Why and When Pausing is Beneficial in Quantum Annealing

Huo Chen[1,2,*] and Daniel A. Lidar[1,2,3,4]

[1] *Department of Electrical Engineering, University of Southern California, Los Angeles, California 90089, USA*
[2] *Center for Quantum Information Science & Technology, University of Southern California, Los Angeles, California 90089, USA*
[3] *Department of Chemistry, University of Southern California, Los Angeles, California 90089, USA*
[4] *Department of Physics and Astronomy, University of Southern California, Los Angeles, California 90089, USA*

(Received 11 May 2020; accepted 9 July 2020; published 31 July 2020)

Recent empirical results using quantum-annealing hardware have shown that midanneal pausing has a surprisingly beneficial impact on the probability of finding the ground state for a variety of problems. A theoretical explanation of this phenomenon has thus far been lacking. Here we provide an analysis of pausing using a master-equation framework, and derive conditions for the strategy to result in a success probability enhancement. The conditions, which we identify through numerical simulations and then prove to be sufficient, require that relative to the pause duration the relaxation rate is large and decreasing right after crossing the minimum gap, small and decreasing at the end of the anneal, and is also cumulatively small over this interval, in the sense that the system does not thermally equilibrate. This establishes that the observed success probability enhancement can be attributed to incomplete quantum relaxation, i.e., is a form of beneficial nonequilibrium coupling to the environment.

DOI: 10.1103/PhysRevApplied.14.014100

I. INTRODUCTION

Quantum annealing [1–4] stands out among the multitude of concurrent approaches being developed to explore quantum computing, as having achieved the largest scale to date when measured in terms of the sheer number of controllable qubits. Today's commercial quantum annealers feature thousands of superconducting flux qubits and are being used routinely to test whether this approach can provide a quantum advantage over classical computing [5–11]. While there is no consensus that such an advantage has been demonstrated, there is significant progress on the development of "software" methods that improve quantum-annealing performance. Such methods take advantage of the advanced control capabilities of quantum annealers to implement protocols that result in higher success probabilities, shorter times to solution, faster equilibration, etc. Continued progress in this direction is clearly critical as a complementary approach to improving the underlying hardware by reducing the physical source of noise and decoherence.

Among the various empirical protocols that have been developed to improve the performance of quantum annealing, such as error suppression and correction [12–15] and inhomogeneous driving [16–19], the midanneal-pausing protocol stands out as particularly powerful. Pausing superficially resembles the idea of slowing down near the minimum gap, as in the optimal schedule for the Grover problem [3,20], but the context here is entirely different due to the fact that pausing happens in an open system subject to thermal relaxation. The first study [21] to systematically test this approach empirically using a D-Wave 2000Q device [22], demonstrated a dramatic improvement in the probability of finding the ground state (i.e., the success probability) when an anneal pause was inserted shortly after crossing the minimum gap. Follow-up studies confirmed that pausing is advantageous on different problems such as portfolio-optimization problems [23] and training deep generative machine-learning models [24]. Numerical studies [25,26] of the p-spin model also agree with these empirical results. However, despite a useful qualitative explanation offered for the thermalization mechanism by which pausing improves success probabilities [21], a thorough analysis of the exact mechanism of this phenomenon is still lacking. Here we provide such an analysis, and identify sufficient conditions for pausing to provide an enhancement.

Our analysis is based on a detailed investigation of a quantum two-level system model coupled to an Ohmic bath. The two-level system can be either a single qubit or a multiqubit system whose lowest two energy levels are separated by a large gap from the rest of the spectrum. The analysis builds on tools from the theory of open quantum systems [27–29], specifically master

*huochen@usc.edu

equations appropriate for time-dependent (driven) Hamiltonians [30–35]. Through numerical investigation we identify a set of sufficient conditions, stated in terms of the properties of the relaxation rate along the anneal, and prove a theorem guaranteeing that it is advantageous to pause midanneal. The advantage gained is a higher success probability than is attainable without pausing.

We thus establish, in a rigorous sense, that there exists a nontrivial optimal pausing point under a set of reasonable assumptions. We do not identify the optimal pausing duration, but we do prove that the optimal pausing point occurs after the minimum gap, in accordance with the prior empirical and numerical evidence. This result is stated in Theorem 1 below, which can be summarized as saying that an optimal pausing point exists if the relaxation rate right after the minimum gap is large relative to the pause duration but small at the end of the anneal, decreases right after crossing the minimum gap and also at the end of the anneal, and is also cumulatively small over this interval, in the sense that the system does not fully thermally equilibrate.

The structure of this paper is as follows. In Sec. II we define our model of the two-level system. In the multiqubit case this involves deriving an effective Hamiltonian for the projection to the low-energy subspace of the full Hamiltonian. We introduce a certain parametrization of the gap and the geometric phase that ensures the problem is hard for quantum annealing, in the sense that the success probability is low even on a timescale that is large compared to the inverse of the minimum spectral gap along the anneal. In Sec. III we treat the same model as an open quantum system using master-equation techniques, specifically the Redfield equation with and without the rotating wave approximation, and the adiabatic master equation. Then, in Sec. IV we introduce a pause into the annealing schedule and study its effects. We first demonstrate numerically that an optimal pausing position exists before the end of the anneal, depending on the monotonicity properties of the relaxation rate after the minimum gap is crossed. Building on these observations we then prove a theorem establishing sufficient conditions for the existence of such an optimal pausing point. We conclude in Sec. V, and present additional technical details in the Appendixes.

II. "HARD" SINGLE-QUBIT AND MULTIQUBIT CLOSED-SYSTEM MODELS

In this section we consider two scenarios: single-qubit annealing, and a projected two-level system (TLS) arising from multiqubit annealing. We define a model that makes these problems "hard" for quantum annealing, in the sense of a small success probability even over a timescale that is long compared to the heuristic adiabatic timescale (given by the inverse of the minimum gap along the anneal path).

A. Single qubit

We write the single-qubit annealing Hamiltonian in the form

$$H_S(s) = -\frac{1}{2}\left[A(s)Z + B(s)X\right], \tag{1}$$

where $s = t/t_f$ is the dimensionless instantaneous time, t is the actual time, t_f is the total anneal time, and $A(s)$ and $B(s)$ are the annealing schedules. Note that we permuted the Pauli matrices X and Z of the conventional single-qubit annealing Hamiltonian in order to have the same expression for both the single qubit and projected TLS cases. After transforming to the adiabatic frame [36], the dimensionless interaction picture Hamiltonian becomes

$$\tilde{H}_S(s) = \frac{1}{2}\left[\frac{d\theta}{ds}Y - t_f\,\Omega(s)Z\right], \tag{2}$$

where $\Omega(s)$ and $\theta(s)$ are a reparameterization of the annealing schedules: $A(s) = \Omega(s)\cos\theta(s)$ and $B(s) = \Omega(s)\sin\theta(s)$ (see Appendix A for a detailed explanation). We call $\theta(s)$ the annealing angle and $\dot{\theta}(s)$ the angular progression. The term $\dot{\theta}Y/2$ has its origin as a geometric phase [37]. Loosely, $\Omega(s)$ corresponds to the time-dependent gap and $d\theta/ds$ corresponds to how fast or slow the Hamiltonian changes. For a typical single-qubit annealing process, the boundary condition $\theta(0) = 0$ and $\theta(1) = \pi/2$ needs to be satisfied [noticing that in Eq. (1) we permuted the Pauli X and Z matrices in the standard notation].

B. Projected TLS from a multiqubit model

For general multiqubit annealing, the Hamiltonian is

$$H_S(s) = A(s)H_d + B(s)H_p = \sum_n E_n(s)|n(s)\rangle\langle n(s)|, \tag{3}$$

where $\{|n(s)\rangle\}$ is the instantaneous energy eigenbasis and $\{E_n(s)\}$ are the instantaneous energies. H_d and H_p are the driver and problem Hamiltonian, respectively. Henceforth we assume that $H_S = (H_S)^T$, i.e., that $H_S(s)$ is real for all s. The system density matrix can be written in the instantaneous energy eigenbasis:

$$\rho(s) = \sum_{nm}\rho_{nm}|n\rangle\langle m|. \tag{4}$$

We call the associated matrix $\tilde{\rho} = [\rho_{nm}]$ the density matrix in the adiabatic frame, and show in Appendix B that it obeys the von Neumann equation $\dot{\tilde{\rho}} = -i\left[\tilde{H}, \tilde{\rho}\right]$ with the

effective Hamiltonian

$$\tilde{H} = \begin{pmatrix} t_f E_0 & -i\langle 0|\dot{1}\rangle & \dots \\ i\langle 0|\dot{1}\rangle & t_f E_1 & \dots \\ \vdots & & \ddots \end{pmatrix}. \tag{5}$$

If we truncate the effective Hamiltonian (5) to the lowest two energy levels and shift it by a constant term, we find

$$\tilde{H}_2 = \langle 0|\dot{1}\rangle Y - \frac{t_f \Omega(s)}{2} Z, \tag{6}$$

where $\Omega(s) = E_1 - E_0$ is the energy gap between the lowest two energy levels. We call this the projected TLS Hamiltonian. An alternative way to derive this effective Hamiltonian is via the well-known adiabatic intertwiner (see, e.g., Ref. [38]). This TLS approximation is valid when (i) there is a large gap separating the two-level subspace from higher excited state (where "large" is in the sense of the adiabatic theorem [39]), and (ii) the geometric terms connecting the two-level subspace to the higher levels in the adiabatic frame are negligible [40]. One may also invoke the Schrieffer-Wolff transformation to establish similar conditions [41,42].

Since the annealing Hamiltonian [Eq. (3)] is real, $\langle 0|\dot{1}\rangle$ in Eq. (5) is also real. This effective Hamiltonian is equivalent to Eq. (2) with $\langle 0|\dot{1}\rangle$ playing the role of $\dot{\theta}/2$. Thus, we can define the angular progression as $\dot{\theta} = 2\langle 0|\dot{1}\rangle$ for the projected TLS. Having done so, a general TLS Hamiltonian can also be written as Eq. (2), where the annealing angle θ in the general case does not need to satisfy the same boundary condition as in the single-qubit case.

C. Hard-problem instances from gap and angular-progression considerations

We call an instance "easy" when a high ground-state probability is achieved within an annealing time that is much shorter than the timescale set by the inverse of the minimum gap along the anneal (we refer to this as the "heuristic" adiabatic condition; it is not to be confused with the rigorous adiabatic condition, which provides a sufficient condition for convergence to the ground state [39,43]). Conversely, we call an instance "hard" when the ground-state probability is low for such an annealing time.

Our strategy is to create toy models that share the same features as certain known hard examples for quantum annealing [25,44]. By closely examining those problems, we identify one crucial characteristic they share: a sharp peak in the angular progression appears at the minimum gap, along with a π jump of the annealing angle θ (see Appendix C where these examples are illustrated).

We take a reverse-engineering approach by first specifying the analytic form of the gap $\Omega(s)$ and angular progression $\dot{\theta}(s)$ in the adiabatic frame. The gap is parametrized as a Gaussian in the form

$$\Omega(s) = E_0\left[1 - (1-\Delta)e^{-(s-\mu_g)^2/2\alpha_g^2}\right], \tag{7}$$

where the parameters Δ, μ_g, and α_g, respectively, control the gap size, position, and width. The angular progression can also be chosen as Gaussian

$$\dot{\theta}(s) = Ce^{-(s-\mu_\theta)^2/2\alpha_\theta^2}, \tag{8}$$

with position and width parameter μ_θ, α_θ. The normalization constant C is chosen according to the boundary condition. We discuss both the single-qubit and projected TLS cases.

D. Single qubit

As a consequence of the aforementioned boundary conditions $\theta(0) = 0$ and $\theta(1) = \pi/2$. the normalization constant is

$$C = \sqrt{\frac{\pi}{2}}\frac{1}{2\alpha_\theta}, \tag{9}$$

and the annealing angle resulting from Eq. (8) is

$$\theta(s) = \frac{\pi}{4}\left[\mathrm{erf}\left(\frac{\mu_\theta}{2\alpha_\theta}\right) + \mathrm{erf}\left(\frac{s-\mu_\theta}{\sqrt{2}\alpha_\theta}\right)\right]. \tag{10}$$

In order to ensure the hardness of the problem and monotonic schedules, we need to overlap the peak region of $\dot{\theta}$ and the minimum gap, i.e., $\mu_\theta \approx \mu_g$. In such a region, the diabatic term is much larger than the adiabatic term and the Landau–Zener transition is strong. A choice of such schedules is illustrated in Fig. 1. It is important to note that, in this construction, the hardness of the problem is not solely determined by the minimum gap Δ. Indeed, the rigorous

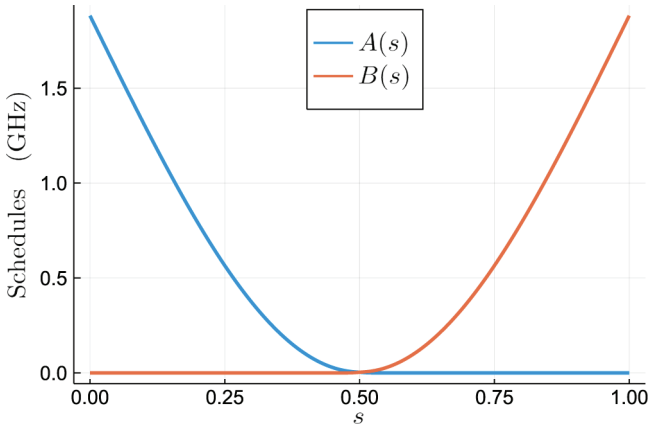

FIG. 1. Example schedules with the following parameter choices: $\mu_\theta = \mu_g = 0.5, \alpha_g = 0.5, E_0 = 15/\pi$ GHz, $\Delta = 0.001$, and $\alpha_\theta = 1/100$.

adiabatic condition involves the derivative of the Hamiltonian as well [39,43]. An example of an easy instance with a small minimum gap is given in Appendix D.

E. Projected TLS

The boundary conditions for the projected TLS are different from the single-qubit case because there is no simple relation between the schedules and the annealing angle. However, a common feature of the small gap instances [25,44] is a localized pulse of angular progression that is present at the minimum gap. Also, this pulse induces a step-function-like π shift of the annealing angle across this region, leading to a near-perfect Landau-Zener transition. The simplest toy model we can construct is to keep the Gaussian form of the gap [Eq. (7)] and angular progression [Eq. (8)] but use a different boundary condition $\theta(0) = 0$ and $\theta(1) = \pi$, which comes from the examination of both the p-spin model [25] and the D-Wave 16-qubit gadget problem [44]. In this case, the normalization constant becomes $C = \sqrt{\pi}/\sqrt{2}\alpha_\theta$.

However, we stress that the core of this construction is the angular-progression pulse at the minimum gap. There is no constraint on $\dot{\theta}(s)$ at other s as long as the system can follow its eigenstates before and after the pulse. In fact, in the problem studied in Ref. [44], the annealing angle first gradually decreases to a nonzero value before the π jump (see Appendix C).

III. OPEN-SYSTEM MODEL

For the open quantum-system model, we directly start with the multiqubit case. We adopt a standard noise model for quantum annealing: each qubit couples to a bosonic bath via a system operator O:

$$H_{\mathrm{SB}} = \sum_{\alpha k} g_{\alpha k} O_\alpha \otimes (b^\dagger_{\alpha k} + b_{\alpha k}) = \sum_\alpha g_\alpha O_\alpha \otimes \mathcal{B}_\alpha. \quad (11)$$

Here g_α and O_α are dimensionless and $\mathcal{B}_\alpha$ has dimensions of energy. The parameters g_α serve as expansion variables, which can later be set to 1. After moving to the adiabatic frame, the system-bath interaction becomes

$$\tilde{H}_{\mathrm{SB}}(s) = t_f \sum_{\alpha mn} g_\alpha O^{mn}_\alpha(s) \otimes \mathcal{B}_\alpha, \quad (12)$$

where

$$O^{mn}_\alpha(s) = \langle m(s)|O_\alpha|n(s)\rangle |m\rangle\langle n|. \quad (13)$$

By defining

$$S_\alpha(s) = \sum_{mn} O^{mn}_\alpha(s), \quad (14)$$

the total projected TLS Hamiltonian in the adiabatic frame can be further simplified as

$$\tilde{H} = \frac{1}{2}\left[\dot{\theta}(s)Y - t_f\,\Omega(s)Z\right] + \sum_\alpha g_\alpha t_f\, S_\alpha(s) \otimes \mathcal{B}_\alpha + H_B. \quad (15)$$

From now on, for conciseness we omit the tilde symbol for adiabatic frame operators. We investigate three approaches for solving the open-system dynamics.

A. Redfield equation

Before proceeding, we define the bath correlation function

$$C_{\alpha\alpha'}(s,s') = \mathrm{Tr}[\mathcal{B}_\alpha(s)\mathcal{B}_{\alpha'}(s')\rho_B] = C^*_\alpha(s',s) \quad (16)$$

in terms of the rotated $\mathcal{B}_\alpha$ operator

$$\mathcal{B}_\alpha(s) = U^\dagger_B(s)\mathcal{B}_\alpha U_B(s), \quad (17)$$

where $U_B(s) = \exp\left[-it_f H_B s\right]$ is the free bath evolution. We call a set of $[\mathcal{B}_\alpha(s)]_\alpha$ independent if $C_{\alpha\alpha'}(s,s') = \delta_{\alpha\alpha'}C_\alpha(s,s')\ \forall\alpha,\alpha'$, and identical if $C_\alpha(s,s') = C(s,s')\ \forall\alpha$.

By assuming $[\mathcal{B}_\alpha(s)]_\alpha$ are independent, the Redfield equation in the adiabatic frame can be shown to be [36]

$$\dot{\rho}_S(s) = -i[H_S(s),\rho_S(s)] - \sum_\alpha (g_\alpha t_f)^2[S_\alpha(s),\Lambda_\alpha(s)\rho_S(s)] + \mathrm{h.c.}, \quad (18)$$

where

$$\Lambda_\alpha(s) = \int_0^s ds' C_\alpha(s,s')U(s,s')S_\alpha(s')U^\dagger(s,s'), \quad (19)$$

and

$$U(s,s') = T_+ \exp\left\{-it_f \int_{s'}^s H_S(s'')\,ds''\right\}, \quad (20)$$

and T_+ denotes time ordering. An important observation is that, after moving to the adiabatic frame, the transformed system Hamiltonian [Eq. (2)] has a different gap than the original one [Eq. (1)], due to the rescaling by t_f. We define the new gap (in energy units) in the adiabatic frame as

$$\Delta(s) = \sqrt{\dot{\theta}^2(s)/t_f^2 + \Omega^2(s)}. \quad (21)$$

B. Redfield equation with rotating wave approximation

From our construction of $\Omega(s)$ and $\dot{\theta}(s)$ [Eqs. (7) and (8)], it follows that at the minimum gap point of $s = \alpha_g$,

$\Delta(s)$ is large. As a consequence, we can safely apply the rotating wave approximation (RWA) with the adiabatic frame Redfield Eq. (18) without worrying about the presence of a small gap. After the RWA, Eq. (19) becomes

$$\begin{aligned}\dot{\rho}_S = &-i[H_S(s) + H_{\rm LS}, \rho_S] \\ &- \Gamma_d \left(\rho_{ba}|b\rangle\langle a| + \rho_{ab}|a\rangle\langle b|\right) \\ &+ \Gamma_t(\rho_{aa} - e^{-\beta\Delta}\rho_{bb})\left(|b\rangle\langle b| - |a\rangle\langle a|\right), \end{aligned} \tag{22}$$

where $\rho_{ab} = \langle a|\rho_S|b\rangle$ with $\{|a\rangle, |b\rangle\}$ being the ground and excited states of $H_S(s)$. All quantities in Eq. (22) are s dependent, and the effective dephasing and thermalization rates Γ_d and Γ_t, respectively, are given by [36,45]

$$\begin{aligned}\Gamma_d(s) = &\frac{t_f}{2}\Gamma_t(s)\left[1 + e^{-\beta\Delta(s)}\right] \\ &+ \frac{t_f}{2}\sum_\alpha \gamma_\alpha(0)(S_\alpha^{aa} - S_\alpha^{bb})^2, \end{aligned} \tag{23a}$$

$$\Gamma_t(s) = t_f \sum_\alpha g_\alpha^2 \gamma_\alpha(\Delta)|S_\alpha^{ab}|^2, \tag{23b}$$

where the projected system-bath coupling operators are

$$S_\alpha^{ab} = \langle a|S_\alpha(s)|b\rangle. \tag{24}$$

The Lamb shift is

$$H_{\rm LS}(s) = \sum_\alpha g_\alpha^2 t_f \left\{\mathcal{S}_\alpha[\Delta(s)]|b\rangle\langle b| + \mathcal{S}_\alpha[-\Delta(s)]|a\rangle\langle a|\right\}. \tag{25}$$

The functions $\gamma_\alpha(\omega)/2$ and $\mathcal{S}_\alpha(\omega)$ are the real and imaginary parts of the noise spectral density (the one-sided Fourier transform of the bath-correlation function).

C. Adiabatic master equation

Outside the peak region of the angular progression, Eq. (22) becomes the adiabatic master equation (AME) [30]. The AME is a special case of Eq. (22). It can be derived by ignoring the geometric part of the Hamiltonian in the unitary part of the Redfield Eq. (18), which holds in the adiabatic limit $t_f \gg 1$. It has the same form as Eq. (22), with $\Delta(s)$ being the physical gap [Eq. (21) with $\dot{\theta} = 0$] and $\{|a\rangle, |b\rangle\}$ being the instantaneous eigenstates.

IV. PAUSING

So far we only considered the case of a linear annealing parameter $s = t/t_f$. In this section we study the effect of including a pause.

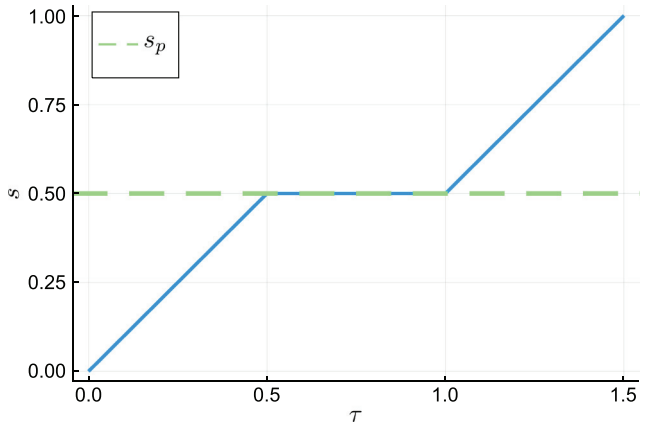

FIG. 2. Example of annealing parameter s against dimensionless time $\tau = t/t_f$. The pause position s_p and duration s_d are chosen as $s_p = 0.5$ and $s_d = 0.5$.

A. Model

To incorporate a pause, let us define $s = s(\tau; s_p, s_d)$ where $\tau = t/t_f$ is the dimensionless time, and where s_p and s_d are the pausing position and pausing duration, respectively. The explicit form of $s(\tau; s_p, s_d)$ is given below, where from now on we suppress the explicit dependence on s_p and s_d for simplicity, and is illustrated in Fig. 2:

$$s(\tau) = \begin{cases} \tau & \tau \le s_p \\ s_p & s_p < \tau \le s_p + s_d \\ \tau - s_d & s_p + s_d < \tau \le 1 + s_d \end{cases}. \tag{26}$$

Note that pausing increases the total annealing time to

$$t_f' = \tau_f t_f, \quad \tau_f = 1 + s_d. \tag{27}$$

To prevent confusion, *henceforth we will denote by $Q(s)$ the original quantity and by $Q(\tau)$ the corresponding paused quantity*, where Q can be any function or operator.

The dimensionless Hamiltonian in the adiabatic frame then becomes

$$\begin{aligned}H(\tau) = &\frac{1}{2}\left[\frac{d\theta}{d\tau}X - t_f'\,\Omega(\tau)Z\right] + t_f' \\ &\times \sum_\alpha g_\alpha S_\alpha(\tau) \otimes \mathcal{B}_\alpha + H_B, \end{aligned} \tag{28}$$

where $d\theta/d\tau$ can be calculated using the chain rule:

$$\frac{d\theta}{d\tau} = \begin{cases} 0 & s_p < \tau \le s_p + s_d \\ d\theta/ds & \text{elsewhere} \end{cases}. \tag{29}$$

B. Numerical results

For our numerical simulations, we use the Gaussian gap [Eq. (7)] and angular progression [Eq. (8)] with two different boundary conditions: $\theta(1) = \pi/2$ and $\theta(1) = \pi$. The

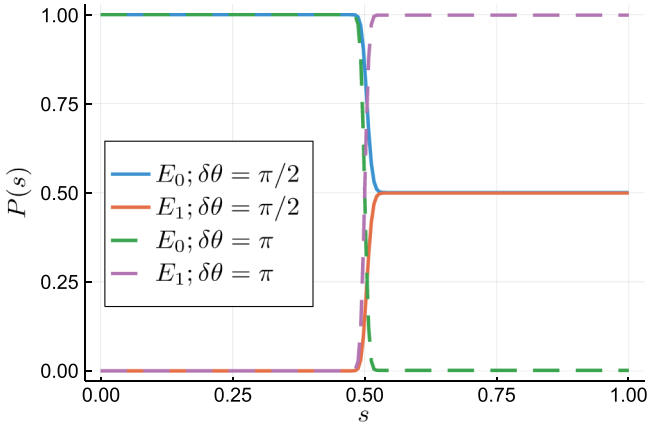

FIG. 3. Populations of instantaneous eigenstates during an anneal with total time $t_f = 100$ ns. $\delta\theta$ denotes the jump of the annealing angle θ across the minimum gap region. In our model, the magnitude of this jump is directly determined by the boundary condition.

other parameters are the same as in Fig. 1. The instantaneous populations during a 100 ns anneal are shown in Fig. 3. We observe that the role of boundary conditions in our setup (with a single minimum gap) is to determine the portion of population transferred to the excited state when traversing the minimum gap.

The simplest $S_\alpha(s)$ we consider is inspired by the single-qubit model with both dephasing and relaxation noise. In this case, $\{O_\alpha\} \equiv \{X, Z\}$ in the interaction Hamiltonian (11). In the adiabatic frame we have $[S_\alpha(s)] \equiv [Z(s), X(s)]$, where

$$Z(s) = \cos[\theta(s)]Z - \sin[\theta(s)]X, \tag{30a}$$

$$X(s) = \cos[\theta(s)]X + \sin[\theta(s)]Z. \tag{30b}$$

Furthermore, we assume the $[\mathcal{B}_\alpha(s)]_\alpha$ are independent and identical to an Ohmic spectral density [30]:

$$\gamma(\omega) = \frac{\eta g^2 \omega e^{-\omega/\omega_c}}{1 - e^{-\beta\omega}}, \tag{31}$$

where ηg^2 is a dimensionless system-bath coupling constant and ω_c is the high-frequency cutoff.

In Fig. 4, results of the three variants of MEs described in Sec. III are compared against the closed-system case. Clearly, the relaxation present in the open-system case drastically increases the ground-state probability. In the parameter region we consider, all three variants of MEs give the same predictions. In principle, the adiabatic and RWA version of the Redfield equation are unreliable in small-gap problems [34]. However, in our construction the small gap and diabatic region is so narrow that any error introduced during that period can be safely ignored. This numerically justifies the delta-function approximation we make in Sec. IV C below.

In addition, numerical results for a pausing schedule of the type shown in Fig. 2 are presented in Fig. 5. In these simulation we fix the pausing duration s_d and investigate the final success probability for different pausing positions s_p. Four primary observations arise from these results: (1) increasing the total annealing time t_f' improves the final success probability. This is because we are neither close to the adiabatic limit (closed system) nor to thermal equilibrium (open system). (2) There is a peak (or more precisely, oscillations) in the success probability if we pause right around the minimum gap. However, for the parameters chosen here this is mainly a closed system effect, which is suppressed in the open-system setting (we show in Sec. IV D below that the choice of the cutoff frequency matters a great deal). The oscillations can be understood as the interference pattern between two different paths leading to

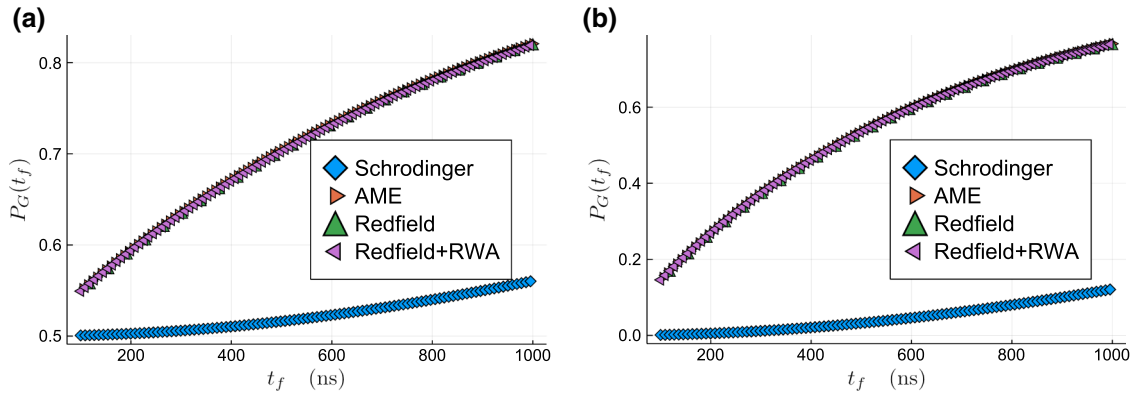

FIG. 4. Success probability (without pausing) calculated in both closed- and open-system settings with boundary conditions: (a) $\theta(1) = \pi/2$; (b) $\theta(1) = \pi$. Open-system simulations are done with all three variants of MEs described in Sec. III, with an Ohmic bath spectral density [Eq. (31)]. Here and below the Ohmic bath parameters are chosen as typical of flux qubits (e.g., Refs. [30,44,46–49]): $2\pi\eta g^2 = 10^{-4}$, $T = 16$ mK, and $\omega_c/2\pi = 4$ GHz. The results of the open-system simulations overlap. Note the different vertical axis scales in (a),(b).

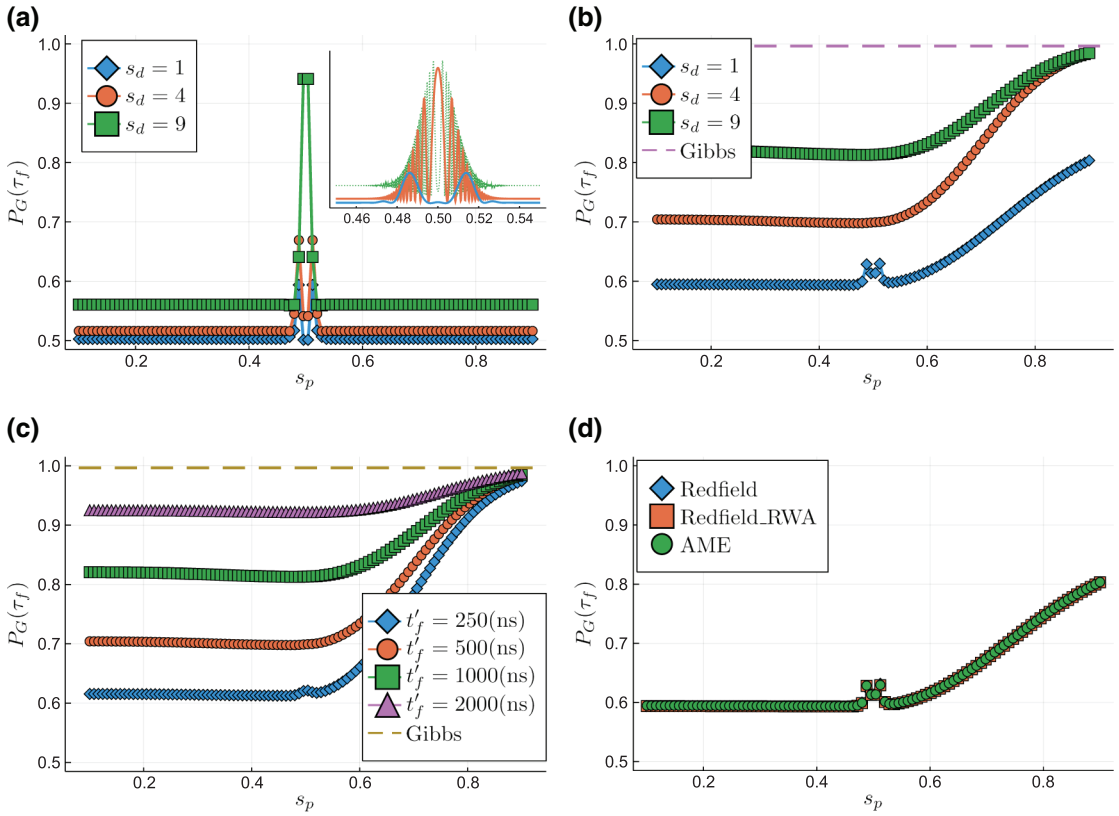

FIG. 5. Success probability for pausing schedules as shown in Fig. 2. The boundary condition for every plot is $\theta(0) = 0$ and $\theta(1) = \pi/2$ [for $\theta(1) = \pi$, the plots are of similar shapes, with the lowest success probability being 0]. For each curve the pausing duration s_d is fixed and the pausing position s_p is varied. (a) Closed system. The inset is an enlargement around $s_p = 0.5$. (b) AME. (c) AME success probabilities for different t'_f values with $s_d = 4$. (d) Final success probabilities from three MEs with $s_d = 1$. For (a),(b), and (d), the original total annealing time t_f is set to 100 ns and the Ohmic bath parameters are chosen as $2\pi\eta g^2 = 10^{-4}$, $T = 16$ mK, and $\omega_c/2\pi = 4$ GHz. The dashed lines in (b),(c) indicate the thermal ground-state population of the final Hamiltonian.

the final ground state [36], because pausing splits the single region of diabatic evolution into two and effectively creates a Mach-Zehnder interferometer [50]. (3) Pausing only helps if it happens after the minimum gap. This phenomenon can be explained by our analytic model presented in Sec. IV C. (4) All the MEs still produce the same results in the presence of pausing, which suggests we may use the AME, since it has the simplest structure for pursuing analytical results.

C. Theoretical analysis

We now present a theoretical analysis in support of the numerical results above, and to address the experimental findings in Ref. [21]. Without loss of generality, we assume $\theta(1) = \pi/2$ in the derivation. The proof also goes through for other boundary conditions. The first assumption we make is the localization of the geometric phase. To be more specific, we assume

$$\dot{\theta}(s) \approx 0 \quad \forall s \notin [\mu_\theta - c\alpha_\theta, \mu_\theta + c\alpha_\theta], \tag{32}$$

where c is a dimensionless constant and $c\alpha_\theta$ measure the effective width of the Gaussian pulse. Later, we take the limit $\alpha_\theta \to 0$. This condition holds for all the hard instances we consider in this work. In the small α_θ limit, the Redfield Eq. (18) can be treated separately inside or outside the region $[\mu_\theta - c\alpha_\theta, \mu_\theta + c\alpha_\theta]$. Within this Landau-Zener (LZ) region, the geometric phase dominates and all the other terms can be considered as a perturbation. With the detailed derivation given in Appendix E, we prove that, in the limit of weak coupling and small α_θ, the evolution across the LZ region can be approximated by a "diabatic pulse" unitary of the following form:

$$U = \begin{pmatrix} \cos(\varphi) & -i\sin(\varphi) \\ -i\sin(\varphi) & \cos(\varphi) \end{pmatrix}, \tag{33}$$

where

$$\varphi = \frac{\pi}{4} e^{-(t_f/t_{\rm ad})^2}, \quad t_{\rm ad} = \frac{\sqrt{2}}{\alpha_\theta \int_0^{\mu_\theta} \Omega(s)\, ds}. \tag{34}$$

In the limit $\alpha_\theta \to 0$ the closed-system adiabatic time scale $t_{\rm ad}$ diverges. This is an approximation to real computational (small-gap) problems where the adiabatic time scale is infinite for practical purposes.

Outside the LZ region, $d\theta/d\tau \approx 0$ and hence the Hamiltonian (28) can be written as

$$H(\tau) = -\frac{t_f'}{2}\Omega(\tau)Z + t_f' \sum_\alpha g_\alpha S_\alpha(\tau) \otimes \mathcal{B}_\alpha + H_B, \quad (35)$$

We also assume the position of the pause is outside the LZ region:

$$s_p \notin [\mu_\theta^-, \mu_\theta^+], \quad \mu_\theta^\pm \equiv \mu_\theta \pm c\alpha_\theta. \quad (36)$$

This is in accordance with the experimental protocol of Ref. [21], where pausing was found to be effective past the position of the avoided crossing, and is explained theoretically below in terms of the absence of a pausing effect before $s = \mu_\theta$.

Following Ref. [31], the AME can be written as two fully decoupled parts, which simplifies the derivation compared to the other master equations considered above:

$$\dot\rho_{00}(\tau) = -\dot\rho_{11}(\tau) = t_f' \Gamma_{01}(\tau)\rho_{11}(\tau) - t_f' \Gamma_{10}(\tau)\rho_{00}(\tau), \quad (37a)$$

$$\dot\rho_{01}(\tau) = -it_f' [\omega_{01}(\tau) + \Sigma_{01}(\tau)]\rho_{01}(\tau) - \xi_{01}(\tau)\rho_{01}(\tau), \quad (37b)$$

where

$$\omega_{01}(\tau) = -\omega_{10}(\tau) = -\Omega(\tau), \quad (38a)$$

$$\Gamma_{01}(\tau) = \sum_\alpha \gamma_\alpha[\Omega(\tau)]|S_\alpha^{01}(\tau)|^2 = e^{\beta\Omega(\tau)}\Gamma_{10}(\tau), \quad (38b)$$

$$\Sigma_{01}(\tau) = -\Sigma_{10}(\tau) = \sum_\alpha |S_\alpha^{01}|^2[\mathcal{S}(\omega_{01}) - \mathcal{S}(\omega_{10})], \quad (38c)$$

$$\xi_{01}(\tau) = \frac{1}{2}(\Gamma_{01} + \Gamma_{10}) + \frac{1}{2}\sum_\alpha \gamma_\alpha(0)|S_\alpha^{00} + S_\alpha^{11}|. \quad (38d)$$

Our primary interest is in the ground-state population [Eq. (37a)]. It can be rewritten as

$$\dot\rho_{00}(\tau) = \Gamma(\tau)\left\{1 - \left[1 + e^{-\beta\Omega(\tau)}\right]\rho_{00}(\tau)\right\}, \quad (39a)$$

$$\Gamma(\tau) \equiv t_f' \Gamma_{01}(\tau), \quad (39b)$$

where $\Gamma(\tau)$ is the dimensionless relaxation rate subject to pausing [51]. The off-diagonal elements of the density matrix decay exponentially with a rate determined by $\xi_{01}(\tau)$. As a consequence, if we start in the ground or thermal state of the initial Hamiltonian, right before the LZ region, the system density matrix will to an exponentially good approximation have only diagonal elements:

$$\rho(\mu_\theta^-) = \begin{bmatrix} P(\mu_\theta^-) & 0 \\ 0 & 1 - P(\mu_\theta^-) \end{bmatrix}. \quad (40)$$

After crossing the LZ region, using Eq. (33) the state becomes

$$\rho(\mu_\theta^+) = \begin{bmatrix} P_\varphi & -i(P_0 - P_1)\sin 2\varphi \\ i(P_0 - P_1)\sin 2\varphi & 1 - P_\varphi \end{bmatrix}, \quad (41a)$$

$$P_0 = P(\mu_\theta^-) = 1 - P_1, \quad (41b)$$

$$P_\varphi = P_0 \cos^2\varphi + P_1 \sin^2\varphi. \quad (41c)$$

Note that P_φ is the ground-state population right after the minimum gap is crossed. We can see from Eqs. (40) and (41) that pausing before μ_θ has no effect on the final results. Therefore, in the following discussion, we assume the pausing position is after the diabatic pulse: $s_p > \mu_\theta$. The solution of Eq. (39a) from μ_θ^+ to $\tau_f = 1 + s_d$ can be written as

$$\rho_{00}(\tau_f) = \exp\left\{-\int_{\mu_\theta^+}^{\tau_f} d\tau \left[1 + e^{-\beta\Omega(\tau)}\right]\Gamma(\tau)\right\} \times \left(P_\varphi + \int_{\mu_\theta^+}^{\tau_f} d\tau\Gamma(\tau) \times \exp\left\{\int_{\mu_\theta^+}^{\tau} d\tau' \left[1 + e^{-\beta\Omega(\tau')}\right]\Gamma(\tau')\right\}\right) \quad (42a)$$

$$= \mathcal{F}_d(\tau_f, s_p, s_d) + \mathcal{F}_g(\tau_f, s_p, s_d), \quad (42b)$$

where

$$\mathcal{F}_d(\tau_f, s_p, s_d) = P_\varphi G(\mu_\theta^+), \quad (43a)$$

$$\mathcal{F}_g(\tau_f, s_p, s_d) = \int_{\mu_\theta^+}^{\tau_f} d\tau\Gamma(\tau)G(\tau), \quad (43b)$$

and where

$$G(\tau) = \exp\left[-\int_{\tau}^{\tau_f} d\tau' X(\tau')\right], \tag{44a}$$

$$X(\tau) = \left[1 + e^{-\beta\Omega(\tau)}\right]\Gamma(\tau). \tag{44b}$$

Note that the functions $\Omega(\tau)$ and $\Gamma(\tau)$ have an implicit dependence on s_p and s_d:

$$\Omega(\tau) = \Omega[s(\tau)], \tag{45a}$$

$$\Gamma(\tau) = (1 + s_d)t_f \sum_{\alpha} \gamma_\alpha\{\Omega[s(\tau)]\}|S_\alpha^{01}[s(\tau)]|^2, \tag{45b}$$

where we combine Eqs. (38b) and (39b). To achieve the maximum success probability, we need to solve the following optimization problem:

$$\underset{\{s_p, s_d\}}{\text{argmax}}\, \mathcal{F}_d(\tau_f, s_p, s_d) + \mathcal{F}_g(\tau_f, s_p, s_d). \tag{46}$$

D. Numerical evidence for an optimal pausing position

First, noticing that the quantities defined in Eqs. (44b), (45a), and (45b) can be functions of either s (unpaused) or τ (paused), to avoid any ambiguity *we henceforth use the notation* $\bar{Q}(s)$ *when the argument is* s. For example, $\bar{\Gamma}(1)$ means $\Gamma(s = 1)$ instead of $\Gamma(\tau = 1)$. Note that this modifies our previous convention of denoting by $Q(s)$ the original quantity and by $Q(\tau)$ the corresponding paused quantity.

Then, we analyze the optimal pausing position s_p^*, for a given pausing duration s_d. We consider an Ohmic bath with different cutoff frequencies. This leads to different behaviors of the dimensionless relaxation rate $\bar{\Gamma}(s)$, from decreasing to nonmonotonic to increasing, as illustrated in Fig. 6. Before proceeding, we emphasize that, as is clear from Eq. (45b), the monotonicity properties of $\bar{\Gamma}(s)$ depend on three different factors:

(a) the gap of the projected TLS;

(b) the projected system-bath coupling operators [Eq. (24)];

(c) the noise spectrum.

In our examples, the first two are fixed by Eqs. (7) and (30), respectively. Thus, at a given temperature, what remains is only the noise spectrum, or more specifically the only tuning parameter for the monotonicity of $\bar{\Gamma}(s)$ is the Ohmic bath cutoff frequency ω_c. The relationship between ω_c and the monotonicity of $\gamma(\omega)$ within the qubit frequency range, which is sufficient to determine the monotonicity of $\bar{\Gamma}(s)$, is not straightforward. Thus the values in Fig. 6 are chosen after numerical investigation. More generally, the first two factors above are of equal importance to the noise spectrum. For example, in the 16-qubit problem of Ref. [44] with pure dephasing couplings, the strength of the projected system-bath coupling operators $\sum_\alpha |S_\alpha^{01}[s(\tau)]|^2$ plays a key role as well: it has a peak around the minimum gap region and decreases to almost zero afterwards.

The corresponding success probabilities as a function of pausing position obtained by numerically solving the AME and via the analytic expression (42) are shown in Fig. 7, and are in excellent agreement. There are two main observations.

(a) An optimal pausing position exists in the middle of the anneal for the cases illustrated in Fig. 6(a) and 6(b) where $\bar{\Gamma}(s)$ is not monotonically increasing.

(b) When $\bar{\Gamma}(s)$ is monotonically increasing after the minimum gap, it is always better to (trivially) pause near the end of anneal.

Guided by these numerical results, we next prove that a nontrivial optimal pausing position exists provided that the dimensionless relaxation rate given in Eq. (39b) (i.e., comparing the actual relaxation rate Γ_{01} to the anneal time t_f') is monotonically decreasing, with respect to s, from the end of the avoided crossing to the end of the anneal.

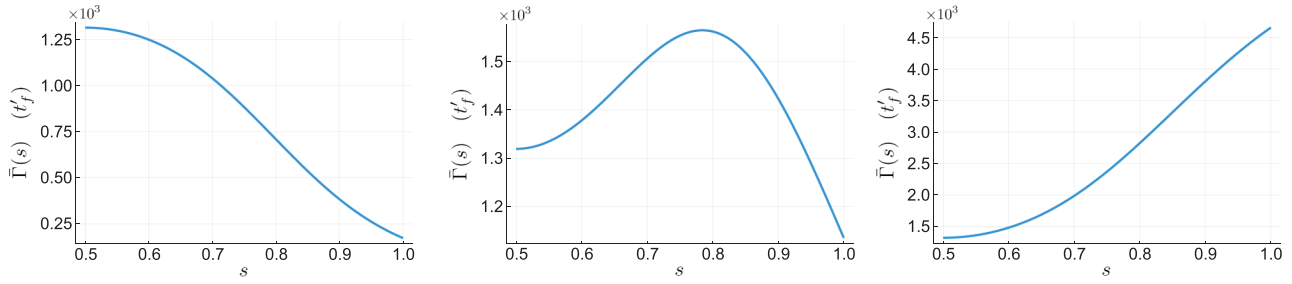

FIG. 6. $\bar{\Gamma}(s)$ [the dimensionless relaxation rate in Eq. (38b) as a function of s, i.e., with τ replaced by s] for an Ohmic bath with different cutoff frequencies: $\omega_c/2\pi = 0.5$ GHz, monotonically decreasing $\bar{\Gamma}(s)$ (left); $\omega_c/2\pi = 1$ GHz, nonmonotonic (middle); $\omega_c/2\pi = 4$ GHz, monotonically increasing $\bar{\Gamma}(s)$ (right).

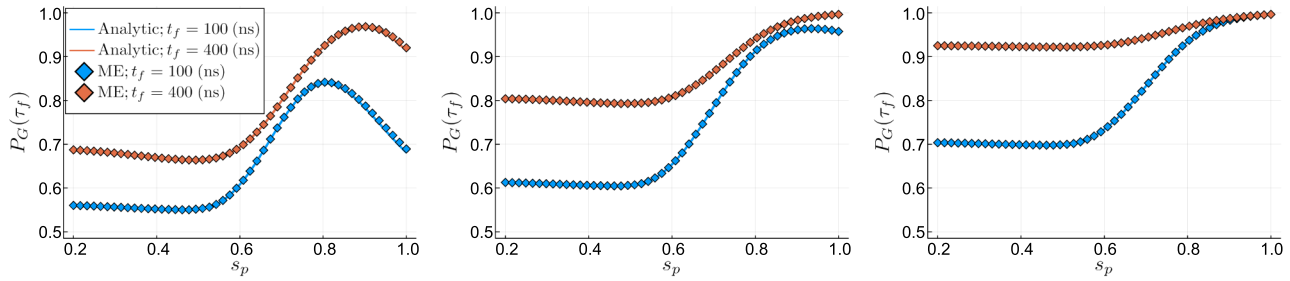

FIG. 7. Success probability versus pausing position for the three Ohmic bath cases shown in Fig. 6, in the same order from left to right. The existence of a maximum in the left and middle panels shows that an optimal pausing exists for the corresponding decay rates. The qubit frequency $\Omega(s)$ [Eq. (2)] changes between 1.88 GHz and 4.77 MHz during the anneal. Other parameters are $\eta g_z^2 = \eta g_y^2 = 10^{-4}/2\pi$ and $T = 16$ mK. The analytic solution refers to Eq. (42). For master-equation simulations, the schedules are chosen according to Fig. 1. The pausing duration is fixed at $s_d = 4$. All three panels share the same legend.

E. Existence of optimal pausing point

In this section we prove the following theorem, which provides sufficient conditions for the existence of a nontrivial optimal pausing position, thus generalizing and formalizing the numerical evidence exhibited above.

Theorem 1 (Optimal pausing point). *Let* $\bar{\Gamma}(s) = (1 + s_d)\, t_f\, \bar{\Gamma}_{01}(s)$ *be the dimensionless relaxation* rate, *let the instantaneous ground-state probability after the minimum gap is crossed be denoted* P_φ [*Eq.* (41c)], *and let* $\bar{P}_{th}(s) = 1/1 + e^{-\beta\bar{\Omega}(s)}$ *be the thermal ground-state probability at s.*

There exists a nontrivial optimal pausing point $s_p^* \in (\mu_\theta^+, 1)$ *for a fixed pausing duration* s_d, *i.e.*,

$$\left.\frac{\partial \rho_{00}}{\partial s_p}\right|_{s_p = s_p^*} = 0, \tag{47}$$

if the following is true.

(1) *The dimensionless relaxation rate decreases at the boundaries of the interval* $[\mu_\theta^+, 1]$, *i.e.,* $\bar{\Gamma}'(\mu_\theta^+) < 0$ *and* $\bar{\Gamma}'(1) < 0$ [52].

(2) *The dimensionless relaxation rate is large right after the minimum* gap : $s_d \bar{\Gamma}(\mu_\theta^+) > c_1$, *where* $c_1 = O(1)$.

(3) *The dimensionless relaxation rate is small at the end of the* anneal : $s_d \bar{\Gamma}(1) < c_2$ *where* $c_2 = O(1)$.

(4) *The ground-state population at the end of the anneal is subthermal:*

$$1 - (1 - P_\varphi) e^{-\int_{\mu_\theta^+}^1 \bar{\Gamma}(s)\, ds} \le \bar{P}_{th}(1).$$

We remark that the existence of a nontrivial optimal pausing point is possible under a substantially broader set of conditions than implied by Theorem 1, as is shown in our proof below. The theorem states a simplified set of conditions for ease of presentation and interpretation. The reader who is not interested in the technical details of the proof may skip ahead to the conclusions in Sec. V.

F. Proof of Theorem 1

The proof follows. Because $\partial\rho_{00}/\partial s_p$ is a continuous function of s_p, a sufficient condition for Eq. (47) is

$$\left.\frac{\partial \rho_{00}}{\partial s_p}\right|_{s_p = \mu_\theta^+} > 0 \quad \text{and} \quad \left.\frac{\partial \rho_{00}}{\partial s_p}\right|_{s_p = 1} < 0. \tag{48}$$

We thus divide the proof into two parts, one for each of these two inequalities.

1. Proof of $\partial\rho_{00}/\partial s_p|_{s_p=\mu_\theta^+} = \partial\mathcal{F}_g/\partial s_p|_{s_p=\mu_\theta^+} + \partial\mathcal{F}_d/s_p|_{s_p=\mu_\theta^+} > 0$

Consider first the $\mathcal{F}_d$ term [Eq. (43a)]:

$$\mathcal{F}_d = P_\varphi \exp\left[-\int_{\mu_\theta^+}^{\tau_f} d\tau X(\tau)\right]. \tag{49}$$

The partial derivative at $s_p = \mu_\theta^+$ is

$$\left.\frac{\partial \mathcal{F}_d}{\partial s_p}\right|_{s_p = \mu_\theta^+} = -P_\varphi s_d \bar{X}'(\mu_\theta^+) G(\mu_\theta^+), \tag{50}$$

where the factor of s_d arises from $\int_{\mu_\theta^+}^{\tau_f} X(\tau)/s_p|_{s_p=\mu_\theta^+}\, d\tau = \bar{X}'(\mu_\theta^+) \int_{\mu_\theta^+}^{\mu_\theta^+ + s_d} 1\, d\tau$. The relation between $\partial X/\partial s_p$ and $\bar{X}'$ is detailed in Appendix F.

Likewise, the partial derivative at $s_p = 1$ is

$$\left.\frac{\partial \mathcal{F}_d}{\partial s_p}\right|_{s_p = 1} = -P_\varphi s_d \bar{X}'(1) G(\mu_\theta^+). \tag{51}$$

As for $\mathcal{F}_g$, again in Appendix F we derive the following identities:

$$\left.\frac{\partial \mathcal{F}_g}{\partial s_p}\right|_{s_p = \mu_\theta^+} = s_d G(\mu_\theta^+) \int_0^1 dx e^{\bar{S}x} \left[\bar{\Gamma}'(\mu_\theta^+) - s_d \bar{\Gamma}(\mu_\theta^+) \bar{X}'(\mu_\theta^+)(1 - x)\right], \quad \bar{S} \equiv s_d \bar{X}(\mu_\theta^+), \tag{52a}$$

$$\left.\frac{\partial \mathcal{F}_g}{\partial s_p}\right|_{s_p=1} = -s_d \bar{X}'(1)\int_{\mu_\theta^+}^{1} d\tau\, G(\tau)\Gamma(\tau) + \int_{1}^{1+s_d} d\tau\, G(\tau)\left[\bar{\Gamma}'(1) - \bar{\Gamma}(1)\bar{X}'(1)(s_d+1-\tau)\right]. \tag{52b}$$

To make further progress we now assume the following.

Assumption 1. *$\bar{\Gamma}(s)$ is decreasing at the boundaries of the interval $[\mu_\theta^+, 1]$ [53], i.e.,*

$$\bar{\Gamma}'(\mu_\theta^+) < 0, \quad \bar{\Gamma}'(1) < 0. \tag{53}$$

Note that $\bar{\Omega}'(\mu_\theta^+) > 0$ since the gap grows for $s > \mu_\theta$ [recall Eq. (7) and that we assume $\mu_\theta \approx \mu_g$]. As a consequence, upon taking the derivative of Eq. (44b) we obtain

$$\bar{X}' = -\beta\bar{\Omega}' e^{-\beta\bar{\Omega}}\bar{\Gamma} + \bar{\Gamma}'(1+e^{-\beta\bar{\Omega}}) \tag{54}$$

and hence

$$\bar{X}'(\mu_\theta^+) \le \bar{\Gamma}'(\mu_\theta^+) < 0. \tag{55}$$

Combining Eqs. (42), (50), and (52a), we thus find the following equivalent sufficient condition for $\partial\rho_{00}/\partial s_p|_{s_p=\mu_\theta^+} > 0$:

$$\bar{\Gamma}'(\mu_\theta^+)\int_0^1 e^{\bar{S}x}\,dx > s_d\bar{\Gamma}(\mu_\theta^+)\bar{X}'(\mu_\theta^+)\int_0^1 e^{\bar{S}x}(1-x)\,dx + P_\varphi\bar{X}'(\mu_\theta^+). \tag{56}$$

Then, by explicitly carrying out the integrals, replacing one factor of $\bar{S}$ by $s_d\bar{X}(\mu_\theta^+)$, and dividing inequality (56) by $\bar{X}'(\mu_\theta^+) < 0$, we have

$$\frac{\bar{\Gamma}'(\mu_\theta^+)}{\bar{X}'(\mu_\theta^+)}\frac{e^{\bar{S}}-1}{\bar{S}} < \frac{\bar{\Gamma}(\mu_\theta^+)}{\bar{X}(\mu_\theta^+)}\left(\frac{e^{\bar{S}}-1}{\bar{S}} - 1\right) + P_\varphi. \tag{57}$$

Let us denote

$$\bar{P}_{\rm th}(s) = \frac{\bar{\Gamma}(s)}{\bar{X}(s)} = \frac{1}{1+e^{-\beta\bar{\Omega}(s)}}, \tag{58a}$$

$$\bar{Q}(s) = \frac{\bar{\Gamma}'(s)}{\bar{X}'(\mu_\theta^+)\bar{P}_{\rm th}(s)}. \tag{58b}$$

We then note that

$$1 - \bar{Q}(\mu_\theta^+) = \frac{\beta\bar{\Omega}' e^{-\beta\bar{\Omega}}\bar{\Gamma}}{\beta\bar{\Omega}' e^{-\beta\bar{\Omega}}\bar{\Gamma} - (1+e^{-\beta\bar{\Omega}})\bar{\Gamma}'} > 0, \tag{59}$$

where the inequality holds at μ_θ^+: we know the gap is increasing so $\bar{\Omega}'(\mu_\theta^+) > 0$ and the numerator is positive, and $\bar{\Gamma}'(\mu_\theta^+) < 0$ by Assumption 1, so the denominator is also positive.

Equation (57) can thus be rewritten as

$$\frac{e^{\bar{S}}-1}{\bar{S}} > \frac{1 - P_\varphi/\bar{P}_{\rm th}(\mu_\theta^+)}{1-Q(\mu_\theta^+)} \equiv x. \tag{60}$$

Note that the function $(e^{\bar{S}}-1)/\bar{S} \ge 1$ and is monotonically increasing for $\bar{S} \ge 0$. Therefore the inequality is automatically satisfied for $x < 1$ by any $\bar{S} > 0$.

Let us denote by $\bar{S}^*$ the solution of inequality (60) replaced by an equality; this transcendental equation has a formal solution in terms of the Lambert-W function [54], i.e., the inverse function of $f(W) = We^W$:

$$\bar{S}^*(x) = -\frac{1}{x}\left[1 + xW_{-1}(-e^{-1/x}/x)\right] = s_d^*\bar{X}(\mu_\theta^+), \tag{61}$$

where $W_{-1}(z)$ is one of the two real branches of $W(z)$ satisfying

$$W_{-1}(z) \le -1 \quad -1/e \le z < 0, \tag{62}$$

with $W_{-1}(-1/e) = -1$. The function $(e^{\bar{S}}-1)/\bar{S}$ is monotonically increasing, so inequality (60) is satisfied for all $\bar{S} > \max(0, \bar{S}^*)$, i.e., for all $s_d > s_d^* = \max[0, \bar{S}^*/\bar{X}(\mu_\theta^+)]$.

We can therefore replace condition (60) with the following.

Assumption 2.

$$s_d\bar{\Gamma}(\mu_\theta^+) > \bar{S}^*(x)\bar{P}_{\rm th}(\mu_\theta^+), \tag{63}$$

where $\bar{P}_{\rm th}(\mu_\theta^+)$ is the thermal ground-state probability at $s = \mu_\theta^+$.

Moreover, using the recursive expression

$$W_{-1}(z) = \ln(-z) - \ln[-W_{-1}(z)], \tag{64}$$

Eq. (61) can be written as

$$\bar{S}^*(x) = \ln x + \ln[-W_{-1}(-e^{-1/x}/x)] < 2\ln(x). \tag{65}$$

The above upper bound is derived in Appendix H.

Thus we can replace Eq. (63) by $s_d\bar{\Gamma}(\mu_\theta^+) > 2\ln(x)$, which grows very mildly, and can for practical purposes be replaced by an $O(1)$ constant. This is how Assumption 2 is stated in Theorem 1.

Note that the case $s_d^* = 0$ arises only when $x < 0$ in Eq. (60) [since then the solution $\bar{S}^*$ given by Eq. (61) is negative]. This, in turn arises when $P_\varphi > \bar{P}_{\rm th}(\mu_\theta^+)$, i.e., when the instantaneous ground-state probability is greater than the thermal ground-state probability, both at $\tau = \mu_\theta^+$. Indeed, this conforms with the expectation this in this case pausing is not advantageous.

2. Proof of $\partial\rho_{00}/\partial s_p|_{s_p=1} = \partial\mathcal{F}_g/\partial s_p|_{s_p=1} + \partial\mathcal{F}_d/\partial s_p|_{s_p=1} < 0$

Note that using Eq. (54) we have $\bar{X}'(1) < 0$ since, by Assumption 1, $\bar{\Gamma}'(1) < 0$, and $\bar{\Omega}'(1) > 0$ since the gap grows at the end of the anneal, as per Eq. (7) (this need not always be the case [55]).

Combining Eqs. (51) and (52b), we thus have [56]

$$\left.\frac{\partial\mathcal{F}_g}{\partial s_p}\right|_{s_p=1} + \left.\frac{\partial\mathcal{F}_d}{\partial s_p}\right|_{s_p=1} = s_d|\bar{X}'(1)|\Bigg\{P_\varphi G(\mu_\theta^+) + \int_{\mu_\theta^+}^1 d\tau G(\tau)\Gamma(\tau) - \left|\frac{\bar{\Gamma}'(1)}{s_d\bar{X}'(1)}\right|\int_1^{\tau_f} d\tau G(\tau) + \bar{\Gamma}(1)\int_1^{\tau_f} d\tau G(\tau)\left[1+\frac{1}{s_d}(1-\tau)\right]\Bigg\}. \tag{66}$$

Therefore it suffices to find a condition under which the expression inside the curly brackets in Eq. (66) is negative. Our strategy for doing so is to replace this expression with a simpler but negative upper bound, and iterating this until we arrive at a conceptually simple final expression.

Now note that for all $\tau \in [1, \tau_f]$:

$$G(\tau) = \exp\left[-\int_\tau^{\tau_f} d\tau' X(\tau')\right] = \exp\left[-(\tau_f - \tau)\bar{X}(1)\right], \tag{67}$$

since when $s_p = 1$ all the schedule-dependent functions are constant for τ in the range $[1, \tau_f]$, as a result of Eq. (26).

Therefore,

$$\int_1^{\tau_f} d\tau G(\tau) = \frac{1 - e^{-s_d\bar{X}(1)}}{\bar{X}(1)}, \tag{68}$$

and

$$\int_1^{\tau_f} d\tau G(\tau)\left[1+\frac{1}{s_d}(1-\tau)\right] \tag{69a}$$

$$= \frac{1 - e^{-s_d\bar{X}(1)}[1 + s_d\bar{X}(1)]}{s_d\bar{X}(1)^2}. \tag{69b}$$

We can find upper bounds involving G by using

$$X(\tau) = (1 + e^{-\beta\Omega(\tau)})\Gamma(\tau) \geq \Gamma(\tau). \tag{70}$$

Thus, using Eqs. (67) and (70) we have

$$\int_{\mu_\theta^+}^1 d\tau G(\tau)\Gamma(\tau) = e^{-s_d\bar{X}(1)}\int_{\mu_\theta^+}^1 ds e^{-\int_\tau^1 d\tau' X(\tau')}\bar{\Gamma}(s) \tag{71a}$$

$$\leq e^{-s_d\bar{X}(1)}\int_{\mu_\theta^+}^1 ds e^{-\int_s^1 ds'\bar{\Gamma}(s')}\bar{\Gamma}(s) \tag{71b}$$

$$= e^{-s_d\bar{X}(1)}\left[1 - e^{-\int_{\mu_\theta^+}^1\bar{\Gamma}(s)\,ds}\right], \tag{71c}$$

and

$$G(\mu_\theta^+) = \exp\left[-\int_{\mu_\theta^+}^1 d\tau X(\tau)\right]\exp\left[-\int_1^{1+s_d} d\tau X(\tau)\right] \tag{72a}$$

$$\leq \exp\left[-\int_{\mu_\theta^+}^1 ds\bar{\Gamma}(s)\right]\exp\left[-s_d\bar{X}(1)\right]. \tag{72b}$$

Next, let us rewrite $|\bar{\Gamma}'(1)/\bar{X}'(1)|$ by using Eq. (54). First, let

$$\epsilon \equiv \frac{\bar{\Gamma}(1)\beta\bar{\Omega}'(1)}{|\bar{\Gamma}'(1)|(1 + e^{\beta\bar{\Omega}(1)})}. \tag{73}$$

In Appendix I, we argue that for spectral densities with an exponential tail (e.g., the Ohmic case we consider) $\epsilon \ll 1$ for sufficiently large $\beta\bar{\Omega}(1)$.

Then, using Assumption 1 again

$$\left|\frac{\bar{\Gamma}'(1)}{\bar{X}'(1)}\right| = \frac{|\bar{\Gamma}'|}{|\bar{\Gamma}'(1 + e^{-\beta\bar{\Omega}}) - \beta\bar{\Omega}'e^{-\beta\bar{\Omega}}\bar{\Gamma}|} \tag{74a}$$

$$= \frac{1}{(1+\epsilon)[1 + e^{-\beta\bar{\Omega}(1)}]} = \frac{1}{1+\epsilon}\bar{P}_{\rm th}(1). \tag{74b}$$

Defining

$$\lambda \equiv s_d\bar{X}(1) = s_d[1 + e^{-\beta\bar{\Omega}(s)}]\bar{\Gamma}(s), \tag{75}$$

we can now combine all these bounds to provide an upper bound on the expression in curly brackets in Eq. (66):

$$\{\cdots\} \leq P_\varphi e^{-\lambda}e^{-\int_{\mu_\theta^+}^1\bar{\Gamma}(s)\,ds} + e^{-\lambda} - e^{-\lambda}e^{-\int_{\mu_\theta^+}^1\bar{\Gamma}(s)\,ds} - \frac{1}{1+\epsilon}\bar{P}_{\rm th}(1)\frac{1-e^{-\lambda}}{\lambda} + \bar{P}_{\rm th}(1)\frac{1 - e^{-\lambda}(1+\lambda)}{\lambda}, \tag{76}$$

an expression we require to be negative. We thus arrive at the sufficient condition

$$\int_{\mu_\theta^+}^1 \bar{\Gamma}(s)\,ds \leq \ln\left[\frac{1 - P_\varphi}{1 - \bar{P}_{\rm th}(1)F(\lambda)}\right], \tag{77}$$

where

$$F(\lambda) = 1 - \frac{e^{\lambda} - 1}{\lambda} \frac{\epsilon}{1+\epsilon}. \tag{78}$$

Since $\bar{\Gamma}(s) \geq 0$ for all s, the bound must be positive to be sensible. Thus the argument of the logarithm must be lower bounded by 1. In order for the bound in Eq. (77) to be positive it is therefore sufficient to require that

$$P_{\varphi}/\bar{P}_{\rm th}(1) < F(\lambda). \tag{79}$$

Without loss of generality, a sufficient condition for Eq. (79) is $\exists c > 1$ such that

$$F(\lambda) > \frac{1}{1+c\epsilon} > \frac{P_{\varphi}}{\bar{P}_{\rm th}(1)}, \tag{80}$$

which is not unreasonable because in practice, we expect $\bar{P}_{\rm th}(1) \sim 1$, $P_{\varphi} \lesssim 0.5$ (for a hard instance) and $\epsilon \ll 1$. Substituting Eq. (78) into Eq. (80), we have

$$\frac{e^{\lambda} - 1}{\lambda} < 1 + \frac{c-1}{1+c\epsilon} \equiv x. \tag{81}$$

Recall that the solution $\lambda^*(x)$ to this inequality considered as an equality is the Lambert-W function [Eq. (61), with λ^* replacing $\bar{S}^*$], and the function $e^{\lambda} - 1/\lambda$ is monotonically increasing. Therefore Eq. (81) is satisfied as long as $\lambda \equiv s_d \bar{X}(1) < \lambda^*$:

Assumption 3.

$$s_d \bar{\Gamma}(1) < \lambda^*(x) < 2\ln(x), \tag{82}$$

where as before we may view λ^* *in practice as an* $O(1)$ *constant. This is how Assumption* 3 *is stated in Theorem* 1.

Finally, using the lower bound (80), Eq. (77) can be replaced with

$$\int_{\mu_\theta^+}^{1} \bar{\Gamma}(s)\, ds \leq \ln\left[\frac{1-P_{\varphi}}{1-\bar{P}_{\rm th}(1)/(1+c\epsilon)}\right]. \tag{83}$$

Rewritten as follows:

Assumption 4.

$$1 - (1-P_{\varphi}) e^{-\int_{\mu_\theta^+}^{1} \bar{\Gamma}(s)\, ds} \leq \frac{1}{1+c\epsilon} \bar{P}_{\rm th}(1) \leq \bar{P}_{\rm th}(1), \tag{84}$$

this can be interpreted as follows: the excited-state population right after crossing the minimum gap is $1 - P_{\varphi}$, and after multiplying this by $\exp\left[-\int_{\mu_\theta^+}^{1} \bar{\Gamma}(s)\, ds\right]$ we have what is left of this excited-state population at the end of the anneal. On the other hand, $\bar{P}_{\rm th}(1)$ is the thermal ground-state population assuming equilibration. In other words, Eq. (84) states that the actual ground-state population reached at the end of the anneal is less than the thermal ground-state population, i.e., the system has not fully equilibrated. This is the version of Assumption 4 given in Theorem 1.

V. CONCLUSIONS

This work establishes numerically as well as analytically, via an open-system analysis of two-level system models, that pausing-induced quantum thermal relaxation can play a positive role in quantum annealing, at least according to the success probability metric. More specifically, we show here that under certain conditions on the relaxation rate after the minimum gap is crossed, the ground-state probability increases when pausing occurs before the end of the anneal. For this to occur the relaxation rate should be decreasing both after the minimum gap is crossed and at the end of the anneal, and cumulatively small over this interval, so that the system does not fully thermally equilibrate. In addition, the pause duration should be large relative to the inverse relaxation rate after the minimum gap is crossed, but small relative to the inverse relaxation rate at the end of the anneal. This provides a set of sufficient conditions relating to nonequilibrium dynamics and incomplete quantum thermal relaxation that explain the improved pause-based performance reported in a series of recent experimental quantum-annealing studies [21,23,24]. The framework we establish also provides tools to solve for the optimal pause position s_p^* [Eq. (47)]. We expect that analytic solutions for s_p^* can be derived in a problem-specific manner with further approximations.

Our results leave open a number of interesting questions for future studies. We do not determine the optimal pause time, nor do we demonstrate that pausing guarantees a quantum speedup. Indeed, computationally meaningful metrics such as the time to solution [5] may not be enhanced due to the extra time cost incurred due to pausing [57], and it is also possible that classical models of quantum annealing, such as the spin-vector Monte Carlo algorithm [58], similarly benefit from pausing [59].

ACKNOWLEDGMENTS

The authors are grateful to Jenia Mozgunov and Humberto Munoz Bauza for useful discussions and feedback. We use the Julia programming [60] and the DifferentialEquations.jl package [61] for all our numerical calculations. The research is based upon work (partially) supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA) and the Defense Advanced Research Projects Agency (DARPA), via the U.S. Army Research Office contract W911NF-17-C-0050. The views and conclusions contained herein are those of the authors and should not be

APPENDIX A: SINGLE-QUBIT ADIABATIC FRAME

We recall how to transform the Hamiltonian $H_S(s) = -\frac{1}{2}[A(s)Z + B(s)X]$ into the adiabatic frame. First, we reparametrize the annealing schedules in terms of the gap $\Omega(s)$ and rotation angle $\theta(s)$:

$$A(s) = \Omega(s)\cos\theta(s), \quad B(s) = \Omega(s)\sin\theta(s). \tag{A1}$$

Then, we rescale it to a dimensionless quantity by a change of variable $s = t/t_f$ in the von Neumann equation

$$\frac{d\rho}{ds} = -i[t_f H_S(s), \rho]. \tag{A2}$$

Finally, we rotate the system with respect to the unitary

$$U = \exp[i\theta(s)Y/2]. \tag{A3}$$

The dimensionless interaction picture Hamiltonian is

$$\begin{aligned}\tilde{H}_S(s) &= t_f\, U^\dagger(s)H_S(s)U(s) - iU^\dagger(s)\frac{\partial}{\partial s}U(s)\\ &= \frac{1}{2}\left[\frac{d\theta}{ds}Y - t_f\,\Omega(s)Z\right].\end{aligned} \tag{A4}$$

The important observation is that in this rotating frame the eigenstates of the Z operator always align with the instantaneous energy eigenstates of the original Hamiltonian. So, it can be thought of as a co-rotating frame of the adiabatic basis. This corotating frame, which we refer to as the adiabatic frame throughout this paper, can be extended to the multiqubit case as described next.

APPENDIX B: MULTIQUBIT ADIABATIC FRAME

Starting from the von Neumann equation (A2), we have

$$\begin{aligned}&\sum_{nm}\dot{\rho}_{nm}|n\rangle\langle m| + \rho_{nm}|\dot{n}\rangle\langle m| + \rho_{nm}|n\rangle\langle n|\dot{m}\\ &\quad = -it_f\sum_{nm}(E_n - E_m)\rho_{nm}|n\rangle\langle m|.\end{aligned} \tag{B1}$$

To derive an effective equation of motion for the density matrix in the adiabatic frame $\tilde{\rho} = [\rho_{nm}]$, we wish to write $|\dot{n}\rangle\langle m|$ and $|n\rangle\langle\dot{m}|$ in $\{|n\rangle\}$ basis. For example, the first term can be written as

$$|\dot{n}\rangle\langle m| = \sum_{n'}\langle n'|\dot{n}\rangle|n'\rangle\langle m|. \tag{B2}$$

It is important to emphasize that $|\dot{n}\rangle$ means the derivative with respect to s of the n'th eigenstate of an s-dependent Hamiltonian and does not obey the Schrödinger equation. We assume that $H_S(s)$ is nondegenerate. We show below that an explicit formula for $\langle m|\dot{n}\rangle$ is given by

$$\langle m|\dot{n}\rangle = \frac{\left\langle m(s)\left|\frac{dH_S(s)}{ds}\right|n(s)\right\rangle}{E_n(s) - E_m(s)}\delta_{mn}, \tag{B3}$$

which directly leads to $\langle m|\dot{n}\rangle = -\langle n|\dot{m}\rangle^*$. Substituting Eq. (B2) into Eq. (B1), we obtain

$$\begin{aligned}\dot{\rho}_{nm} &= -it_f\,(E_n - E_m)\rho_{nm}\\ &\quad - \sum_{n'\neq n}\rho_{n'm}\langle n|\dot{n'}\rangle - \sum_{m'\neq m}\rho_{nm'}\langle\dot{m'}|m\rangle\\ &= -it_f\,(E_n - E_m)\rho_{nm}\\ &\quad - i\left[-i\sum_{n'\neq n}\langle n|\dot{n'}\rangle\rho_{n'm} + i\sum_{m'\neq m}\rho_{nm'}\langle m'|\dot{m}\rangle\right]\end{aligned} \tag{B4}$$

for each ρ_{nm}. Thus, an effective Hamiltonian satisfying $\dot{\tilde{\rho}} = -i[\tilde{H}, \tilde{\rho}]$ for $\tilde{\rho} = [\rho_{nm}]$ is the one given in Eq. (5) in the main text [where we assume that $H_S(s)$ is real].

Let us now prove Eq. (B3). Writing the system Hamiltonian in its eigenbasis as

$$H_S(s) = \sum_n E_n(s)|n\rangle\langle n|, \tag{B5}$$

we derive the expression for the geometric term $\langle m|\dot{n}\rangle$ under the assumption that $H_S(s)$ is nondegenerate and real. Taking the derivative of $H_S|n\rangle = E_n|n\rangle$ with respect to s and multiplying both sides by $\langle m|$, we have

$$\langle m|\dot{H}_S|n\rangle - \dot{E}_n\langle m|n\rangle = (E_n - E_m)\langle m|\dot{n}\rangle. \tag{B6}$$

If $m \neq n$ and $|m\rangle$, $|n\rangle$ are nondegenerate, the above expression reduces to Eq. (B3). For the case where $m = n$, Eq. (B6) becomes

$$\langle n|\dot{H}_S|n\rangle = \dot{E}_n. \tag{B7}$$

On the other hand, we can also take the derivative of $\langle n|H_S|n\rangle = E_n$, which leads to

$$\langle\dot{n}|H_S|n\rangle + \langle n|\dot{H}_S|n\rangle + \langle n|H_S|\dot{n}\langle = \dot{E}_n. \tag{B8}$$

By cancelling out $\langle n|\dot{H}_S|n\rangle$ and $\dot{E}_n$ and noticing that $\langle n|\dot{n}\rangle$ is real, we can deduce that $\langle n|\dot{n}\rangle = 0$. It is important to note

that the eigenvector $|n\rangle$ is only uniquely determined up to a constant factor of ± 1. As a result, we need to implement a continuous constraint

$$\lim_{\Delta s \to 0} \langle n(s) | n(s + \Delta s) \rangle = 1 \tag{B9}$$

to ensure the continuity of the geometric term.

This result can be extended to a general complex-valued Hamiltonian. In this case, the eigenvectors of the Hamiltonian are uniquely determined up to a constant of unit modulus. However, by enforcing the continuity condition (B9), we have an analytic $|n(s)\rangle$ that also satisfies $\langle n|\dot{n}\rangle = 0$.

A method to calculate $\langle m|\dot{n}\rangle$ for degenerate Hamiltonian is provided in Ref. [62]. A special case that is not discussed in this reference is when two or more states $|m\rangle$ and $|n\rangle$ become degenerate in a closed interval $s \in [s_a, s_b]$. In such a case, we can still obtain a pair of orthogonal states $|m(s_a)\rangle$ and $|n(s_a)\rangle$ by enforcing the continuity condition (B9) across the boundary. Within the interval $\langle m|\dot{n}\rangle$s are usually 0 in practice. A sufficient condition for this is

$$\lim_{\Delta s \to 0} \langle m(s) | n(s + \Delta s) \rangle / \Delta s = 0 \tag{B10}$$

for all $s \in [s_a, s_b]$. By expanding $|n(s + \Delta s)\rangle$ as a Taylor series in Δs, Eq. (B10) reduces to

$$\lim_{\Delta s \to 0} \langle m|n\rangle / \Delta s + \langle m|\dot{n}\rangle + O(\Delta s) = 0, \tag{B11}$$

which implies $\langle m|\dot{n}\rangle = 0$. One example of condition (B10) is when the transverse field becomes zero during the anneal and the problem Hamiltonian has degenerate excited states.

APPENDIX C: ANNEALING-ANGLE BEHAVIOR FOR PREVIOUSLY STUDIED SMALL-GAP QUANTUM-ANNEALING EXAMPLES

Here we provide a brief look at the annealing-angle aspect of two previously studied quantum-annealing examples with small gaps. In the following examples, all the results are produced with a linear schedule instead of the D-Wave schedule used in the references.

The first example, shown in Figs. 8(a) and 8(b), is the p-spinmodel [25]. The angular progression is localized around $s = 0.483$. Across the region, there is a π jump of the annealing angle.

The second example, shown in Fig. 8(c), is the D-Wave 16-qubit gadget problem [44]. Unlike the p-spin model, the annealing angle in this case does not stay zero during the first half of the anneal. There is still a sharp π jump across the minimum gap region.

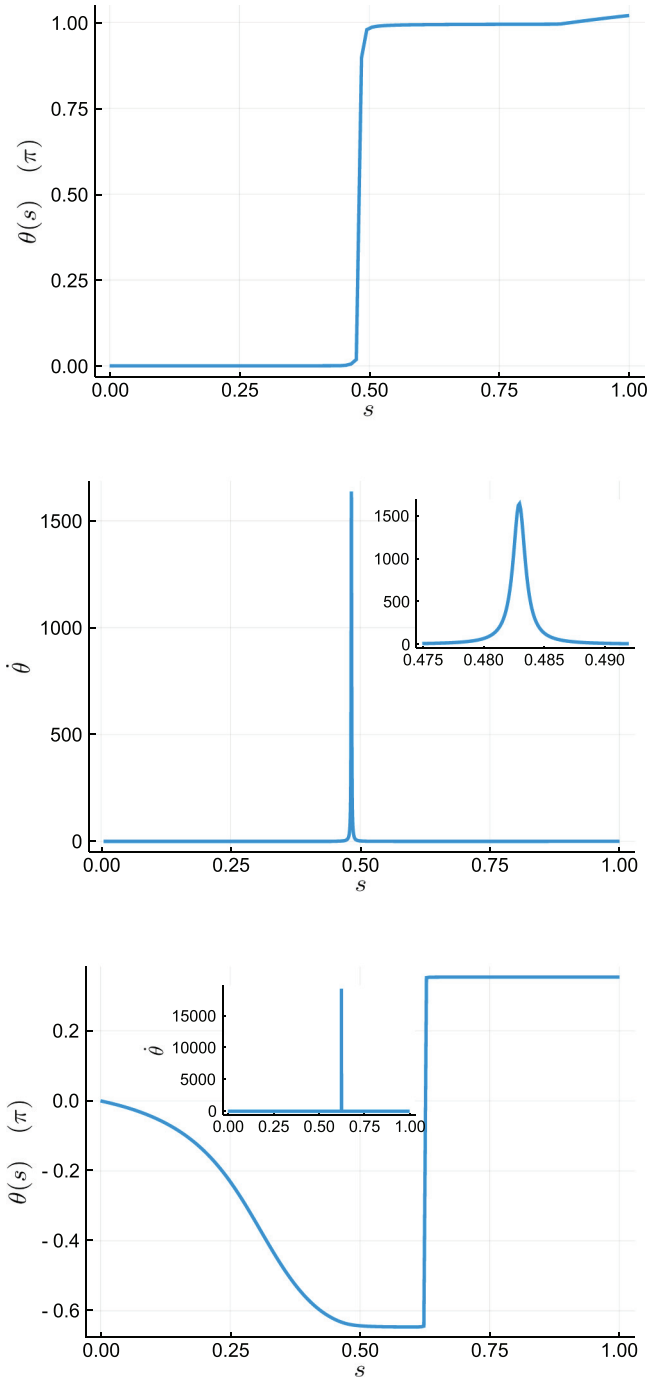

FIG. 8. (a) Annealing angle for the p-spin model with $n = 20$, $p = 19$ [25]. The inset is an enlargement around the peak of angular progression, shown in (b). In (c) we show the annealing angle for the D-Wave 16-qubit gadget [44]. The inset illustrates the angular progression.

APPENDIX D: EASY PROBLEM WITH SMALL GAP: CONSTANT ANGULAR PROGRESSION

The simplest example we can construct of an easy problem with a small gap is one with the same small gap structure as in Eq. (7) but a constant $\dot{\theta}(s) = \pi/2$. If we

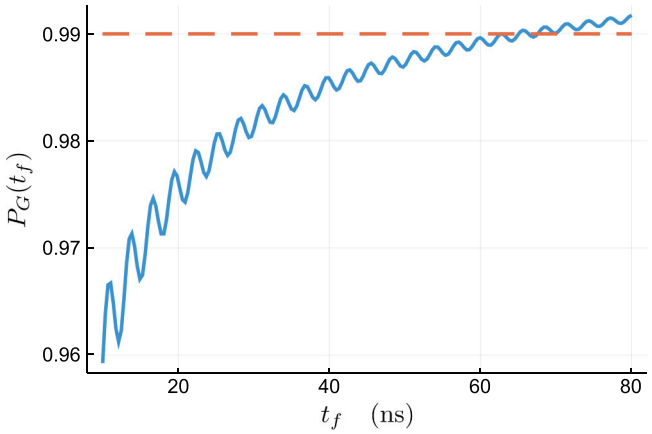

FIG. 9. Closed-system success probability versus total annealing time. The gap is chosen according to Eq. (7) with parameters $\mu_g = 0.5$, $\alpha_g = 0.5$, $E_0 = 15/\pi$ GHz and $\Delta = 0.001$.

define the adiabatic time scale $t_{\rm ad}$ as the point where

$$P_G(t) \geq 0.99 \quad \forall t \geq t_{\rm ad}, \tag{D1}$$

we find in the numerical example shown in Fig. 9, that the adiabatic time scale is much smaller than the inverse gap

$$t_{\rm ad} \ll h/(E_0\Delta)^2 \approx 142\,502 \text{ (ns)}, \tag{D2}$$

where

$$h = \max_{s\in[0,1]} \|\dot{H}(s)\|, \tag{D3}$$

and $\|\cdot\|$ is the operator norm [for the Frobenius norm we instead find 201 528 (ns)].

APPENDIX E: LOCAL PERTURBATION AROUND THE NONADIABATIC TRANSITION

Our analysis in this appendix closely mirrors Ref. [36], with some modifications. We start with the Redfield Eq. (18) and define Liouville operators

$$\mathcal{L}_A = i\frac{t_f\,\Omega(s)}{2}[Z,\cdot], \tag{E1a}$$

$$\mathcal{L}_G = -i\frac{\dot{\theta}}{2}[Y,\cdot], \tag{E1b}$$

$$\mathcal{L}_R = -\sum_\alpha (g_\alpha t_f)^2[S_\alpha(s), \Lambda_\alpha(s)\cdot] + \text{h.c.}, \tag{E1c}$$

which represents the adiabatic, geometric, and Redfield parts in the ME. Now we present a perturbation method for the evolution across the region $[\mu_\theta - c\alpha_\theta, \mu_\theta + c\alpha_\theta]$. First we rotate the equation with respect to the adiabatic parts of the Hamiltonian and denote the resulting Liouville operators as $\mathcal{L}_{\tilde{G}}$ and $\mathcal{L}_{\tilde{R}}$.

Using any unitarily invariant norm, such as the operator norm, we can bound

$$\|\mathcal{L}_{\tilde{R}}(\rho)\| \leq 4\sum_\alpha (g_\alpha t_f)^2 \|S_\alpha(s)\|\|\Lambda_\alpha(s)\|\|\rho\|_1 \tag{E2a}$$

$$\leq \sum_\alpha (2g_\alpha t_f\,\|S_\alpha\|)^2 \int_0^s ds'|C_\alpha(s,s')|, \tag{E2b}$$

where we use $\|\rho\|_1 = 1$ (trace norm), and we make use of Eq. (19). We define $g = \max_\alpha g_\alpha$ and note that we can always choose a normalization such that $S_\alpha \leq 1$. Then

$$\|\mathcal{L}_{\tilde{R}}(\rho)\| \leq (2gt_f)^2 \int_0^s ds'|C(s,s')|. \tag{E3}$$

Noting that the correlation function is translation-invariant

$$C(s,s') = C(s-s'), \tag{E4}$$

the bound (E3) can be further simplified by using the following inequality:

$$\int_0^s ds'|C(s,s')| < \int_0^\infty ds|C(s)|. \tag{E5}$$

Defining

$$\frac{1}{\tau_{\rm SB}} = t_f \int_0^\infty |C(s)|\,ds, \tag{E6}$$

where $\tau_{\rm SB}$ can be interpreted as the fastest system decoherence time scale [34], the final expression becomes

$$\|\mathcal{L}_{\tilde{R}}(\rho)\| \leq 4g^2 t_f/\tau_{\rm SB}. \tag{E7}$$

This bound allows us to rigorously establish conditions under which it is safe to drop the dissipative part of the evolution.

Next, using the bound above, we write down the evolution operator and its first-order Magnus expansion:

$$\mathcal{E}(\mu_\theta + c\alpha_\theta, \mu_\theta - c\alpha_\theta) \tag{E8a}$$

$$= \mathcal{T}\exp\left[-i\int_{\mu_\theta - c\alpha_\theta}^{\mu_\theta + c\alpha_\theta} \mathcal{L}_{\tilde{G}}(\tau) + \mathcal{L}_{\tilde{R}}(\tau)\,d\tau\right] \tag{E8b}$$

$$= \exp\left[-i\int_{\mu_\theta - c\alpha_\theta}^{\mu_\theta + c\alpha_\theta} \mathcal{L}_{\tilde{G}}(\tau)\,d\tau + O(4\alpha_\theta t_f g^2/\tau_{\rm SB})\right]. \tag{E8c}$$

We make use of the formal Magnus expansion of the superoperator in going from line (E8b) to line (E8c), wherein

the lowest-order term in the time-ordered propagator is the argument of line (E8b), and the order term is a commutator of the two Liouville operators. As long as $\alpha_\theta t_f g^2/\tau_{\rm SB} \ll 1$, which requires either weak coupling or small α_θ (as confirmed for the two examples mentioned in Appendix C), we may ignore the dissipative part due to $\mathcal{L}_{\tilde{R}}$. After dropping this term we are left with

$$\exp\left[-i\int_{\mu_\theta-c\alpha_\theta}^{\mu_\theta+c\alpha_\theta}\mathcal{L}_{\tilde{G}}(\tau)\,d\tau\right]\times\rho=\tilde{U}\rho\tilde{U}^\dagger, \tag{E9}$$

where the unitary is

$$\begin{aligned}&\tilde{U}(\mu_\theta+c\alpha_\theta,\mu_\theta-c\alpha_\theta)\\&\quad=\exp\left[-\frac{i}{2}\int_{\mu_\theta-c\alpha_\theta}^{\mu_\theta+c\alpha_\theta}\dot{\theta}(s)\tilde{Y}(s)\,ds\right],\end{aligned} \tag{E10}$$

and $\tilde{Y}(s)$ is Y in the interaction picture generated by $\mathcal{L}_A$ [Eq. (E1a)]:

$$\begin{aligned}\tilde{Y}(s)&=U_A^\dagger(s)YU_A(s)\\&=-ie^{-it_f\int_0^s\Omega(s')\,ds'}S_+ +\text{h.c.},\end{aligned} \tag{E11}$$

where $S_+=(\sigma^x+i\sigma^y)/2$. Substituting $\tilde{Y}(s)$ into Eq. (E10), we have

$$\tilde{U}(\mu_\theta+c\alpha_\theta,\mu_\theta-c\alpha_\theta)=\exp\{-\phi S_+ +\text{h.c.}\}, \tag{E12}$$

where

$$\phi=\frac{1}{2}\int_{\mu_\theta-c\alpha_\theta}^{\mu_\theta+c\alpha_\theta}\dot{\theta}e^{-it_f\int_0^s\Omega(s')\,ds'}\,ds. \tag{E13}$$

Because $\dot{\theta}$ is highly localized within the integral limit, we can further simplify the expression as

$$\phi=\frac{1}{2}\int_{-\infty}^{\infty}\dot{\theta}e^{-it_f\int_0^{\mu_\theta}\Omega(s')\,ds'}\,ds=\frac{1}{2}\int_{-\infty}^{\infty}\dot{\theta}e^{-it_f\mu_\theta\tilde{\Omega}s}\,ds, \tag{E14}$$

where

$$\tilde{\Omega}=\frac{\int_0^{\mu_\theta}\Omega(s')\,ds'}{\mu_\theta}. \tag{E15}$$

Using Eq. (8), together with the boundary condition $\theta(1)=\pi/2$, the integral in Eq. (E14) can be carried out, yielding

$$\phi=\frac{\pi}{4}e^{-(t_f/t_{\rm ad})^2+i\mu_\theta t_f\tilde{\Omega}}, \tag{E16}$$

where $t_{\rm ad}=\sqrt{2}/(\alpha_\theta\mu_\theta\tilde{\Omega})$ [Eq. (34)]. Finally, we can write down the matrix form of the unitary (E12):

$$\tilde{U}=\exp(-\phi S_+ +\text{h.c.}) \tag{E17a}$$

$$=\begin{bmatrix}\cos(|\phi|) & -\sin(|\phi|)e^{i\mu_\theta t_f\tilde{\Omega}}\\ \sin(|\phi|)e^{-i\mu_\theta t_f\tilde{\Omega}} & \cos(|\phi|)6\end{bmatrix} \tag{E17b}$$

$$=U_A^\dagger(\mu_\theta)\begin{bmatrix}\cos(|\phi|) & -\sin(|\phi|)\\ \sin(|\phi|) & \cos(|\phi|)\end{bmatrix}U_A(\mu_\theta). \tag{E17c}$$

If we rotate the unitary back into the adiabatic frame, it becomes Eq. (33).

As a final remark, in the limit of $\alpha_\theta\to 0$, the unitary nonadiabatic transition has the same effects as an ideal beam splitter:

$$U\to\frac{\sqrt{2}}{2}(I-iY). \tag{E18}$$

APPENDIX F: DERIVATION OF Eq. (52)

Taking the partial derivative of Eq. (43b) with respect to s_p, we have

$$\frac{\partial\mathcal{F}_g}{\partial s_p}=\int_{\mu_\theta^+}^{\tau_f}d\tau\left\{\frac{\partial\Gamma}{\partial s_p}G+\Gamma\frac{\partial G}{\partial s_p}\right\} \tag{F1a}$$

$$=\int_{\mu_\theta^+}^{\tau_f}d\tau G\left\{\frac{\partial\Gamma}{\partial s_p}-\Gamma\int_\tau^{\tau_f}\frac{\partial X}{\partial s_p}\,d\tau'\right\}. \tag{F1b}$$

Denoting either $\Gamma(\tau)$ or $X(\tau)$ by $F(\tau)$, their partial derivatives with respect to s_p are given by (for a proof see Appendix G):

$$\frac{\partial F(\tau)}{\partial s_p}=\begin{cases}0 & \tau\le s_p\\ \bar{F}'(s_p) & s_p<\tau\le s_p+s_d\\ 0 & s_p+s_d<\tau\le 1+s_d\end{cases}, \tag{F2}$$

where $\bar{F}'(s_p)=\partial\bar{F}(s)/\partial s|_{s=s_p}$. This can also be thought of as the chain rule

$$\frac{\partial F(\tau)}{\partial s_p}=\frac{d\bar{F}(s)}{ds}\frac{\partial s(\tau)}{\partial s_p}, \tag{F3}$$

where $\partial s(\tau)/\partial s_p$ follows by differentiating Eq. (26)

$$\frac{\partial s(\tau)}{\partial s_p} = \begin{cases} 0 & \tau \le s_p \\ 1 & s_p < \tau \le s_p + s_d \\ 0 & s_p + s_d < \tau \le 1 + s_d \end{cases}. \tag{F4}$$

Let us consider the two end points of Eq. (F1b), $\partial\mathcal{F}_g/\partial s_p|_{s_p=\mu_\theta^+}$ and $\partial\mathcal{F}_g/\partial s_p|_{s_p=1}$.

1. Derivation of Eq. (52a)

Consider $\partial\mathcal{F}_g/\partial s_p|_{s_p=\mu_\theta^+}$. This integral can be simplified by realizing that both $\partial\Gamma/\partial s_p|_{s_p=\mu_\theta^+}(\tau)$ and $\partial X/\partial s_p|_{s_p=\mu_\theta^+}(\tau)$ are rectangular functions within the pausing region $[\mu_\theta^+, \mu_\theta^+ + s_d]$ [Eq. (F2)]. As a result, $\partial\Gamma/\partial s_p|_{\mu_\theta^+}(\tau)$ and $\int_\tau^{\tau_f} \partial X/\partial s_p|_{\mu_\theta^+}(\tau')\,d\tau'$ are zero when $\tau > \mu_\theta^+ + s_d$. The former follows directly from the definition and the latter is because the integrand is zero in the entire region of integration. This allows us to change the integration limit from τ_f to $\mu_\theta^+ + s_d$ in Eq. (F1b) and write

$$\begin{aligned} &\frac{\partial\mathcal{F}_g}{\partial s_p}\bigg|_{s_p=\mu_\theta^+} \\ &= \int_{\mu_\theta^+}^{\mu_\theta^+ + s_d} d\tau\, G(\tau) \left(\frac{\partial\Gamma}{\partial s_p}\bigg|_{\mu_\theta^+} - \Gamma \int_\tau^{\tau_f} \frac{\partial X}{\partial s_p}\bigg|_{\mu_\theta^+} d\tau' \right) \\ &= \int_{\mu_\theta^+}^{\mu_\theta^+ + s_d} d\tau\, G(\tau) \left[\bar{\Gamma}'(\mu_\theta^+) \right. \\ &\quad \left. -\bar{\Gamma}(\mu_\theta^+)\bar{X}'(\mu_\theta^+)(s_d + \mu_\theta^+ - \tau) \right]. \end{aligned} \tag{F5}$$

To obtain the second equality above, we explicitly carried out the integration

$$\begin{aligned} \int_\tau^{\tau_f} \frac{\partial X(\tau')}{\partial s_p}\bigg|_{s_p=\mu_\theta^+} d\tau' &= \bar{X}'(\mu_\theta^+) \int_\tau^{\mu_\theta^+ + s_d} 1\, d\tau' \\ &= \bar{X}'(\mu_\theta^+)(s_d + \mu_\theta^+ - \tau). \end{aligned} \tag{F6}$$

Using Eq. (44a) and noticing that $X(\tau')$ is constant for $\tau' \in [\mu_\theta^+, \tau]$ with $\tau \in [\mu_\theta^+, \mu_\theta^+ + s_d]$ (the pausing region), we have

$$\begin{aligned} G(\tau)\big|_{s_p=\mu_\theta^+} &= \exp\left[\left(\int_{\mu_\theta^+}^{\tau} - \int_{\mu_\theta^+}^{1+s_d} \right) X(\tau')\, d\tau' \right] \\ &= \exp\left[\bar{X}(\mu_\theta^+)(\tau - \mu_\theta^+) \right] G(\mu_\theta^+). \end{aligned} \tag{F7}$$

Thus Eq. (F5) can be further simplified with a change of variable $\tau = s_d x + \mu_\theta^+$, upon which we arrive at Eq. (52a):

$$\begin{aligned} \frac{\partial\mathcal{F}_g}{\partial s_p}\bigg|_{s_p=\mu_\theta^+} &= s_d G(\mu_\theta^+) \int_0^1 dx e^{\bar{S}x} \\ &\quad \times \left[\bar{\Gamma}'(\mu_\theta^+) - s_d \bar{\Gamma}(\mu_\theta^+)\bar{X}'(\mu_\theta^+)(1-x) \right], \end{aligned}$$

where $\bar{S} = s_d \bar{X}(\mu_\theta^+)$.

2. Derivation of Eq. (52b)

Consider $\partial\mathcal{F}_g/\partial s_p|_{s_p=1}$. The expression at this end point can similarly be obtained by splitting the integral into two parts $\int_{\mu_\theta^+}^{\tau_f} d\tau = \int_{\mu_\theta^+}^{1} d\tau + \int_1^{1+s_d} d\tau$ [recall, per Eq. (27), that $\tau_f = 1 + s_d$]. The integrands of the first integral $\int_{\mu_\theta^+}^1 d\tau$ satisfy

$$\frac{\partial\Gamma(\tau)}{\partial s_p}\bigg|_{s_p=1} = 0 \tag{F8a}$$

$$\begin{aligned} \int_\tau^{\tau_f} \frac{\partial X(\tau')}{\partial s_p}\bigg|_{s_p=1} d\tau' &= \bar{X}'(1) \int_1^{\tau_f} 1\, d\tau' \\ &= s_d \bar{X}'(1). \end{aligned} \tag{F8b}$$

Equation (F8b) follows from the same reasoning as Eq. (F6). Equation (F8a) follows from the chain rule applied to Eq. (45b), which gives $\partial s(\tau)/\partial s_p|_{s_p=1} = 0$ [Eq. (F4)] as an overall prefactor since the integration over τ goes up to $\tau = 1$.

Again using Eqs. (F6) and (F4), the integrands of the second integral $\int_1^{1+s_d} d\tau$ satisfy

$$\frac{\partial\Gamma(\tau)}{\partial s_p}\bigg|_{s_p=1} = \bar{\Gamma}'(1) \tag{F9a}$$

$$\begin{aligned} \int_\tau^{\tau_f} \frac{\partial X(\tau')}{\partial s_p}\bigg|_{s_p=1} d\tau' &= \bar{X}'(1) \int_\tau^{\tau_f} 1, d\tau' \\ &= \bar{X}'(1)(s_d + 1 - \tau). \end{aligned} \tag{F9b}$$

Combining these results into Eq. (F1b), the final expression becomes Eq. (52b):

$$\begin{aligned} \frac{\partial\mathcal{F}_g}{\partial s_p}\bigg|_{s_p=1} &= -s_d \bar{X}'(1) \int_{\mu_\theta^+}^1 d\tau\, G(\tau)\Gamma(\tau) \\ &\quad + \int_1^{1+s_d} d\tau\, G(\tau) \\ &\quad \times \left[\bar{\Gamma}'(1) - \bar{\Gamma}(1)\bar{X}'(1)(s_d + 1 - \tau) \right]. \end{aligned}$$

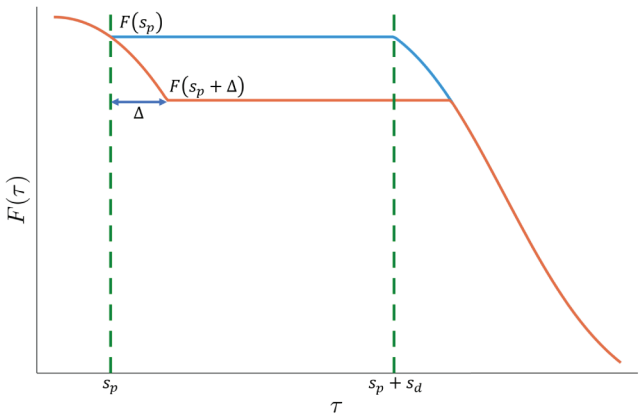

FIG. 10. Graphical proof of Eq. (F2). The partial derivative is obtained in the limit of $\Delta \to 0$.

APPENDIX G: PROOF OF Eq. (F2)

To prove Eq. (F2), we start from the definition of the partial derivative:

$$\frac{\partial F}{\partial s_p} = \lim_{\Delta\to 0} \frac{F(\tau, s_p+\Delta, s_d) - F(\tau, s_p, s_d)}{\Delta}. \tag{G1}$$

From Fig. 10, we see that in the limit $\Delta \to 0$

$$F(\tau, s_p+\Delta, s_d) = \begin{cases} F[s(\tau)] & \tau < s_p \\ F(s_p+\Delta) & s_p < \tau < s_p + s_d \\ F[s(\tau)] & s_p + s_d < \tau < 1 + s_d \end{cases}, \tag{G2}$$

so

$$\frac{\partial F}{\partial s_p} = \lim_{\Delta\to 0} \frac{F(s_p+\Delta) - F(s_p)}{\Delta} = F'(s=s_p) \tag{G3}$$

for $s_p < \tau < s_p + s_d$ and zero elsewhere.

APPENDIX H: PROOF OF Eq. (65)

To prove Eq. (65) we use

$$f(x) = x + W_{-1}(-e^{-1/x}/x) > 0 \quad \forall x > 1, \tag{H1}$$

which can be proved with $f(1) = 0$ and

$$f'(x) = 1 + \frac{W_{-1}(-e^{-1/x}/x)}{1 + W_{-1}(-e^{-1/x}/x)} \frac{1-x}{x^2} \tag{H2a}$$

$$= \frac{|1+W_{-1}|x^2 - |W_{-1}|x + |W_{-1}|}{|1+W_{-1}|x^2} \tag{H2b}$$

$$> 0 \quad \forall x > 1. \tag{H2c}$$

Line (H2a) is obtained using the derivative formula

$$\frac{dW(z)}{dz} = \frac{W(z)}{z[1+W(z)]} \tag{H3}$$

and line (H2b) comes from Eq. (62). To prove line (H2c), we consider the quadratic function in the numerator of line (H2b), whose roots are

$$x_\pm = -\frac{|W_{-1}| \pm \sqrt{-3|W_{-1}|^2 + 4|W_{-1}|}}{2|1+W_{-1}|}. \tag{H4}$$

For $|W_{-1}| > 4/3$, the discriminant is smaller than zero. For $1 < |W_{-1}| \le 4/3$, the discriminant lies within $[0, 1)$ so both roots are negative. In both cases, the quadratic function is positive for $x > 1$, which leads to line (H2c).

APPENDIX I: DEMONSTRATION THAT $\epsilon \ll 1$

To show that $\epsilon \ll 1$ for $\epsilon \equiv \bar{\Gamma}(1)\beta\bar{\Omega}'(1)/|\bar{\Gamma}'(1)|(1+e^{\beta\bar{\Omega}(1)})$ [Eq. (73)], we start with the expression for $\bar{\Gamma}(s)$ [Eq. (45b)]. Because $\sum_\alpha |S_\alpha^{01}(s)|^2 = 1$ for the system-bath coupling operators in our model Eq. (30), we obtain

$$\bar{\Gamma}(s) = \bar{\Gamma}[\bar{\Omega}(s)] = (1+s_d) t_f \gamma[\bar{\Omega}(s)], \tag{I1}$$

where we have made the gap dependence of $\bar{\Gamma}(s)$ explicit.

Next, ϵ can be simplified as

$$\epsilon = \frac{\bar{\Omega}'(1)\bar{\Gamma}(1)}{|\bar{\Gamma}'(1)|} \frac{\beta}{1+e^{\beta\bar{\Omega}(1)}} \tag{I2a}$$

$$= \frac{\gamma[\bar{\Omega}(1)]}{\left|\frac{d\gamma}{d\bar{\Omega}}[\bar{\Omega}(1)]\right|} \frac{\beta}{1+e^{\beta\bar{\Omega}(1)}}, \tag{I2b}$$

where we use the chain rule to write $\bar{\Gamma}'(s) = (1+s_d) t_f \bar{\Omega}'(s)(d\gamma/d\bar{\Omega})$.

Any spectral density with an exponential high-frequency cutoff $e^{-\omega/\omega_c}$, such as the Ohmic bath [Eq. (31)], will result in

$$\frac{\gamma[\bar{\Omega}(1)]}{\left|\frac{d\gamma}{d\bar{\Omega}}[\bar{\Omega}(1)]\right|} \sim \omega_c \quad \text{for } \omega > \omega_c. \tag{I3}$$

Thus, as long as $\bar{\Omega}(1) > \omega_c, 1/\beta$ we find that ϵ is exponentially small in the final energy gap $\bar{\Omega}(1)$.

We plot $\log_{10}\epsilon$ for different temperatures and cutoff frequencies for the Ohmic case in Fig. 11, and confirm that for reasonable parameters indeed $\epsilon \ll 1$.

This argument fails if $\bar{\Gamma}'(s^*) = 0$ for $s^* \in [\mu_\theta^+, 1]$. For such cases, we only need to shift the end point from 1 to s^*. Then the optimal pausing position will be in the interval $[\mu_\theta^+, s^*]$.

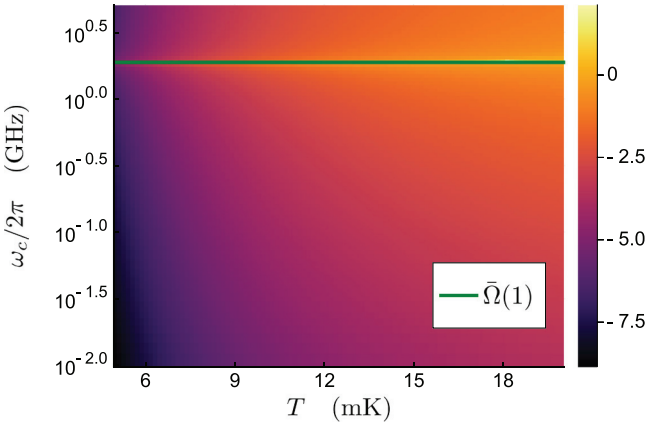

FIG. 11. Heat map of $\log_{10}\epsilon$ for different ω_c and T values for the Ohmic bath spectrum [Eq. (31)], covering the entire parameter region shown in Fig. 6. The gap $\bar{\Omega}(1) \approx 1.88$ GHz is chosen in accordance with parameters used in Fig. 1. The maximum of ϵ appears near the line $\omega_c = 2\pi\bar{\Omega}(1)$.

[1] Tadashi Kadowaki and Hidetoshi Nishimori, Quantum annealing in the transverse Ising model, Phys. Rev. E **58**, 5355 (1998).

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

[3] Tameem Albash and Daniel A. Lidar, Adiabatic quantum computation, Rev. Mod. Phys. **90**, 015002 (2018).

[4] Philipp Hauke, Helmut G. Katzgraber, Wolfgang Lechner, Hidetoshi Nishimori, and William D. Oliver, Perspectives of quantum annealing: Methods and implementations, Rep. Prog. Phys. **83** (2020).

[5] Troels F. Rønnow, Zhihui Wang, Joshua Job, Sergio Boixo, Sergei V. Isakov, David Wecker, John M. Martinis, Daniel A. Lidar, and Matthias Troyer, Defining and detecting quantum speedup, Science **345**, 420 (2014).

[6] Vasil S. Denchev, Sergio Boixo, Sergei V. Isakov, Nan Ding, Ryan Babbush, Vadim Smelyanskiy, John Martinis, and Hartmut Neven, What is the Computational Value of Finite-Range Tunneling? Phys. Rev. X **6**, 031015 (2016).

[7] Andrew D. King *et al.*, Observation of topological phenomena in a programmable lattice of 1,800 qubits, Nature **560**, 456 (2018).

[8] R. Harris *et al.*, Phase transitions in a programmable quantum spin glass simulator, Science **361**, 162 (2018).

[9] Tameem Albash and Daniel A. Lidar, Demonstration of a Scaling Advantage for a Quantum Annealer Over Simulated Annealing, Phys. Rev. X **8**, 031016 (2018).

[10] Salvatore Mandrà and Helmut G. Katzgraber, A deceptive step towards quantum speedup detection, Quant. Sci. Technol. **3**, 04LT01 (2018).

[11] Alex Mott, Joshua Job, Jean-Roch Vlimant, Daniel Lidar, and Maria Spiropulu, Solving a Higgs optimization problem with quantum annealing for machine learning, Nature **550**, 375 (2017).

[12] Kristen L. Pudenz, Tameem Albash, and Daniel A. Lidar, Error-corrected quantum annealing with hundreds of qubits, Nat. Commun. **5**, 3243 (2014).

[13] Walter Vinci, Tameem Albash, and Daniel A. Lidar, Nested quantum annealing correction, npj Quant. Inf. **2**, 16017 (2016).

[14] Walter Vinci and Daniel A. Lidar, Scalable effective-temperature reduction for quantum annealers via nested quantum annealing correction, Phys. Rev. A **97**, 022308 (2018).

[15] Adam Pearson, Anurag Mishra, Itay Hen, and Daniel A. Lidar, Analog errors in quantum annealing: Doom and hope, npj Quant. Inf. **5**, 107 (2019).

[16] Trevor Lanting, Andrew D. King, Bram Evert, and Emile Hoskinson, Experimental demonstration of perturbative anticrossing mitigation using nonuniform driver hamiltonians, Phys. Rev. A **96**, 042322 (2017).

[17] Juan I. Adame and Peter L. McMahon, Inhomogeneous driving in quantum annealers can result in orders-of-magnitude improvements in performance, Quant. Sci. Technol. **5**, 035011 (2020).

[18] Ting-Jui Hsu, Quantum annealing with anneal path control: Application to 2-sat problems with known energy landscapes, Commun. Comput. Phys. **26**, 928 (2019).

[19] Sheir Yarkoni, Hao Wang, Aske Plaat, and Thomas Bäck, in *Quantum Technology and Optimization Problems*, edited by Sebastian Feld and Claudia Linnhoff-Popien (Springer International Publishing, Cham, 2019), p. 157.

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

[21] Jeffrey Marshall, Davide Venturelli, Itay Hen, and Eleanor G. Rieffel, Power of Pausing: Advancing Understanding of Thermalization in Experimental Quantum Annealers, Phys. Rev. Appl. **11**, 044083 (2019).

[22] D-Wave Systems Inc., The D-Wave 2000Q Quantum Computer Technology Overview (2018), https://www.dwavesys.com/sites/default/files/D-Wave%202000Q%20Tech%20Collateral_1029F.pdf.

[23] Davide Venturelli and Alexei Kondratyev, Reverse quantum annealing approach to portfolio optimization problems, Quant. Mach. Intell. **1**, 17 (2019).

[24] Walter Vinci, Lorenzo Buffoni, Hossein Sadeghi, Amir Khoshaman, Evgeny Andriyash, and Mohammad H. Amin, A path towards quantum advantage in training deep generative models with quantum annealers, arXiv:1912.02119 [quant-ph] (2019).

[25] G. Passarelli, V. Cataudella, and P. Lucignano, Improving the quantum annealing of the ferromagnetic p-spin model through pausing, arXiv:1902.06788 (2019).

[26] Gianluca Passarelli, Ka-Wa Yip, Daniel A. Lidar, Hidetoshi Nishimori, and Procolo Lucignano, Reverse quantum annealing of the p-spin model with relaxation, Phys. Rev. A **101**, 022331 (2020).

[27] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press, Oxford, 2002).

[28] Robert Alicki and K. Lendi, *Quantum Dynamical Semigroups and Applications* (Springer Science & Business Media, Berlin Heidelberg, 2007).

[29] Daniel A. Lidar, Lecture notes on the theory of open quantum systems, arXiv:1902.00967 (2019).

[30] Tameem Albash, Sergio Boixo, Daniel A. Lidar, and Paolo Zanardi, Quantum adiabatic Markovian master equations, New J. Phys. **14**, 123016 (2012).

[31] Tameem Albash and Daniel A. Lidar, Decoherence in adiabatic quantum computation, Phys. Rev. A **91**, 062320 (2015).

[32] Anatoly Yu Smirnov and Mohammad H. Amin, Theory of open quantum dynamics with hybrid noise, New J. Phys. **20**, 103037 (2018).

[33] Roie Dann, Amikam Levy, and Ronnie Kosloff, Time-dependent Markovian quantum master equation, Phys. Rev. A **98**, 052129 (2018).

[34] Evgeny Mozgunov and Daniel Lidar, Completely positive master equation for arbitrary driving and small level spacing, Quantum **4**, 227 (2020).

[35] Frederik Nathan and Mark S. Rudner, Universal Lindblad equation for open quantum systems, arXiv:2004.01469 [cond-mat.mes-hall] (2020).

[36] Humberto Munoz-Bauza, Huo Chen, and Daniel Lidar, A double-slit proposal for quantum annealing, npj Quant. Inf. **5**, 51 (2019).

[37] Walter Vinci and Daniel A. Lidar, Non-stoquastic Hamiltonians in quantum annealing via geometric phases, npj Quant. Inf. **3**, 38 (2017).

[38] A. T. Rezakhani, D. F. Abasto, D. A. Lidar, and P. Zanardi, Intrinsic geometry of quantum adiabatic evolution and quantum phase transitions, Phys. Rev. A **82**, 012321 (2010).

[39] Sabine Jansen, Mary-Beth Ruskai, and Ruedi Seiler, Bounds for the adiabatic approximation with applications to quantum computation, J. Math. Phys. **48**, 102111 (2007).

[40] Walter Vinci and Daniel A. Lidar, Non-stoquastic hamiltonians in quantum annealing via geometric phases, npj Quant. Inf. **3**, 38 (2017).

[41] Sergey Bravyi, David P. DiVincenzo, and Daniel Loss, Schrieffer–Wolff transformation for quantum many-body systems, Ann. Phys. **326**, 2793 (2011).

[42] Gioele Consani and Paul A. Warburton, Effective Hamiltonians for interacting superconducting qubits: Local basis reduction and the Schrieffer–Wolff transformation, New J. Phys. **22**, 053040 (2020).

[43] Daniel A. Lidar, Ali T. Rezakhani, and Alioscia Hamma, Adiabatic approximation with exponential accuracy for many-body systems and quantum computation, J. Math. Phys. **50**, 102106 (2009).

[44] N. G. Dickson *et al.*, Thermally assisted quantum annealing of a 16-qubit problem, Nat. Commun. **4**, 1903 (2013).

[45] The expression we arrive at here for Γ here is slightly different from Ref. [36] since here the RWA is done in the adiabatic frame, while in Ref. [36] the RWA is done in an additional rotating frame.

[46] Fei Yan, Simon Gustavsson, Archana Kamal, Jeffrey Birenbaum, Adam P. Sears, David Hover, Ted J. Gudmundsen, Danna Rosenberg, Gabriel Samach, S. Weber, Jonilyn L. Yoder, Terry P. Orlando, John Clarke, Andrew J. Kerman, and William D. Oliver, The flux qubit revisited to enhance coherence and reproducibility, Nat. Commun. **7**, 12964 (2016).

[47] C. M. Quintana *et al.*, Observation of Classical-Quantum Crossover of $1/f$ Flux Noise and its Paramagnetic Temperature Dependence, Phys. Rev. Lett. **118**, 057702 (2017).

[48] S. Novikov, R. Hinkey, S. Disseler, J. I. Basham, T. Albash, A. Risinger, D. Ferguson, D. A. Lidar, and K. M. Zick, in *2018 IEEE International Conference on Rebooting Computing (ICRC)* (IEEE, McLean, VA, USA, 2018), p. 1–7.

[49] Mostafa Khezri, Jeffrey A. Grover, James I. Basham, Steven M. Disseler, Huo Chen, Sergey Novikov, Kenneth M. Zick, and Daniel A. Lidar, Anneal-path correction in flux qubits, arXiv:2002.11217 [quant-ph] (2020).

[50] William D. Oliver, Yang Yu, Janice C. Lee, Karl K. Berggren, Leonid S. Levitov, and Terry P. Orlando, Mach-Zehnder interferometry in a strongly driven superconducting qubit, Science **310**, 1653 (2005).

[51] Recall our convention of denoting by $Q(s)$ the original quantity and by $Q(\tau)$ the corresponding paused quantity. In the case of $\Gamma(s)$ this includes the paused anneal time t_f', so in that sense it represents a mixed quantity.

[52] Note that henceforth, for notational simplicity, we use the prime symbol for the derivative with respect to s throughout this work.

[53] These two assumptions are deduced from our observations of examples we study numerically. In principle, one could choose a different set of assumptions and perform a similar analysis. The final result will be different from Theorem 1.

[54] Eric W. Weisstein, Lambert W-Function. From MathWorld–A Wolfram Web Resource, https://mathworld.wolfram.com/LambertW-Function.html.

[55] Boris Altshuler, Hari Krovi, and Jérémie Roland, Anderson localization makes adiabatic quantum optimization fail, Proc. Natl. Acad. Sci. **107**, 12446 (2010).

[56] Note that if $\bar{X}'(1) > 0$, the RHS of Eq. (66) is automatically negative, since the prefactor $|\bar{X}'(1)|$ comes from $-\bar{X}'(1)$, and the $-|\bar{X}'(1)|$ inside the curly brackets becomes $\bar{X}'(1)$, so every term inside these brackets is positive. In this case Assumptions 3 and 4 of Theorem 1 can be dropped. However, $\bar{X}'(1) < 0$ in our model.

[57] Zoe Gonzalez Izquierdo, Shon Grabbe, Stuart Hadfield, Jeffrey Marshall, Zhihui Wang, and Eleanor Rieffel, Ferromagnetically shifting the power of pausing, arXiv:2006.08526 (2020).

[58] Seung Woo Shin, Graeme Smith, John A. Smolin, and Umesh Vazirani, How "quantum" is the D-Wave machine? arXiv:1401.7087 (2014).

[59] Jeffrey Marshall and Tameem Albash (private communication).

[60] J. Bezanson, A. Edelman, S. Karpinski, and V. Shah, Julia: A fresh approach to numerical computing, SIAM Rev. **59**, 65 (2017).

[61] Christopher Rackauckas and Qing Nie, DifferentialEquations.jl – A performant and feature-rich ecosystem for solving differential equations in Julia, J. Open Res. Softw. **5**, 15 (2017).

[62] Alan L. Andrew and Roger C. E. Tan, Computation of derivatives of repeated eigenvalues and the corresponding eigenvectors of symmetric matrix pencils, SIAM J. Matrix Anal. Appl. **20**, 78 (1998).