Quantum annealing showing an exponentially small success probability despite a constant energy gap with polynomial energy

Hiroshi Hayasaka [1,*] Takashi Imoto,[1,†] Yuichiro Matsuzaki,[1,2,‡] and Shiro Kawabata [1,2,§]

[1]*Global Research and Development Center for Business by Quantum-AI Technology (G-QuAT), National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1, Umezono, Tsukuba, Ibaraki 305-8568, Japan*

[2]*NEC-AIST Quantum Technology Cooperative Research Laboratory, National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1, Umezono, Tsukuba, Ibaraki 305-8568, Japan*

Quantum annealing (QA) is a method for solving combinatorial optimization problems. We can estimate the computational time for QA using what is referred to as the adiabatic condition derived from the adiabatic theorem. The adiabatic condition consists of two parts: an energy gap and a transition matrix. Most past studies have focused on the relationship between the energy gap and computational time. The success probability of QA is considered to decrease exponentially owing to the exponentially decreasing energy gap at the first-order phase-transition point. In this study, through a detailed analysis of the relationship between the energy gap, transition matrix, and computational cost during QA, we propose a general method for constructing counterintuitive models wherein QA with a constant annealing time fails despite a constant energy gap, based on polynomial energy. We assume that the energy of the total Hamiltonian is at most $\Theta(L)$, where L is the number of qubits. In our formalism, we choose a known model that exhibits an exponentially small energy gap during QA, and modify the model by adding a specific penalty term to the Hamiltonian. In the modified model, the transition matrix in the adiabatic condition becomes exponentially large as the number of qubits increases, while the energy gap remains constant. Moreover, we achieve a quadratic speedup, for which the upper bound for improvement in the adiabatic condition is determined by the polynomial energy. For concrete examples, we consider the adiabatic Grover search and the ferromagnetic p-spin model. In these cases, with the addition of the penalty term, although the success probability of QA on the modified models becomes exponentially small despite a constant energy gap, we are able to achieve a success probability considerably higher than that of conventional QA. Moreover, in concrete examples, we numerically show the scaling of the computational cost is quadratically improved compared to the conventional QA. Our findings pave the way for a better understanding of QA performance.

I. INTRODUCTION

Quantum annealing (QA) is used to solve combinatorial optimization problems with quantum properties [1–7]. The solutions to combinatorial optimization problems can be embedded in the ground state of an Ising Hamiltonian [8–10]. On the other hand, quantum fluctuation can be induced using the Hamiltonian of the transverse magnetic field. By gradually decreasing the transverse magnetic field while gradually increasing the Ising Hamiltonian, we can obtain the ground state of the Ising Hamiltonian if what is known as the adiabatic condition is satisfied [11–17]. This condition is expressed as follows:

$$\eta = \frac{|\langle E_1(t)|\partial_t \hat{H}_{\mathrm{QA}}(t)|E_0(t)\rangle|}{\Delta_{\mathrm{QA}}(t)^2} \ll 1, \tag{1}$$

where $\hat{H}_{\mathrm{QA}}(t)$, $|E_0(t)\rangle$, $|E_1(t)\rangle$, and $\Delta_{\mathrm{QA}}(t)$ are a QA Hamiltonian, the ground state, the first excited state, and the energy gap between these states, respectively. Throughout this paper, we refer to the middle term of Eq. (1) as an adiabatic-condition term. The energy gap and computational complexity are considered to be related to each other; hence, this probable relationship has been extensively studied [18–24]. Several methods for estimating and controlling the energy gap have been proposed to improve the performance of QA [25–30].

A phase transition can occur if there is competition between the quantum fluctuations and magnetic interactions. In this case, the energy gap vanishes at the point of the phase transition when the system size approaches infinity. In statistical mechanics, the order of systems is characterized by the order parameter. In a second-order phase transition, the order parameter is continuous, whereas in a first-order phase transition, it exhibits discontinuity at a certain time. If a first-order phase transition occurs, it is generally supposed that the energy gap becomes exponentially small at the phase-transition point, with some exceptions [31–33]. When QA is used to solve hard problems (that cannot be efficiently solved using classical algorithms), a first-order phase transition typically occurs. For example, such a phase transition occurs

*Contact author: hayasaka.hiroshi@aist.go.jp

†Contact author: takashi.imoto@aist.go.jp

‡Contact author: matsuzaki.yuichiro@aist.go.jp, present address: ymatsuzaki872@g.chuo-u.ac.jp

§Contact author: s-kawabata@aist.go.jp, present address: kawabata@hosei.ac.jp

when exact cover and database-search problems are solved via QA [34,35]. Therefore, the adiabatic-condition term in (1) is considered to increase exponentially at the first-order phase-transition point.

In the realm of quantum many-body systems, determining whether the energy gap is closed or open, i.e., gapless or gapped, in the thermodynamic limit is a notorious problem itself [36]. However, modifying the energy gap through Hamiltonian rescaling is possible, even when the energy gap is exponentially small. While such an energy rescaling improves the adiabatic condition term by the factor of the energy gap, this does not inherently reduce a computational cost of QA. Using the quantum speed limit [37–41], we can evaluate the computational cost required to obtain the solution of QA with a probability above a certain threshold [42,43]. The cost is given as follows:

$$Q = \int_0^T \sigma[\hat{H}_{QA}(t), |\Psi(t)\rangle]dt, \qquad (2)$$

where $\sigma[\hat{H}, |\Psi\rangle] = [\langle\Psi|\hat{H}^2|\Psi\rangle - \langle\Psi|\hat{H}|\Psi\rangle^2]^{1/2}$. If we use an energy rescaling for QA Hamiltonian such as $e^L H_{QA}(t)$, the cost is not changed (see Appendix A). In other words, the cost required to obtain the solution of QA with a probability above a certain threshold cannot be improved by energy rescaling. Hence, we do not use energy rescaling in this paper. Moreover, throughout our study, we assume that the QA Hamiltonian $\hat{H}_{QA}(t)$ is bounded; as such, $\Theta(1) \leqslant \|\hat{H}_{QA}(t)\|_{op} \leqslant \Theta(L)$, where $\|\cdot\|_{op}$ denotes the operator norm.

The performance of QA seems to be often determined by the energy gap. A notable exception arises in the case of perturbed Hamming weight problems, where the transition matrix (the numerator of the adiabatic-condition term) is more relevant than the energy gap, thus affecting the performance of QA [44,45]. In this type of problem, the transition matrix scales most proportionally to a polynomial in the system size. Consequently, the calculation can be finished within polynomial time, thus preventing catastrophic failures in QA.

In this paper, we propose a systematic method for constructing counterintuitive models wherein QA with a constant annealing time fails despite a constant energy gap. We demonstrate that the cause of such failure is actually not the energy gap but, rather, the exponential increase in the size of the transition matrix in the adiabatic-condition term in Eq. (1). The key idea of our proposal is to add a penalty term $\hat{H}_{pena}(t)$ to the QA Hamiltonian, which does not change the eigenstate of the Hamiltonian but changes the eigenvalue. We analytically show that when we add such a penalty term, the transition matrix becomes exponentially large, while the energy gap remains constant. Furthermore, we demonstrate that we can achieve a quadratic speedup of QA by adding the penalty term to the QA Hamiltonian with polynomial energy, i.e., $\|\hat{H}_{QA}(t) + \hat{H}_{pena}(t)\|_{op} \leqslant \Theta(L)$, which cannot be reproduced via energy rescaling. We also numerically perform QA on the Grover search and the ferromagnetic p-spin model. The success probability of QA on these models becomes exponentially small as we increase the problem size L despite an energy gap that scales as $\Theta(L^0)$ during QA. However, the success probability on our modified model is considerably higher than that of conventional QA. Moreover, we numerically show

that the scaling of the computational cost is quadratically improved compared to the conventional QA.

The remainder of our paper is organized as follows. In Sec. II, we review QA and the adiabatic Grover search. In Sec. III, we introduce the general framework for constructing a case in which the transition matrix increases exponentially. In Sec. IV, for our first example, we demonstrate an application of our general theory on the adiabatic Grover search. In Sec. V, for our second example, we show a numerical analysis on the ferromagnetic p-spin model. Finally, in Sec. VI, we present our conclusion.

II. QUANTUM ANNEALING AND ADIABATIC GROVER SEARCH

In this section, we review QA and adiabatic Grover search.

A. Quantum annealing

In QA, the total Hamiltonian is defined as follows [4,5]:

$$\hat{H}_{QA}(t) = \frac{t}{T}\hat{H}_p + \left(1 - \frac{t}{T}\right)\hat{H}_d, \qquad (3)$$

where $\hat{H}_p$ and $\hat{H}_d$ are the problem Hamiltonian and driver Hamiltonian, respectively. We prepare a ground state of $\hat{H}_d$, and let this state evolve via the total Hamiltonian. If an initial state at $t = 0$ evolves in a sufficiently large T to satisfy the adiabatic condition, we can obtain the ground state of $\hat{H}_p$ at $t = T$.

B. Adiabatic Grover search

Let us consider the problem of searching for a specific element in a database composed of N elements. On a classical computer, we need to check, on average, half the elements to find the target element. To reduce this effort, Grover proposed a quantum search algorithm that requires only $\sqrt{N}$ evaluations to find the target element [46]. To search the database, we can adopt an adiabatic algorithm called the adiabatic Grover search [4,7,47]. In the adiabatic Grover search, the problem Hamiltonian is defined by

$$\hat{H}_p = \hat{I} - |m\rangle\langle m|. \qquad (4)$$

Here, the solution to be found is denoted by $|m\rangle$, which is represented by the computational basis ($|\uparrow\rangle$ and $|\downarrow\rangle$). The number of qubits is L, and thus, the dimension of the Hilbert space is 2^L. Meanwhile, the driver Hamiltonian is defined by

$$\hat{H}_d = \hat{I} - |++\cdots+\rangle\langle++\cdots+|, \qquad (5)$$

where $|+\rangle$ is the eigenstate of the Pauli matrix $\hat{\sigma}_x$, i.e., $|+\rangle = (|\uparrow\rangle + |\downarrow\rangle)/\sqrt{2}$. Because the adiabatic Grover search can be block diagonalized, we can analytically obtain the eigenvalues and eigenstates by diagonalizing a two-by-two matrix. Using the desired state $|m\rangle$ and its orthogonal state $|m^\perp\rangle$, the total Hamiltonian $\hat{H}_{QA}(t)$ can effectively be described to be

$$\hat{H}_{QA}(t) = \frac{1}{2}\hat{I} - \frac{\Delta_{QA}(t)}{2}\cos\theta\, \tilde{\sigma}^z - \frac{\Delta_{QA}(t)}{2}\sin\theta\, \tilde{\sigma}^x, \qquad (6)$$

where we define Pauli matrices represented on the bases of $|m\rangle$ and $|m^\perp\rangle$ as $\tilde{\sigma}^z \equiv |m\rangle\langle m| - |m^\perp\rangle\langle m^\perp|$ and $\tilde{\sigma}^x \equiv |m\rangle\langle m^\perp| + |m^\perp\rangle\langle m|$. Here, $\cos\theta(t)$, $\sin\theta(t)$, and $\Delta_{\text{QA}}(t)$ are given by

$$\cos\theta(t) = \frac{1}{\Delta_{\text{QA}}(t)}[1 - 2(1 - t/T)(1 - 2^{-L})], \quad (7)$$

$$\sin\theta(t) = \frac{2}{\Delta_{\text{QA}}(t)}(1 - t/T)\sqrt{2^{-L}(1 - 2^{-L})}, \quad (8)$$

$$\Delta_{\text{QA}}(t) = \sqrt{\left(1 - \frac{2t}{T}\right)^2 + \frac{2^{-L+2}t}{T}\left(1 - \frac{t}{T}\right)}, \quad (9)$$

where we choose the branch of $\text{Tan}^{-1}(x)$, which is given by $-\pi/2 \leqslant \text{Tan}^{-1}(x) \leqslant \pi/2$, and define $\theta(t) = \text{Tan}^{-1}[\sin\theta(t)/\cos\theta(t)]$. The ground state $|E_0(t)\rangle$ and first excited state $|E_1(t)\rangle$ of the Hamiltonian (6) are as follows:

$$|E_0(t)\rangle = \cos\frac{\theta(t)}{2}|m\rangle + \sin\frac{\theta(t)}{2}|m^\perp\rangle, \quad (10)$$

$$|E_1(t)\rangle = -\sin\frac{\theta(t)}{2}|m\rangle + \cos\frac{\theta(t)}{2}|m^\perp\rangle. \quad (11)$$

The energies of these states are given by

$$E_0(t) = \tfrac{1}{2}[1 - \Delta_{\text{QA}}(t)], \quad (12)$$

$$E_1(t) = \tfrac{1}{2}[1 + \Delta_{\text{QA}}(t)]. \quad (13)$$

Thus, the energy gap is $E_1(t) - E_0(t) = \Delta_{\text{QA}}(t)$. Based on Eq. (9), the energy gap scales to $\Theta(2^{-L/2})$ at $t = T/2$. The numerator of the adiabatic-condition term (1) for the adiabatic Grover search is $\Theta(L^0)$ [47]. The annealing time T should be scaled to $\Theta(2^L)$ to satisfy the adiabatic condition. This implies that the time required to find the solution using the adiabatic Grover search is the same as that required for a classical search. However, Roland and Cerf showed that a quadratic speedup, similar to that for Grover's algorithm, can be attained by choosing an optimal scheduling function [47].

III. GENERAL FRAMEWORK

In this section, we introduce a general method for constructing models on which QA with a constant annealing time has an exponentially small success probability despite a constant energy gap. Let us define a system Hamiltonian as $\hat{H}(t) = \hat{H}_{\text{QA}}(t) + \hat{H}_{\text{pena}}(t)$ and the energy gap of $\hat{H}(t)$ as $\Delta(t)$, where $H_{\text{pena}}(t)$ denotes a Hamiltonian added to the QA Hamiltonian to improve the performance. Throughout our study, we assume that the system Hamiltonian and QA Hamiltonian satisfy $\|\hat{H}(t)\|_{\text{op}} \leqslant \Theta(L)$ and $\Theta(1) \leqslant \|\hat{H}_{\text{QA}}(t)\|_{\text{op}} \leqslant \Theta(L)$, respectively. Here, it is worth mentioning that a physical Hamiltonian is usually assumed to be extensive, $\|\hat{H}(t)\|_{\text{op}} = \|\hat{H}_{\text{QA}}(t)\|_{\text{op}} = \Theta(L)$ [14,48,49]. However, to consider a Hamiltonian without a physical background, such as the adiabatic Grover search, we adopt assumptions, $\|\hat{H}(t)\|_{\text{op}} \leqslant \Theta(L)$ and $\Theta(1) \leqslant \|\hat{H}_{\text{QA}}(t)\|_{\text{op}} \leqslant \Theta(L)$. We also assume that the transition matrix $|\langle E_1(t)|\partial_t\hat{H}_{\text{QA}}(t)|E_0(t)\rangle|$ scales polynomially with L and that the energy gap of the QA Hamiltonian $\Delta_{\text{QA}}(t)$ becomes exponentially smaller as L increases. Subsequently, we show that by adding a specific penalty term to the QA Hamiltonian, we can systematically construct a model with a constant energy gap, where the transition matrix becomes exponentially large as L increases.

The adiabatic-condition term can be expressed in the following form:

$$\eta = \frac{|\langle E_1(t)|\partial_t\hat{H}_{\text{QA}}(t)|E_0(t)\rangle|}{\Delta_{\text{QA}}(t)^2}$$

$$= \left[\frac{|\langle E_1(t)|\partial_t\hat{H}_{\text{QA}}(t)|E_0(t)\rangle|}{\Delta_{\text{QA}}(t)}\right]\frac{1}{\Delta_{\text{QA}}(t)}. \quad (14)$$

The expression between the square brackets on the right-hand side of Eq. (14) yields

$$\frac{|\langle E_1(t)|\partial_t\hat{H}_{\text{QA}}(t)|E_0(t)\rangle|}{\Delta_{\text{QA}}(t)} = |\langle E_1(t)|d/dt|E_0(t)\rangle|. \quad (15)$$

Let us consider the following penalty term [to be added to the Hamiltonian $\hat{H}_{\text{QA}}(t)$]:

$$\hat{H}_{\text{pena}}(t) = \sum_{i=0}^{N-1} C_i(t)|E_i(t)\rangle\langle E_i(t)|, \quad (16)$$

where $N = 2^L$ denotes the dimension of the Hilbert space, and $|E_i(t)\rangle$ is the eigenstate of $\hat{H}_{\text{QA}}(t)$. Time-dependent coefficients $C_i(t)$ play a role in shifting the eigenenergies $E_i(t)$. To maintain a constant energy gap during QA, we choose $C_i(t)$ as follows:

$$C_i(t) = E_i(t = 0) - E_i(t). \quad (17)$$

Here, notably, if the operator norm of the QA Hamiltonian is bounded by $\Theta(L)$, this penalty term is also bounded by $\Theta(L)$ (see Appendix B). From Eq. (15), we show that $\frac{|\langle E_1(t)|\partial_t\hat{H}_{\text{QA}}(t)|E_0(t)\rangle|}{\Delta_{\text{QA}}(t)}$ remains unchanged by the addition of a penalty term to the QA Hamiltonian, because the penalty term is diagonal in the energy eigenstate basis. This implies that if we add the penalty term to open an exponentially small energy gap, the transition matrix becomes exponentially large. For example, let us consider a QA model where the scaling of $|\langle E_1(t)|\partial_t\hat{H}_{\text{QA}}(t)|E_0(t)\rangle|$ [$\Delta_{\text{QA}}(t)$] is given by $\Theta(L^0)$ [$\Theta(2^{-L/2})$]. In this case, as will be shown later, by adding the penalty term (16), we change the energy gap from $\Delta_{\text{QA}}(t) = \Theta(2^{-L/2})$ to $\Delta(t) = \Theta(L^0)$, while changing the transition matrix from $|\langle E_1(t)|\partial_t\hat{H}_{\text{QA}}(t)|E_0(t)\rangle| = \Theta(L^0)$ to $|\langle E_1(t)|\partial_t\hat{H}(t)|E_0(t)\rangle| = \Theta(2^{L/2})$. Notably, if we multiply an exponentially large factor by the QA Hamiltonian, e.g., $\hat{H}(t) = 2^{L/2}\hat{H}_{\text{QA}}(t)$, we obtain similar results. However, the operator norm of such a rescaled Hamiltonian is $\|\hat{H}(t)\|_{\text{op}} = \Theta(2^{L/2})$, which violates our assumption of $\|\hat{H}(t)\|_{\text{op}} \leqslant \Theta(L)$.

From Eqs. (14) and (15), clearly, the improvement of the adiabatic condition due to the addition of operators for the QA Hamiltonian, such that the energy eigenstates are unchanged, is by a factor of $1/\Delta_{\text{QA}}(t)$. To achieve optimal improvement of the adiabatic condition, where η is minimized, we can choose the coefficient $C_i(t)$ such that the energy of the ground state is set to be 0, and the energies of all excited states are set to the largest eigenenergy of $\hat{H}_{\text{QA}}(t)$. Thus, the optimal penalty term $\hat{H}_{\text{pena}}^{(\text{opt})}(t)$ is given by

$$\hat{H}_{\text{pena}}^{(\text{opt})}(t) = \sum_{i=0}^{N-1} C_i^{(\text{opt})}(t)|E_i(t)\rangle\langle E_i(t)|, \quad (18)$$

where $C_{i=0}^{(\text{opt})}(t) = -E_0(t)$ and $C_{i\neq 0}^{(\text{opt})}(t) = -E_i(t) + \max_{1\leqslant j\leqslant N-1}|E_j(t)|$. In addition, we set the QA Hamiltonian to be extensive, i.e., $\|\hat{H}_{\text{QA}}(t)\|_{\text{op}} = \Theta(L)$ [14]. In this case, the factor $1/\Delta(t)$ is scaled to $\Theta(L^{-1})$ because $\Delta(t) = \max_{1\leqslant i\leqslant N-1}|E_i(t)| = \|\hat{H}(t)\|_{\text{op}}$ for the Hermitian operator. Notably, although we considered an adiabatic condition term for the transition from the ground state to the first excited state, we could also consider transitions to other excited states. Let us define an adiabatic condition for the other excited states as $\eta_j = \frac{|\langle E_j(t)|\partial_t \hat{H}_{\text{QA}}(t)|E_0(t)\rangle|}{(E_j(t)-E_0(t))^2}$ for $2\leqslant j\leqslant 2^L - 1$. Using our modified Hamiltonian, we minimize not only η but all η_j. Although Eq. (18) is the optimal choice for the penalty term, we use Eqs. (16) and (17) to maintain consistency hereafter.

Let us compare our penalty term with the counterdiabatic driving [50–52]. Similar to our penalty term, the counterdiabatic term typically contains the instantaneous eigenstate of the QA Hamiltonian. With the counterdiabatic term, exact adiabatic dynamics can be achieved in a shorter computational time than with traditional QA. However, if we adopt the counterdiabatic term, passing through a first-order phase-transition point during QA will require an exponentially large amount of energy (see Appendix C). By contrast, using our penalty term, we can achieve a quadratic speedup over the conventional QA with polynomial energy, i.e., $\|\hat{H}_{\text{QA}}(t) + \hat{H}_{\text{pena}}(t)\|_{\text{op}} \leqslant \Theta(L)$.

IV. ADIABATIC GROVER SEARCH WITH PENALTY TERM

For our first example, we first present an analytical investigation of the scaling of the adiabatic-condition term on the adiabatic Grover search with the penalty term (16). Subsequently, we numerically show that the success probability of QA decreases exponentially for both our modified and conventional Hamiltonians. We also show, however, that the success probability of our approach is higher than that of the conventional one.

A. Scaling of adiabatic-condition term

We consider the Hamiltonian (6). The penalty term is given by

$$\hat{H}_{\text{pena}}(t) = -E_0(t)|E_0(t)\rangle\langle E_0(t)| + \{1 - E_1(t)\}|E_1(t)\rangle\langle E_1(t)|, \quad (19)$$

where we use Eqs. (10) and (11). Meanwhile, the total Hamiltonian is defined as

$$\hat{H}(t) = \hat{H}_{\text{QA}}(t) + \hat{H}_{\text{pena}}(t)$$
$$= \tfrac{1}{2}\hat{I} - \tfrac{1}{2}\cos\theta(t)\,\tilde{\sigma}^z - \tfrac{1}{2}\sin\theta(t)\,\tilde{\sigma}^x. \quad (20)$$

We can easily calculate the energy gap in this model (20) to be $\Delta(t) = 1$. The energy gap in our model does not depend on the number of qubits L. Notably, the difference between Eq. (20) and Eq. (6) is the prefactor of $\Delta_{\text{QA}}(t)$ in front of $\cos\theta(t)$ and $\sin\theta(t)$. Thus, except for $\hat{I}/2$, our penalty term simply multiplies the original Hamiltonian (6) by an exponentially large factor $[1/\Delta_{\text{QA}}(t)]$ in the subspaces of $|m\rangle$ and $|m^\perp\rangle$. However, $\|\hat{H}(t)\|_{\text{op}}$ remains as $\Theta(1)$, even with the penalty term.
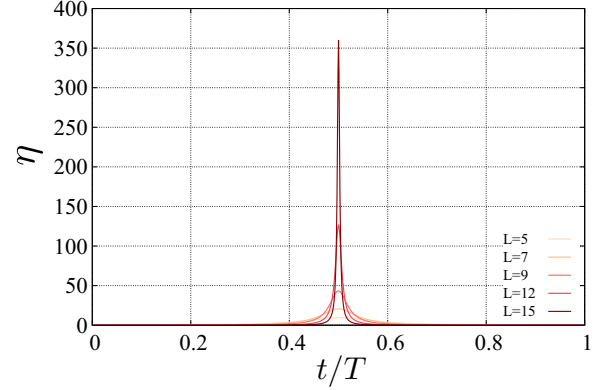

FIG. 1. Plot of adiabatic-condition term η for the adiabatic Grover search as a function of normalized time t/T, where L is the number of qubits.

Figure 1 shows the adiabatic-condition term as a function of t. At $t = T/2$, the adiabatic-condition term becomes exponentially larger as L increases. We then analyze the scaling of the adiabatic-condition term. From Eqs. (10), (11), and (15), we obtain

$$|\langle E_1(t)|d/dt|E_0(t)\rangle|$$
$$= \frac{\sqrt{(1 - 2^{-L})2^{-L}}}{T\left[1 + 4\left(\frac{t}{T}\right)^2(1 - 2^{-L}) - 4\frac{t}{T}(1 - 2^{-L})\right]}. \quad (21)$$

At $t = T/2$, Eq. (21) yields

$$|\langle E_1(t)|d/dt|E_0(t)\rangle| = \frac{1}{T}\sqrt{2^L - 1} = \Theta(2^{L/2}). \quad (22)$$

Given that the energy gap is always constant owing to the penalty term, the scaling of the adiabatic-condition term is $\eta = \Theta(2^{L/2})$ at $t = T/2$. This indicates that the adiabatic-condition term is improved by a factor of Δ_{QA}^{-1} compared to without the penalty term. The improvement of $1/\Delta_{\text{QA}}(t)$ is reminiscent of what is achieved using optimized scheduling on the adiabatic Grover search [47]. However, further improvements can be achieved by selecting an optimized schedule for the adiabatic Grover search with the penalty term (see Appendix D) [7,47].

Based on Eqs. (15) and (22), the transition matrix becomes

$$\langle E_1(t)|\partial_t \hat{H}(t)|E_0(t)\rangle = \Theta(2^{L/2}). \quad (23)$$

Hence, as expected from our general framework, the divergence of the adiabatic-condition term in Fig. 1 stems from the exponentially large transition matrix (23).

Quantum many-body systems often exhibit point gap closing owing to quantum phase transitions [18,53–55]. In the adiabatic Grover search, the competition between the driver Hamiltonian and problem Hamiltonian causes a first-order quantum phase transition from the paramagnetic to ferromagnetic phase. The energy gap at this phase-transition point becomes exponentially small as we increase the size L and vanishes at the thermodynamic limit $L \to \infty$. Conversely, the energy gap in our model with the penalty term does not close at the thermodynamic limit, which appears to avoid a first-order phase transition with the addition of the penalty
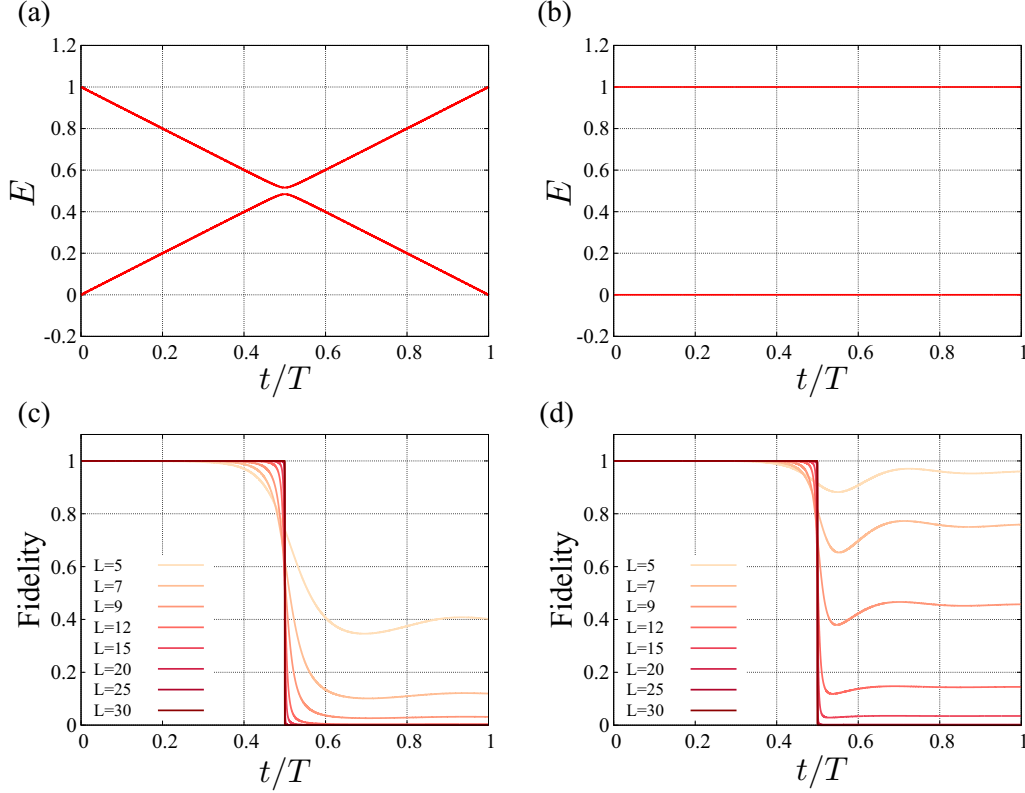
FIG. 2. Energy spectrum and fidelity for the adiabatic Grover search as functions of t/T. Energy diagram in $L = 10$ (a) without penalty term and (b) with penalty term. Fidelity (c) without penalty term and (d) with penalty term. Annealing time $T = 20$.

term. To check whether a first-order phase transition occurs, we analyze the total magnetization as an order parameter. This analysis is essentially equivalent to the mean-field analysis of a p-spin model for $p \to \infty$ [55]. Meanwhile, we directly obtain the magnetization using the ground state of our model (20); this is in contrast to that done in Ref. [55]. The total magnetization of the ground state is given by

$$\langle E_0(t)| \sum_i^L \hat{\sigma}_i^z |E_0(t)\rangle = \cos^2\frac{\theta(t)}{2} - \frac{1}{2^L - 1}\sin^2\frac{\theta(t)}{2}, \quad (24)$$

where $\hat{\sigma}^z$ is the z component of the Pauli matrix, $\hat{\sigma}^z = |\uparrow\rangle\langle\uparrow| - |\downarrow\rangle\langle\downarrow|$. The occurrence of a first-order phase transition can be shown as a discontinuity of magnetization at $t = T/2$ within the thermodynamic limit. We let the thermodynamic limit be $L \to \infty$ in Eq. (24):

$$\lim_{L\to\infty}\left(\cos^2\frac{\theta(t)}{2} - \frac{1}{2^L - 1}\sin^2\frac{\theta(t)}{2}\right) = \frac{1}{2} + \frac{1}{2}\lim_{L\to\infty}\cos\theta(t)$$

$$= \frac{1}{2} - \frac{1 - 2(t/T)}{2|1 - 2(t/T)|}. \quad (25)$$

Thus, we obtain

$$\lim_{t\to\frac{T}{2}-0}\lim_{L\to\infty}\langle E_0(t)| \sum_i^L \hat{\sigma}_i^z |E_0(t)\rangle = 0, \quad (26)$$

$$\lim_{t\to\frac{T}{2}+0}\lim_{L\to\infty}\langle E_0(t)| \sum_i^L \hat{\sigma}_i^z |E_0(t)\rangle = 1. \quad (27)$$

Therefore, in our model, a first-order phase transition occurs at $t = T/2$ even if the gap is $\Theta(L^0)$. Notably, there are some counterexamples in which the gap does not become exponentially small at the first-order phase-transition point [31–33].

B. Numerical analysis

To investigate the effects of the exponential increase in size of the transition matrix of QA, we perform numerical calculations. Figures 2(a) and 2(b) show the energy spectra without and with the penalty term, respectively. Without the penalty term, the energy gap becomes minimal at $t = T/2$. By contrast, with the penalty term, the energy gap remains constant and is not affected by the number of qubits, L, as shown in Fig. 3(a). We numerically solve the Schrödinger equation and obtain the state $|\Psi_0(t)\rangle$. We show the fidelity as a function of t in Figs. 2(c) and 2(d), where we define fidelity as $|\langle\Psi_0(t)|E_0(t)\rangle|^2$. As L increases, the fidelity decreases more rapidly around $t = T/2$ owing to a nonadiabatic transition to the first excited state.

Figures 3(a) and 3(b) show the scaling of the energy gap at $t = T/2$ and the fidelity at $t = T$, respectively. Without the penalty term, the energy gap becomes exponentially small. As mentioned in Sec. II, the scaling of the transition matrix is $\Theta(L^0)$ without the penalty term. Therefore, the exponential decrease in the fidelity [bottom line in Fig. 3(b)] originates from the gap closing at $t = T/2$. By contrast, with the penalty term, the energy gap is not affected by L. Thus, the decrease in fidelity [top line in Fig. 3(b)] originates from the exponential increase in size of the transition matrix. Notably,
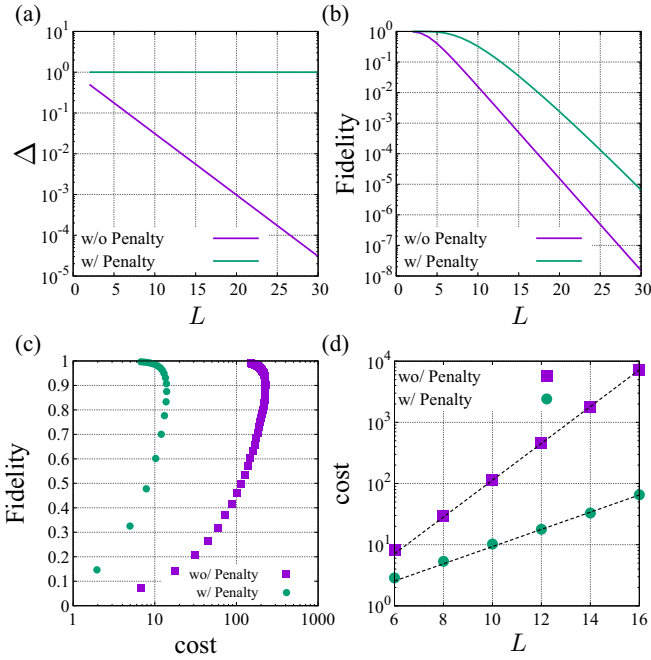
FIG. 3. Scaling of (a) energy gap at $t = T/2$ and (b) fidelity at $t = T$ for the adiabatic Grover search. Purple (bottom) and green (top) lines correspond to cases without and with penalty term, respectively. In (a) and (b), annealing time is $T = 20$. (c) The fidelity and corresponding computational cost where the system size is $L = 10$. Here, we use $T = 10n$ ($T = 100m$) for $1 \leqslant n \leqslant 20, n \in \mathbb{N}$ ($1 \leqslant m \leqslant 60, m \in \mathbb{N}$) in the case with (without) penalty term. (d) The computational cost to obtain a fidelity higher than 0.5. We plot the cost against the number of qubits to identify the scaling behavior. In (c) and (d), purple (square) and green (circle) dots correspond to cases without and with penalty term, respectively.

according to Fig. 3(b), the fidelity for the model with the penalty term is higher than that for the model without the penalty term. This behavior is consistent with our general framework, demonstrating that the adiabatic-condition term with the penalty term is smaller than that without the penalty term. Figure 3(c) shows the computational cost defined in Eq. (2). Here, after we choose a value of T, we calculate the fidelity and computational cost to be plotted. We repeat this process for several T, and plot these values with and without penalty term. A nonmonotonic behavior of the fidelity against the cost is observed, and we explain the origin in Appendix E. From Fig. 3(c), we can estimate the required cost to achieve a certain value of fidelity. Such a cost with the penalty term is smaller than that without the penalty term. Figure 3(d) shows the scaling of the cost against the number of qubits. By solving a time-dependent Schrödinger equation, we can plot the fidelity against the annealing time T. As we increase T, the fidelity also increases. Let us consider the smallest T where the corresponding fidelity is more than or equal to 0.5. We plot the cost for such an annealing time T in Fig. 3(d). To identify the scaling behavior, we use a fitting function as $\alpha 2^{\beta L}$. Without the penalty term, we obtain $\alpha = 0.111$ and $\beta = 0.999$. On the other hand, with the penalty term, we obtain $\alpha = 0.365$ and $\beta = 0.467$. Thus, the cost with the penalty

the penalty term is quadratically smaller than that without the penalty term.

V. FERROMAGNETIC p SPIN MODEL

For our second example, we apply our theory to the ferromagnetic p-spin model [18,55–57]. The problem Hamiltonian is defined as

$$\hat{H}_p = -L \left(\frac{1}{L} \sum_{i=1}^{L} \hat{\sigma}_i^z \right)^p. \tag{28}$$

We use the transverse field as the driver Hamiltonian, which is given by $\hat{H}_d = -\sum_{i=1}^{L} \hat{\sigma}_i^x$. The Hamiltonian of QA is defined as follows:

$$\hat{H}_{QA}(t) = \frac{t}{T} \hat{H}_p + \left(1 - \frac{t}{T} \right) \hat{H}_d. \tag{29}$$

The penalty term is given by

$$\hat{H}_{pena}(t) = \sum_{i=0}^{2S} \{ E_i(t = 0) - E_i(t) \} |E_i(t)\rangle \langle E_i(t)|, \tag{30}$$

where S is the maximum angular momentum and $|E_i\rangle$ (E_i) is the eigenstate (eigenenergy) of $\hat{H}_{QA}(t)$. We numerically diagonalize $\hat{H}_{QA}(t)$ each time and construct the penalty term. The total Hamiltonian is given by

$$\hat{H}(t) = \hat{H}_{QA}(t) + \hat{H}_{pena}(t). \tag{31}$$

The components of the total spin operator are determined to be $\hat{S}^z = \sum_{i=1}^{L} \hat{\sigma}_i^z / 2$ and $\hat{S}^x = \sum_{i=1}^{L} \hat{\sigma}_i^x / 2$. As the total spin operator $\hat{S}$ is conserved, the initial state, belonging to the maximal spin sector, remains in this sector throughout the time evolution. Therefore, we consider only the subspace of $2S = L$.

The ferromagnetic p-spin model with a transverse field for $p = 2$ ($p \geqslant 3$) has a second-order (first-order) phase-transition point in QA [18,55–57]. Throughout our study, we set $p = 5$. Figures 4(a) and 4(b) show the energy spectra for $L = 40$ without and with the penalty term, respectively. The minimum energy gap is at $t/T \sim 0.463$ [Fig. 4(a)]. We numerically solve the Schrödinger equation. Then, in Figs. 4(c) and 4(d), we plot the fidelity for the model without and with the penalty term, respectively. In the case with the penalty term, the energy gap is always constant, as shown in Fig. 4(b). However, at $t/T \sim 0.463$, the fidelity decreases rapidly as we increase the number of qubits [Fig. 4(d)]. We also plot the scaling of the minimum energy gap and the fidelity at $t = T$ in Figs. 5(a) and 5(b), respectively. Clearly, these results are consistent with the predictions of our general framework, i.e., despite the constant energy gap in the case with the penalty term, the fidelity decays exponentially owing to the exponential increase in size of the transition matrix. In addition, the fidelity of our approach is considerably higher than that of conventional QA. Similar to what we did in Sec. IV, we plot the computational cost defined in Eq. (2) against the fidelity in Fig. 5(c). It is clearly shown that, in the presence of penalty terms, the cost is reduced compared to the case without penalty terms. We also show the scaling of the cost to achieve a fidelity greater than 0.5 in Fig. 5(d). We employ the same fitting function as in Sec. IV. Without the
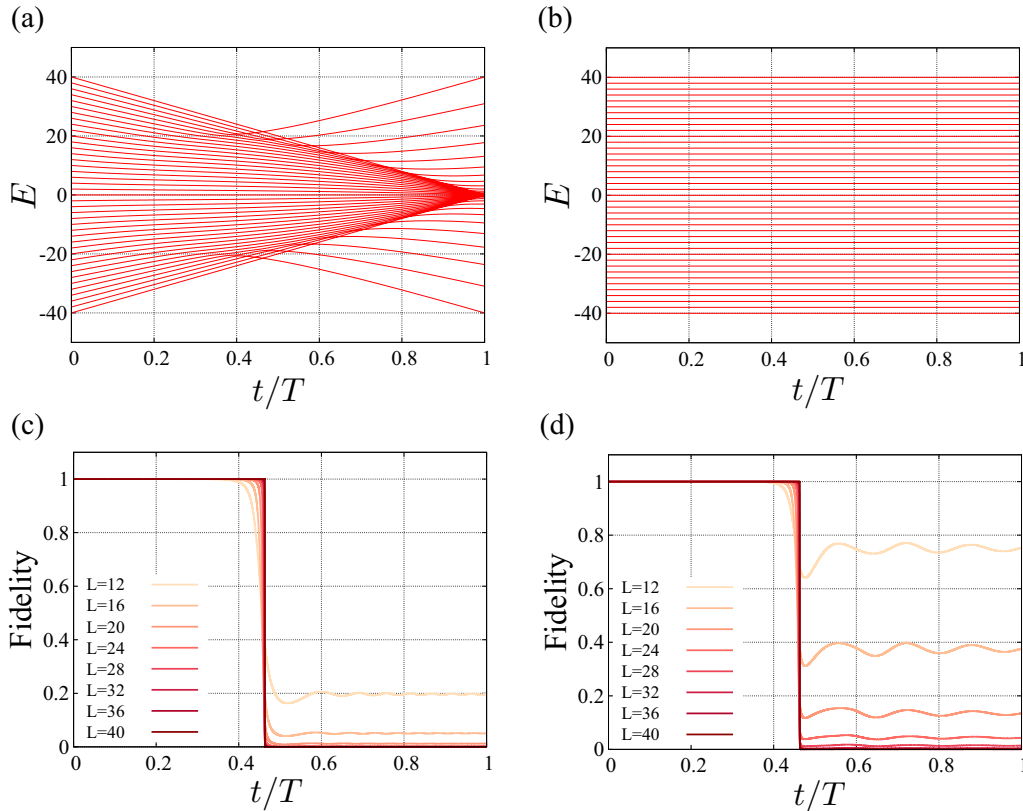
FIG. 4. Energy spectrum and fidelity as functions of t/T for the ferromagnetic p-spin model. Energy diagram in $L = 40$ (a) without penalty term and (b) with penalty term. Fidelity (c) without penalty term and (d) with penalty term. Annealing time $T = 20$, $p = 5$.

penalty term, we obtain $\alpha = 0.990$ and $\beta = 0.530$, whereas in the presence of penalty terms, we obtain $\alpha = 1.08$ and $\beta = 0.265$. Thus, the cost with the penalty term is quadratically smaller than that without the penalty term. It should be noted that we ignore the cost required for diagonalization of the QA Hamiltonian to construct the penalty term while we observe an improvement in the computational cost [Eq. (2)].

We compare our penalty term with the nonstoquastic Hamiltonian in the p-spin model [18]. Although the scaling of the energy gap in the p-spin model can be changed from exponentially to polynomially decreasing by adding a nonstoquastic Hamiltonian, the non-stoquastic Hamiltonian does not lead to an exponential increase in the size of the transition matrix (see Appendix F). This is due to the change in not only the energy gap but also the energy eigenstate. In contrast to the nonstoquastic Hamiltonian, the modification to open the exponentially small energy gap necessarily leads to an exponential increase in the size of the transition matrix as long as we use our method.

VI. CONCLUSION

In conclusion, we propose a general method for constructing models on which QA with a constant annealing time fails despite a constant energy gap. In accordance with our framework, we choose a known model that exhibits an exponentially small energy gap during QA and add a penalty term to the Hamiltonian. In the modified model, the transition

matrix in the adiabatic-condition term becomes exponentially larger as the number of qubits increases, while the energy gap remains constant. Moreover, we demonstrated that with our penalty term, the scaling of the adiabatic condition is improved from the inverse of squared to the energy gap to the inverse of the energy gap. We emphasize that adding the penalty term does not lead to an exponential increase in the operator norm of the Hamiltonian. Therefore, we can conclude that by using our penalty term, we can achieve quadratic speedup with polynomial energy.

Based on our framework, we investigated two models as concrete examples: the adiabatic Grover search and the ferromagnetic p-spin model. For the adiabatic Grover search, we analytically showed that the transition matrix becomes exponentially large and that the magnetization exhibits discontinuity, i.e., a first-order phase transition occurs, although the energy gap is always constant in QA. Moreover, for both the adiabatic Grover search and the ferromagnetic p-spin model, we numerically showed that the success probability decays exponentially owing to the exponential increase in size of the transition matrix. In addition, we showed that the success probability with our penalty term is considerably higher than that of the conventional approach. Furthermore, we numerically showed that the scaling of the computational cost is quadratically improved compared to the conventional QA.

The computational speed in QA is generally considered to be limited by the energy gap, which corresponds to the
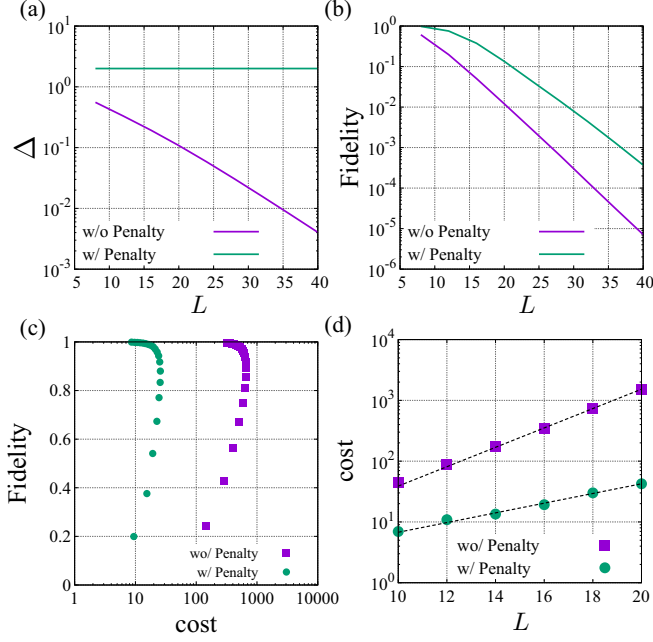
FIG. 5. Scaling of (a) minimum of energy gap and (b) fidelity at $t = T$ for the ferromagnetic p-spin model. Purple (bottom) and green (top) lines correspond to cases without and with penalty term, respectively. In (a) and (b), we set $T = 20$ and $p = 5$. (c) The fidelity and corresponding computational cost where the system size is $L = 16$. Here, we use $T = 10n$ ($T = 100m$) for $1 \leqslant n \leqslant 20, n \in \mathbb{N}$ ($1 \leqslant m \leqslant 20, m \in \mathbb{N}$) in the case with (without) penalty term. (d) The computational cost to obtain a fidelity higher than 0.5. We plot the cost against the number of qubits to identify the scaling behavior. In (c) and (d), purple (square) and green (circle) dots correspond to cases without and with penalty term, respectively.

denominator of the adiabatic-condition term. Our findings prove otherwise and will lead to a deeper understanding of QA performance.

APPENDIX A: HAMILTONIAN RESCALING FOR COMPUTATIONAL COST

We explain the effect of the rescaled Hamiltonian given by $e^L \hat{H}_{\mathrm{QA}}(t)$ for the computational cost. Let us consider a time evolution by a Hamiltonian $\hat{H}_{\mathrm{QA}}(t')$ from a time $t' = 0$ to $t' = t$. We assume that $\hat{H}_{\mathrm{QA}}(t)$ is described as (3). The state is described as follows:

$$|\Psi(t)\rangle = \mathcal{T} \exp\left[-i \int_0^t \hat{H}_{\mathrm{QA}}(t')dt'\right]|\Psi(0)\rangle, \quad (A1)$$

where $\mathcal{T}$ is the time-ordering operator. Now, we also consider rescaled Hamiltonian and time. More specifically, we consider a time evolution by a Hamiltonian of $\hat{H}'_{\mathrm{QA}}(t) = e^L(\frac{t}{e^{-L}T}\hat{H}_p + (1 - \frac{t}{e^{-L}T})\hat{H}_d)$ from a time $t' = 0$ to $t' = e^{-L}t$ as follows:

$$|\Psi'(e^{-L}t)\rangle = \mathcal{T} \exp\left[-i \int_0^{e^{-L}t} \hat{H}'_{\mathrm{QA}}(t')dt'\right]|\Psi(0)\rangle. \quad (A2)$$

By changing the variable in Eq. (A2), $t' = e^{-L}\tilde{t}$, we obtain

$$|\Psi'(e^{-L}t)\rangle$$
$$= \mathcal{T} \exp\left[-i \int_0^t \left(\frac{e^{-L}\tilde{t}}{e^{-L}T}\hat{H}_p + \left(1 - \frac{e^{-L}\tilde{t}}{e^{-L}T}\right)\hat{H}_d\right)d\tilde{t}\right]|\Psi(0)\rangle$$
$$= \mathcal{T} \exp\left[-i \int_0^t \hat{H}_{\mathrm{QA}}(\tilde{t})d\tilde{t}\right]|\Psi(0)\rangle$$
$$= |\Psi(t)\rangle. \quad (A3)$$

From Eq. (A3), the time evolution of the state $|\Psi'(e^{-L}t)\rangle$ by $\hat{H}'_{\mathrm{QA}}(t)$ is the same as that of $|\Psi(t)\rangle$ by $\hat{H}_{\mathrm{QA}}(t)$. Thus, it is clearly shown that the Hamiltonian rescaling can shorten the annealing time from T to Te^{-L}.

The cost using the rescaling Hamiltonian is given by

$$Q_{\mathrm{res}} = \int_0^{e^{-L}T} \sigma[\hat{H}'_{\mathrm{QA}}(t'), |\Psi'(t')\rangle]dt'. \quad (A4)$$

We replace time t' to $t' = e^{-L}t$, then we have

$$Q_{\mathrm{res}} = \int_0^T \sigma[\hat{H}'_{\mathrm{QA}}(e^{-L}t), |\Psi'(e^{-L}t)\rangle]e^{-L}dt$$
$$= \int_0^T \sigma[\hat{H}_{\mathrm{QA}}(t), |\Psi(t)\rangle]dt$$
$$= Q. \quad (A5)$$

Thus, the rescaling for the Hamiltonian does not change the cost.

APPENDIX B: OPERATOR NORM OF PENALTY TERM

In Appendix B, we show that the operator norm of the penalty term (16) is bounded by $\Theta(L)$ as follows:

$$\|\hat{H}_{\mathrm{pena}}(t)\|_{\mathrm{op}} = \left\|\sum_{i=0}^{N-1}(E_i(0) - E_i(t))|E_i(t)\rangle\langle E_i(t)|\right\|_{\mathrm{op}}$$
$$\leqslant \left\|\sum_{i=0}^{N-1} E_i(0)|E_i(t)\rangle\langle E_i(t)|\right\|_{\mathrm{op}}$$
$$+ \left\|\sum_{i=0}^{N-1} E_i(t)|E_i(t)\rangle\langle E_i(t)|\right\|_{\mathrm{op}}$$
$$\leqslant \max_{0\leqslant i\leqslant N-1}|E_i(0)| + \|\hat{H}_{\mathrm{QA}}(t)\|_{\mathrm{op}}. \quad (B1)$$

Because $\hat{H}_{\mathrm{QA}}(0)$ is Hermitian, we can show that $\max_{0\leqslant i\leqslant N-1}|E_i(0)|$ is equal to $\|\hat{H}_{\mathrm{QA}}(0)\|_{\mathrm{op}}$. Thus, if the QA Hamiltonian is bounded by $\Theta(L)$ in the interval $0 \leqslant t \leqslant T$, the penalty term is also bounded by $\Theta(L)$.
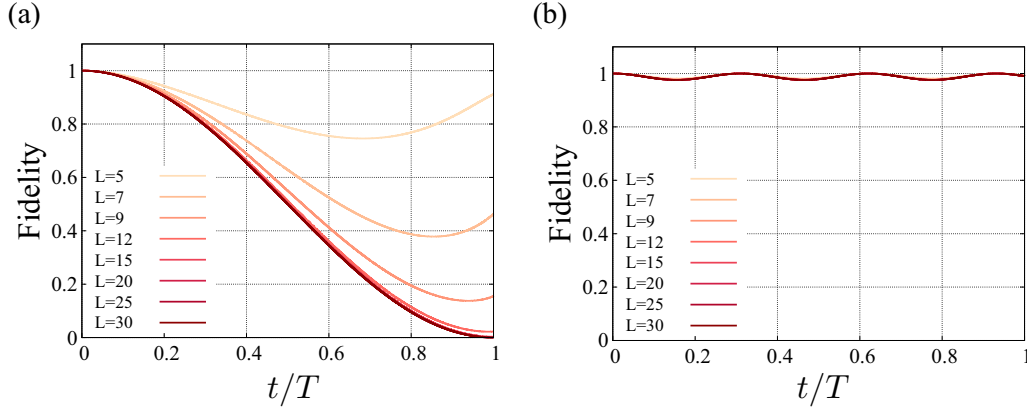
FIG. 6. Fidelity (a) without penalty term and (b) with penalty term as functions of t/T. Annealing time $T = 20$.

APPENDIX C: OPERATOR NORM OF COUNTERDIABATIC TERM

In Appendix C, we show that the operator norm of the counterdiabatic term becomes exponentially large at the first-order phase-transition point. The counterdiabatic term is given by [50–52]

$$\hat{H}_{cd}(t)$$
$$= i\sum_n (|\partial_t E_n(t)\rangle\langle E_n(t)| - \langle E_n(t)|\partial_t E_n(t)\rangle|E_n(t)\rangle\langle E_n(t)|).$$
(C1)

The operator norm of the counterdiabatic term is given by

$$\|\hat{H}_{cd}(t)\|_{op} = \max_{\||\Psi\rangle\|=1} \sqrt{\langle\Psi|\hat{H}_{cd}^\dagger(t)\hat{H}_{cd}(t)|\Psi\rangle}$$
$$= \sqrt{\sum_{n\neq m} |\langle E_m(t)|\partial_t E_n(t)\rangle|^2},$$
(C2)

where $\|\cdot\|$ denotes vector norm. As shown in Sec. III, as the system size increases, $|\langle E_m(t)|\partial_t E_n(t)\rangle|$ increases exponentially at the first-order phase-transition point. Thus, the operator norm of the counterdiabatic term becomes exponentially large at the first-order phase-transition point.

APPENDIX D: OPTIMAL SCHEDULING FUNCTION FOR THE ADIABATIC GROVER SEARCH WITH PENALTY TERM

Although we chose linear scheduling in the QA Hamiltonian [Eq. (3)] for our present study, we can use a more general QA Hamiltonian that uses a nonlinear scheduling (but monotonously increasing) function $f(t)$ as follows:

$$\hat{H}_{QA} = f(t)\hat{H}_p + [1 - f(t)]\hat{H}_d.$$
(D1)

For Eq. (D1), the adiabatic condition can be modified as follows:

$$\eta_{gen} \equiv \left|\frac{df(s)}{ds}\right| \frac{|\langle E_1(s)|d\hat{H}(s)/df|E_0(s)\rangle|}{\Delta(t)^2} \ll 1,$$
(D2)

where s denotes the normalized time t/T. We will optimize the scheduling function to minimize η_{gen}. We show that

through optimization, we can achieve an exponential speedup. From Eq. (D2), we obtain

$$\left|\frac{df}{ds}\right| \leqslant \epsilon \frac{T\Delta(f)^2}{|\langle E_1(t)|d\hat{H}(f)/df|E_0(t)\rangle|}$$
$$= \epsilon \frac{T\Delta(f)}{|\langle E_1(t)|d/df|E_0(t)\rangle|} \quad (\epsilon \ll 1).$$
(D3)

The scheduling function is optimized if equality holds in Eq. (D3). We have

$$\frac{df}{ds} = c\epsilon T(|\langle E_1(t)|d/df|E_0(t)\rangle|)^{-1}.$$
(D4)

Here, $\Delta(t)$ is now $\Theta(L^0)$; and thus, we set it to c. Substituting Eq. (21) into Eq. (D4), we obtain

$$\frac{ds}{df} = \frac{1}{c\epsilon T} \frac{\sqrt{(1-x)x}}{1 + 4f^2(1-x) - 4f(1-x)}$$
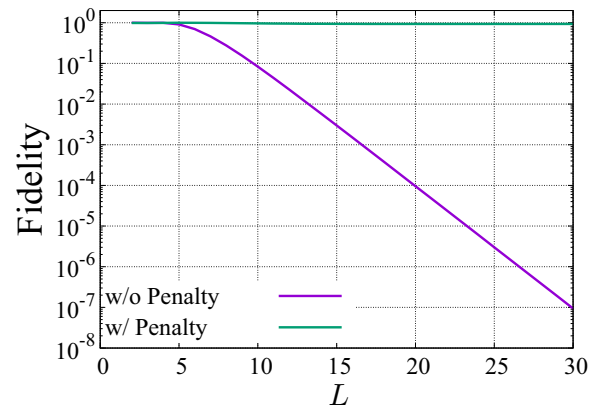$$= \frac{1}{c\epsilon T} \frac{\sqrt{(1-x)x}}{1 - 4(1-x)f(1-f)},$$
(D5)

FIG. 7. Scaling of fidelity at $t = T$. Purple (bottom) and green (top) lines denote the cases without and with penalty term, respectively. Annealing time $T = 20$.
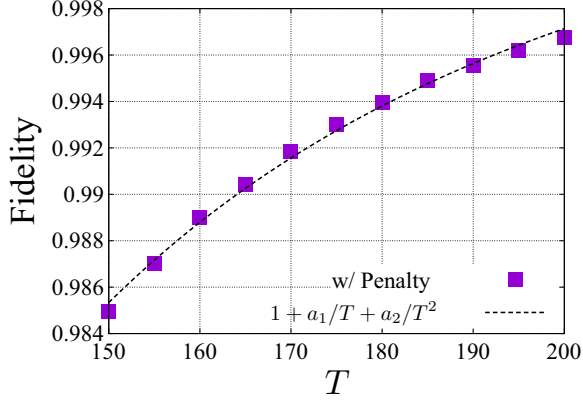
FIG. 8. Fidelity for the adiabatic Grover search with penalty term as functions of T. The system size is $L = 10$.

for which we define $x = 2^{-L}$. We solve this differential equation at $s(f = 0) = 0$ and obtain

$$s = \frac{1}{2c\epsilon T}\left[\text{Tan}^{-1}\sqrt{\frac{1-x}{x}} + \text{Tan}^{-1}\left((2f-1)\sqrt{\frac{1-x}{x}}\right)\right]. \tag{D6}$$

By solving this equation in reverse, we obtain the optimized scheduling function:

$$f(s) = \frac{1}{2} + \frac{1}{2}\sqrt{\frac{x}{1-x}}\tan\left[(2s-1)\text{Tan}^{-1}\sqrt{\frac{1-x}{x}}\right]. \tag{D7}$$

Considering the limit $x \to 0$ at $f(s = 1) = 1$, we have

$$T = \frac{1}{c\epsilon}\text{Tan}^{-1}\sqrt{\frac{1-x}{x}} \to \frac{\pi}{2c\epsilon}. \tag{D8}$$

Thus, the annealing time T does not depend on L. This finding can be intuitively understood as follows. While the penalty term provides a factor of $2^{L/2}$ to the adiabatic condition, optimization of the scheduling function also provides a factor of $2^{L/2}$ to it. Thus, in total, we obtain an improvement by a factor of 2^L.
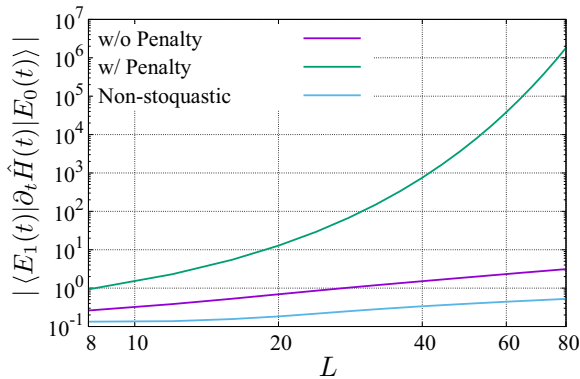

FIG. 9. Maximum of transition matrix for the ferromagnetic p-spin model. Purple (middle), green (top), and blue (bottom) lines denote cases without and with penalty term and with nonstoquastic term, respectively. Annealing time $T = 20$.

To validate our analytical results, we performed numerical simulations. Figures 6 and 7 show the time dependence of the fidelity and scaling of the fidelity at $t = T$, respectively. The annealing time was set to $T = 20$. As expected from Eq. (D8), when the penalty term and optimized scheduling are used, the fidelity at $t = T$ becomes $\Theta(L^0)$.

APPENDIX E: COST VS FIDELITY IN TWO-LEVEL SYSTEM

To explain a nonmonotonic behavior of the fidelity as a function of the cost in Fig. 3(c), we consider a simple case where the system is in a superposition of the ground and first excited states. Let us consider a state as follows:

$$|\Psi(t)\rangle = \frac{1}{\sqrt{1+|\epsilon|^2}}|E_0(t)\rangle + \frac{\epsilon}{\sqrt{1+|\epsilon|^2}}|E_1(t)\rangle. \tag{E1}$$

Using this state, the variance becomes

$$\langle\Psi(t)|\hat{H}^2(t)|\Psi(t)\rangle - \langle\Psi(t)|\hat{H}(t)|\Psi(t)\rangle^2$$

$$= \frac{1}{1+|\epsilon|^2}\left[E_0^2(t) + |\epsilon|^2 E_1^2(t)\right]$$

$$- \left(\frac{1}{1+|\epsilon|^2}\right)^2\left[E_0^2(t) + 2|\epsilon|^2 E_0(t)E_1(t) + |\epsilon|^4 E_1^2(t)\right]. \tag{E2}$$

For simplicity, let us consider the QA Hamiltonian with a penalty term. In this case, we can set $E_0(t) = 0$ and $E_1(t) = 1$. We obtain the standard deviation as follows:

$$\sigma = \frac{|\epsilon|}{1+|\epsilon|^2}. \tag{E3}$$

Since the fidelity is given by $F = |\langle E_0(t)|\Psi(t)\rangle|^2 = 1/(1+|\epsilon|^2)$, the standard deviation becomes

$$\sigma = \sqrt{F(1-F)}. \tag{E4}$$

To proceed with further calculations, we assume that an asymptotic form of the fidelity F is given by $F = 1 + a_1/T + a_2/T^2$. Also, we assume that the cost is given by $T\sqrt{F(1-F)}$ for simplicity. Then, we have

$$Q = \sqrt{-a_1^2 - a_2 - a_1 T} + \mathcal{O}(T^{-1/2}). \tag{E5}$$

To determine parameters a_1 and a_2, we use $F = 1 + a_1/T + a_2/T^2$ as a fitting function to reproduce the numerical results of the fidelity against the annealing time in the adiabatic Grover search with penalty term (Fig. 8). We obtain $a_1 = 4.31$ and $a_2 = -976$. Since a_1 is positive, the cost becomes a decreasing function of T in the region with large T.

APPENDIX F: COMPARISON OF PENALTY TERM WITH NONSTOQUASTIC HAMILTONIAN

In this Appendix, we compare the transition matrix with the penalty term and the nonstoquastic Hamiltonian in the p-spin model. The following antiferromagnetic interaction is introduced as the nonstoquastic Hamiltonian [18]:

$$\hat{H}_{\text{NS}} = L\left(\frac{1}{L}\sum_{i=1}^{L}\hat{\sigma}_i^x\right)^2. \tag{F1}$$

We use a QA Hamiltonian with the nonstoquastic Hamiltonian given as follows:

$$\hat{H}(t) = \frac{t}{T}\{\lambda\hat{H}_p + (1-\lambda)\hat{H}_{NS}\} + \left(1 - \frac{t}{T}\right)\hat{H}_d, \quad (F2)$$

where $\hat{H}_p$ is given by Eq. (28) and $\hat{H}_d = -\sum_{i=1}^{L}\hat{\sigma}_i^x$. To avoid a first-order phase transition, we set $p = 5$ and $\lambda = 0.1$ [18]. In Fig. 9, we show the scaling of the transition matrix in cases with and without our penalty term, and with the nonstoquastic

Hamiltonian. It can be observed that the nonstoquastic Hamiltonian does not lead to an exponential increase in the size of the transition matrix, even with an exponentially wide gap at the first-order phase-transition point. According to our general framework, opening the gap using the penalty term leads to an exponential increase in the size of the transition matrix. In contrast to our penalty term, the nonstoquastic Hamiltonian (F1) changes not only the eigenenergy but also the energy eigenstate.

[1] B. Apolloni, C. Carvalho, and D. De Falco, Quantum stochastic optimization, Stoch. Process. Appl. **33**, 233 (1989).

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

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

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

[5] E. Farhi, J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren, and D. Preda, A quantum adiabatic evolution algorithm applied to random instances of an NP-complete problem, Science **292**, 472 (2001).

[6] A. Das and B. K. Chakrabarti, Colloquium: Quantum annealing and analog quantum computation, Rev. Mod. Phys. **80**, 1061 (2008).

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

[8] F. Barahona, On the computational complexity of Ising spin glass models, J. Phys. A: Math. Gen. **15**, 3241 (1982).

[9] A. Lucas, Ising formulations of many NP problems, Front. Phys. **2**, (2014).

[10] W. Lechner, P. Hauke, and P. Zoller, A quantum annealing architecture with all-to-all connectivity from local interactions, Sci. Adv. **1**, e1500838 (2015).

[11] T. Kato, On the adiabatic theorem of quantum mechanics, J. Phys. Soc. Jpn. **5**, 435 (1950).

[12] A. Messiah, *Quantum Mechanics*, Dover Books on Physics (Dover, New York, 2014).

[13] S. Jansen, M.-B. Ruskai, and R. Seiler, Bounds for the adiabatic approximation with applications to quantum computation, J. Math. Phys. **48**, 102111 (2007).

[14] S. Morita and H. Nishimori, Mathematical foundation of quantum annealing, J. Math. Phys. **49**, 125210 (2008).

[15] M. H. S. Amin, Consistency of the adiabatic theorem, Phys. Rev. Lett. **102**, 220401 (2009).

[16] Y. Kimura and H. Nishimori, Rigorous convergence condition for quantum annealing, J. Phys. A: Math. Theor. **55**, 435302 (2022).

[17] R. MacKenzie, E. Marcotte, and H. Paquette, Perturbative approach to the adiabatic approximation, Phys. Rev. A **73**, 042104 (2006).

[18] Y. Seki and H. Nishimori, Quantum annealing with antiferromagnetic fluctuations, Phys. Rev. E **85**, 051112 (2012).

[19] Y. Seki and H. Nishimori, Quantum annealing with antiferromagnetic transverse interactions for the Hopfield model, J. Phys. A: Math. Theor. **48**, 335301 (2015).

[20] Y. Susa, T. Imoto, and Y. Matsuzaki, Nonstoquastic catalyst for bifurcation-based quantum annealing of the ferromagnetic p-spin model, Phys. Rev. A **107**, 052401 (2023).

[21] H. Nishimori and K. Takada, Exponential enhancement of the efficiency of quantum annealing by non-stoquastic Hamiltonians, Frontiers in ICT **4**, 2 (2017).

[22] Y. Susa, Y. Yamashiro, M. Yamamoto, and H. Nishimori, Exponential speedup of quantum annealing by inhomogeneous driving of the transverse field, J. Phys. Soc. Jpn. **87**, 023002 (2018).

[23] Y. Susa, Y. Yamashiro, M. Yamamoto, I. Hen, D. A. Lidar, and H. Nishimori, Quantum annealing of the *p*-spin model under inhomogeneous transverse field driving, Phys. Rev. A **98**, 042326 (2018).

[24] Y. Susa and H. Nishimori, Performance enhancement of quantum annealing under the Lechner–Hauke–Zoller scheme by non-linear driving of the constraint term, J. Phys. Soc. Jpn. **89**, 044006 (2020).

[25] Y. Matsuzaki, H. Hakoshima, K. Sugisaki, Y. Seki, and S. Kawabata, Direct estimation of the energy gap between the ground state and excited state with quantum annealing, Jpn. J. Appl. Phys. **60**, SBBI02 (2021).

[26] T. Imoto, Y. Seki, Y. Matsuzaki, and S. Kawabata, Quantum annealing with twisted fields, New J. Phys. **24**, 113009 (2022).

[27] Y. Mori, S. Kawabata, and Y. Matsuzaki, How to experimentally evaluate the adiabatic condition for quantum annealing, Sci. Rep. **14**, 8177 (2024).

[28] B. F. Schiffer, J. Tura, and J. I. Cirac, Adiabatic spectroscopy and a variational quantum adiabatic algorithm, PRX Quantum **3**, 020347 (2022).

[29] T. Imoto, Y. Seki, and Y. Matsuzaki, Quantum annealing with symmetric subspaces, arXiv:2209.09575.

[30] T. Kadowaki and H. Nishimori, Greedy parameter optimization for diabatic quantum annealing, Phil. Trans. R. Soc. A **381**, 20210416 (2023).

[31] G. G. Cabrera and R. Jullien, Role of boundary conditions in the finite-size Ising model, Phys. Rev. B **35**, 7062 (1987).

[32] C. R. Laumann, R. Moessner, A. Scardicchio, and S. L. Sondhi, Quantum adiabatic algorithm and scaling of gaps at first-order quantum phase transitions, Phys. Rev. Lett. **109**, 030502 (2012).

[33] J. Tsuda, Y. Yamanaka, and H. Nishimori, Energy gap at first-order quantum phase transitions: an anomalous case, J. Phys. Soc. Jpn. **82**, 114004 (2013).

[34] B. Altshuler, H. Krovi, and J. Roland, Anderson localization makes adiabatic quantum optimization fail, Proc. Natl. Acad. Sci. USA **107**, 12446 (2010).

[35] A. P. Young, S. Knysh, and V. N. Smelyanskiy, First-order phase transition in the quantum adiabatic algorithm, Phys. Rev. Lett. **104**, 020502 (2010).

[36] T. S. Cubitt, D. Perez-Garcia, and M. M. Wolf, Unde-cidability of the spectral gap, Nature (London) **528**, 207 (2015).

[37] L. Mandelstam and I. Tamm, The uncertainty relation between energy and time in non-relativistic quantum mechanics, J. Phys. (USSR) **9**, 249 (1945).

[38] L. Vaidman, Minimum time for the evolution to an orthogonal quantum state, Am. J. Phys. **60**, 182 (1992).

[39] N. Margolus and L. B. Levitin, The maximum speed of dynamical evolution, Physica D **120**, 188 (1998).

[40] S. Deffner and E. Lutz, Energy–time uncertainty relation for driven quantum systems, J. Phys. A: Math. Theor. **46**, 335302 (2013).

[41] S. Deffner and S. Campbell, Quantum speed limits: from Heisenberg's uncertainty principle to optimal quantum control, J. Phys. A: Math. Theor. **50**, 453001 (2017).

[42] K. Suzuki and K. Takahashi, Performance evaluation of adiabatic quantum computation via quantum speed limits and possible applications to many-body systems, Phys. Rev. Res. **2**, 032016(R) (2020).

[43] T. Hatomura and K. Takahashi, Controlling and exploring quantum systems by algebraic expression of adiabatic gauge potential, Phys. Rev. A **103**, 012220 (2021).

[44] B. W. Reichardt, The quantum adiabatic optimization algorithm and local minima, in *Proceedings of the Thirty-Sixth Annual ACM Symposium on Theory of Computing STOC '04* (Association for Computing Machinery, New York, NY, 2004), pp. 502–510.

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

[46] L. K. Grover, Quantum mechanics helps in searching for a needle in a haystack, Phys. Rev. Lett. **79**, 325 (1997).

[47] J. Roland and N. J. Cerf, Quantum search by local adiabatic evolution, Phys. Rev. A **65**, 042308 (2002).

[48] M. Kac, G. Uhlenbeck, and P. Hemmer, On the van der Waals theory of the vapor-liquid equilibrium. I. Discussion of a one-dimensional model, J. Math. Phys. **4**, 216 (1963).

[49] A. Campa, T. Dauxois, and S. Ruffo, Statistical mechanics and dynamics of solvable models with long-range interactions, Phys. Rep. **480**, 57 (2009).

[50] M. Demirplak and S. A. Rice, Adiabatic population transfer with control fields, J. Phys. Chem. A **107**, 9937 (2003).

[51] M. V. Berry, Transitionless quantum driving, J. Phys. A: Math. Theor. **42**, 365303 (2009).

[52] D. Guéry-Odelin, A. Ruschhaupt, A. Kiely, E. Torrontegui, S. Martínez-Garaot, and J. G. Muga, Shortcuts to adiabaticity: Concepts, methods, and applications, Rev. Mod. Phys. **91**, 045001 (2019).

[53] G. E. Santoro, R. Martoňák, E. Tosatti, and R. Car, Theory of quantum annealing of an Ising spin glass, Science **295**, 2427 (2002).

[54] B. Heim, T. F. Rønnow, S. V. Isakov, and M. Troyer, Quantum versus classical annealing of Ising spin glasses, Science **348**, 215 (2015).

[55] T. Jörg, F. Krzakala, J. Kurchan, A. C. Maggs, and J. Pujos, Energy gaps in quantum first-order mean-field–like transitions: The problems that quantum annealing cannot solve, Europhys. Lett. **89**, 40004 (2010).

[56] M. Filippone, S. Dusuel, and J. Vidal, Quantum phase transitions in fully connected spin models: An entanglement perspective, Phys. Rev. A **83**, 022327 (2011).

[57] V. Bapst and G. Semerjian, On quantum mean-field models and their quantum annealing, J. Stat. Mech.: Theory Exp. (2012) P06007.