Energetic Perspective on Rapid Quenches in Quantum Annealing

Adam Callison[,1,*] Max Festenstein,[1,2] Jie Chen[,2] Laurentiu Nita,[2] Viv Kendon[,2] and
Nicholas Chancellor[2,†]

[1]*Blackett Laboratory, Imperial College London, London SW7 2BW, United Kingdom*

[2]*Department of Physics and Joint Quantum Centre Durham-Newcastle, Durham University, South Road,
Durham DH1 3LE, United Kingdom*

There are well-developed theoretical tools to analyze how quantum dynamics can solve computational problems by varying Hamiltonian parameters slowly, near the adiabatic limit. On the other hand, there are relatively few tools to understand the opposite limit of rapid quenches, as used in quantum annealing and (in the limit of infinitely rapid quenches) in quantum walks. In this paper, we develop several tools that are applicable in the rapid-quench regime. Firstly, we analyze the energy expectation value of different elements of the Hamiltonian. From this, we show that monotonic quenches, where the strength of the problem Hamiltonian is consistently increased relative to fluctuation (driver) terms, will yield a better result on average than random guessing. Secondly, we develop methods to determine whether dynamics will occur locally under rapid-quench Hamiltonians and identify cases where a rapid quench will lead to a substantially improved solution. In particular, we find that a technique we refer to as "preannealing" can significantly improve the performance of quantum walks. We also show how these tools can provide efficient heuristic estimates for Hamiltonian parameters, a key requirement for practical application of quantum annealing.

I. INTRODUCTION

Quantum computing using continuous-time evolution has gained much interest in recent years. This includes adiabatic quantum computing (AQC) [1], quantum annealing [2,3], and continuous-time quantum walks (QWs) [4]. Optimization tasks are a natural application for quantum computing in this setting and have been explored in many diverse fields, including traditional computer science [5–7], decoding communications [8], finance [9–11], error correction of quantum memories [12], scheduling [13–15], computational biology [16], flight gate assignment [17], air traffic management [18], and hydrology [19]. This is partially due to advances in the theoretical foundations of adiabatic quantum computing, including proofs that it is universal in certain settings [20,21], improved versions of the adiabatic condition [22–24], and an extension of the adiabatic theorem to open systems [25]. For a comprehensive review of these and other advances, see

Ref. [26]. More recently, it was shown by Hastings [27] that even when no sign problem exists, there is a superpolynomial oracle separation between adiabatic quantum computing and classical computing. Other recent advances have come from new ways to map problems; for instance, the methods of encoding more connected graphs than the native hardware connections using parity, often called "parity AQC" [28–31]. These provide an alternative to the more traditional minor embedding techniques [32,33] and may be easier to implement experimentally.

In this paper, we focus on the coherent regime of operation, for which the effects of thermal dissipation and decoherence can be ignored. Such a regime could be experimentally reached by reducing noise, implementing quantum error correction [34–42], or quenching on a timescale that is much faster than the decoherence time. A complementary approach to reduce noise is to implement dynamics that reduce or eliminate the interaction between the system and its environment through quantum interference effects, known as dynamical decoupling [42–44]. Although current superconducting quantum annealing hardware operates in a dissipative regime [45], quantum annealing has been implemented in atomic settings, where coherence is easier to maintain than in superconducting circuits [46], and progress has been made to reduce noise in superconducting circuit settings [47]. There have been experimental implementations of simple forms of error

correction in quantum annealing [40,48–51], and efforts have been made to circumvent experimental limitations on quench rates in superconducting systems [52].

In a fully coherent regime, the dynamics are straightforward to model theoretically, since they can be described by a set of qubits (two-state quantum systems) under the action of a Hamiltonian, evolving according to the Schrödinger equation. Conventionally, the Hamiltonian for this evolution is written as the sum of a problem Hamiltonian H_{prob}, which is diagonal in the computational basis and encodes the classical problem being solved, and a driver Hamiltonian H_{driver}, which implements quantum dynamics to explore the solution space. We use two equivalent forms for the total Hamiltonian. First,

$$H_{AB}(t) = A(t)H_{driver} + B(t)H_{prob}, \qquad (1)$$

where $A(t)$ and $B(t)$ are positive, time-dependent control functions. However, typically the crucial feature is what happens to the ratio of driver strength to problem strength $A(t)/B(t)$ as the algorithm progresses. Therefore, we define an alternative parametrization of the Hamiltonian, up to an overall (time-dependent) scaling factor $B(t)$, as

$$H_{\Gamma}(t) = \Gamma(t)H_{driver} + H_{prob}, \qquad (2)$$

where there is a single control function $\Gamma(t) > 0$ for the ratio $A(t)/B(t)$. Since Eqs. (1) and (2) are equivalent up to a rescaling of the time parameter, results for one form of the Hamiltonian will generalize to results for the other. We use both forms, choosing the most convenient for the specific problem or example.

Hamiltonians of the form of Eqs. (1) and (2), which begin with $A(t) > 0$ and $B(t) = 0$ and end with $A(t) = 0$ and $B(t) > 0$, or equivalently begin with $\Gamma(t) \gg 1$ and end with $\Gamma(t) = 0$, are used for most types of continuous-time quantum computing. When such a protocol is run on a much shorter timescale than required for adiabatic quantum computing, we call this a "rapid quench."

The simplest form of continuous-time quantum computing in the coherent regime is continuous-time QW introduced in Refs. [4,53], in which the control functions are time independent and set so that $\Gamma(t) = \gamma$, where γ is a constant hopping rate. This can be viewed as the limit of an infinitely fast quench, in which $B(0)$ jumps from zero to $A(0)/\gamma$ at $t = 0$ and $A(t_f)$ drops to zero at the final time t_f. The other pure state continuous-time quantum computing that is commonly considered is AQC introduced in Ref. [1], for which the control functions $A(t)$ and $B(t)$ are varied slowly from $A(0) = 1$ and $B(0) = 0$ to $A(t_f) = 0$ and $B(t_f) = 1$. By the adiabatic theorem of quantum mechanics, this achieves a success probability (probability of finding the ground state of the problem Hamiltonian H_{prob}) that approaches 1 as $t_f \rightarrow \infty$. For a review of AQC, see Ref. [26]. For a thorough discussion of the relationship between AQC and QW,

see the introductions in Refs. [54,55]. The fully coherent regime has provable quantum speedups in the case of both AQC and QW. For instance, unstructured search, the continuous-time analog of Grover's search, can yield the same speedup in the AQC [56] and QW [53] settings as the gate-based counterpart. It is possible to interpolate between these two techniques while preserving the speedup [54].

For problems that are closer to real-world optimization, theoretical studies have focused mostly on AQC [26], likely because the adiabatic theorem provides a general way to show that such algorithms could, in principle, succeed with high probability. While theoretically tractable, the adiabatic regime is difficult to reach experimentally and contains some counterintuitive effects in the deep adiabatic regime [57–59]. Solving NP-Hard problems (i.e problems at least as hard as any in the non-deterministic polynomial time complexity class NP) adiabatically will at most obtain a polynomial speed up (assuming P $\neq$ NP). Since AQC requires the system to remain coherent throughout, an exponentially long run time requires an exponentially long coherence time, which is experimentally challenging for near-term quantum computing. When the run time is limited by a constant or mildly scaling coherence time, such an algorithm could solve the problem only with an exponentially low probability and therefore requires exponentially many repeats to succeed with high probability. This approach, however, is a valid one for problems other than search. Recent numerical results on spin glasses using QW show favorable scaling from many short run repeats [55]. It has also been numerically demonstrated that rapid quenches can be superior to long quenches for AQC-like algorithms [60]. Recently, Crosson and Lidar [61] made an important contribution to the theory of quantum annealing outside the adiabatic limit by introducing diabatic quantum annealing, which formalizes ideas described in Refs. [62,63]. Diabatic quantum annealing relies on a generalization of the adiabatic theorem from Ref. [64] and describes a class of quantum annealing algorithms in which the amplitude is restricted to a low-energy part of the Hamiltonian spectrum.

Finally, for single-shot, high success probability algorithms for NP-hard problems, achieving even a polynomial speedup typically requires setting, with exponential precision, the control functions to values that lead to exponential small gaps in the Hamiltonian spectrum. This was shown to be necessary for unstructured search in Refs. [54,56,65] and for the random energy model [66] in Ref. [55]. This requirement is problematic, as there are no general methods for determining where these gaps occur and because such precise control settings can be difficult to achieve in real hardware. Recent work by Chakraborty *et al.* [67] demonstrates that some of the fine-tuning requirements in unstructured search can be avoided by formulating the Hamiltonian differently; it is unclear whether

this approach would extend to the random energy model of Ref. [66].

Given the near-term importance of methods that can succeed with limited coherence time, in this paper, we develop mathematical tools to increase our understanding of how computation is achieved in both the rapid-quench regime and quantum walks. These tools are important not only for theoretical understanding of when adiabatic algorithms and rapid quenches will be effective but also for choosing parameters for the Hamiltonians used. While some theoretical arguments [55,68] can be made for why QW with short run times seems to perform well, a theoretical understanding of rapid quenches with time-dependent Hamiltonians, but far from the regime where the adiabatic theorem applies, is essentially missing.

It was recently shown numerically [69] that the optimal protocol for solving problems often involves an annealing step, as opposed to bang-bang controls where driver and problem Hamiltonians are not active simultaneously. Previous theoretical work based on Pontryagin's minimum principle showed that optimal control patterns would always take the bang-bang form [70], but that these controls would sometimes require unphysically switching between the driver and problem Hamiltonians an infinite number of times in a finite time span. The work by Brady *et al.* [69] is restricted to cases with a finite number of "bangs" and found that in this more realistic setting protocols involving annealing may be superior.

We begin in Sec. II with some numerical examples to illustrate the performance gains that can be obtained from well-chosen rapid quenches in quantum annealing. This provides motivation to understand why rapid quenches work and how to exploit the effects more systematically. We then analyze the energy flow between different quantum states, altering the expectation values of driver and problem terms in the Hamiltonian, as laid out in Sec. III. Next we provide a general set of conditions (essentially requiring that quenches be monotonic) under which rapid quenches will preferentially seek out high-quality solutions. We augment this analysis by studying the transitions between different computational basis states, to deduce the level of dynamics that will occur, in Sec. IV, and we apply our tools to different problem settings, including discussing the conditions for general optimization problems to yield a significant level of dynamics. In Sec. V we show how the tools developed here can be used to construct heuristics for setting the parameters for continuous-time quantum walks and rapid quenches. Section VI provides details of our numerical methods, and we summarize and discuss our results in Sec. VII.

II. RAPID QUENCH EXAMPLES

As motivation for our theoretical tools, we start with three illustrative examples showing the power of rapid

quenches to solve problems. For simplicity and concreteness, we focus on monotonic quenches; that is, quenches for which the control parameter $\Gamma(t') \leq \Gamma(t)$ for all $t' > t$.

A. Two-stage quantum walk

This is a minimal modification to the time-independent continuous-time quantum walk. It consists of two time-independent stages of evolution separated by an infinitely fast quench. Because each stage is effectively a continuous-time quantum walk, we refer to this as a "two-stage quantum walk." We use a simple transverse field driver Hamiltonian

$$H_{\text{driver}} = n\mathbb{1} - \sum_{j=1}^{n} \hat{X}_j, \qquad (3)$$

where $\mathbb{1}$ is the identity operator and $\hat{X}_j$ is the Pauli $\hat{X}$ operator acting in qubit j. Instead of using a constant control function ($\Gamma(t) = \gamma$), we use the time-dependent schedule

$$\Gamma(t) = \begin{cases} \gamma_1, & 0 < t < t_1, \\ \gamma_2, & t_1 < t < (t_1 + t_2), \end{cases} \qquad (4)$$

which consists of two consecutive evolution stages with two different time-independent Hamiltonians. Each of these stages is effectively a quantum walk, although the second stage uses nonstandard starting conditions as its initial state is the final state of the first stage. The standard initial state is the equal superposition of all basis states, $|\psi_0\rangle = 2^{-n/2} \sum_j |j\rangle$, chosen because it is the ground state of the driver Hamiltonian and also represents our ignorance of which basis state is the solution to the problem. The schedules we use for the two-stage quantum walks are shown in Fig. 1 for each of our three examples. As dis-
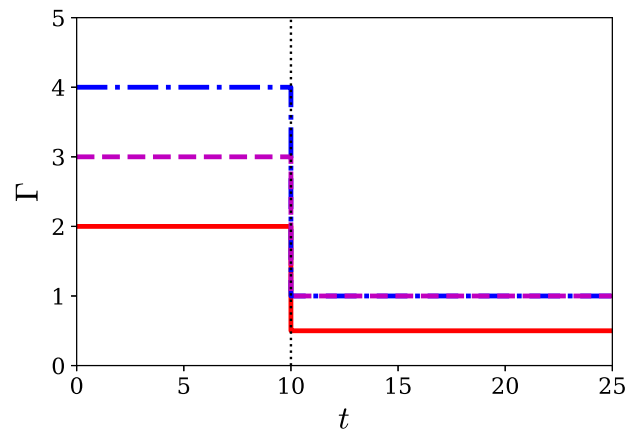

FIG. 1. The annealing schedule for the two-stage quantum walks in Fig. 2 (solid red line), Fig. 3 (dot-dashed blue line), and Fig. 4 (dashed magenta line). In all cases the step occurs at $t = 10$ (dotted black line).

cussed in Ref. [55], a quantum walk can be understood from an energetic perspective according to a mechanism referred to there as the "energy conservation mechanism." Being time independent, quantum walks conserve the total energy of the system. To show the effect of changing the hopping rate γ part way through the walk, thus disrupting the energy conservation, our first example is a simple two-qubit problem Hamiltonian

$$H_{\text{prob}}^{(2Q)} = -\hat{Z}_1\hat{Z}_2 - \frac{1}{2}\hat{Z}_1, \tag{5}$$

where $\hat{Z}_j$ is the Pauli $\hat{Z}$ operator acting on qubit j. We start the system at $t = 0$ in the state $|\psi_0\rangle = \frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle)$, the two-qubit ground state of the driver Hamiltonian H_{driver} in Eq. (3). To simplify the notation, we define $\langle H_{\text{prob}}\rangle_{\psi(t)} \equiv \langle\psi(t)|H_{\text{prob}}|\psi(t)\rangle$, the instantaneous expectation value of the problem Hamiltonian with respect to the state $|\psi(t)\rangle$ at time t. Likewise, $\langle H_{\text{driver}}\rangle_{\psi(t)} \equiv \langle\psi(t)|H_{\text{driver}}|\psi(t)\rangle$ for the driver Hamiltonian. We have the total energy $E_{\Gamma}(t) = \Gamma(t)\langle H_{\text{driver}}\rangle + \langle H_{\text{prob}}\rangle$.

Figure 2 (top) shows that the expectation value $\langle H_{\text{driver}}\rangle$ for the transverse field is zero initially ($t = 0$). As in Ref. [55], the energy conservation mechanism then decreases the expectation value of the problem Hamiltonian at the expense of increasing the expectation value of the driver Hamiltonian. When the instantaneous quench is performed, the problem Hamiltonian expectation value is unchanged, but the driver Hamiltonian expectation value [and therefore the total energy expectation value $E_{\Gamma}(t)$] is reduced. As the minimum eigenvalue of H_{driver} is zero, the

total energy expectation value $E_{\Gamma}(t)$ acts as an effective upper bound on $\langle H_{\text{prob}}\rangle_{\psi(t)}$. The net effect is that even if all of the energy stored in the transverse field were returned to the problem Hamiltonian, its expectation value would still be less than it was at the beginning of the algorithm. What actually happens, however, is that the transverse field is able to capture even more of the energy, thereby reducing the problem Hamiltonian expectation value further and increasing the average probability of finding the ground state; see Fig. 2 (bottom).

A more realistic problem is the Sherrington-Kirkpatrick (SK) spin-glass [71] ground-state problem investigated in Ref. [55]. This has the problem Hamiltonian

$$H_{\text{prob}}^{(\text{SK})} = -\frac{1}{2}\sum_{(a\neq b)=0}^{n-1} J_{ab}\hat{Z}_a\hat{Z}_b - \sum_{b=0}^{n-1} h_b\hat{Z}_b, \tag{6}$$

where the couplings J_{ab} and fields h_b are drawn independently from the normal distribution $\mathcal{N}(0, \sigma^2)$ with mean 0 and variance σ^2.

Figure 3 shows a two-stage quantum walk performed on a nine-qubit Sherrington-Kirkpatrick Hamiltonian [73] from the public repository in Ref. [72] associated with Ref. [55]. In the setting of this larger problem, the fluctuations after each stage of the quantum walk are smaller relative to the dynamical range than in the two-qubit case, a very early sign of the approach to the thermodynamic limit. Apart from this, the behavior is qualitatively similar to that of the two-qubit toy model $H_{\text{prob}}^{(2Q)}$ of Eq. (5) shown in Fig. 2 and produces a significant increase in the probability of finding the ground state.
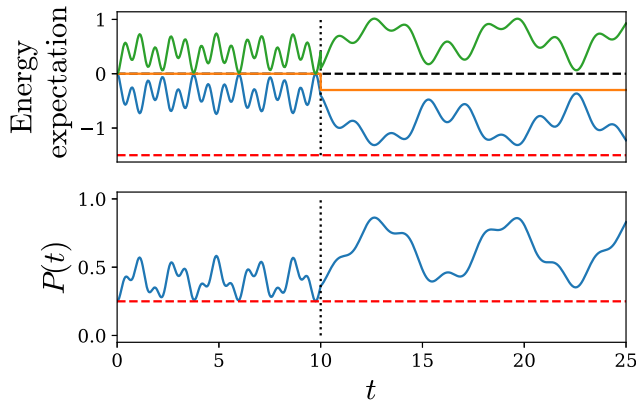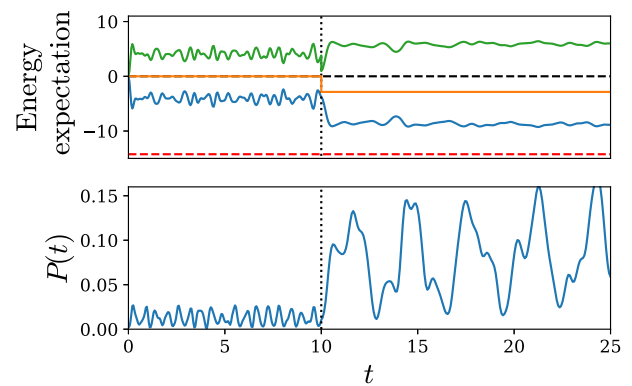

FIG. 2. Two-stage quantum walk using the Hamiltonian in Eq, (5) with $\gamma_1 = 2$ and $\gamma_2 = \frac{1}{2}$. The instantaneous quench occurs at time $t_1 = 10$ (vertical dotted line). The top panel shows energy expectation values $E_{\Gamma} = \Gamma\langle H_{\text{driver}}\rangle + \langle H_{\text{prob}}\rangle$ (gold line), $\Gamma\langle H_{\text{driver}}\rangle$ (green line), and $\langle H_{\text{prob}}\rangle$ (blue line). Also shown are a guide for the eye at zero energy (dashed black line) and the minimum eigenvalue of H_{prob} (dashed red line). The bottom panel shows the probability $P(t)$ of being in the ground state of H_{prob} at time t (blue line) and the probability of random guessing (dashed red line).

FIG. 3. Two-stage quantum walk on a nine-qubit Sherrington-Kirkpatrick spin glass, ID code ovcjhwbhtcpcvwicoxpdpvjzqo-jril from the public repository in Ref. [72] with $\gamma_1 = 4$ and $\gamma_2 = 1$. The instantaneous quench occurs at time $t_1 = 10$, (vertical dotted line). The top panel shows energy expectations $E_{\Gamma} = \Gamma\langle H_{\text{driver}}\rangle_{\psi(t)} + \langle H_{\text{prob}}\rangle_{\psi(t)}$ (gold line), $\Gamma\langle H_{\text{driver}}\rangle_{\psi(t)}$ (green line), and $\langle H_{\text{prob}}\rangle_{\psi(t)}$ (blue line). Also shown are a guide for the eye at zero energy (dashed black line) and the minimum eigenvalue of H_{prob} (dashed red line). The bottom panel shows the probability $P(t)$ of being in the ground state of H_{prob} at time t (blue line).

B. Biased two-stage quantum walk

We introduce a biased driver Hamiltonian, similar to the one used in Refs. [74,75]. We formulate our biased driver Hamiltonian slightly differently as

$$H_{\text{bias}}(g, \theta) = n\mathbb{1} - \sum_{i=1}^{n} \left[\cos(\theta)\hat{X}_i + g_i \sin(\theta)\hat{Z}_i \right], \quad (7)$$

where $g_i \in \{-1, 1\}$ is a candidate (or guess) solution, and takes the value 1 if the ith bit of the guess solution is 0 and the value -1 if it is 1. The certainty of the guess is parametrized by $0 \leq \theta \leq \pi/2$; if $\theta = 0$, the guess goes unused and the driver reduces to a transverse field of Eq. (3). In the other extreme, if $\theta = \pi/2$, then the ground state of $H_{\text{bias}}(g, \theta = \pi/2)$ is the candidate solution and there are no dynamics. The ground state of the biased driver Hamiltonian has zero energy for all allowed values of θ and g, and is a tensor product of spin states that are each antiparallel to the fields in Eq. (7); this state is used as the initial state. For simplicity, in this example we consider only biasing toward the optimal solution (i.e., correct guesses), and we use the same nine-qubit SK spin glass as in the previous subsection.

As Fig. 4 shows, the effect of biasing toward the optimal solution is to lower the initial values of E_Γ and $\langle H_{\text{prob}} \rangle_{\psi(t)}$; biasing toward a well-chosen guess effectively gives the algorithm a "head start" with respect to energy expectation values. This is qualitatively similar to what happens at
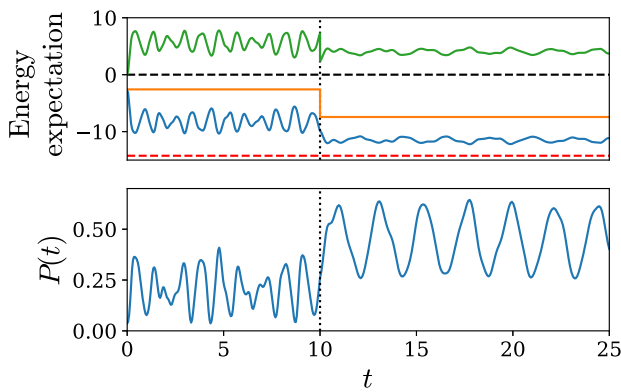

FIG. 4. Biased two-stage quantum walk on a nine-qubit Sherrington-Kirkpatrick spin glass, ID code ovcjhwb-htcpcvwicoxpdpvjzqojril from the public repository in Ref. [72] with $\gamma_1 = 3$ and $\gamma_2 = 1$, using a biased driver [Eq. (7)], biased toward the optimal solution of H_{prob} using $\theta = \pi/8$. The instantaneous quench occurs at time $t_1 = 10$ (vertical dotted line). The top panel shows energy expectations $E_\Gamma = \Gamma \langle H_{\text{driver}} \rangle_{\psi(t)} + \langle H_{\text{prob}} \rangle_{\psi(t)}$ (gold line), $\Gamma \langle H_{\text{driver}} \rangle_{\psi(t)}$ (green line), and $\langle H_{\text{prob}} \rangle_{\psi(t)}$ (blue line). Also shown are a guide for the eye at zero energy (dashed black line) and the minimum eigenvalue of H_{prob} (dashed red line). The bottom panel shows the probability $P(t)$ of being in the ground state of H_{prob} at time t (blue line).

the beginning of the second stage of the two-stage quantum walk, except that the driver energy $\langle H_{\text{bias}}(g, \theta) \rangle_{\psi(t)}$ starts at exactly zero rather than having some initial energy left over from a previous stage. The bias increases the initial-stage success probability by a factor of 10 compared with the unbiased walk in Fig. 3, while the second stage again provides a (further) factor of 3 increase. This biased two-stage quantum walk example provides proof of concept that the mechanism we describe can be leveraged on top of a biased search. A thorough analysis of biased (single-stage) quantum walks as a subroutine for hybrid quantum-classical computing is to be published [76].

C. Preannealed quantum walk

Our final example is again in two stages, but this time the first stage is a quantum annealing stage and the second stage is a quantum walk that starts from the point where the annealing stops. The motivating intuition is that the initial time-dependent annealing stage will prepare an initial state for the quantum walk that has a lower average problem energy $\langle H_{\text{prob}} \rangle_{\psi(t)}$ than the usual uniform superposition state. If performed too slowly, such a quench will put the system into its instantaneous ground state, by the adiabatic theorem of quantum mechanics, and there will be no quantum walk dynamics. If performed too rapidly, the state will not evolve much during the annealing stage and the resulting quantum walk will be similar to one without a preannealing stage. However, if the annealing is performed at an intermediate rate, it leads to significant quantum walk dynamics, starting from a lower problem Hamiltonian expectation value $\langle H_{\text{prob}} \rangle_{\psi(t)}$.

Using the H_{AB} parametrization defined in Eq. (1), we consider preannealing with a quadratic schedule for a time t_1, and then a steady-state quantum walk afterward; specifically, we define the schedule

$$A(t) = \begin{cases} \gamma \left[1 + \left(\dfrac{t}{t_1} - 1 \right)^2 \right], & 0 \leq t \leq t_1, \\ \gamma, & t_1 < t \leq (t_1 + t_2), \end{cases}$$
$$(8)$$

$$B(t) = \begin{cases} \left[1 - \left(\dfrac{t}{t_1} - 1 \right)^2 \right], & 0 \leq t \leq t_1 \\ 1, & t_1 < t \leq (t_1 + t_2), \end{cases} \quad (9)$$

which is plotted in Fig. 5 for the values of t_1 we use. Using the same nine-qubit SK problem as before, with its optimal γ value of approximately 1.004, we show the results for three different values of t_1 in Fig. 6. Preannealing both decreases the average problem expectation value $\langle H_{\text{prob}} \rangle_{\psi(t)}$ and increases the success probability, but causes the peak values to be reached more slowly. In the longest preannealing with $t_1 = 4$, the success probability undergoes small-amplitude, approximately sinusoidal,
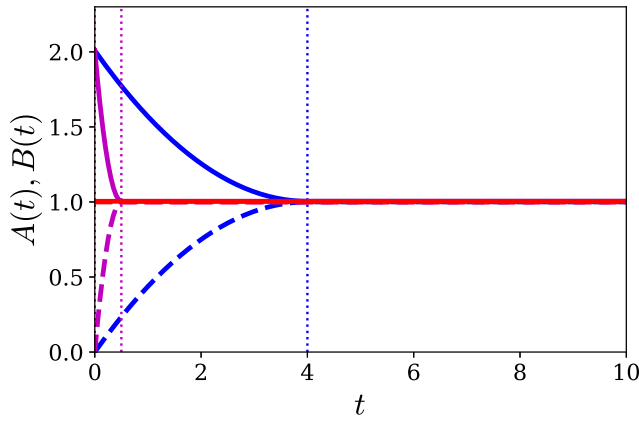
FIG. 5.   Schedule $A(t)$ (solid lines) and $B(t)$ (dashed lines) of a preannealed quantum walk using $\gamma \approx 1.004$ and $t_1 = 4$ (vertical dotted blue line), $t_1 = 0.5$ (vertical dotted magenta line), and $t_1 = 0$ (red line; pure QW).

oscillations, suggesting that the dynamics are dominated by a two-level subspace. For $t_1 = 0.5$ and $t_1 = 0$, the oscillations are less structured, indicating that more than two energy levels are playing a nontrivial role in the dynamics. The increases in the success probability seen in Fig. 6 are relatively modest for this example. To determine the typical increase in success probability due to preannealing, we use all 10 000 Sherrington-Kirkpatrick instances from Ref. [72] at each size from $n = 5$ to $n = 11$ and compare the quantum walk success probability averaged over the quantum walk stage using 20 different linearly spaced preannealing times up to $t_1 = 4$. In Fig. 7 (top),
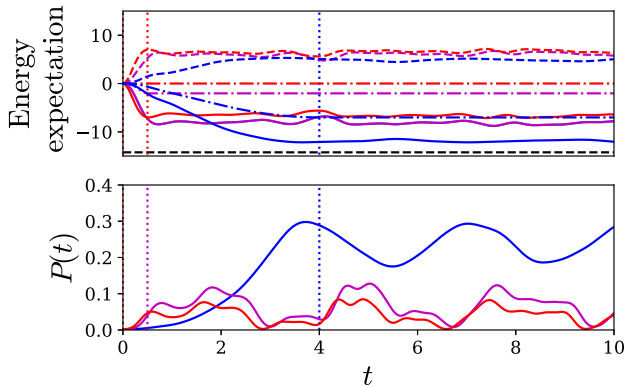

FIG. 6.   Preannealing performed on a nine-qubit Sherrington-Kirkpatrick spin glass, ID code ovcjhwbhtcpcvwicoxpdpvjzqo-jril from Ref. [72], for preannealing times $t_1 = 4$ (blue lines), $t_1 = 0.5$ (magenta lines), and $t_1 = 0$ (red lines; i.e., pure quantum walk). Dotted lines show when the preannealing ends. The top panel shows expectation values $E_\Gamma(t) = (A/B)\langle H_{\mathrm{driver}}\rangle_{\psi(t)} + \langle H_{\mathrm{prob}}\rangle_{\psi(t)}$ (dot-dashed lines), $\langle H_{\mathrm{prob}}\rangle_{\psi(t)}$ (solid lines), and $\frac{A}{B}\langle H_{\mathrm{driver}}\rangle_{\psi(t)}$ (dashed colored lines). The dashed black line indicates the minimum eigenvalue of H_{prob}. The bottom panel shows the success probability $P(t)$ to be in the lowest eigenstate of H_{prob} at time t.
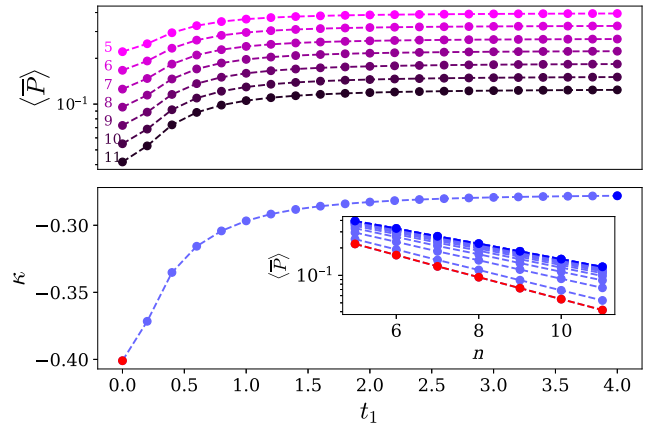

FIG. 7.   The top panel shows the success probability $\langle \overline{P} \rangle$ for $n = 5$ to $n = 11$ for 21 different linearly spaced preannealing times from $t_1 = 0$ to $t_1 = 4$; darker magenta color indicates higher n. All data are averaged over all 10 000 Sherrington-Kirkpatrick instances from Ref. [72] at each size. The bottom panel shows the scaling exponent κ for a model where $p_{\mathrm{success}} \propto 2^{\kappa n}$ extracted from the linear fit on log-linear axes for different preannealing times in the inset. The inset shows the scaling of success probability versus n for the same t_1 values, with $t_1 = 0$ in red and $t_1 = 4$ in dark blue (same color coding as the main figure).

we see that the success probability increases with preannealing time, up to a plateau, and the relative effect of preannealing becomes larger as n increases. To quantify this effect, we calculate the scaling exponent at each preannealing time by fitting a linear model on log-linear axes. We find a scaling exponent κ such that the success probability $p_{\mathrm{success}} \propto 2^{\kappa n}$. The fitted values of κ are plotted in Fig. 7 (bottom). As the inset in Fig. 7 (bottom) shows, the success probability is modeled well by a simple exponential function, as in Ref. [55]. We find that preannealing significantly improves the scaling from $\kappa = -0.418$ for a pure quantum walk, in agreement with Ref. [55], to a maximum of $\kappa \approx -0.278$. It is, of course, an open question whether this scaling will continue to problem sizes that are of practical interest, but the lack of visible finite-size effects in Fig. 7 suggests that it might. Since very fast quenches can be experimentally challenging to implement, although methods are being explored [52], determining the effects of quenching at a finite rate is of practical importance. Our results show that such quenches are potentially a better strategy than trying to speed up or slow down to approach QW or adiabatic extremes.

III. ENERGY REDISTRIBUTION MECHANISM

In all the examples in Sec. II, we observe that the total energy expectation value $E_\Gamma(t)$ never increases during a rapid quench and that $E_\Gamma(t)$ serves as an upper bound to the

problem expectation value $\langle H_{\text{prob}}\rangle_{\psi(t)}$, assuming that the ground state of H_{driver} is arranged to be at zero energy [the identity term in Eq. (3) ensures this]. In this section, we formalize these observations into a mechanism that we refer to as the "energy redistribution mechanism." Our analysis extends the energy conservation mechanism described in Ref. [55] and recapped in Appendix A 1 (similar arguments are also made by Hastings [68]) to quenches where the Hamiltonian is not time invariant and therefore total energy is not conserved.

Consider a closed-system quantum annealing schedule on a system with a Hamiltonian $H(t)$ defined by Eq. (2):

$$H(t) = \Gamma(t)H_{\text{driver}} + H_{\text{prob}}. \qquad (10)$$

We show that (for duration $t_f \geq 0$) the energy expectation value with respect to the problem Hamiltonian at the end is never higher than at the initial time $t = 0$,

$$\langle \psi(t_f) \mid H_{\text{prob}} \mid \psi(t_f)\rangle \leq \langle \psi(0) \mid H_{\text{prob}} \mid \psi(0)\rangle, \quad (11)$$

provided the following conditions are satisfied:

(1) Initial ground state. The initial state $|\psi(t=0)\rangle$ is a ground state of the driver Hamiltonian H_{driver}.
(2) Positivity. The control function is non-negative: $\Gamma(t) \geq 0$ for all t.
(3) Monotonicity. The control function is monotonically decreasing: $\Gamma(t) \geq \Gamma(t')$ for all $t' > t$.

Condition 1 is simply that the system is initially prepared in the ground state of the driver Hamiltonian. This condition is necessary for AQC, and is also standard for QW. Condition 2 prevents pathological behavior where the driver spectrum is effectively inverted by taking negative values of the control function $\Gamma(t)$. This condition is satisfied in all traditional AQC and QW settings. Condition 3 is that the quench is monotonic; this condition excludes methods such as reverse annealing, both the dissipatively driven form proposed in Ref. [77] and implemented on D-Wave Systems devices [78] and the similar coherent method proposed in Ref. [79] that is sometimes also referred to as "reverse annealing." The biased driver Hamiltonian proposed in Refs. [74,75] is compatible with condition 3. Our results do not rely on the adiabatic theorem, and the control function $\Gamma(t)$ does not need to be a continuous function.

Without loss of generality, the driver Hamiltonian H_{driver} can be chosen such that its ground-state eigenvalue (and hence its expectation value with the initial state) is zero, $\langle \psi(0) \mid H_{\text{driver}} \mid \psi(0)\rangle = 0$. In other words, we impose semidefiniteness on H_{driver} by defining its ground state $|\psi(0)\rangle$ to have eigenvalue 0. Let

$$E_\Gamma(t) = \langle \psi(t) \mid H_\Gamma(t) \mid \psi(t)\rangle \qquad (12)$$

be the expectation value of the energy at time t. Then it follows immediately from condition 1 that at time $t = 0$,

$$E_\Gamma(0) = \langle \psi(0) \mid H_{\text{prob}} \mid \psi(0)\rangle. \qquad (13)$$

Furthermore, it follows from conditions 1 and 2 that at any later time $t > 0$,

$$E_\Gamma(t) \geq \langle \psi(t) \mid H_{\text{prob}} \mid \psi(t)\rangle, \qquad (14)$$

since $\langle \psi(t) \mid H_{\text{driver}} \mid \psi(t)\rangle \geq \langle \psi(0) \mid H_{\text{driver}} \mid \psi(0)\rangle = 0$ can only increase from the ground-state initial energy.

It can be shown that the energy expectation value $E_\Gamma(t)$ defined in Eq. (12) decreases monotonically in time; that is,

$$E_\Gamma(t') \leq E_\Gamma(t) \text{ for all } t, t' \ (t' > t). \qquad (15)$$

To see this, consider the discretized approximation to the evolution

$$|\psi_k^{(q)}\rangle = \mathcal{T} \prod_{k'=k}^{1} \exp\left[-iH_\Gamma\left(\frac{k't_f}{q}\right)\frac{t_f}{q}\right]|\psi(0)\rangle \qquad (16)$$

for $1 \leq k \leq q$ and where the symbol $\mathcal{T}$ is added to emphasize that the time order of the product must be preserved, since the Hamiltonians at different times are noncommuting. This discretized approximation becomes exact in the limit $q \to \infty$. The evolution of a quantum system under the time-dependent Hamiltonian given in Eq. (1) from time $t = 0$ to time $t = t_f$ from the initial state $|\psi(0)\rangle$ is broken down as follows: The initial state is evolved under the constant Hamiltonian $H(t_f/q)$ for time t_f/q to produce a state $|\psi_1^{(q)}\rangle$ that then evolves under the constant Hamiltonian $H(2t_f/q)$ for time t_f/q and so on, until a final state $|\psi_q^{(q)}\rangle$ is reached. Then, in the limit, $|\psi(t_f)\rangle = \lim_{q\to\infty}|\psi_q^{(q)}\rangle$. This kind of discretization can be thought of as an extension of the Suzuki-Trotter decomposition [80,81] and is therefore sometimes informally referred to as Trotterization. In the same manner, we can define a discretized version of the energy expectation value as

$$E_{\Gamma,k}^{(q)} = \Gamma\left(\frac{k't_f}{q}\right)\langle \psi_k^{(q)} \mid H_{\text{driver}} \mid \psi_k^{(q)}\rangle +$$
$$\langle \psi_k^{(q)} \mid H_{\text{prob}} \mid \psi_k^{(q)}\rangle. \qquad (17)$$

During each time-independent evolution step, the energy expectation value $E_{\Gamma,k}^{(q)}$ is conserved. Furthermore, since by definition H_{driver} is positive semidefinite and $\Gamma((k+1)t_f/q) \leq \Gamma(kt_f/q)$ (by conditions 2 and 3), it

follows that

$$E_{\Gamma,k+1}^{(q)} \leq E_{\Gamma,k}^{(q)}. \tag{18}$$

Repeated application of this inequality results in the more useful inequality

$$E_{\Gamma,q}^{(q)} \leq E_{\Gamma,1}^{(q)}. \tag{19}$$

Since $E_{\Gamma,1}^{(q)}$ is the energy during the whole of the first evolution step, it follows that

$$E_{\Gamma,1}^{(q)} = E_{\Gamma}(t=0). \tag{20}$$

Furthermore, we have

$$\lim_{q \to \infty} E_{\Gamma,q}^{(q)} = E_{\Gamma}(t_f), \tag{21}$$

which means

$$E_{\Gamma}(t_f) \leq E_{\Gamma}(0). \tag{22}$$

Since this equation holds for all $t_f > 0$, we have shown that $E_{\Gamma}(t)$ monotonically decreases with t, and Eq. (15) is proven.

Taken together, the statements in Eqs. (13), (14), and (15) imply

$$\langle \psi(t=0) \mid H_{\text{prob}} \mid \psi(t=0) \rangle \geq$$
$$\langle \psi(t=t_f) \mid H_{\text{prob}} \mid \psi(t=t_f) \rangle \tag{23}$$

for final time t_f. In other words, the energy expectation with respect to the problem Hamiltonian can only decrease compared with the initial state. If the energy expectation of the problem Hamiltonian decreases, then the probability of measuring low-energy states increases.

The result in Eq. (23) holds for quenches, parameterized with the single control function $\Gamma(t)$, in the form of Eq. (2). However, since the control function $\Gamma(t)$ is identified with the ratio $A(t)/B(t)$ of control functions for quenches in the form of Eq. (1), the result in Eq. (23) follows automatically for quenches in $A(t)$, $B(t)$ form, except for when $B(0) = 0$, when $\Gamma(0)$ is not well defined. In Appendix A 2, we extend quenches to the case where $B(0) = 0$, with the additional condition that the driver Hamiltonian H_{driver} has a finite gap between its ground-state and first-excited-state manifolds (which is automatically true for all Hamiltonians on Hilbert spaces of finite dimension).

The key result is that for quenches where the control function $\Gamma(t)$ decreases monotonically, the energy expectation value of the problem Hamiltonian H_{prob} cannot be higher than its initial value. Put another way, on average, a monotonic quench can never perform worse than random guessing. This result is important for two reasons. Firstly,

although not being harmful to average solution quality is a rather weak statement, it applies very generally to a broad class of algorithms. Secondly, and more importantly, this result can be built upon to determine control functions that can provide a significant improvement, which is important for algorithm design. To do this, we need to combine the result in this section with criteria for when the transfer of amplitude between computational basis states will be significant, which we obtain in the next section.

IV. ENSURING SIGNIFICANT DYNAMICS

In Sec. II, we showed examples of a quantum quench giving significantly better performance than pure quantum walks. In this section, we consider theoretically how a significant improvement can occur. We know from Sec. III that dynamics will never be detrimental; this means that if dynamics occur, in general they will be beneficial. What remains is to determine the circumstances in which significant dynamics will occur.

A. Quantifying the strength of short-time dynamics

In the analytical solutions for unstructured search in a continuous-time setting [53,56], the method involves analyzing the dynamics in a two-dimensional subspace. To obtain significant dynamics in this setting, the hopping rate γ or schedule functions $A(t)$, $B(t)$ must carefully balance the relative strengths of the driver and problem Hamiltonians such that the off-diagonal terms in the two-dimensional subspace are maximized. Motivated by this, but being interested in shorter timescales, we instead investigate local subspaces spanned by a pair of basis states. To analyze whether significant dynamics will occur, we perform a similar analysis to characterize how strong the transitions are to locally redistribute amplitude. If these are large for most of the transitions mediated by the driver, then the system will generate a high level of dynamics on a short timescale; otherwise, it will not, although dynamics may still occur on longer timescales.

As we want a measure of dynamics that can be efficiently estimated at all sizes, we analyze individual pairs of computational basis states connected by the driver to determine whether significant transfer occurs between them, assuming the rest of the system remains in its initial state. For classical problems in the setting we are considering, the problem Hamiltonian is diagonal in the computational basis and hence all of its subspaces are too. Consider two basis states $|j\rangle$ and $|k\rangle$ connected by the driver (i.e., $\langle j | H_{\text{driver}} | k \rangle \neq 0$) and define an effective two-level system Hamiltonian

$$H_{\Gamma}^{(jk)}(t) = \Gamma(t) H_{\text{driver}}^{(jk)} + H_{\text{prob}}^{(jk)} \tag{24}$$

with the local problem Hamiltonian $H_{\text{prob}}^{(jk)}$ defined as

$$H_{\text{prob}}^{(jk)} = \begin{pmatrix} E^{(j)} & 0 \\ 0 & E^{(k)} \end{pmatrix}, \qquad (25)$$

where $E^{(j)} = \langle j | H_{\text{prob}} | j \rangle$ is the energy of computational basis state $|j\rangle$ with respect to the problem Hamiltonian (similarly for k) and with the local driver Hamiltonian $H_{\text{driver}}^{(jk)}$ defined as

$$H_{\text{driver}}^{(jk)} = \begin{pmatrix} \langle j | H_{\text{driver}} | j \rangle & \langle j | H_{\text{driver}} | k \rangle \\ \langle k | H_{\text{driver}} | j \rangle & \langle k | H_{\text{driver}} | k \rangle \end{pmatrix}. \qquad (26)$$

The extent to which the local subspace Hamiltonian $H_\Gamma^{(jk)}(t)$ can transfer amplitude between the basis states $|j\rangle$ and $|k\rangle$ can be characterized by comparing the off-diagonal energy scale with the diagonal one. We define a local transfer coefficient, which takes values $0 \leq T^{(jk)} \leq 1$, as

$$T^{(jk)} = \mathcal{R}\left(\Gamma(t) H_{\text{driver}}^{(jk)}, H_{\text{prob}}^{(jk)} \right) \qquad (27)$$

$$\equiv \frac{2\Gamma(t) | \langle k | H_{\text{driver}} | j \rangle |}{2\Gamma(t) | \langle k | H_{\text{driver}} | j \rangle | + |\Delta_{jk}|}, \qquad (28)$$

where

$$\Delta_{jk} = \left\{ \Gamma(t) \langle j | H_{\text{driver}} | j \rangle + E^{(j)} \right\} - \left\{ \Gamma(t) \langle k | H_{\text{driver}} | k \rangle + E^{(k)} \right\} \qquad (29)$$

is the difference between the diagonal elements in the diagonal basis of the problem Hamiltonian.

Similarly, as implied by the energy redistribution mechanism described in Sec. III, transfer between driver eigenstates is also important. To capture this, we define a local driver coefficient $D^{(jk)}$ by transforming the local subspace Hamiltonian $H_\Gamma^{(jk)}(t)$ into the diagonal basis of the local driver Hamiltonian $H_{\text{driver}}^{(jk)}$ and writing an expression similar to Eq. (28). That is,

$$D^{(jk)} = \mathcal{R}\Big[U^{(jk)\dagger} H_{\text{prob}}^{(jk)} U^{(jk)},$$

$$\Gamma(t) U^{(jk)\dagger} H_{\text{driver}}^{(jk)} U^{(jk)} \Big], \qquad (30)$$

where $U^{(jk)}$ is a unitary such that $U^{(jk)\dagger} H_{\text{driver}}^{(jk)} U^{(jk)}$ is diagonal.

It is easily shown that for unbiased drivers such as Eq. (3), the local driver coefficient $D^{(jk)}$ and the local transfer coefficient $T^{(jk)}$ are related by $D^{(jk)} = 1 - T^{(jk)}$. This makes it clear there is a trade-off between the two quantities to obtain significant dynamics under the combined

Hamiltonian. We quantify the overall level of amplitude transfer we expect by the product of the transfer and driver coefficients $T^{(jk)}$ and $D^{(jk)}$, which we call the "dynamic coefficient":

$$\chi^{(jk)} = T^{(jk)} D^{(jk)}. \qquad (31)$$

For unbiased drivers, since $D^{(jk)} = 1 - T^{(jk)}$ and $0 \leq D^{(jk)}, T^{(jk)} \leq 1$, it follows that $\chi^{(jk)}$ satisfies $0 \leq \chi^{(jk)} \leq 0.25$.

The dynamic coefficient $\chi^{(jk)}$ captures the level of algorithmically useful local dynamics experienced by the system. In particular, if $\Gamma \gg 1$, then the driver Hamiltonian dominates and the problem Hamiltonian H_{prob} will have little effect on the dynamics of the system. Since the initial state is the ground state of the driver Hamiltonian, the dynamics are driven by the much smaller problem Hamiltonian on short timescales. This limit is captured by the dynamic coefficient, as $D^{(jk)} \approx 0$, and hence $\chi^{(jk)} \approx 0$. In the opposite extreme, if $\Gamma \ll 1$, then the problem Hamiltonian dominates, but since it is diagonal, the dynamics will consist almost entirely of phase rotations in the computational basis, and the amplitudes will change very little. This limit is captured by the transfer coefficient, as $T^{(jk)} \approx 0$, and hence $\chi^{(jk)} \approx 0$.

To characterize the level of dynamics in the entire system, we can simply take a mean value of $\chi^{(jk)}$ over the values of j and k that correspond to a nonzero off-diagonal element in H_{driver}. That is, we define the average dynamic coefficient

$$\bar{\chi} = \langle \chi^{(jk)} \rangle_{jk}, \qquad (32)$$

where $\langle \cdot \rangle_{jk}$ represents the mean over all pairs of computational basis states j, k connected by the driver Hamilton H_{driver}. Although Eq. (32) cannot be exactly calculated efficiently, it should, in general, be possible to approximate it efficiently (up to additive error) by sampling. This follows from the fact that the values of $\chi^{(jk)}$ are bounded $0 \leq \chi^{(jk)} \leq 0.25$, and therefore the error $\delta\bar{\chi}$ can be reduced to the range this value can take, multiplied by the statistical noise in the sample, which scales as the square root of the number of samples; that is,

$$\delta\bar{\chi} \sim \frac{0.25}{N_{\text{samples}}^{1/2}}. \qquad (33)$$

As the average dynamic coefficient $\bar{\chi}$ is calculated by considering only those states that are directly connected by the driver Hamiltonian H_{driver}, it naturally captures only the the fastest quantum dynamics that are present in the system. For example, in the case of the transverse field driver from Eq. (3), $\bar{\chi}$ depends only on transitions between states that differ by a single bit flip, which will typically be happening much faster than those that involve two or more bit flips.

The minimum gap between the ground state and the first excited state of the total Hamiltonian is often used in the adiabatic limit of quantum annealing as an indication of the computational difficulty of different parts of the annealing. Although inspired by the analytical solution to the search problem, where balancing the driver and problem Hamiltonians corresponds to this minimum gap, we have no reason to expect the local Hamiltonians balanced by maximizing $\bar{\chi}$ to also locate the global minimum gap, except in special cases. Figure 8 shows a comparison, for two different $n = 9$ SK instances, between the average dynamic coefficient $\bar{\chi}$ (solid red line, right axis) and the gap between the ground state and first excited state (solid blue line, left axis). The quantities are plotted against the schedule parameter $s(t)$ in the AQC-like paramaterization $A(t) = 1 - s(t), B(t) = s(t)$. The maximum of $\bar{\chi}$ and the
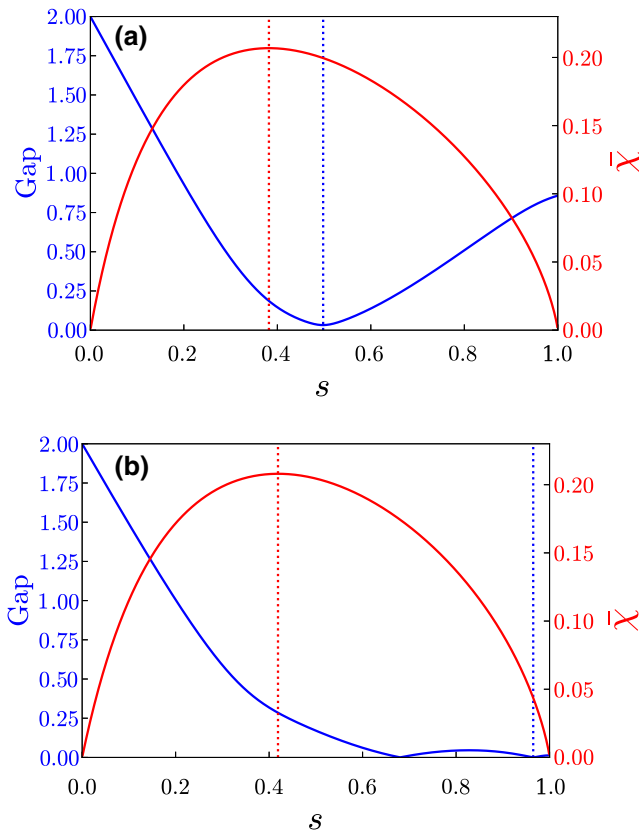

FIG. 8.   Comparison, for two different $n = 9$ SK instances, between the average dynamic coefficient $\bar{\chi}$ (solid red line, right axis) and the gap between the ground state and the first excited state (solid blue line, left axis). The quantities are plotted against the schedule parameter s in the AQC-like paramaterization $A(t) = 1 - s(t), B(t) = s(t)$. The maximum of $\bar{\chi}$ and the minimum gap are indicated by the dotted red and blue lines, respectively. (a) The instance with ID code ovcjhwb-htcpcvwicoxpdpvjzqojril from Ref. [72], used in Figs. 3, 4, and 6. (b) The $n = 9$ instance from Ref. [72] (ID code cpahzp-paxangdnisyqutdbbjlkqamc) with the smallest minimum gap.

minimum gap are indicated by the dotted red and blue lines, respectively. As in these examples, it is typical in SK spin glasses for the maximum $\bar{\chi}$ value to appear significantly before the minimum gap (i.e., closer to the driver end of the schedule). This could be related to the fact that the smallest gaps occur in a spin-glass phase in which dynamics are expected to be much slower, as described in Ref. [82] and discussed in relation to the SK problem in Ref. [55]. Transitioning slowly through the minimum gap is important for the long timescales of adiabatic quantum computing, but it is not necessarily related to what is needed for maximizing the success probability for shorter run times. Away from the adiabatic limit, there are different mechanisms at play, as highlighted by Wong and Meyer [83] and discussed elsewhere [54,55].

B. Analytical bounds on $\bar{\chi}$

Equipped with the definition of the average dynamic coefficient $\bar{\chi}$, we can investigate when it is possible to find a value of $\Gamma(t)$ such that $\bar{\chi}$ is large enough for significant short-time dynamics to be generated. For simplicity, we restrict ourselves to the unbiased driver case, when the local driver coefficient $D^{(jk)}$ and local transfer coefficient $T^{(jk)}$ are related by $D^{(jk)} = 1 - T^{(jk)}$. In this case, the local dynamic coefficient $\chi^{(jk)}$ can be written in terms of the driver strength $\Gamma(t)$ and a single scaled gap parameter $\zeta_{jk} = \frac{|\Delta_{jk}|}{2|\langle k|H_{\text{driver}}|j\rangle|}$ as

$$\chi^{(jk)} = \frac{\zeta_{jk}/\Gamma(t)}{[1 + \zeta_{jk}/\Gamma(t)]^2}. \tag{34}$$

If we write p_ζ for the probability density function that governs the distribution of ζ_{jk} in the particular problem and driver Hamiltonians under consideration, then it can be shown that the maximum value attained by the average dynamic coefficient $\bar{\chi}$ for any choice of driver strength $\Gamma(t)$ has a lower bound that can be stated formally as

$$\max_{\Gamma}(t)(\bar{\chi})$$
$$\geq \max_{0<c<1}\left[\frac{1-c}{(2-c)^2}\left(1 - \frac{1}{c^2}\frac{\mu_2(p_\zeta)}{\mu_1^2(p_\zeta)}\right)\right], \tag{35}$$

where $\mu_1(p_\zeta)$ $(\mu_2(p_\zeta))$ is the first (second) central moment of the distribution governed by the probability density function p_ζ. This bound is obtained by choosing the specific driver strength $\Gamma = \mu_1(p_\zeta)$ (i.e., the mean of the rescaled local gaps), which is not necessarily optimal but serves to produce a nontrivial lower bound. We give the proof of the formal lower bound (35) in Appendix B 2.

It is illuminating to look at the shape of this bound, which can be easily computed numerically for any given value of the ratio of moments $\mu_2(p_\zeta)/\mu_1^2(p_\zeta)$. The bound is plotted for the interesting range of the ratio of moments
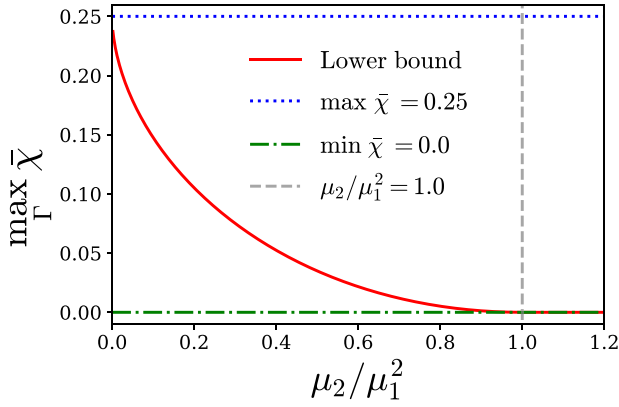
FIG. 9. Semianalytical lower bound (solid red line) on $\bar{\chi}$ as a function of the ratio of moments $\mu_2(p_\zeta)/\mu_1^2(p_\zeta)$ of the distribution, governed by p_ζ, of the rescaled energy gaps ζ_{jk} between computational basis states connected by the driver Hamiltonian H_{driver}. Also shown are minimum (0.0, dot-dashed green line) and maximum (0.25, dotted blue line) possible values of $\bar{\chi}$. The lower bound is nontrivial for $\mu_2(p_\zeta)/\mu_1^2(p_\zeta) < 1.0$ (left of dashed gray line) and is trivially zero otherwise.

in Fig. 9. It can be seen that the lower bound is nontrivial when $\mu_2(p_\zeta)/\mu_1^2(p_\zeta) < 1.0$ but is trivially zero otherwise. This shows that there is a continuous range where $\bar{\chi}$ is bounded away from zero, and hence dynamics will definitely happen on short timescales, even for nonoptimal choices of $\Gamma(t)$. This bound is, in general, far from tight but still allows us to produce some interesting examples. We next illustrate the calculation of $\bar{\chi}$ and the lower bound in Eq. (35) for some specific cases.

C. Example: two-qubit system

As a simple example, consider the problem Hamiltonian

$$H_{\text{prob}}^{(2Q)} = -\hat{Z}_1\hat{Z}_2 - \frac{1}{2}\hat{Z}_1$$

as defined in Eq. (5), with a transverse field driver as defined in Eq. (3). For this problem Hamiltonian, there are four two-level subspaces connected by the driver, $|00\rangle \leftrightarrow |10\rangle$, $|00\rangle \leftrightarrow |01\rangle$, $|10\rangle \leftrightarrow |11\rangle$, and $|01\rangle \leftrightarrow |11\rangle$, with $|\Delta_{jk}| = 3, 2, 2, 1$, respectively We can thus calculate $\bar{\chi}$ exactly:

$$\bar{\chi} = \frac{1}{4}\left(\frac{3/2\Gamma}{(1+3/2\Gamma)^2} + \frac{1/2\Gamma}{(1+1/2\Gamma)^2}\right.$$
$$\left. + 2\frac{2/2\Gamma}{(1+2/2\Gamma)^2}\right)$$

$$= \frac{\Gamma}{2}\left(\frac{3}{(3+2\Gamma)^2} + \frac{1}{(1+2\Gamma)^2}\right.$$
$$\left. + \frac{4}{(2+2\Gamma)^2}\right), \tag{36}$$

where the time dependence in $\Gamma(t)$ is omitted for clarity. To obtain the maximum value of $\bar{\chi}$, we need to maximize the expression in Eq. (36) with respect to Γ. This is done most easily numerically, giving $\max_\Gamma(\bar{\chi}) \approx 0.241$ for $\Gamma \approx 0.941$. From comparison with the bound in Eq. (35), the first moment of p_ζ is $\mu_1(p_\zeta) = 1$, while the second moment is $\mu_2(p_\zeta) = 0.125$. On the basis of the ratio $\frac{\mu_2(p_\zeta)}{\mu_1^2(p_\zeta)} = 0.125$, we obtain the lower bound $\max_\Gamma \bar{\chi} \gtrsim 0.135$. This is just more than half the actual value but holds for any Hamiltonian with the same moments of the distribution.

D. Example: Sherrington-Kirkpatrick spin glass

We consider the Sherrington-Kirkpatrick spin-glass problem Hamiltonian given in Eq. (6). We take the driver Hamiltonian H_{driver} to be the transverse field defined in Eq. (3). Because of the promising results found in Ref. [55] for solving this problem with quantum walks, as well as for the more general quenches presented in Sec. II, we expect intuitively that it should be generally possible to find values of Γ for which the average dynamic coefficient $\bar{\chi}$ takes an appreciable value.

The transverse field driver connects pairs of states j, k that differ only by a single bit flip. Thus, it can be seen from Eq. (6) that for all such pairs, the energy difference can be written as

$$\Delta_{jk} = -\sum_{b\neq a} s_{ab}^{(j)} J_{ab} - 2s_a^{(j)} h_a, \tag{37}$$

where a is the index of the spin that is flipped between states $|j\rangle$ and $|k\rangle$, the sum runs over b, which indexes the other spins, $s_{ab}^{(j)}$ is the eigenvalue (± 1) of the operator $\hat{Z}_a\hat{Z}_b$ on the state $|j\rangle$, and $s_a^{(j)}$ is the eigenvalue (± 1) of the operator $\hat{Z}_a$ on the state $|j\rangle$. The gap Δ_{jk} in Eq. (37) is a sum of normally distributed variables with mean 0, and so Δ_{jk} is itself a normally distributed variable with mean 0 and can be shown to have a standard deviation $\varsigma = \sqrt{2(n+1)}\sigma$, where n is the number of spins (qubits). Then, since $\langle k|H_{\text{driver}}|j\rangle = 1$ for the unbiased transverse field driver, the scaled gap ζ_{jk} is distributed according to the half-normal distribution with probability density function

$$p_\zeta(\zeta) = \frac{1}{\varsigma\sqrt{2\pi}}\exp\left(-\frac{\zeta^2}{8\varsigma^2}\right), \quad \zeta \geq 0. \tag{38}$$

For this distribution, it can be shown that the ratio of moments is

$$\frac{\mu_2(p_\zeta)}{\mu_1^2(p_\zeta)} = \frac{1 - 2/\pi}{\frac{2}{\pi}}$$

$$\approx 0.571, \tag{39}$$

which we emphasize is independent of the width ς of the distribution of the scaled gap ζ_{jk}. For this value of the ratio, the lower bound shown in Fig. 9 is

$$\max_\Gamma \bar{\chi} \gtrsim 0.03. \tag{40}$$

While this value is small compared with the maximum possible value of $\bar{\chi} = 0.25$, which is not unexpected for a hard problem (NP hard), we emphasize that it is independent of the width ς of the distribution of the scaled gap ζ_{jk} and thus does not scale with the system size. Bounding $\bar{\chi}$ away from zero for all sizes proves that dynamics will occur over short timescales for suitable control parameters, thus providing evidence that the scaling found in Ref. [55] may continue to useful problem sizes.

E. Example: unstructured search

As a contrasting example, we consider the problem of unstructured search on n qubits, in which a single computational basis state $|m\rangle$, out of the total $N = 2^n$ basis states, is marked by being given a lower energy. The Hamiltonian for this problem is

$$H_{\text{search}} = \mathbb{1} - 2|m\rangle\langle m|, \tag{41}$$

and again we take the driver Hamiltonian H_{driver} to be the transverse field defined in Eq. (3). While unstructured search is a well-known example with a provable quantum advantage, the algorithms that yield this advantage all involve coherent operations on timescales of order $\sqrt{N} = 2^{n/2}$ rather than the short-time dynamics we are discussing in this paper. Therefore, we would intuitively not expect the lower bound in Eq. (35) to be large in this case.

Of the $n2^{n-1}$ total off-diagonal matrix element pairs in the transverse field driver, only n of these will connect a pair of computational basis states with nonzero energy difference, having energy difference $\Delta_{jk} = 2$, with the remaining $n2^{n-1} - n$ pairs having zero energy difference $\Delta_{jk} = 0$. Therefore, the distribution of scaled gaps ζ_{jk} can be written as

$$p_\zeta(\zeta) = \frac{n}{n2^{n-1}}\delta(\zeta - 1) + \left(1 - \frac{n}{n2^{n-1}}\right)\delta(\zeta). \tag{42}$$

Calculating the first and second central moments of this distribution gives

$$\mu_1(p_\zeta) = \frac{1}{2^{n-1}}, \tag{43}$$

$$\mu_2(p_\zeta) = \frac{1}{2^{n-1}} - \left(\frac{1}{2^{n-1}}\right)^2, \tag{44}$$

and so the relevant ratio of moments is

$$\frac{\mu_2(p_\zeta)}{\mu_1^2(p_\zeta)} = 2^{n-1} - 1. \tag{45}$$

Looking at the plot of the lower bound in Fig. 9, we can see that for unstructured search the bound is trivially zero for all $n > 1$ We can also calculate the exact value using Eq. (34). For each of the $n2^{n-1} - n$ pairs of states j, k with $|\Delta_{jk}| = 0$, $\chi^{(jk)} = 0$ for all Γ. For the remaining n pairs of states j, k with $|\Delta_{jk}| = 1$, the choice of driver strength $\Gamma = 1.0$ will maximize $\chi^{(jk)} = 0.25$. Thus, the average dynamic coefficient for unstructured search is

$$\bar{\chi} = \frac{1}{2^{n-1}} \times 0.25$$

$$= \frac{1}{2^{n+1}}, \tag{46}$$

which tends toward the lower bound of zero in the limit as $n \to \infty$.

This tells us that, for search, most two-level subspaces do not exhibit dynamics and probability enhancement of the marked state can happen only through finely tuned control. For an adiabatic algorithm, this is achieved by slowly adjusting the Hamiltonian within a precise range so that the system can follow a very delicate path, whereas for quantum walk, this is achieved by reaching a finely tuned resonance between the marked state and the rest of a symmetric subspace of the Hilbert space. While interpolations between these two extremes are possible [54], all of the interpolated algorithms also rely on dynamics of a two-level system with a gap proportional to $\sqrt{N} = 2^{n/2}$. In such a system, significant dynamics cannot occur on the timescales of rapid quenches, $O(1)$ or $O(\text{poly}(n))$.

V. USING DYNAMICS TO FIND HEURISTIC QUENCH PARAMETERS

As mentioned in Sec. IV, the average dynamic coefficient $\bar{\chi}$ can, in general, be efficiently estimated by sampling. In this section, we show via two practical examples that this estimate can be used to develop heuristic methods for setting the control function $\Gamma(t)$, or equivalently $A(t)$ and $B(t)$, for a rapid quench in both quantum walk and quantum annealing settings. In both cases, we use the unbiased transverse field driver Hamiltonian defined in Eq. (3). First, we consider the quantum walk algorithm, starting with a simplified example of a two-qubit system. We then develop a heuristic for the Sherrington-Kirkpatrick spin glass and show that it performs almost as well as the numerically fine-tuned heuristic described

in Ref. [55], without needing any fine-tuning. Second, we develop a simple heuristic method for defining a schedule for a time-dependent rapid quench, also applied to the Sherrington-Kirkpatrick spin glass, that outperforms a linear ramp.

In all the examples discussed in this section, we compute the average dynamic coefficient $\bar{\chi}$ numerically using all nonzero j-k pairs rather than estimating it by sampling such pairs. This is computationally easy to do at these problem sizes and allows us to separate the effectiveness of the heuristic from errors due to sampling.

A. Heuristic hopping rate for a quantum walk

For a quantum walk, the average dynamic coefficient $\bar{\chi}$ is a function of the hopping rate chosen $\Gamma(t) = \gamma$. Informed by the result in Sec. III that dynamics will typically be useful, it follows that by maximizing $\bar{\chi}$ we can obtain a heuristic hopping rate γ_χ that should ensure significant dynamics occur over short timescales. For the two-qubit Hamiltonian from Eq. (5), Fig. 10 shows how the average success probability within 100 dimensionless time units P_{100} varies with γ. For this two-qubit system, we can exactly calculate $\bar{\chi}$ (see Sec. IV C), as shown in Fig. 10. The maximum value of $\bar{\chi}$ gives a value for γ_χ that is a good-quality estimate for the value of γ_{opt}. Using bisection and a numerically calculated derivative, we find that $\gamma_\chi \approx 0.864$, while the peak of P_{100} occurs at a slightly lower value of γ. Since the peak of P_{100} is quite broad, the discrepancy between γ_χ and γ_{opt} reduces P_{100} only by a small amount, as can be seen in Fig. 10.

To test how well this heuristic hopping rate works for a more realistic example, we numerically calculate γ_χ for each instance of size $5 \leq n \leq 15$ of the spin-glass problems from Ref. [72]. This is done by our performing a bisection optimi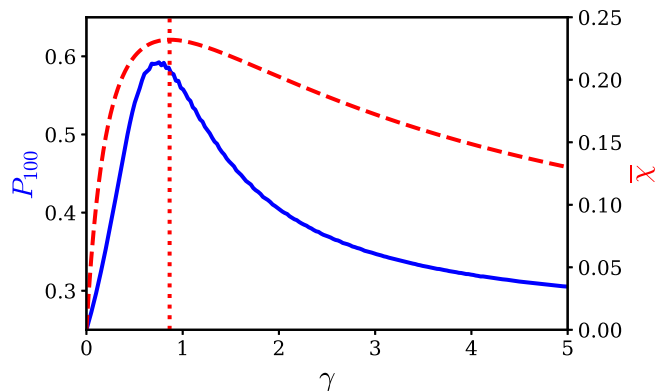zation to maximize the value of $\bar{\chi}$ as a function of γ for each instance. Following the methods in Ref. [55], we perform a short-time quantum walk and calculate the success probability P_{short}, which is time averaged over a short run time. Averaging over all instances of a given size, we obtain the average short-time success probability

$$P_{short} = \int_{12.5/\sqrt{n}}^{17.5/\sqrt{n}} dt P(t), \qquad (47)$$

defined in Ref. [55]), for measuring the problem ground state. This is shown (red line) for each size in Fig. 11. Included for comparison (blue line) are the results from Ref. [55] obtained with the fine-tuned heuristic γ_{heur} defined there, using properties of the eigenvalue distribution for the spin-glass problem Hamiltonian. It can be seen that despite γ_{heur} being numerically fine-tuned specifically for the Sherrington-Kirkpatrick spin-glass problem, it performs only marginally better than the general method we use here. Fitting the data produces $\langle P_{short} \rangle \sim O(N^{(-0.411 \pm 0.002)})$ for γ_{heur} compared with $\langle P_{short} \rangle \sim O(N^{(-0.425 \pm 0.001)})$ for γ_χ. The eigenvalue distribution used in Ref. [55] would not generally be available to calculate γ for real problems; this comparison shows that using $\bar{\chi}$ is a viable method for determining a useful value for γ in this case.

For the small size instances we are using, we use all the values of $\chi^{(jk)}$ to calculate the average in the definition of $\bar{\chi}$ in Eq. (32). We can show that the error in γ_χ due to sampling a subset of $\chi^{(jk)}$ values stays manageable for larger sizes. Consider a small error $\delta\gamma$ in γ. Doing a Taylor

FIG. 10. Average success probability P_{100} between $t = 0$ and $t = 100$ (solid blue line) calculated on the basis of 10 000 independent random points within this range and $\bar{\chi}$ (dashed red line) versus γ for the two-qubit system given in Eq. (5). The vertical dotted line indicates the value of γ_χ.
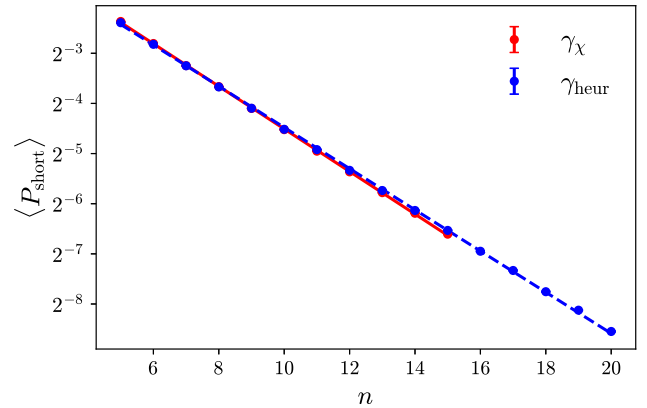

FIG. 11. Log-linear plot of average short-time success probability $\langle P_{short} \rangle$ against number of qubits n for quantum walks on the spin-glass dataset from Ref. [72] using the heuristic hopping rate γ_χ derived for each instance by optimization of the average dynamic coefficient $\bar{\chi}$ (red line). Also shown for comparison is $\langle P_{short} \rangle$ obtained with the fine-tuned heuristic hopping rate γ_{heur} (blue line) described in Ref. [55].

expansion of $\bar{\chi}(\gamma)$ around its peak value $\bar{\chi}_{max}$ gives

$$\delta\bar{\chi} = \bar{\chi}_{max} - \bar{\chi}(\gamma_\chi + \delta\gamma)$$

$$= -(\delta\gamma)^2 \left.\frac{\partial^2 \bar{\chi}(\gamma)}{\partial\gamma^2}\right|_{\gamma=\gamma_\chi} + O((\delta\gamma)^3), \quad (48)$$

where γ_χ is the value of γ our heuristic would find [84] with the exact $\bar{\chi}_{max}$. Using the sampling error in $\bar{\chi}$ from Eq. (33) and rearranging, we obtain

$$\delta\gamma \propto N_{sample}^{-1/4} \left(-\left.\frac{\partial^2 \bar{\chi}(\gamma)}{\partial\gamma^2}\right|_{\gamma=\gamma_\chi}\right)^{-1/2}. \quad (49)$$

This is a general expression that can be used for any problem Hamiltonian. For the Sherrington-Kirkpatrick spin glass, we can use the distribution of the scaled gaps from Eq. (38) and the definition of $\chi^{(j,k)}$ from Eq. (34) to obtain the average value of $\bar{\chi}(\gamma)$ for SK instances, $\langle\bar{\chi}\rangle(\gamma)$:

$$\langle\bar{\chi}\rangle(\gamma)$$

$$= \frac{1}{\varsigma\sqrt{2\pi}} \int_0^\infty d\zeta \exp\left(-\frac{\zeta^2}{8\varsigma^2}\right) \frac{\zeta/\gamma}{(1+\zeta/\gamma)^2}. \quad (50)$$

Making the substitution $z = \zeta/(2\sqrt{2}\varsigma)$ to remove the ς dependence in the exponential and differentiating twice with respect to γ, we obtain

$$\frac{\partial^2}{\partial\gamma^2}\langle\bar{\chi}\rangle(\gamma)$$

$$= \frac{8}{\varsigma^2\sqrt{2\pi}} \int_0^\infty dz\, 2z \exp\left(-z^2\right) \frac{(\gamma/\varsigma) - 4\sqrt{2}z}{[(\gamma/\varsigma) + 2\sqrt{2}z]^4}. \quad (51)$$

This needs to be evaluated at $\gamma = \gamma_\chi$, at the peak of $\langle\bar{\chi}\rangle(\gamma)$, which doing the substitution $z = \zeta/(2\sqrt{2}\varsigma)$ in Eq. (50) shows occurs at a fixed value of γ/ς. Hence, the scaling with n of the double derivative at $\gamma = \gamma_\chi$ is determined solely by the ς^{-2} prefactor in Eq. (51). Recalling from Sec. IV D that $\varsigma = \sqrt{2(n+1)}$ for these SK spin glasses and putting it back into Eq. (49), we have

$$\delta\gamma \propto N_{sample}^{-1/4}(n+1)^{1/2}. \quad (52)$$

The peak in the success probability as a function of γ is very broad for SK spin glasses, and the width of this peak decreases as $1/n$ (determined numerically [55]). Combined with Eq. (52), this means the sampling rate to calculate $\bar{\chi}$ needs to increase by a poly(n) factor as n increases in order to determine γ_χ to sufficient accuracy. Since n corresponds to the number of qubits, this can be done efficiently.

B. Heuristic schedule for quantum annealing

For a time-dependent rapid quench of the form $H_{AB}(t)$ defined in Eq. (1) and total duration t_f, a common choice of control functions, inspired by the adiabatic algorithm, is $A(t) = 1 - s(t)$ and $B(t) = s(t)$, where $s(t)$ is a schedule function with boundary conditions $s(0) = 0$ and $s(t_f) = 1$. In the absence of any knowledge of where along the schedule useful computation can happen, the schedule function is often set to be the linear function $s(t) = t/t_f$. The average dynamic coefficient $\bar{\chi}$ provides a measure of the level of dynamics at each point along the schedule. Intuition gained from Sec. III suggests that the linear schedule can, in general, be improved by spending less time in regions where $\bar{\chi}$ is small and more time in regions where $\bar{\chi}$ is large. A straightforward way to do this is to choose $ds/dt \propto 1/\bar{\chi}$ [the constant of proportionality is set by the boundary conditions $s(0) = 0$ and $s(T) = 1$]. We approximate such a schedule for a typical nine-qubit Sherrington-Kirkpatrick
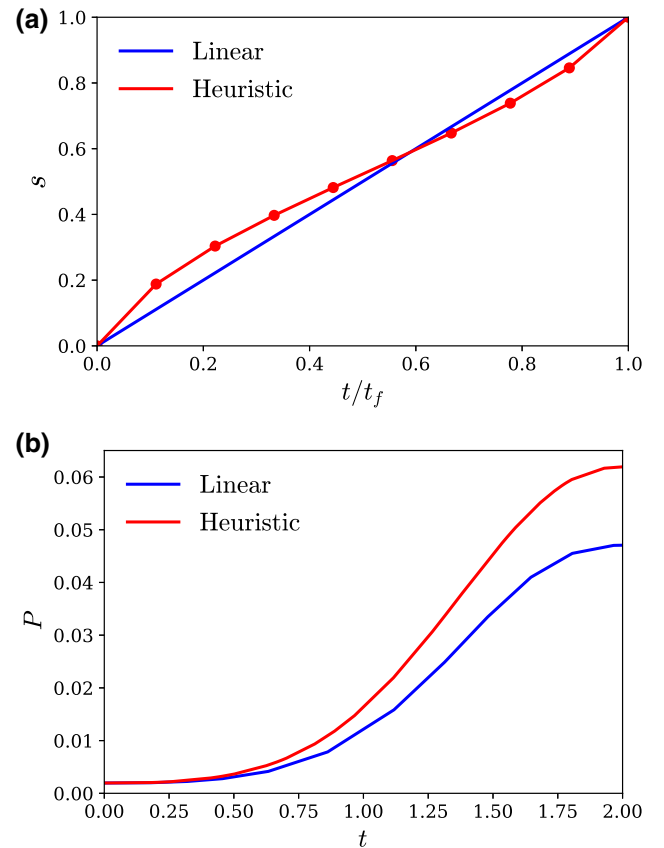
FIG. 12. (a) A heuristic quench schedule of duration $t_f = 2.0$ (red line) derived from the average dynamic coefficient $\bar{\chi}$ for a typical nine-qubit spin-glass instance from Ref. [72]. For comparison, a linear schedule (also of duration $t_f = 2.0$) is also shown. (b) The instantaneous success probability $P(t)$ for measuring the problem ground state for each time t as the quench progresses along the heuristic schedule (red line) and the linear schedule (blue line).

spin-glass instance, as shown in Fig. 12(a) (red line). We do this by fixing the value of the points marked by circles according to $\Delta s \propto \Delta t / \bar{\chi}$, subject to the boundary conditions, and then linearly interpolating between them. A linear schedule $s(t) = t / t_f$ (blue line) is also shown for comparison. Figure 12(b) shows the instantaneous success probability $P(t)$ for measuring the problem ground state as the quench progresses along the heuristic schedule (red line) and the linear schedule (blue line) for a quench duration of $t_f = 2$. It can be seen that the simple heuristic we use here results in a significant increase in success probability at the end of the schedule. We check sufficiently many of the instances to determine that this level of increase is typical for this size of problem and total time duration $t_f = 2$. Further increases may be achievable by varying t_f or choosing a different function of $\bar{\chi}$ for ds/dt.

VI. NUMERICAL METHODS

Numerical simulation and optimization are used extensively throughout this work, as much of the analysis we perform is not analytically tractable. The simulations are performed and the plots are produced with use of the PYTHON language [85], aided extensively by the NumPy [86], SciPy [87], quimb, [88], and Matplotlib [89] libraries. We also use the IPython interpreter [90] and Jupyter notebook system [91]. MATLAB was used for some early numerical experiments, but not for any results that directly appear in this paper.

The numerical optimization used to produce Figs. 9–12, as well as the curve fitting used in Figs. 7 and 11, is performed using the optimization tools in SciPy [87].

The Sherrington-Kirkpatrick spin-glass instances in the data repository at Ref. [72] are used extensively. In any cases where a single example Sherrington-Kirkpatrick spin-glass instance is used, it is the instance ovcjhwbhtcpcvwicoxpdpvjzqojril. The plot of average short-time success probability $\langle P_{\text{short}} \rangle$ against number of spins n in Fig. 11 uses all of the Sherrington-Kirkpatrick spin-glass instances in the repository.

VII. SUMMARY AND FURTHER WORK

In this paper, we generalize and extend work begun in Ref. [55] to time-varying quantum annealing schedules. Callison *et al.* [55] provide numerical evidence for the ability of quantum walks to solve NP-hard problems using many repeats of short runs. This strategy scales better than quantum search by exploiting the correlations in the problem Hamiltonian. The energy conservation mechanism identified in Ref. [55] explains how energy-conserving quantum walks can find lower-energy states with better than guessing probability. In Sec. III, we generalize the energy conservation mechanism to an energy redistribution mechanism that holds for all monotonic quenches that

start in the ground state of the driver Hamiltonian and have non-negative control functions. This thus includes a wide range of quantum annealing protocols used in both theoretical and experimental work. The improvements leveraged by time-varying rapid quenches can be considerable, as we illustrate in Sec. II.

To generate significant energy redistribution, there needs to be significant dynamics driving the system away from the initial state. To characterize the dynamics, in Sec. IV we define the average local dynamic coefficient that balances the contributions from both the driver Hamiltonian and the problem Hamiltonian. This allows the control functions in the Hamiltonian to be optimized for fast dynamics and provides a very general way to estimate good values to use for specific problems. For the spin-glass data [92], we show in Fig. 11 that such estimates are almost as good as the numerically optimized values used in Ref. [55]. We verify in Sec. IV E that our average local dynamic coefficient correctly predicts that the search problem will not have significant dynamics on short timescales. The average dynamic coefficient we define is one way to capture the local dynamics in a quantum annealing Hamiltonian system; doubtless there are other formulations that would serve equally well. In the transverse Ising setting, it focuses on single spin flips, which intuitively are likely to provide the fastest dynamics. Settings with driver Hamiltonians applying multiple spin flips (see, e.g., Refs. [53,56]) may prove less favorable for obtaining fast dynamics, a worthwhile direction for future investigations.

Taken together, the energy redistribution mechanism and the average dynamic coefficient are powerful tools for understanding, designing, and optimally controlling rapid-quench quantum annealing algorithms. We provide a simple example of how to do this to good effect for annealing schedules in Sec. V B, and in Sec. V A verify that it is both efficient and effective for estimating hopping rates for quantum walks on spin glasses. While adiabatic quantum computing and quantum walk search have long had theoretical underpinnings, this represents a significant step in understanding how to exploit quantum annealing schedules run for short times. For current state-of-the-art noisy quantum computers, short run times are a big advantage over the long coherence times required for adiabatic quantum computing, or quantum walk search.

We show that our tools apply to the biased drivers proposed in Refs. [74,75], which provide a method of incorporating prior information into annealing schedules. This can produce significant improvements, as we illustrate in Sec. II B. On the other hand, reverse annealing schedules, both as proposed in Refs. [79,93,94] and discussed in Ref. [77] and as implemented in the latest D-Wave Systems devices [78] are by definition not monotonic, so the tools and mechanisms identified here cannot be applied. Since reverse annealing is a powerful tool, extending our

results to nonmonotonic cases is an important direction for further research.

APPENDIX A: PROOF: MONOTONIC QUENCHES DO NO WORSE THAN GUESSING

1. Energy conservation mechanism

In this appendix, we recap the special case presented in Refs. [55,68] for time-independent controls. Quantum walks can be viewed as a closed-system annealing protocol with a discontinuous schedule [54]. For QW, when formulated in terms of Eq. (1) $A(t)$ and $B(t)$ are constant, independent of time. This picture however does not follow the convention of how annealing protocols are formulated, where the system starts in the ground state of the initial Hamiltonian and the driver is completely absent at the end of the annealing. Following such a convention is important, for instance, to define an interpolation between annealing protocols and QW, as was done in Ref. [54]. To define QW as an annealing protocol in which $A(0) = B(t_f) = 1$ and $A(t_f) = B(0) = 0$, we can write $A(t) = \gamma \Theta(t_f - t + \epsilon)$ and $B(t) = \Theta(t - \epsilon)$, where Θ is the Heaviside Θ function, $\Theta(a > 0) = 1$, $\Theta(a < 0) = 0$, $\Theta(a = 0) = \frac{1}{2}$, and take the limit where $\epsilon \to 0$.

Since the initial state $|\psi(t=0)\rangle$ is a ground state of the driver Hamiltonian H_{driver}, it follows immediately that the expectation value of the driver Hamiltonian is at its lowest at $t = 0$–that is, $\langle \psi(t) | H_{\text{driver}} | \psi(t) \rangle \geq \langle \psi(t = 0) | H_{\text{driver}} | \psi(t = 0) \rangle$–since the expectation value of the driver Hamiltonian H_{driver} for any quantum state cannot be less than that of the ground state.

The total energy expectation as a function of time can be written as

$$E(t) = \langle \psi(t) | \gamma H_{\text{driver}} + H_{\text{prob}} | \psi(t) \rangle$$
$$= \gamma \langle H_{\text{driver}} \rangle_{\psi(t)} + \langle H_{\text{prob}} \rangle_{\psi(t)}, \quad \text{(A1)}$$

where the notation $\langle . \rangle_\psi$ is used to denote the expectation value with respect to the state $|\psi\rangle$. Since energy is conserved for $0 < t < t_f$, it follows that for $\epsilon \to 0$, $E(\epsilon) =$

$E(t_f - \epsilon)$, and therefore

$$\gamma \langle H_{\text{driver}} \rangle_{\psi(t=0)} + \langle H_{\text{prob}} \rangle_{\psi(t=0)}$$
$$= \gamma \langle H_{\text{driver}} \rangle_{\psi(t_f)} + \langle H_{\text{prob}} \rangle_{\psi(t_f)}. \quad \text{(A2)}$$

Rearranging terms, and recalling that $\psi(t = 0)$ is the ground state of H_{driver} and $\gamma \geq 0$, we observe that

$$\langle H_{\text{prob}} \rangle_{\psi(t_f)} - \langle H_{\text{prob}} \rangle_{\psi(t=0)}$$
$$= \gamma [\langle H_{\text{driver}} \rangle_{\psi(t=0)} - \langle H_{\text{driver}} \rangle_{\psi(t_f)}] \leq 0, \quad \text{(A3)}$$

and therefore $\langle H_{\text{prob}} \rangle_{\psi(t_f)} \leq \langle H_{\text{prob}} \rangle_{\psi(t=0)}$. Since $\psi(t = 0)$ is not an eigenstate of the full Hamiltonian, some dynamics are guaranteed to happen, and thus there will be times $t > 0$ when $\langle H_{\text{prob}} \rangle_{\psi(t)}$ is strictly less than $\langle H_{\text{prob}} \rangle_{\psi(t=0)}$.

2. Energy redistribution mechanism in the case of $B(t) \to 0$: divergence of Γ

The result in Sec. III is that the inequality (23) holds for any quench with a Hamiltonian in the form of Eq. (2) that satisfies the three conditions listed in Sec. III. We now consider quenches with a Hamiltonian in the form of Eq. (1). Any Hamiltonian in the form of Eq. (1) with $B(0) > 0$ can be put in the form of Eq. (2) by identifying the ratio $A(t)/B(t)$ with $\Gamma(t)$ and rescaling by a factor of $1/B(t)$, which can be formally compensated for by rescaling time by a factor of $B(t)$. Thus, the inequality (23) holds also for any quench with a Hamiltonian in the form of Eq. (1) with $B(0) > 0$ and that otherwise satisfies the three conditions listed in Sec. III. Here we show that this can be extended to the case where $B(0) = 0$.

In the case that $B(0) = 0$, consider the modified Hamiltonians

$$H'_{\text{driver}} = H_{\text{driver}} - \frac{\epsilon}{A(0)} H_{\text{prob}}, \quad \text{(A4)}$$

$$H'_{\text{prob}} = H_{\text{prob}} \quad \text{(A5)}$$

and the modified control functions

$$A'(t) = A(t), \quad \text{(A6)}$$

$$B'(t) = B(t) + \frac{A(t)}{A(0)} \epsilon$$
$$= B(t) \left(1 + \Gamma(t) \frac{\epsilon}{A(0)} \right), \quad \text{(A7)}$$

where $\epsilon \ll 1$. It can be seen that that total Hamiltonian is unchanged,

$$H'_{A,B}(t) \equiv A'(t) H'_{\text{driver}} + B'(t) H'_{\text{prob}}$$
$$= A(t) H_{\text{driver}} + B(t) H_{\text{prob}}, \quad \text{(A8)}$$

but we have

$$B'(0) = \epsilon. \tag{A9}$$

We define

$$\Gamma'(t) \equiv \frac{A'(t)}{B'(t)}. \tag{A10}$$

$$\Gamma'(t) = \frac{\Gamma(t)}{\left(1 + \Gamma(t)\frac{\epsilon}{A(0)}\right)}. \tag{A11}$$

It can be immediately seen that $\Gamma'(t)$ is non-negative if $\Gamma(t)$ is non-negative, and so condition 2 is satisfied. Furthermore,

$$\frac{d\Gamma'(t)}{d\Gamma(t)} = \frac{1}{\left(1 + \Gamma(t)\frac{\epsilon}{A(0)}\right)^2}. \tag{A12}$$

Thus, $\Gamma'(t)$ is monotonically decreasing if $\Gamma(t)$ is monotonically decreased, and so condition 3 is satisfied.

If we were to start the protocol in the state $|\psi'_{\text{GS}}\rangle$, a ground state of H'_{driver}, condition 1 would be satisfied and the result would be proven. However, the original protocol we are considering starts in the state $|\psi(0)\rangle$, a ground state of H_{driver}. Applying first-order perturbation theory in ϵ to H'_{driver}, we find that H'_{driver} has a ground state

$$|\psi'_{\text{GS}}\rangle = |\psi(0)\rangle + O\left(\frac{\epsilon}{A(0)\Delta}\right)|\psi_\perp\rangle, \tag{A13}$$

where $|\psi_\perp\rangle$ is a normalized state vector orthogonal to $|\psi(0)\rangle$ and Δ is the energy gap between the ground-state and first-excited-state manifolds of the actual driver Hamiltonian H_{driver}. Thus, assuming the driver Hamiltonian H_{driver} is not gapless (which is automatically true for all Hamiltonians on Hilbert spaces of finite dimension), the inequality in Eq. (23) is satisfied in the limit as $\epsilon \to 0$.

APPENDIX B: LOWER BOUND ON THE AVERAGE DYNAMIC COEFFICIENT

1. Bound on probabilities in a range based on second moment

Here we prove a useful bound that is applied in the following subsection. Assume that the distribution $p(x)$ has a finite second moment

$$\mu_2(p) = \int_{-\infty}^{\infty} dx\, p(x)[x - \mu_1(p)]^2, \tag{B1}$$

where

$$\mu_1(p) = \int_{-\infty}^{\infty} dx\, p(x)x \tag{B2}$$

is the first moment (mean). Let us choose some values $x_{\max} > x_{\min}$ such that $\mu_1(p) = \frac{1}{2}(x_{\max} + x_{\min})$. The distribution $q(x) = \frac{1}{2}\delta(x_{\min} - \epsilon) + \frac{1}{2}\delta(x_{\max} + \epsilon)$ has the minimum possible second moment while having no support in the interval $[x_{\min}, x_{\max}]$, where δ is the Dirac δ distribution. In the limit $\epsilon \to 0$, the second moment of this distribution is $\mu_2(q) = \frac{1}{4}(x_{\max} - x_{\min})^2$. Thus, if $\mu_2(p) < \mu_2(q)$, then $p(x)$ must have some support within the range $[x_{\min}, x_{\max}]$. In particular, because the second moment $\mu_2(p)$ can be lower bounded as

$$\begin{aligned}
\mu_2 &= \int_{-\infty}^{\infty} dx\, p(x)[x - \mu_1(p)]^2 \\
&\geq \int_{-\infty}^{x_{\min}} dx\, p(x)[x - \mu_1(p)]^2 \\
&\quad + \int_{x_{\max}}^{\infty} dx\, p(x)[x - \mu_1(p)]^2 \\
&\geq \mu_2(q)\left(\int_{-\infty}^{x_{\min}} dx\, p(x) + \int_{x_{\max}}^{\infty} dx\, p(x)\right) \\
&= \mu_2(q)\left(1 - \int_{x_{\min}}^{x_{\max}} dx\, p(x)\right),
\end{aligned}$$

the probability for x to be in the interval $[x_{\min}, x_{\max}]$ can also be lower bounded as

$$\begin{aligned}
\int_{x_{\min}}^{x_{\max}} dx\, p(x) &\geq 1 - \frac{\mu_2(p)}{\mu_2(q)} \\
&= 1 - \frac{4\mu_2(p)}{(x_{\max} - x_{\min})^2}. \tag{B3}
\end{aligned}$$

2. A simple lower bound

Let $\zeta_{jk} = \frac{|\Delta_{jk}|}{2|\langle k|H_{\text{drive}}|j\rangle|}$ and let $\eta_{jk} = \zeta_{jk}/\Gamma$. Furthermore, let p_ζ and p_η be probability density functions that govern the distribution of the values ζ_{jk} and η_{jk}, respectively, over a set of problem instances. Let $\mu_1(p)$ and $\mu_2(p)$ refer to the first and second moments, respectively, of a distribution governed by the probability density function p.

The dynamic coefficient is

$$\chi^{(jk)} = \frac{\eta_{jk}}{(1 + \eta_{jk})^2}, \tag{B4}$$

so we consider the function

$$f(x) = \frac{x}{(1 + x)^2}, \tag{B5}$$

where $x > 0$.

Let x be distributed according to the probability density function p_η. We know that the expectation value $\langle f(x)\rangle_x$ is

then

$$
\begin{aligned}
\langle f(x)\rangle_x &= \int_0^\infty dx p_\eta(x) f(x) \\
&= \int_{x_{\min}}^{x_{\max}} dx p_\eta(x) f(x) + \int_0^{x_{\min}} dx p_\eta(x) f(x) \\
&\quad + \int_{x_{\max}}^\infty dx p_\eta(x) f(x) \\
&= P_\eta(x_{\min} < x < x_{\max}) \langle f(x)\rangle_{x_{\min}}^{x_{\max}} \\
&\quad + P_\eta(x \geq x_{\min}) \langle f(x)\rangle_0^{x_{\min}} \\
&\quad + P_\eta(x_{\max} \geq x) \langle f(x)\rangle_{x_{\max}}^\infty,
\end{aligned} \tag{B6}
$$

where $x_{\max} > x_{\min}$, $P_\eta(\ldots)$ is the probability of its argument being true if η is distributed according to p_η, and $\langle f(x)\rangle_a^b$ is the expectation value of $f(x)$ if x is distributed according to a (renormalized) version of p_η with all support on $x < a$ and $x > b$ removed. As $f(x)$ is positive for all $x > 0$, we get the lower bound on $\langle f(x)\rangle_x$,

$$
\begin{aligned}
\langle f(x)\rangle_x &\geq P_\eta(x_{\min} < x < x_{\max}) \langle f(x)\rangle_{x_{\min}}^{x_{\max}} \\
&> P_\eta(x_{\min} < x < x_{\max}) \min_{x_{\min} < x < x_{\max}} f(x).
\end{aligned}
$$

Since $f(x)$ is also convex, we know that

$$
\begin{aligned}
&\min_{x_{\min} < x < x_{\max}} f(x) \\
&= \min\left[f(x_{\min}), f(x_{\max}) \right].
\end{aligned} \tag{B7}
$$

Let the interval $[x_{\min}, x_{\max}]$ be of width $2c$ (for some $c > 0$) and centered at the mean $\mu_1(p_\eta)$, thereby also constraining $c < \mu_1(p_\eta)$. That is, $x_{\min} = \mu_1(p_\eta) - c$ and $x_{\max} = \mu_1(p_\eta) + c$, and we must find out which of $f(\mu_1(p_\eta) - c)$ and $f(\mu_1(p_\eta) + c)$ is smaller. To do this, we consider under what conditions it is true that

$$
f(\mu_1(p_\eta) - c) < f(\mu_1(p_\eta) + c). \tag{B8}
$$

It can be shown that Eq. (B8) is true when

$$
c^2 > \mu_1^2(p_\eta) - 1. \tag{B9}
$$

This inequality means that when the mean $\mu_1(p_\eta)$ is greater than 1, the truth of the inequality in Eq. (B8) depends on the value of c, but for $\mu_1(p_\eta) \leq 1$, it is always true. Therefore, if we choose $\Gamma = \mu_1(p_\zeta)$ (the mean of the distribution of ζ rather than η), then we have $\mu_1(p_\eta) = 1$, which means

the inequality $f(1 - c) < f(1 + c)$ is always true (where now $0 < c < 1$), and consequently

$$
\begin{aligned}
&\min_{(1-c)<x<(1+c)} f(x) = f(1 - c), \\
&\langle f(x)\rangle_{1-c}^{1+c} > f(1 - c), \\
&\langle f(x)\rangle_x > P_\eta(1 - c < x < 1 + c) f(1 - c).
\end{aligned} \tag{B10}
$$

Since $\zeta_{jk} \equiv \Gamma \eta_{jk} = \mu_1(p_\zeta) \eta_{jk}$, we have

$$
\begin{aligned}
&P_\eta(1 - c < x < 1 + c) \\
&= P_\zeta(\mu_1(p_\zeta)(1 - c) < x < \mu_1(p_\zeta)(1 + c)), \tag{B11}
\end{aligned}
$$

where $P_\zeta(\ldots)$ is the probability of its argument being true if x is distributed according to p_ζ.

Applying the result in Appendix B 1, we have

$$
\begin{aligned}
&P_\zeta(\mu_1(p_\zeta)(1 - c) < x < \mu_1(p_\zeta)(1 + c)) \\
&\geq 1 - \frac{4\mu_2(p_\zeta)}{\left\{ [\mu_1(p_\zeta)(1 + c)] - [\mu_1(p_\zeta)(1 - c)] \right\}^2} \\
&= 1 - \frac{1}{c^2} \frac{\mu_2(p_\zeta)}{\mu_1^2(p_\zeta)}. \tag{B12}
\end{aligned}
$$

Putting this all together gives

$$
\max_\Gamma \bar\chi \geq f(1 - c) \left(1 - \frac{1}{c^2} \frac{\mu_2(p_\zeta)}{\mu_1^2(p_\zeta)} \right) \tag{B13}
$$

$$
= \frac{1 - c}{(2 - c)^2} \left(1 - \frac{1}{c^2} \frac{\mu_2(p_\zeta)}{\mu_1^2(p_\zeta)} \right). \tag{B14}
$$

While this inequality gives a valid lower bound on $\bar\chi$, the greatest lower bound can be written as

$$
\begin{aligned}
&\max_\Gamma \bar\chi \\
&\geq \max_{0<c<1} \left[\frac{1 - c}{(2 - c)^2} \left(1 - \frac{1}{c^2} \frac{\mu_2(p_\zeta)}{\mu_1^2(p_\zeta)} \right) \right], \tag{B15}
\end{aligned}
$$

which can be found numerically for any given value of $\mu_2(p_\zeta)/\mu_1^2(p_\zeta)$ by optimization over the parameter c. This is plotted in Fig. 9.

[1] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser, Quantum computation by adiabatic evolution, arXiv:quant-ph/0001106 (2000). .

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

[4] Edward Farhi and Sam Gutmann, Quantum computation and decision trees, Phys. Rev. A **58**, 915 (1998).

[5] N. Chancellor, S. Zohren, P. A. Warburton, S. C. Benjamin, and S. Roberts, A direct mapping of max k-sat and high order parity checks to a chimera graph, Sci. Rep. **6**, 37107 (2016).

[6] Vicky Choi, Adiabatic quantum algorithms for the NP-complete Maximum-Weight Independent set, Exact Cover and 3SAT problems, arXiv:quant-ph/1004.2226 (2010).

[7] Vicky Choi, Different adiabatic quantum optimization algorithms for the NP-complete exact cover and 3SAT problems, Quantum Inf. Comput. **11**, 638 (2011).

[8] Nicholas Chancellor, Szilard Szoke, Walter Vinci, Gabriel Aeppli, and Paul A. Warburton, Maximum-entropy inference with a programmable annealer, Sci. Rep. **6**, 22318 (2016).

[9] Michael Marzec, *Portfolio Optimization: Applications in Quantum Computing* (John Wiley & Sons, Inc., Hoboken, NJ, 2016), p. 73.

[10] Román Orús, Samuel Mugel, and Enrique Lizaso, Forecasting financial crashes with quantum computing, Phys. Rev. A **99**, 060301 (2019).

[11] Davide Venturelli and Alexei Kondratyev, Reverse quantum annealing approach to portfolio optimization problems, (2019). ISSN 2524-4914. https://doi.org/10.1007/s42484-019-00001-w.

[12] Joschka Roffe, Stefan Zohren, Dominic Horsman, and Nicholas Chancellor, Decoding quantum error correction with ising model hardware, arxiv:quant-ph/1903.10254 (2019).

[13] Davide Venturelli, Dominic J. J. Marchand, and Galo Rojo, Quantum Annealing Implementation of Job-Shop Scheduling, quant-ph/1506.08479 (2000). https://arxiv.org/abs/quant-ph/1506.08479.

[14] A. Crispin and A. Syrichas, in *2013 IEEE International Conference on Systems, Man, and Cybernetics, October* (2013), p. 3523. doi:10.1109/SMC.2013.601. http://doi.org/10.1109/SMC.2013.601.

[15] Tony T. Tran, Minh N. Do, Eleanor G. Rieffel, Jeremy Frank, Zhihui Wang, Bryan O'Gorman, Davide Venturelli, and J. Christopher Beck, in *SOCS* (2016). https://www.aaai.org/ocs/index.php/SOCS/SOCS16/paper/view/1395.

[16] Alejandro Perdomo-Ortiz, Neil Dickson, Marshall Drew-Brook, Geordie Rose, and Alán Aspuru-Guzik, Finding low-energy conformations of lattice protein models by quantum annealing, Sci. Rep. **2**, 571 (2012).

[17] Tobias Stollenwerk, Elisabeth Lobe, and Martin Jung, in *Quantum Technology and Optimization Problems*, edited by Sebastian Feld and Claudia Linnhoff-Popien (Cham, Springer International Publishing, 2019), p. 99.

[18] T. Stollenwerk, B. O'Gorman, D. Venturelli, S. Mandrà, O. Rodionova, H. Ng, B. Sridhar, E. G. Rieffel, and R. Biswas, Quantum annealing applied to de-conflicting optimal trajectories for air traffic management, IEEE Trans. Intell. Transp. Syst. **21**, 285 (2020).

[19] Daniel O'Malley, An approach to quantum-computational hydrologic inverse analysis, Sci. Rep. **8**, 6919 (2018).

[20] Julia Kempe, Alexei Kitaev, and Oded Regev, The complexity of the local hamiltonian problem, SIAM J. Comput. **35**, 1070 (2006).

[21] Dorit Aharonov, Daniel Gottesman, Sandy Irani, and Julia Kempe, The power of quantum systems on a line, Commun. Math. Phys. **287**, 41 (2009).

[22] M. H. S. Amin, Consistency of the Adiabatic Theorem, Phys. Rev. Lett. **102**, 220401 (2009).

[23] Donny Cheung, Peter Høyer, and Nathan Wiebe, Improved error bounds for the adiabatic approximation, J. Phys. A: Math. Theor. **44**, 415302 (2011).

[24] 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).

[25] Lorenzo Campos Venuti, Tameem Albash, Daniel A. Lidar, and Paolo Zanardi, Adiabaticity in open quantum systems, Phys. Rev. A **93**, 032118 (2016).

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

[27] M. B. Hastings, The power of adiabatic quantum computation with no sign problem, quant-ph/2005.03791 (2020).

[28] Wolfgang Lechner, Philipp Hauke, and Peter Zoller, A quantum annealing architecture with all-to-all connectivity from local interactions, Sci. Adv. **1** (2015).

[29] Andrea Rocchetto, Simon C. Benjamin, and Ying Li, Stabilisers as a design tool for new forms of Lechner-Hauke-Zoller annealer, Sci. Adv. **2** (2016).

[30] Martin Leib, Peter Zoller, and Wolfgang Lechner, A transmon quantum annealer: Decomposing many-body Ising constraints into pair interactions, Quantum Sci. Technol. **1**, 015008 (2016).

[31] Nicholas Chancellor, Stefan Zohren, and Paul A. Warburton, Circuit design for multi-body interactions in superconducting quantum annealing systems with applications to a scalable architecture, npj Quantum Inf. **3**, 21 (2017).

[32] Vicky Choi, Minor-embedding in adiabatic quantum computation: I. The parameter setting problem, Quantum Inf. Proc. **7**, 193 (2008).

[33] Vicky Choi, Minor-embedding in adiabatic quantum computation: II. Minor-universal graph design, Quantum Inf. Proc. **10**, 343 (2011).

[34] Stephen P. Jordan, Edward Farhi, and Peter W. Shor, Error-correcting codes for adiabatic quantum computation, Phys. Rev. A **74**, 052322 (2006).

[35] Mohan Sarovar and G. J. Milburn, Continuous quantum error correction by cooling, Phys. Rev. A **72**, 012306 (2005).

[36] Kevin C. Young, Robin Blume-Kohout, and Daniel A. Lidar, Adiabatic quantum optimization with the wrong Hamiltonian, Phys. Rev. A **88**, 062314 (2013).

[37] Mohan Sarovar and Kevin C. Young, Error suppression and error correction in adiabatic quantum computation: Non-equilibrium dynamics, New J. Phys. **15**, 125032 (2013).

[38] C. Daniel Freeman, Mohan Sarovar, C. M. Herdman, and K. B. Whaley, Stable quantum memories with limited measurement, Phys. Rev. A **98**, 032322 (2018).

[39] J. Atalaya, S. Zhang, M. Y. Niu, A. Babakhani, H. C. H. Chan, J. Epstein, and K. B. Whaley, Continuous quantum error correction for evolution under time-dependent hamiltonians, arXiv:2003.11248 (2020).

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

[41] Adam D. Bookatz, Edward Farhi, and Leo Zhou, Error suppression in hamiltonian-based quantum computation using energy penalties, Phys. Rev. A **92**, 022317 (2015).

[42] Daniel A. Lidar, Towards Fault Tolerant Adiabatic Quantum Computation, Phys. Rev. Lett. **100**, 160506 (2008).

[43] Gerardo A. Paz-Silva, A. T. Rezakhani, Jason M. Dominy, and D. A. Lidar, Zeno Effect for Quantum Computation and Control, Phys. Rev. Lett. **108**, 080501 (2012).

[44] Gregory Quiroz and Daniel A. Lidar, High-fidelity adiabatic quantum computation via dynamical decoupling, Phys. Rev. A **86**, 042333 (2012).

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

[46] Hannes Bernien, Sylvain Schwartz, Alexander Keesling, Harry Levine, Ahmed Omran, Hannes Pichler, Soonwon Choi, Alexander S. Zibrov, Manuel Endres, Markus Greiner, Vladan Vuletić, and Mikhail D. Lukin, Probing many-body dynamics on a 51-atom quantum simulator, Nature **551**, 579 EP (2017).

[47] D-Wave Systems Inc. Probing mid-band and broad-band noise in lower-noise d-wave 2000q fabrication stacks. https://www.dwavesys.com/sites/default/files/14-1034A-A_Probling_noise_in_LN_2000Q_fabrication_stacks.pdf, 2019. Accessed: 2019-11-07.

[48] Kristen L. Pudenz, Tameem Albash, and Daniel A. Lidar, Quantum annealing correction for random ising problems, Phys. Rev. A **91**, 042302 (2015).

[49] Walter Vinci, Tameem Albash, Gerardo Paz-Silva, Itay Hen, and Daniel A. Lidar, Quantum annealing correction with minor embedding, Phys. Rev. A **92**, 042310 (2015).

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

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

[52] Trevor Lanting, Next Generation of QA hardware, 2018. AQC 2018 https://www.youtube.com/watch?v=05ovPNxmfjE.

[53] Andrew M. Childs and Jeffrey Goldstone, Spatial search by quantum walk, Phys. Rev. A **70**, 022314 (2004).

[54] James G. Morley, Nicholas Chancellor, Sougato Bose, and Viv Kendon, Quantum search with hybrid adiabatic–quantum-walk algorithms and realistic noise, Phys. Rev. A **99**, 022339 (2019).

[55] Adam Callison, Nicholas Chancellor, Florian Mintert, and Viv Kendon, Finding spin glass ground states using quantum walks, New J. Phys. **21**, 123022 (2019).

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

[57] Nathan Wiebe and Nathan S. Babcock, Improved error-scaling for adiabatic quantum evolutions, New J. Phys. **14**, 013024 (2012).

[58] Lorenzo Campos Venuti and Daniel A. Lidar, Error reduction in quantum annealing using boundary cancellation: Only the end matters, Phys. Rev. A **98**, 022315 (2018).

[59] M. R. Passos, M. M. Taddei, and R. L. de Matos Filho, Error-run-time trade-off in the adiabatic approximation beyond scaling relations, Ann. Phys. **418**, 168172 (2020).

[60] Elizabeth Crosson, Edward Farhi, Cedric Yen-Yu Lin, Han-Hsuan Lin, and Peter Shor, Different Strategies for Optimization Using the Quantum Adiabatic Algorithm, quant-ph/1401.7320 (2014).

[61] E. J. Crosson and D. A. Lidar, Prospects for quantum enhancement with diabatic quantum annealing, quant-ph/2008.09913 (2020).

[62] Siddharth Muthukrishnan, Tameem Albash, and Daniel A. Lidar, Tunneling and Speedup in Quantum Optimization for Permutation-Symmetric Problems, Phys. Rev. X **6**, 031010 (2016).

[63] Hitoshi Katsuda and Hidetoshi Nishimori, Nonadiabatic quantum annealing for one-dimensional trasverse-field ising model, cond-mat.stat-mech/1303.6045 (2013).

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

[65] Andrew M. Childs, Enrico Deotto, Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Andrew J. Landahl, Quantum search by measurement, Phys. Rev. A **66**, 032314 (2002).

[66] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Daniel Nagaj, How to make the quantum adiabatic algorithm fail, Int. J. Quantum Inf. **6**, 503 (2008).

[67] Shantanav Chakraborty, Leonardo Novo, and Jérémie Roland, Finding a marked node on any graph by continuous-time quantum walk. arXiv:1807.05957, (2018).

[68] Matthew B. Hastings, Duality in quantum quenches and classical approximation algorithms: Pretty good or very bad, Quantum **3**, 201 (2019).

[69] Lucas T. Brady, Christopher L. Baldwin, Aniruddha Bapat, Yaroslav Kharkov, and Alexey V. Gorshkov, Optimal protocols in quantum annealing and qaoa problems, arXiv:2003.08952 (2020).

[70] Zhi-Cheng Yang, Armin Rahmani, Alireza Shabani, Hartmut Neven, and Claudio Chamon, Optimizing Variational Quantum Algorithms Using Pontryagin's Minimum Principle, Phys. Rev. X **7**, 021027 (2017).

[71] David Sherrington and Scott Kirkpatrick, Solvable Model of a Spin-Glass, Phys. Rev. Lett. **35**, 1792 (1975).

[72] N. Chancellor, A. Callison, V. Kendon, and F. Mintert, Finding spin-glass ground states using quantum walks [dataset], 2019. https://doi.org/10.15128/r21544bp097. Data archive at Durham University, UK, doi:10.15128/r21544bp09710.15128/r21544bp097.

[73] We chose instance ovcjhwbhtcpcvwicoxpdpvjzqojril and used it throughout this paper for all single SK problem examples.

[74] Qian-Heng Duan, Shuo Zhang, Wei Wu, and Ping-Xing Chen, An alternative approach to construct the initial

hamiltonian of the adiabatic quantum computation, Chin. Phys. Lett. **30**, 010302 (2013).

[75] Tobias Graß, Quantum Annealing with Longitudinal Bias Fields, Phys. Rev. Lett. **123**, 120501 (2019).

[76] Laurentiu Nita, Jie Chen, Matthew Walsh, Adam Callison, and Nicholas Chancellor, Hybrid quantum/classical computation with biased quantum walks (2021, to be published).

[77] Nicholas Chancellor, Modernizing quantum annealing using local searches, New J. Phys. **19**, 023024 (2017).

[78] D-Wave Systems Inc. Reverse quantum annealing for local refinement of solutions. https://www.dwavesys.com/sites/default/files/14-1018A-A_Reverse_Quantum_Annealing_for_Local_Refinement_of_Solutions.pdf, 2019. Accessed: 2019-05-03.

[79] Alejandro Perdomo-Ortiz, Salvador E. Venegas-Andraca, and Alán Aspuru-Guzik, A study of heuristic guesses for adiabatic quantum computation, Quantum Inf. Process. **10**, 33 (2011).

[80] H. F. Trotter, On the product of semi-groups of operators, Proc. Am. Math. Soc. **10**, 545 (1959).

[81] Masuo Suzuki, Improved Trotter-like formula, Phys. Lett. A **180**, 232 (1993).

[82] Sergey Knysh, Zero-temperature quantum annealing bottlenecks in the spin-glass phase, Nat. Commun. **7**, 12370 EP (2016).

[83] Thomas G. Wong and David A. Meyer, Irreconcilable difference between quantum walks and adiabatic quantum computing, Phys. Rev. A **93**, 062313 (2016).

[84] The first order term is absent, and the second order term is negative, because we are sampling at the maximum, and we implicitly assume that there are no other peaks in $\bar{\chi}$ that take similar values. If $\partial^2 \bar{\chi}(\gamma)/\partial\gamma^2\big|_{\gamma=\gamma_\chi}$ vanishes, then the next nonzero even derivative should be used.

[85] Guido Van Rossum and Fred L. Drake, *Python Language Reference Manual* (Network Theory United Kingdom, 2003).

[86] Travis E. Oliphant, *A Guide to NumPy* (Trelgol Publishing USA, 2006), Vol. 1.

[87] Eric Jones, Travis Oliphant, and Pearu Peterson *et al.*, SciPy: Open source scientific tools for Python, 2001. http://www.scipy.org/. [Online; accessed January 22, 2021].

[88] Johnnie Gray, Quimb: A python package for quantum information and many-body calculations, J. Open Source Software **3**, 819 (2018).

[89] John D. Hunter, Matplotlib: A 2D graphics environment, Comput. Sci. Eng. **9**, 90 (2007).

[90] Fernando Pérez and Brian E. Granger, IPython: A system for interactive scientific computing, Comput. Sci. Eng. **9**, 21 (2007).

[91] Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica Hamrick, Jason Grout, Sylvain Corlay, Paul Ivanov, Damián Avila, Safia Abdalla, and Carol Willing, in *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, edited by F. Loizides and B. Schmidt (IOS Press, 2016), p. 87.

[92] Nicholas Chancellor, Fluctuation guided search in quantum annealing, arXiv:2009.06335 (2020). To appear in Physical Review A.

[93] Masaki Ohkuwa, Hidetoshi Nishimori, and Daniel A. Lidar, Reverse annealing for the fully connected p-spin model, Phys. Rev. A **98**, 022314 (2018).

[94] Yu Yamashiro, Masaki Ohkuwa, Hidetoshi Nishimori, and Daniel A. Lidar, Dynamics of reverse annealing for the fully connected p-spin model, Phys. Rev. A **100**, 052321 (2019).