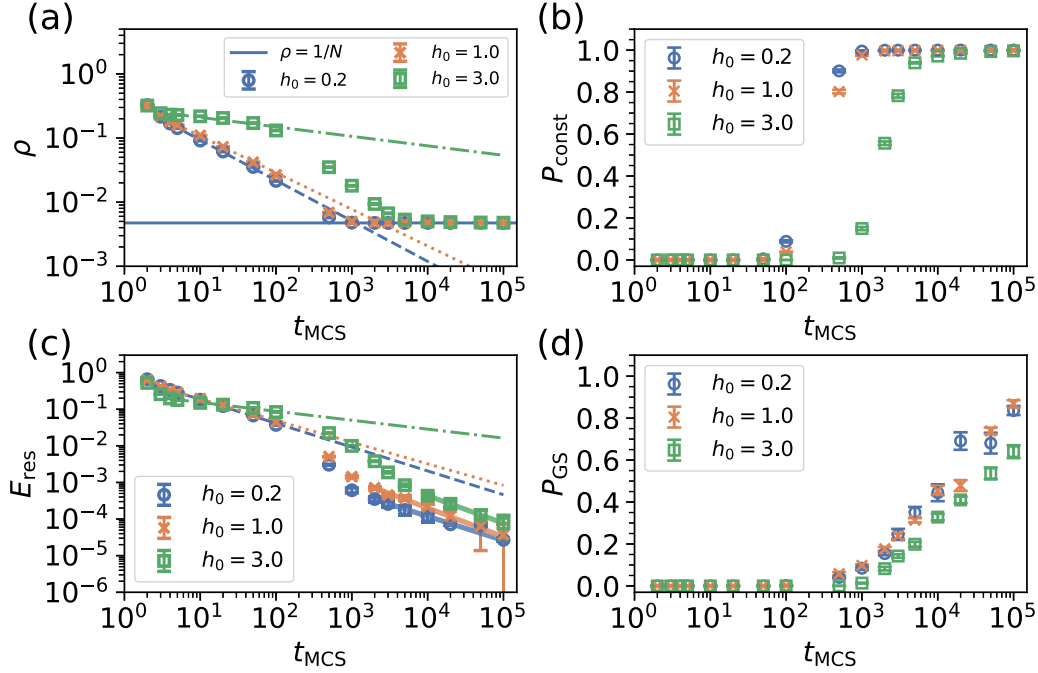


FIG. 10. Dependence of the observables obtained by SQA on the annealing time for different values of  $h_0$ . Both axes are the same as those in Fig. 4.

classical Ising configuration. The coefficients of the problem and transverse field are linearly controlled as

$$\frac{A(s)}{2} = 1 - s = 1 - \frac{t}{t_{\text{MCS}} - 1}, \quad (22)$$

$$\frac{B(s)}{2} = s = \frac{t}{t_{\text{MCS}} - 1}, \quad (23)$$

where  $t = 0, \dots, t_{\text{MCS}} - 1$ . We set the effective temperature  $\beta/M = 1$  and  $M = 10^3$ . We utilize a single-spin update with the Metropolis rule. We sweep the spin configurations  $MN$  times for each Monte Carlo step. We perform SQA 20 times independently. As shown in Ref. [58], the final spin configurations in Trotter replicas are utilized to compute the observables.

### 1. Kink density

Figure 10 shows the results obtained from SQA. In Fig. 10(a) the kink density does not depend on the value of  $h_0$  in the shortest-annealing-time region  $2 \leq t_{\text{MCS}} \leq 3$ . This behavior is similar to that of the coherent QA in Fig. 6(a). In the short-annealing-time region  $4 \leq t_{\text{MCS}} \leq 100$ , the polynomial decay of  $\rho$  can be seen. Except for  $h_0 = 3.0$ , SQA has the largest exponent in annealing protocols for discrete-value optimization, as shown in Table I. The exponents of  $\rho$  become smaller as the height of the energy barrier increases. The tendency is shared by other annealing protocols except for the coherent QA. For  $h_0 = 1.0$  and  $3.0$ , SQA requires less annealing time to satisfy the single-domain-wall condition as compared to SA, and thus classically simulated quantum fluctuations are more effective than thermal fluctuations in the higher barrier case. This result reflects the difference in the dynamics of both methods and can also be understood from the behaviors of  $P_{\text{const}}$  in Fig. 10(b). For  $h_0 = 0.2$ , we cannot

observe a clear difference between SA and SQA. Compared with the coherent QA, the dependence of  $\rho$  on  $h_0$  appears. The dynamics of SQA is fundamentally different from the dynamics of coherent QA [59] and may be regarded to be rather closer to the D-Wave behavior and SA.

### 2. Single-domain-wall condition

Figure 10(b) shows that few samples satisfy the single-domain-wall condition in the short-time region, similarly to the cases of SA and coherent QA in Figs. 9(b) and 6(b). The condition  $P_{\text{const}} \approx 1$  holds in the long-annealing-time region  $t_{\text{MCS}} \geq 10^4$  including the case of  $h_0 = 3.0$ . This behavior is different from the D-Wave and SA in Figs. 4 and 9. No suppression of  $P_{\text{const}}$  for  $h_0 = 3.0$  is observed for SQA, which is similar to coherent QA.

### 3. Residual energy

Figure 10(c) illustrates that the residual energy decreases polynomially in the short-annealing-time region. The range of fitting is the same as that in Fig. 10(a) to extract the values in Table I. As seen in this table, as the height of the energy barrier increases, the exponents of SQA decrease. The exponents of SQA are larger than those of SA. The result indicates that SQA can solve the problem more efficiently than SA. For  $h_0 = 3.0$ , the exponent of SQA is smaller than that of the coherent QA. Although SQA can simulate some aspects of quantum effects, the higher-energy barrier degrades the performance, unlike the coherent QA. In the long-annealing-time region  $10^3 \leq t_{\text{MCS}} \leq 10^4$ , the polynomial decay of  $E_{\text{res}}$  can be observed. As seen in Table II, SQA can solve the problem more efficiently than the D-Wave and SA since SQA gives larger values of the exponent than the D-Wave and SA. In our experiments, the exponential decay of  $E_{\text{res}}$  as in coherent



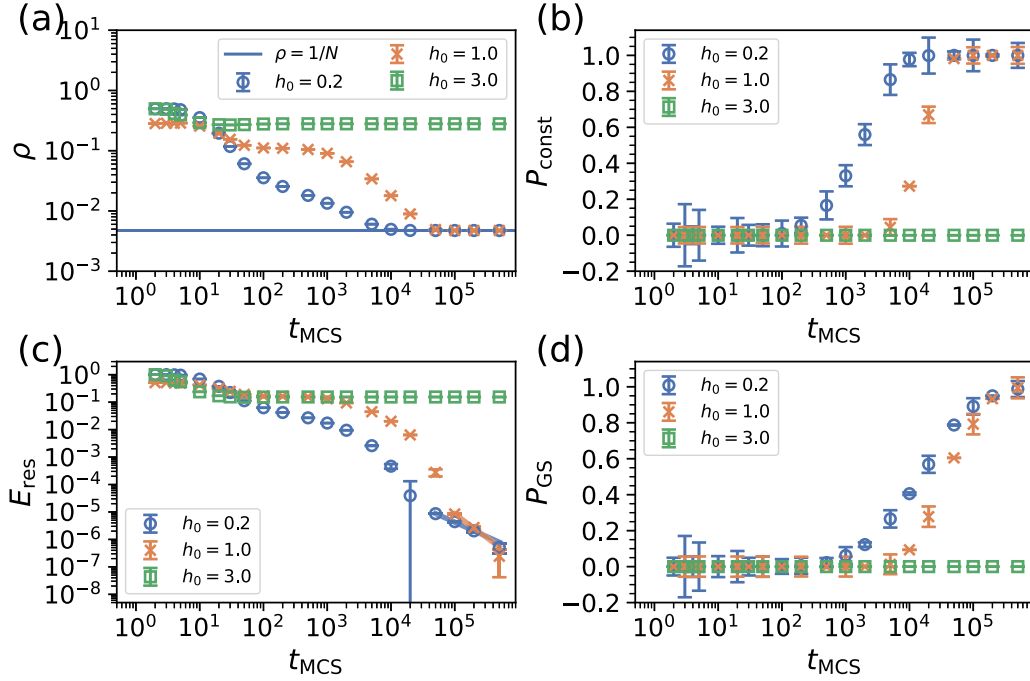


FIG. 11. Dependence of the observables attained by SVMC on the annealing time for different values of  $h_0$ . Both axes are the same as those in Fig. 4.

quantum annealing is not observed in SQA. To see the behavior, we may need to investigate longer annealing-time regions where  $P_{\text{GS}} \approx 1$  holds.

#### 4. Ground-state probability

Figure 10(d) shows that the success probability becomes smaller as the height of the energy barrier increases in the long-annealing-time region  $t_{\text{MCS}} \geq 10^4$ . Since the values of  $P_{\text{GS}}$  obtained by SQA are larger than those by SA in the region  $t_{\text{MCS}} \geq 2 \times 10^3$ , the effect of the energy barrier is weakened by classically simulated quantum fluctuations.

#### 5. Summary

The observables obtained by SQA depend on the value of  $h_0$ , which is different from the case of coherent QA. This attributes to the fact that the dynamics of SQA does not follow the Schrödinger dynamics but is based on classical Monte Carlo updates under thermal fluctuations. Compared with SA, SQA suffers less performance degradation under the higher-energy barrier, possibly by partly simulating tunneling effects [60].

#### E. Spin-vector Monte Carlo

Spin-vector Monte Carlo is a classical algorithm proposed to explain the outputs from the D-Wave device [40]. In SVMC, a quantum spin is replaced by a classical rotor of unit length with a continuous angle variable  $\theta_i \in [0, \pi]$  as

$$\hat{\sigma}_i^z \rightarrow \cos \theta_i, \quad \hat{\sigma}_i^x \rightarrow \sin \theta_i. \quad (24)$$

The total Hamiltonian can then be described as

$$H = \frac{B(s)}{2} H_0 - \frac{A(s)}{2} \sum_{i=1}^N \sin \theta_i, \quad (25)$$

$$H_0 = \lambda \sum_{i=1}^N h_i \cos \theta_i - J \sum_{i=1}^{N-1} \cos \theta_i \cos \theta_{i+1} + h(\cos \theta_1 - \cos \theta_N). \quad (26)$$

The annealing schedules of  $A(s)/2$  and  $B(s)/2$  are the same as those in SQA. The temperature in the Monte Carlo process is set in our simulation to  $T = 10^{-5}$  and the number of all spin updates is written as  $N t_{\text{MCS}}$ . Each continuous angle is initialized to  $\theta_i = \pi/2$  and spins are randomly chosen and updated with the Metropolis rule as  $\theta_i \rightarrow \theta'_i \in [0, \pi]$ . In addition, we apply the transverse-field-dependent (TFD) update [61,62], which is proposed to capture the freezing phenomena in the D-Wave device. In the TFD update, the current angle is updated as  $\theta_i \rightarrow \theta_i + \epsilon_i(s)$ , where  $\epsilon_i(s)$  is the random variable dependent on the annealing parameter and the energy coefficients,

$$\epsilon_i(s) \in \left[ -\min\left(1, \frac{A(s)}{B(s)}\right)\pi, \min\left(1, \frac{A(s)}{B(s)}\right)\pi \right]. \quad (27)$$

The range of updating the angle gradually decreases in the late annealing process  $A(s) < B(s)$ . We execute SVMC ten times independently and sample  $10^3$  spin configurations for each run.

#### 1. Kink density

Figure 11 shows the results of SVMC. In Fig. 11(a), nonlinear behaviors of  $\rho$  are observed. In the region of very short annealing time  $2 \leq t_{\text{MCS}} \leq 3$ , the kink density does not depend on the values of  $h_0$ . The dependence of  $\rho$  on the values of  $h_0$  appears for  $t_{\text{MCS}} \geq 3$ . For  $h_0 = 3.0$ , the kink density saturates at a larger value than that of SA in the region  $t_{\text{MCS}} \geq 50$ . This result represents that SVMC is more affected by the high-energy barrier than SA.

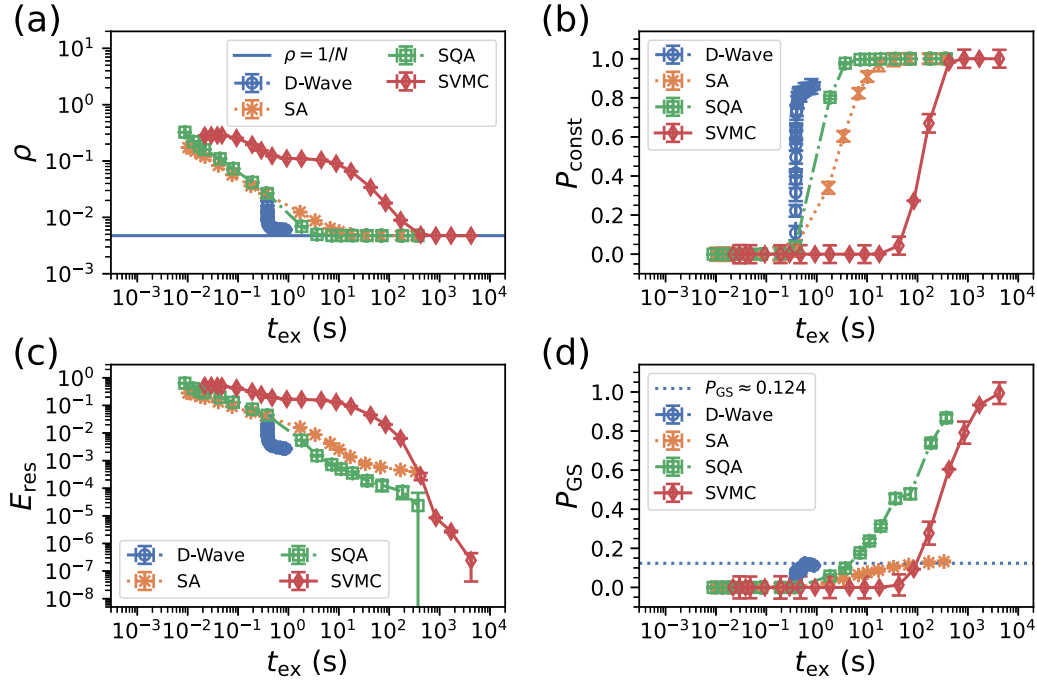


FIG. 12. Dependence of the observables on the execution time for different annealing protocols with  $k = 0.5$ ,  $w_0 = 0.2$ , and  $h_0 = 1.0$ . Both axes are the same as those in Fig. 4.

## 2. Single-domain-wall condition

Figure 11(b) shows that most samples satisfy the single-domain-wall condition  $t_{\text{MCS}} \geq 10^5$  for  $h_0 = 0.2$  and  $1.0$ . Compared with other annealing protocols, SVMC needs more annealing time to satisfy  $P_{\text{const}} \approx 1$ . The result indicates that the dynamics of SVMC is slower than that of other annealing protocols.

## 3. Residual energy

As in  $\rho$  in Figs. 11(a) and 11(c) illustrates that nonlinear behavior of  $E_{\text{res}}$  emerges and the saturation of  $E_{\text{res}}$  for  $h_0 = 3.0$  appears. In the long-annealing-time region, the polynomial decay of  $E_{\text{res}}$  can be observed. The fitting regions are  $5 \times 10^4 \leq t_{\text{MCS}} \leq 5 \times 10^5$  for  $h_0 = 0.2$  and  $10^5 \leq t_{\text{MCS}} \leq 5 \times 10^5$  for  $h_0 = 1.0$ . In these regions,  $P_{\text{const}} \approx 1$  holds, as seen in Fig. 11(b). As seen in Table II, the exponents of SVMC are the largest among all annealing protocols. Therefore, SVMC can reach the ground states faster than other annealing protocols in the long-time region. Compared with SQA in the region where the single-domain-wall condition holds, the values of  $E_{\text{res}}$  obtained by SVMC are smaller for  $h_0 = 0.2$  and  $1.0$ .

## 4. Ground-state probability

Figure 11(d) demonstrates that the ground-state probability starts to increase more slowly than SA and SQA. The higher-energy barrier suppresses the increase of  $P_{\text{GS}}$ . At a later stage, SVMC produces a larger value of  $P_{\text{GS}}$  than SQA does for  $h_0 = 0.2$  and  $1.0$  in the region where  $P_{\text{const}} \approx 1$  holds.

## 5. Summary

Values of observables in SVMC depend on the value of  $h_0$  as in SA and SQA. The dynamics of SVMC is more

affected by the higher-energy barrier. This is a consequence of the fact that the dynamics of SVMC is activated by thermal fluctuations in the semiclassical potential function. Although the behaviors of the observables are unique to SVMC and are different from those obtained by the D-Wave, the performance of SVMC is the best in the region where  $P_{\text{const}} \approx 1$  holds except for the coherent QA. This result suggests that the dynamics of SVMC is more effective than other classical annealing protocols for the Rastrigin function in the space where only a single domain wall exists.

Considering SVMC as a practical solver for continuous-function optimization, a smaller absolute value of the execution time is required, not just larger scaling exponents. In the next section we analyze the dependence of observables on the absolute execution time.

## F. Actual execution time

Finally, we show the dependence of observables on the actual execution time in seconds for different annealing protocols in Fig. 12. We use the same data as those presented throughout this section (Sec. IV) for  $k = 0.5$ ,  $w_0 = 0.2$ , and  $h_0 = 1.0$ . The definition of  $t_{\text{ex}}$  is the same as that defined in Sec. III B. Since the TEBD is the deterministic algorithm and different from the other annealing protocols based on the Monte Carlo method, we avoid showing the TEBD data for direct comparison in this section.

Figure 12(a) illustrates that SQA reaches  $\rho = 1/N$  most quickly among different annealing protocols.<sup>6</sup> Although the

<sup>6</sup>Note that the D-Wave becomes stuck at a value larger than  $\rho = 1/N$  as shown in Fig. 4(a), though this behavior is not clearly seen in the scale of Fig. 12(a).

D-Wave reaches the lowest value of  $\rho$  around  $0.4 \text{ s} \leq t_{\text{ex}} \leq 0.9 \text{ s}$ , its value of  $\rho$  remains suboptimal in the longer-time region due to the noise and other imperfections, as observed more clearly in Fig. 4(a).

Around the moderately-short-time region  $0.4 \text{ s} \leq t_{\text{ex}} \leq 0.9 \text{ s}$ ,  $P_{\text{const}}$  reaches close to 1 by the D-Wave, as shown in Fig. 12(b). If the goal is to satisfy the condition of a single domain wall as quickly as possible in the actual time of execution, the D-Wave is the choice.

Figure 12(c) shows that the D-Wave has the smallest value of  $E_{\text{res}}$  around  $0.4 \text{ s} \leq t_{\text{ex}} \leq 0.9 \text{ s}$ . Around  $2 \text{ s} \leq t_{\text{ex}} \leq 370 \text{ s}$ , SQA performs best. From the analysis of the exponents in Table II, SVMC surpasses SQA in the long-time region, and the difference between SQA and SVMC widens with time. To satisfy the single-domain-wall condition, SVMC needs more time [see Figs. 12(a) and 12(b)]. However, once the condition is satisfied, the residual energy decreases most rapidly with SVMC.

Figure 12(d) indicates that the success probability of the D-Wave has the largest value around  $0.4 \text{ s} \leq t_{\text{ex}} \leq 0.9 \text{ s}$ ,  $P_{\text{GS}} \approx 0.124$ , as indicated by the horizontal line. The D-Wave reaches this value most quickly among all annealing protocols. However, it shows saturation at around this finite value and the performance hardly improves with time. In Fig. 2 we saw that the ground-state probabilities obtained by BH and DE are large  $P_{\text{GS}} > 0.9$  even in the longest-time region  $10 \text{ s} \leq t_{\text{ex}} \leq 100 \text{ s}$ . As far as our implementation is concerned, other annealing protocols, i.e., SA, SQA, and SVMC, are similarly less effective than the global optimization algorithms for continuous-variable optimization since the former do not achieve such high values in this long-time region.

Overall, in the short-time window of  $0.4 \text{ s} \leq t_{\text{ex}} \leq 0.9 \text{ s}$ , the D-Wave is the best choice. At around the intermediate time  $t_{\text{ex}} \approx 2 \text{ s}$ , SQA is competitive. Beyond this time region, dedicated global optimization algorithms for continuous variables, i.e., DE and BH, keep improving with time, whereas classical metaheuristics for discrete optimization remain relatively mediocre. The D-Wave stays stagnant. It is nevertheless expected from the data of TEBD that the ideal QA without noise is likely to be most competitive, if realized at the hardware level, also in the intermediate- and long-time regions.

## V. CONCLUSION

We have performed extensive benchmarks of QA and classical optimization algorithms for the continuous-variable Rastrigin function in one dimension with a rugged energy landscape. By using the domain-wall encoding, the continuous-variable problem was represented by a one-dimensional Ising chain.

We first compared the D-Wave with classical algorithms for continuous-variable optimization with local or global updates and demonstrated quantitatively that, in the case of a higher-energy barrier, the performance of the D-Wave is better or comparable to that of the local optimization algorithms in a limited region of execution time and is comparable to the global optimization algorithms in the same time region. For longer execution times, the global algorithms show outstanding performance, whereas the D-Wave remains stagnant. For

the potential function with lower-energy barrier, no clear advantage of the D-Wave was observed. While the performance of the global optimization algorithms kept improving with increasing execution time, no improvements were identified in the D-Wave.

For a better understanding of the performance of QA, in its hardware realization of the D-Wave as well as in an ideal setting without noise, we next compared QA with protocols for discrete-variable optimization, i.e., TEBD for coherent QA, SA, SQA, and SVMC. The observed values of physical quantities show a clear dependence on the height of the energy barrier except for TEBD for coherent QA. This is natural for SA, SQA, and SVMC because these classical algorithms make explicit use of thermal noise to go over the barrier. In the case of the D-Wave, its barrier-height dependence implies that the process realized in the D-Wave is explicitly affected by thermal noise. Particularly in the case of the highest barrier with  $h_0 = 3.0$ , we found no sign of the successful solution by the D-Wave, whereas coherent QA by TEBD reached the ground state with high probability for relatively long annealing time with little to no dependence on the barrier height. This result suggests that the performance degradation of the D-Wave at long annealing time may be significantly reduced by improvements in the hardware to reduce noise and possibly other imperfections as demonstrated in recent experiments of quantum simulation [47,63]. Stated otherwise, quantum annealing has the potential to become quite competitive even for continuous-variable problems under the domain-wall encoding. This is achieved by quantum tunneling effects, as implied by the independence of the barrier height observed in the TEBD data. Classical algorithms, including those for continuous-variable optimization, would struggle to find a solution if the barrier were to become even higher than we tested, while QA, if realized coherently, would be much less affected by the height. This conclusion has become possible by the analysis of the continuous-variable problem with an explicit and easy control of the barrier height, which is generally difficult to achieve in problems expressed directly by the Ising model. In other words, we have demonstrated the effectiveness of coherent quantum tunneling effects explicitly and quantitatively in an optimization problem. Simulated quantum annealing may be regarded as a quantum-inspired classical algorithm that partly takes into account quantum effects. Nevertheless, the comparison of data for TEBD in Fig. 6 and SQA in Fig. 10 suggests that a complete realization of quantum effects in QA is necessary for better results.

This paper represents a step toward systematic and quantitative studies of continuous-variable optimization by QA in comparison with a series of well-established classical algorithms. In particular, the present one-dimensional problem with a periodic structure of the barrier is one of the simplest examples of continuous-variable optimization. The global optimization algorithm DE based on the genetic algorithm may find it more difficult to reach the correct solution for aperiodic potentials than for the periodic Rastrigin function. Also, higher-dimensional problems with the complex energy landscape as studied qualitatively in Refs. [25–27], which are generally difficult to analyze by TEBD, may be an interesting target of subsequent systematic studies.

## ACKNOWLEDGMENT

The work of H.N. was based on Project No. JPNP16007 commissioned by the New Energy and Industrial Technology Development Organization.

## APPENDIX A: CLASSICAL ALGORITHMS FOR CONTINUOUS-VARIABLE OPTIMIZATION

In this Appendix we explain in some detail the classical optimization algorithms used for benchmarking. We consider the optimization problem as

$$\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}), \quad (\text{A1})$$

where the objective function is defined by  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ . The implementation of all classical algorithms for benchmarking utilizes SCIPY [42].

### 1. Nelder-Mead method

The Nelder-Mead (NM) method is commonly used for unconstrained continuous-optimization problems [32,64]. The NM method does not need any information on derivatives and utilizes only the observed functional value. In the NM method, we hold a simplex constructed from the convex hull of  $n + 1$  points in  $\mathbb{R}^n$ . We iteratively update the simplex by removing the worst functional point and substituting another better point for it. The candidate is selected by the following process: reflecting, expanding, or contracting the simplex along the line connecting the worst point with the centroid of the other points [65]. If the solution cannot be improved anymore, the simplex is shrunk as all other points get closer to the current best solution. The detailed algorithm of the NM method is explained in Refs. [64,65].

### 2. Conjugate gradient descent method

The conjugate gradient decent (CGD) method is an iterative algorithm with the gradient of the objective function [65]. The original algorithm was designed to solve a linear system of equations [33]. The extended algorithm for nonlinear functions was proposed by Fletcher and Reeves [66]. In the CGD method, we set the initial point  $\mathbf{x}^{(1)} \in \mathbb{R}^n$  and its direction  $\mathbf{d}^{(1)} = -\nabla f(\mathbf{x}^{(1)})$  and update the current point  $\mathbf{x}^{(t)}$  as follows:  $\mathbf{x}^{(t+1)} \leftarrow \mathbf{x}^{(t)} + \alpha^{(t)} \mathbf{d}^{(t)}$ , where  $\alpha^{(t)}$  is the step length obtained via a line search for minimizing  $f(\mathbf{x}^{(t)})$  in the direction  $\mathbf{d}^{(t)}$ . The current direction is calculated by using the gradient as  $\mathbf{d}^{(t+1)} \leftarrow -\nabla f(\mathbf{x}^{(t+1)}) + \beta^{(t+1)} \mathbf{d}^{(t)}$ . In the Fletcher-Reeves formulations, the hyperparameter  $\beta^{(t+1)}$  is determined by  $\beta^{(t+1)} \leftarrow \nabla f(\mathbf{x}^{(t+1)})^T \nabla f(\mathbf{x}^{(t+1)}) / \nabla f(\mathbf{x}^{(t)})^T \nabla f(\mathbf{x}^{(t)})$ . We iterate the above procedure until the algorithm converges to a local minimum or the number of iterations reaches the maximum iteration.

### 3. Basin hopping method

The basin hopping (BH) method is a widely used global optimization algorithm for nonlinear multimodal optimization problems, where the local optimization algorithm often fails to find the global minimum due to the rugged energy landscape [34]. The BH method is combined with the gradient-based

local optimization and the perturbation based on Monte Carlo random walks with the Metropolis rule [67,68]. Starting from an initial point  $\mathbf{x}^{(1)}$ , we apply the local optimization to map a current point  $\mathbf{x}^{(t)}$  to the nearest minima  $\mathbf{y}^{(t)}$  in the local optimization process. Next we add a perturbation to the current solution as  $\mathbf{x}_{\text{new}} = \mathbf{y}^{(t)} + \boldsymbol{\epsilon}$ , where  $\boldsymbol{\epsilon}$  is a random coordinate generated from the uniform distribution  $U(a, b)$ . We perform the local optimization once again and attain another local minimum  $\mathbf{y}_{\text{new}}$ . Following the Metropolis rule, we accept the proposed point as  $\mathbf{y}^{(t+1)} = \mathbf{y}_{\text{new}}$  or reject it as  $\mathbf{y}^{(t+1)} = \mathbf{y}^{(t)}$  with the probability  $\min(1, \exp\{-[f(\mathbf{y}_{\text{new}}) - f(\mathbf{y}^{(t)})]/T\})$ , where  $T$  is the temperature to adjust the scale of the energy. The above iteration continues until  $t = t_{\text{max}}$ . In our experiments, we utilize the default setting  $a = -0.5$ ,  $b = 0.5$ , and  $T = 1$  and adopt the CGD method in the local optimization process.

## 4. Differential evolution method

The differential evolution (DE) method is a stochastic search algorithm for minimizing real-valued global optimization problems based on the evolutionary strategy [35,69]. In DE, we initialize a population vector for each generation  $G$  as  $\mathbf{x}_i^G \in \mathbb{R}^n$  ( $i = 1, \dots, p$ ;  $G = 0, \dots, G_{\text{max}}$ ), where  $p$  is the number of populations. The population vector is generated from the uniform distribution to cover the entire parameter space. Next we create a new population vector with different ones as  $\tilde{\mathbf{x}}_i^{G+1} = \mathbf{x}_{r_1}^G + F(\mathbf{x}_{r_2}^G - \mathbf{x}_{r_3}^G)$ , where  $r_1, r_2$ , and  $r_3$  are indices of the population determined by the crossover rule and  $F > 0$  is the amplification factor. The new trial vector  $\mathbf{u}_i^{G+1}$  can be obtained by mixing the target vector  $\mathbf{x}_i$  and the new population vector  $\tilde{\mathbf{x}}_i^{G+1}$  with the crossover rule [70]. If the new trial vector has a lower objective function value than the target vector, the new trial vector supersedes the target vector and is inherited by the next generation. The above process is called the selection. The algorithm stops at  $G = G_{\text{max}}$  or the mean objective function value over all populations becomes lower than the threshold  $\epsilon$ . In our experiments, we adopt the best-two-bin strategy [35,69] and set  $\epsilon = 10^{-7}$ . The number of populations  $G_{\text{max}}$  corresponds to the maximum number of iterations  $t_{\text{max}}$  in the main text. The other hyperparameters are used in the default settings.

## APPENDIX B: ADDITIONAL DATA FOR QA

In this Appendix we show some additional data from the D-Wave and consider the problem with  $k = 0.5$ ,  $w_0 = 0.2$ , and  $h_0 = 1.0$ . We compute the observables as illustrated in Sec. III B. The experimental setting of the D-Wave is almost the same as that in Fig. 4.

### 1. Comparison of D-Wave 2000Q and D-Wave Advantage

Figure 13 illustrates the dependence of observables on the annealing time for a few different quantum annealers. In addition to the DW\_2000Q\_6 device, we utilize the Advantage\_system 4.1, Advantage\_system 6.1, and Advantage2\_prototype 1.1 devices and compare their performance.

Regardless of the type of annealers, the polynomial decay of  $\rho$  can be observed in the short-annealing-time region  $1 \mu\text{s} \leq t_a \leq 10 \mu\text{s}$ , as shown in Fig. 13(a). The fitting



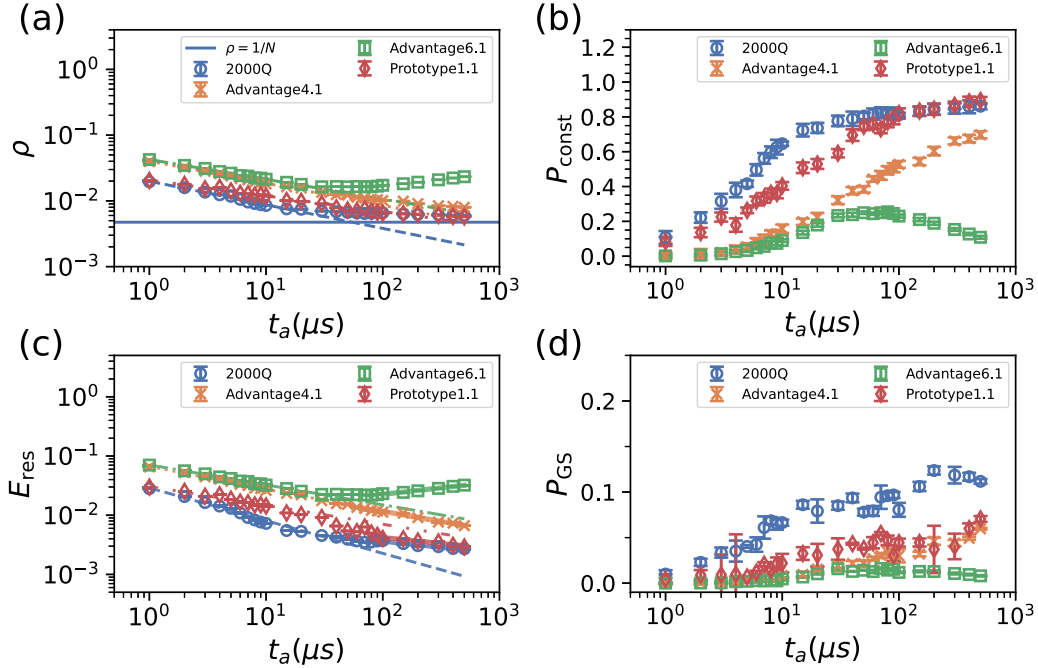


FIG. 13. Dependence of the observables on the annealing time for the different quantum annealers. Both axes are the same as those in Fig. 4.

exponents are summarized in Table III. The largest exponent of  $\rho$  is obtained from the D-Wave 2000Q. Except for the D-Wave Advantage 6.1, the kink density gradually decreases as we increase the annealing time. The kink density obtained by the D-Wave Advantage 6.1 shows behavior similar to that obtained by the D-Wave Advantage 4.1 in the region  $1 \mu\text{s} \leq t_a \leq 10 \mu\text{s}$  and starts to increase at  $t_a = 30 \mu\text{s}$ . This behavior may be due to decoherence. In the long-annealing-time region  $t_a \geq 100 \mu\text{s}$ , no device reaches the target  $\rho = 1/N$ . In this region, the D-Wave 2000Q takes values similar to those of the D-Wave Advantage2 Prototype 1.1.

Figure 13(b) illustrates that the D-Wave 2000Q has the largest  $P_{\text{const}}$  in the region  $1 \mu\text{s} \leq t_a \leq 100 \mu\text{s}$ . The D-Wave Advantage2 Prototype 1.1 takes values of  $P_{\text{const}}$  similar to those obtained by the D-Wave 2000Q in the region  $100 \mu\text{s} \leq t_a \leq 500 \mu\text{s}$ . In the long-annealing-time region, most samples obtained by these two devices satisfy the single-domain-wall condition. Both devices yield a larger value of  $P_{\text{const}}$  than the D-Wave Advantage 4.1 and 6.1.

Figure 13(c) shows that the residual energy decays polynomially irrespective of the types of quantum annealers in the short-annealing-time region, as also seen in Fig. 13(a). Table III shows that the largest exponent of  $E_{\text{res}}$  can be

obtained by the D-Wave 2000Q. In our experiments, the D-Wave 2000Q gives the lowest value of  $E_{\text{res}}$  among all devices. In the long-annealing-time region  $t_a \geq 100 \mu\text{s}$ , we can see a polynomial decay of  $E_{\text{res}}$ . The fitting exponents obtained by the D-Wave 2000Q and the D-Wave Advantage2 prototype 1.1 have values similar to those in Table III. The result indicates that the qualitative behaviors of the two devices show little differences. The D-Wave Advantage 4.1 has the largest exponent among all devices, from which one may expect Advantage 4.1 to yield smaller values of  $E_{\text{res}}$  than other models if the device would operate with minimal noise in the longer-annealing-time region.

Figure 13(d) demonstrates that the D-Wave 2000Q gives the largest value of  $P_{\text{GS}}$ . Although the D-Wave 2000Q and the D-Wave Advantage2 Prototype 1.1 give similar exponents in  $E_{\text{res}}$  in the long-annealing-time region from Table III, the performance gap appears in  $P_{\text{GS}}$ . This result implies that the D-Wave Advantage2 Prototype 1.1 may reach low-energy states, but not necessarily the ground states.

The observable values obtained by different quantum annealers show different behaviors. We generally conclude that the D-Wave 2000Q is suitable for optimizing the Rastrigin function for the present problem setting.

TABLE III. Fitting exponents of the observables for the different quantum annealers. The  $E_{\text{res}}$  (short) represents the residual energy in the short-annealing-time region  $1 \leq t_a \leq 10$ . The  $E_{\text{res}}$  (long) is the residual energy in the long-annealing-time region  $100 \leq t_a \leq 500$ .

Parameter \ Machine	2000Q	Advantage4.1	Advantage6.1	Prototype1.1
$\rho$	$0.359 \pm 0.011$	$0.337 \pm 0.008$	$0.301 \pm 0.005$	$0.239 \pm 0.015$
$E_{\text{res}}$ (short)	$0.553 \pm 0.025$	$0.383 \pm 0.012$	$0.338 \pm 0.006$	$0.319 \pm 0.021$
$E_{\text{res}}$ (long)	$0.201 \pm 0.02$	$0.341 \pm 0.036$	$-0.218 \pm 0.012$	$0.201 \pm 0.031$

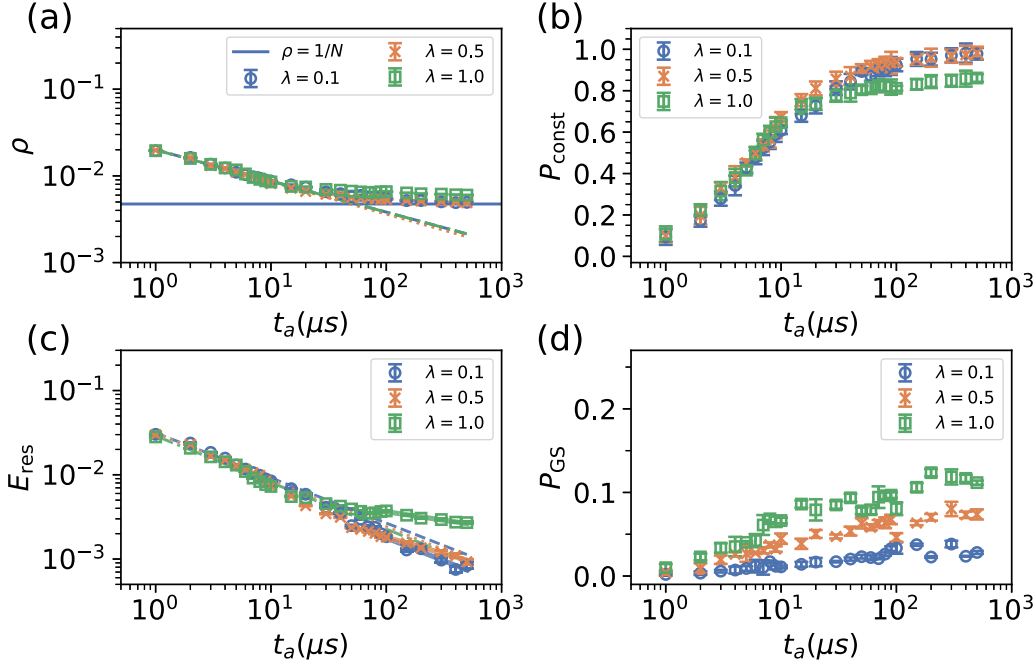


FIG. 14. Dependence of the observables attained by the D-Wave on the annealing time for different values of  $\lambda$ . Both axes are the same as those in Fig. 4.

**2. Dependence of the observables for different values of  $\lambda$**

Figure 14 shows the dependence of the observables on the annealing time for different values of  $\lambda$ . All data are obtained from the D-Wave 2000Q. As in the main text, we refer to the D-Wave 2000Q as the D-Wave. The data for  $\lambda = 1.0$  are the same as those used in Fig. 4. In Fig. 14(a) the polynomial decay of  $\rho$  can be seen in the short-annealing-time region  $1 \mu s \leq t_a \leq 10 \mu s$  for different values of  $\lambda$ . The fitting exponents are summarized in Table IV. In this region, the coefficient  $\lambda$  hardly affects the behavior of the system. On the other hand, in the long-annealing-time region  $t_a \geq 100 \mu s$ , the saturation seen in  $\rho$  for  $\lambda = 1.0$  does not occur in  $\lambda = 0.1$  and  $0.5$ . The small  $\lambda$  weakens the effect of the problem term in Eq. (10) and enhances the effect of the constraint term in Eq. (8).

Figure 14(b) demonstrates that the value of  $\lambda$  hardly affects  $P_{\text{const}}$  in the short-annealing-time region. Similarly to the kink density in Fig. 14(a), the effect of  $\lambda$  can be seen in the long-annealing-time region  $t_a \geq 100 \mu s$ . Premature finite saturation of  $P_{\text{const}}$  appears for  $\lambda = 1.0$ . The number of samples that satisfy the single-domain-wall condition increases by adjusting  $\lambda$ . In the long-annealing-time region, the effect of quantum fluctuation becomes weaker due to the decoherence. In this region, thermal fluctuations affect the system. Since small

values of  $\lambda$  largely reduce the height of the energy barrier and the lower-energy barrier is favorable for thermal fluctuations as shown in Fig. 9, the number of samples that satisfy the single-domain-wall condition increases with decreasing  $\lambda$  in this region.

Figure 14(c) shows that the polynomial decay of  $E_{\text{res}}$  can be observed in the short-annealing-time region  $1 \mu s \leq t_a \leq 10 \mu s$ . From Table IV it can be seen that the value of  $\lambda$  hardly affects the exponent of  $E_{\text{res}}$  as in the kink density. In the long-annealing-time region  $t_a \geq 100 \mu s$ , the residual energy decreases polynomially and the fitting exponents are given in Table IV(b). A smaller  $\lambda$  yields a larger exponent. This result implies that a smaller  $\lambda$  leads to lower-energy states and reduces the hardness of optimization due to the higher-energy barrier.

Figure 14(d) illustrates that larger values of  $\lambda$  give larger  $P_{\text{GS}}$ . This result is in contrast to the behavior of  $E_{\text{res}}$  in

TABLE IV. Fitting exponents of the observables for different values of  $\lambda$ . The fitting region is the same as that used in Table III.

$\lambda$	$\rho$	$E_{\text{res}}$ (short)	$E_{\text{res}}$ (long)
0.1	$0.359 \pm 0.011$	$0.536 \pm 0.026$	$0.545 \pm 0.073$
0.5	$0.372 \pm 0.011$	$0.553 \pm 0.024$	$0.392 \pm 0.019$
1.0	$0.359 \pm 0.011$	$0.553 \pm 0.025$	$0.201 \pm 0.02$

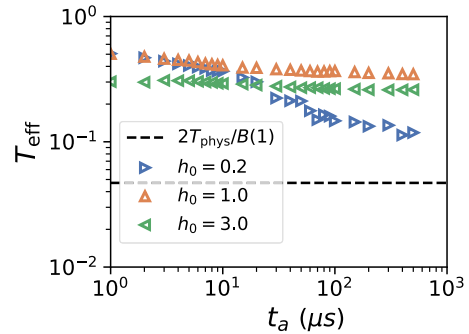


FIG. 15. Effective temperature obtained from the data of the D-Wave as a function of annealing time. The dashed horizontal line represents the physical temperature of the D-Wave.

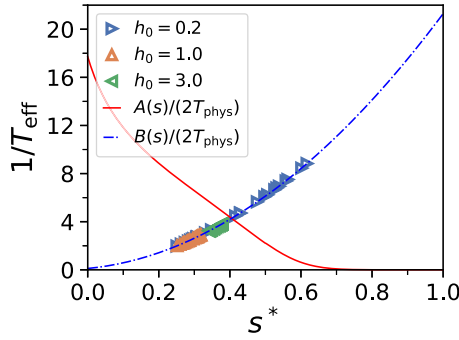


FIG. 16. Relation between the inverse effective temperature and freeze-out point. The lines represent the annealing schedule functions used to determine  $s^*$  from  $T_{\text{eff}}$ .

Fig. 14(c). Smaller values of  $\lambda$  decrease the values of the magnetic field physically encoded into the D-Wave. Since the bit width to represent the problem Hamiltonian is limited in the D-Wave, the actual implemented value in the D-Wave may be different. Due to this discretization error, the actual energy landscape slightly changes. Because the success probability is more affected by variation of the energy landscape than the residual energy, the success probability for the smaller  $\lambda$  is degraded.

### APPENDIX C: COMPARISON OF THE DATA OF THE D-WAVE 2000Q AND THERMAL EQUILIBRIUM STATE

The D-Wave is a device to reproduce the system behavior under the time-dependent Hamiltonian  $H(s)$  in Eq. (2). However, as the hardware is an open system, its dynamics is affected by thermal noise. To describe such dynamics, the hypothesis of quasistatic time evolution and freeze-out phenomenon has been proposed [71]. In this hypothesis, during the annealing process, the density matrix of the system at a given time is approximately equal to the density matrix of the thermal equilibrium state of the Hamiltonian at that time. The dynamics freezes at a time denoted by  $s^*$ , when the relaxation time to equilibrium becomes much larger than the timescale of annealing, and the density matrix of the final result becomes  $\exp[-H(s^*)/T_{\text{phys}}]$ , where  $T_{\text{phys}}$  denotes the

physical device temperature that is about 13.5 mK. If the strength of the transverse field at the freeze-out point,  $A(s^*)$ , is much smaller than the classical part of the Hamiltonian,  $B(s^*)$ , then the density matrix of the system can be approximated as

$$\rho_{\text{sys}}(s^*) \approx \frac{e^{-[B(s^*)/2]H_0/T_{\text{phys}}}}{\text{Tr}(e^{-[B(s^*)/2]H_0/T_{\text{phys}}})}. \quad (\text{C1})$$

This can be interpreted as the density matrix at the thermal equilibrium state of  $H_0$  with a temperature of  $2T_{\text{phys}}/B(s^*)$ . By determining this rescaled temperature from the output samples of the D-Wave, it is possible to estimate the freeze-out point  $s^*$ . It should be noted that if the obtained effective temperature value is too high, the estimation of  $s^*$  may become inaccurate and it would be appropriate to estimate the temperature using  $H(s^*)$  instead of  $H_0$ .

We now discuss the consistency between the experimental results of the D-Wave in this study and the hypothesis of quasistatic time evolution and freeze-out. The effective temperature  $T_{\text{eff}}$  is defined as the value obtained by solving the equation  $E(T) = \langle H_0 \rangle$  for  $T$ , where  $E(T)$  is the internal energy of the thermal equilibrium state of  $H_0$  at temperature  $T$  and  $\langle H_0 \rangle$  is the energy expectation value of the data of the D-Wave. The annealing-time dependence of the effective temperature from the experimental results of the D-Wave is shown in Fig. 15. For  $h_0 = 0.2$ , the effective temperature decreases toward  $T_{\text{phys}}$ , while for  $h_0 = 1.0$  and  $3.0$ , the effective temperature saturates at a constant value several times higher than  $2T_{\text{phys}}/B(1)$ . Regarding  $T_{\text{eff}}$  as  $2T_{\text{phys}}/B(s^*)$ , the freeze-out point  $s^*$  is estimated as the value obtained by solving the equation  $2T_{\text{phys}}/B(s) = T_{\text{eff}}$  for  $s$ . The relation between  $s^*$  and  $1/T_{\text{eff}}$  is shown in Fig. 16 and the annealing schedule function  $B(s)$  used to determine  $s^*$  and the schedule function of the transverse field  $A(s)$  are also shown. For  $h_0 = 0.2$ , at  $t_a = 500$   $\mu\text{s}$  the freeze-out point  $s^*$  reaches approximately 0.6, where  $B(s^*) \gg A(s^*)$ , indicating that the classical term is dominant in  $H(s^*)$ . For  $h_0 = 1.0$  and  $3.0$ ,  $s^*$  reaches approximately 0.35 and 0.4, respectively. For these  $s^*$  values,  $B(s^*) \approx A(s^*)$  and  $B(s^*) < A(s^*)$ , respectively, which may lead to the deviation of the data of the D-Wave from the thermal equilibrium of  $H_0$ .

To examine how closely the D-Wave data matches a thermal equilibrium value, we compare the data of the D-Wave

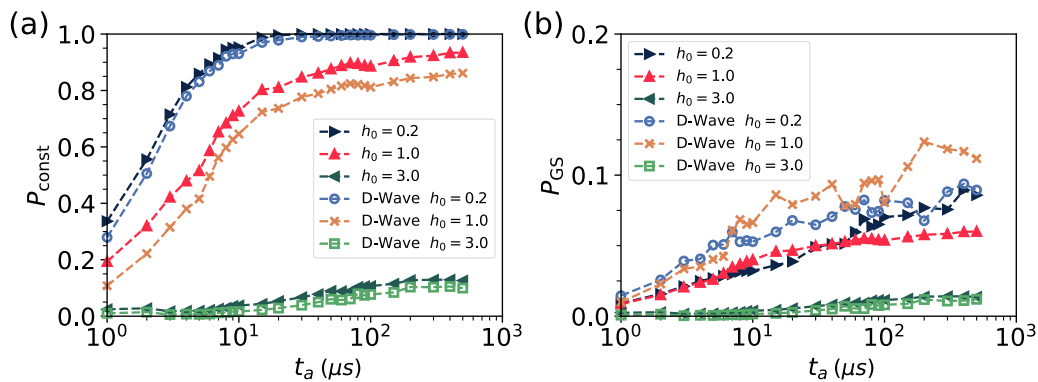


FIG. 17. Comparison between data of the D-Wave (shown by symbols with lines) and the thermal equilibrium of  $H_0$  at the corresponding effective temperature (shown by closed triangles). (a) Constraint satisfaction probability and (b) ground-state probability as functions of annealing time.

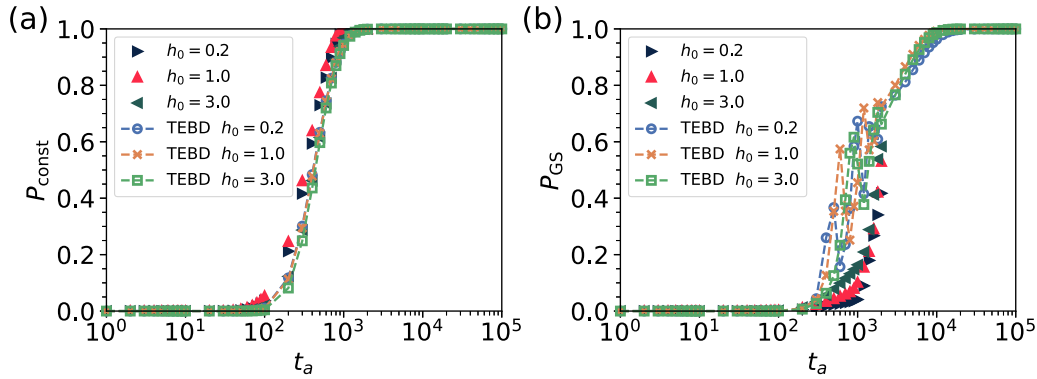


FIG. 18. Comparison between data of coherent QA obtained by TEBD (shown by symbols with lines) and the thermal equilibrium of  $H_0$  at the corresponding effective temperature (shown by closed triangles). (a) Constraint satisfaction probability and (b) ground-state probability as functions of annealing time. The annealing time  $t_a$  is in units of  $J$ .

and the observables of the thermal equilibrium of  $H_0$  at the obtained effective temperature as shown in Fig. 17. For  $h_0 = 0.2$  and  $t_a > 10$ , where  $s^* > 0.5$ , the ground-state probability  $P_{GS}$  and the constraint satisfaction probability  $P_{const}$  are close to the values of the thermal equilibrium state, implying that the hypothesis of quasistatic time evolution is valid. On the other hand, in the region where  $s^* < 0.5$  for all  $h_0$  values,  $P_{GS}$  of the D-Wave is larger than that of the thermal equilibrium. Furthermore,  $P_{const}$  of the D-Wave is smaller than that of thermal equilibrium. This result suggests that in the case of  $s^* < 0.5$ , the hypothesis of quasistatic time evolution may not hold or that the final state of the annealing is affected by the transverse field.

A similar analysis of the effective temperature is performed for the results of TEBD, as shown in Fig. 18. As in the case of the D-Wave,  $P_{GS}$  is larger and  $P_{const}$  is smaller than the values of the corresponding thermal equilibrium state. This result

suggests that it may be generally observed that the single-kink state probability is larger and the ground-state probability is smaller in nonequilibrium states compared to the thermal equilibrium states at the corresponding effective temperature.

In summary, the annealing-time dependence of the effective temperature varies significantly depending on  $h_0$ . For  $h_0 = 0.2$ , the effective temperature decreases toward the hardware temperature, while it saturates at higher temperatures for  $h_0 = 1.0$  and  $3.0$ . Moreover, the value of effective temperature changes nonmonotonically with  $h_0$ . The mechanism of this nontrivial dependence of the effective temperature on  $h_0$  remains to be clarified. In the region where the freeze-out point is small and not in thermal equilibrium of  $H_0$ , the behaviors of  $P_{GS}$  and  $P_{const}$  exhibit similar characteristics to coherent QA, although the underlying mechanism remains unclear. For a deeper understanding of the results, analysis taking the effects of thermal noise into account is necessary.

- [1] T. Kadowaki and H. Nishimori, *Phys. Rev. E* **58**, 5355 (1998).
- [2] E. Farhi, J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren, and D. Preda, *Science* **292**, 472 (2001).
- [3] G. E. Santoro, R. Martoňák, E. Tosatti, and R. Car, *Science* **295**, 2427 (2002).
- [4] G. E. Santoro and E. Tosatti, *J. Phys. A: Math. Gen.* **39**, R393 (2006).
- [5] A. Das and B. K. Chakrabarti, *Rev. Mod. Phys.* **80**, 1061 (2008).
- [6] S. Morita and H. Nishimori, *J. Math. Phys.* **49**, 125210 (2008).
- [7] S. Tanaka, R. Tamura, and B. K. Chakrabarti, *Quantum Spin Glasses, Annealing and Computation* (Cambridge University Press, Cambridge, 2017).
- [8] T. Albash and D. A. Lidar, *Rev. Mod. Phys.* **90**, 015002 (2018).
- [9] P. Hauke, H. G. Katzgraber, W. Lechner, H. Nishimori, and W. D. Oliver, *Rep. Prog. Phys.* **83**, 054401 (2020).
- [10] E. K. Grant and T. S. Humble, *Oxford Research Encyclopedia of Physics* (Oxford University Press, Oxford, 2020), pp. 1–26.
- [11] E. J. Crosson and D. A. Lidar, *Nat. Rev. Phys.* **3**, 466 (2021).
- [12] M. W. Johnson, P. Bunyk, F. Maibaum, E. Tolkacheva, A. J. Berkley, E. M. Chapple, R. Harris, J. Johansson, T. Lanting, I. Perminov, E. Ladizinsky, T. Oh, and G. Rose, *Supercond. Sci. Technol.* **23**, 065004 (2010).
- [13] A. J. Berkley, M. W. Johnson, P. Bunyk, R. Harris, J. Johansson, T. Lanting, E. Ladizinsky, E. Tolkacheva, M. H. S. Amin, and G. Rose, *Supercond. Sci. Technol.* **23**, 105014 (2010).
- [14] R. Harris, M. W. Johnson, T. Lanting, A. J. Berkley, J. Johansson, P. Bunyk, E. Tolkacheva, E. Ladizinsky, N. Ladizinsky, T. Oh, F. Cioata, I. Perminov, P. Spear, C. Enderud, C. Rich, S. Uchaikin, M. C. Thom, E. M. Chapple, J. Wang, B. Wilson *et al.*, *Phys. Rev. B* **82**, 024511 (2010).
- [15] 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.*, *Nat. Commun.* **4**, 1903 (2013).
- [16] 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.*, *Phys. Rev. X* **4**, 021041 (2014).
- [17] C. McGeoch and P. Farré, Advantage Processor Overview, D-Wave Report No. 14-1058A-A, 2022 (unpublished).
- [18] A. B. Finnila, M. A. Gomez, C. Sebenik, C. Stenson, and J. D. Doll, *Chem. Phys. Lett.* **219**, 343 (1994).



- [19] L. Stella, G. E. Santoro, and E. Tosatti, *Phys. Rev. B* **72**, 014303 (2005).
- [20] S. Shinomoto and Y. Kabashima, *J. Phys. A: Math. Gen.* **24**, L141 (1991).
- [21] E. M. Inack and S. Pilati, *Phys. Rev. E* **92**, 053304 (2015).
- [22] Y. W. Koh and H. Nishimori, *Phys. Rev. A* **105**, 062435 (2022).
- [23] J. M. Dieterich and B. Hartke, *Appl. Math.* **3**, 1552 (2012).
- [24] N. Chancellor, *Quantum Sci. Technol.* **4**, 045004 (2019).
- [25] S. Abel, N. Chancellor, and M. Spannowsky, *Phys. Rev. D* **103**, 016008 (2021).
- [26] S. Abel and M. Spannowsky, *PRX Quantum* **2**, 010349 (2021).
- [27] S. Abel, A. Blance, and M. Spannowsky, *Phys. Rev. A* **106**, 042607 (2022).
- [28] A. Perdomo-Ortiz, S. E. Venegas-Andraca, and A. Aspuru-Guzik, *Quantum Inf. Process.* **10**, 33 (2011).
- [29] M. Ohkuwa, H. Nishimori, and D. A. Lidar, *Phys. Rev. A* **98**, 022314 (2018).
- [30] Y. Yamashiro, M. Ohkuwa, H. Nishimori, and D. A. Lidar, *Phys. Rev. A* **100**, 052321 (2019).
- [31] S. Arai, M. Ohzeki, and K. Tanaka, *Phys. Rev. Res.* **3**, 033006 (2021).
- [32] J. A. Nelder and R. Mead, *Comput. J.* **7**, 308 (1965).
- [33] M. R. Hestenes and E. Stiefel, *J. Res. Natl. Bur. Stand.* **49**, 409 (1952).
- [34] R. H. Leary, *J. Global Optim.* **18**, 367 (2000).
- [35] R. Storn and K. Price, *J. Global Optim.* **11**, 341 (1997).
- [36] G. Vidal, *Phys. Rev. Lett.* **93**, 040502 (2004).
- [37] S. R. White and A. E. Feiguin, *Phys. Rev. Lett.* **93**, 076401 (2004).
- [38] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, *Science* **220**, 671 (1983).
- [39] M. Suzuki, *Prog. Theor. Phys.* **56**, 1454 (1976).
- [40] S. W. Shin, G. Smith, J. A. Smolin, and U. Vazirani, [arXiv:1401.7087](https://arxiv.org/abs/1401.7087).
- [41] D-Wave Ocean software documentation, <https://docs.ocean.dwavesys.com/en/stable/index.html> (2023).
- [42] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey *et al.*, *Nat. Methods* **17**, 261 (2020).
- [43] Geekbench browser, <https://browser.geekbench.com/> (2023).
- [44] V. S. Denchev, S. Boixo, S. V. Isakov, N. Ding, R. Babbush, V. Smelyanskiy, J. Martinis, and H. Neven, *Phys. Rev. X* **6**, 031015 (2016).
- [45] B. Tasseff, T. Albash, Z. Morrell, M. Vuffray, A. Y. Lokhov, S. Misra, and C. Coffrin, [arXiv:2210.04291](https://arxiv.org/abs/2210.04291).
- [46] K. Nishimura, H. Nishimori, and H. G. Katzgraber, *Phys. Rev. A* **102**, 042403 (2020).
- [47] A. D. King, S. Suzuki, J. Raymond, A. Zucca, T. Lanting, F. Altomare, A. J. Berkley, S. Ejtemaee, E. Hoskinson, S. Huang, E. Ladizinsky, A. J. R. MacDonald, G. Marsden, T. Oh, G. Poulin-Lamarre, M. Reis, C. Rich, Y. Sato, J. D. Whittaker, J. Yao *et al.*, *Nat. Phys.* **18**, 1324 (2022).
- [48] P. Weinberg, M. Tylutki, J. M. Rönkkö, J. Westerholm, J. A. Åström, P. Manninen, P. Törmä, and A. W. Sandvik, *Phys. Rev. Lett.* **124**, 090502 (2020).
- [49] T. Hamasaki and H. Nishimori, *J. Phys. Soc. Jpn.* **73**, 1490 (2004).
- [50] S. Suzuki and M. Okada, *Interdisc. Inf. Sci.* **13**, 49 (2007).
- [51] U. Schollwöck, *Ann. Phys. (NY)* **326**, 96 (2011).
- [52] T. W. B. Kibble, *J. Phys. A: Math. Gen.* **9**, 1387 (1976).
- [53] W. H. Zurek, *Nature (London)* **317**, 505 (1985).
- [54] J. Dziarmaga, *Phys. Rev. Lett.* **95**, 245701 (2005).
- [55] S. Suzuki and M. Okada, *J. Phys. Soc. Jpn.* **74**, 1649 (2005).
- [56] T. Zaborniak and R. de Sousa, *IEEE Trans. Quantum Eng.* **2**, 1 (2021).
- [57] N. Chancellor, P. J. D. Crowley, T. Durić, W. Vinci, M. H. Amin, A. G. Green, P. A. Warburton, and G. Aeppli, *npj Quantum Inf.* **8**, 73 (2022).
- [58] M. S. Könz, W. Lechner, H. G. Katzgraber, and M. Troyer, *PRX Quantum* **2**, 040322 (2021).
- [59] Y. Bando and H. Nishimori, *Phys. Rev. A* **104**, 022607 (2021).
- [60] Z. Jiang, V. N. Smelyanskiy, S. V. Isakov, S. Boixo, G. Mazzola, M. Troyer, and H. Neven, *Phys. Rev. A* **95**, 012322 (2017).
- [61] T. Albash and J. Marshall, *Phys. Rev. Appl.* **15**, 014029 (2021).
- [62] Y. Bando, K.-W. Yip, H. Chen, D. A. Lidar, and H. Nishimori, *Phys. Rev. Appl.* **17**, 054033 (2022).
- [63] A. D. King, J. Raymond, T. Lanting, R. Harris, A. Zucca, F. Altomare, A. J. Berkley, K. Boothby, S. Ejtemaee, C. Enderud, E. Hoskinson, S. Huang, E. Ladizinsky, A. J. R. MacDonald, G. Marsden, R. Molavi, T. Oh, G. Poulin-Lamarre, M. Reis, C. Rich *et al.*, *Nature (London)* **617**, 61 (2023).
- [64] F. Gao and L. Han, *Comput. Optim. Appl.* **51**, 259 (2012).
- [65] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. (Springer, New York, 2006).
- [66] R. Fletcher and C. M. Reeves, *Comput. J.* **7**, 149 (1964).
- [67] M. Iwamatsu and Y. Okabe, *Chem. Phys. Lett.* **399**, 396 (2004).
- [68] Z. Luo, B. Olson, I. Hashmi, K. Molloy, and A. Shehu, *Adv. Artif. Intell.* **2012**, 674832 (2012).
- [69] R. Storn, in *Proceedings of the North American Fuzzy Information Processing Society, Berkeley, 1996*, edited by M. H. Smith, M. A. Lee, J. Keller, and J. Yen (IEEE, Piscataway, 1996), pp. 519–523.
- [70] S. Das and P. N. Suganthan, *IEEE Trans. Evol. Comput.* **15**, 4 (2011).
- [71] M. H. Amin, *Phys. Rev. A* **92**, 052323 (2015).