SAND2021-3375C

Behavior of Analog Quantum Algorithms

Lucas T. Brady,[1,2,*] Lucas Kocia,[3] Przemek Bienias,[1,2] Aniruddha Bapat,[1,2] Yaroslav Kharkov,[1,2] and Alexey V. Gorshkov[1,2]

[1] *Joint Center for Quantum Information and Computer Science, NIST/University of Maryland, College Park, Maryland 20742, USA*
[2] *Joint Quantum Institute, NIST/University of Maryland, College Park, Maryland 20742, USA*
[3] *Sandia National Laboratories, Livermore, California 94550, USA*

(Dated: March 5, 2021)

Analog quantum algorithms are formulated in terms of Hamiltonians rather than unitary gates and include quantum adiabatic computing, quantum annealing, and the Quantum Approximate Optimization Algorithm (QAOA). These algorithms are promising candidates for near-term quantum applications, but they often require fine tuning via the annealing schedule or variational parameters. In this work we connect all these algorithms to the optimal analog procedure. Notably, we explore how the optimal procedure approaches a smooth adiabatic procedure but with a superposed oscillatory pattern that can be explained in terms of the interactions between the ground state and first excited state. Furthermore, we provide numeric and analytic evidence that QAOA emulates this optimal procedure with the length of each QAOA layer equal to the period of the oscillatory pattern. Furthermore, the ratios of the QAOA bangs is determined by the smooth, non-oscillatory part of the optimal procedure. We provide arguments for these phenomena in terms of the product formula expansion of the optimal procedure. All of this shows that these analog algorithms are all different limits and approximations of the same optimal quantum protocol. In the appendix, we present a new algorithm for better approximating the optimal protocol using the insights from the main body.

I. INTRODUCTION

Analog quantum algorithms come in a variety of forms from Adiabatic Quantum Computing (AQC) [1] and Quantum Annealing (QA) [2] to variational algorithms such as the Quantum Approximate Optimization Algorithm (QAOA) [3]. Analog quantum algorithms are particularly relevant in the NISQ [4] era where they are capable of running effectively on small scale devices [**?**].

All these analog quantum algorithms use the same basic ingredients but combined in different ways that obfuscate the connections between these algorithms. AQC relies on the adiabatic theorem [5] to ensure the the state is transferred from the ground state of an initial Hamiltonian to a final Hamiltonian in a time that scales as an inverse polynomial of the spectral gap. Quantum annealing is a broader algorithm that allows for non-adiabatic effects, and this has recently led to the field of Diabatic Quantum Annealing [6] that explicitly uses excitations to solve problems faster, the problem being how to control and utilize these excitations.

QAOA relies on a different framework as a variational algorithm using a bang-bang structure to solve the problem by optimizing the lengths of Hamiltonian pulses in a quantum-classical hybrid loop. While QAOA was inspired by a Trotterization of AQC and Quantum Annealing, in practice, it performs differently [7, 8]. Results on how QAOA relates to other analog quantum algorithms and how it scales with the number of variational parameters have remained sparse. The numeric results of [7, 8] show that QAOA variational parameters fall along certain curves as the depth of the circuit increases. These curves superficially resemble a Trotterization of an annealing path, but the Trotter steps are large enough to invalidate normal Trotter error arguments. There is also numerical evidence from [7] that these curves, when interpreted as annealing paths, demonstrate properties of diabatic speedups.

More recently techniques from optimal control theory [9] have been applied to analog quantum algorithms [10–14], specifically in the context of the variational approach of QAOA. These optimal control techniques were applied to the more generalized problem of analog quantum algorithms in [14]. There techniques were developed to numerically find the optimal annealing protocol. This optimal protocol takes on a bang-anneal-bang format with guaranteed bangs at the beginning and end that become vanishingly smaller as the allowed time for the protocol increases. In the middle, the protocol often takes on an annealing like form with a smooth control function. This bang-anneal-bang form is designed to be the optimal annealing schedule for a given time, disregarding any concern for adiabaticity.

This paper focuses on analytically proving the connections between all these algorithms. The core focus will be on how QAOA emulates the optimal annealing curve. Furthermore, that optimal annealing curve in the long time and large circuit depth limit is just an optimized adiabatic path, similar to the annealing schedule of [23] that is optimized based off the spectral gap condition that ensures adiabaticity. Therefore, the asymptotic curves discovered for QAOA in [7, 8] are derived from the opti-

*Electronic address: Lucas.Brady@nist.gov

mal adiabatic path of the system. In the short time and low circuit depth limit, the optimal annealing curve and QAOA are still connected and begin to resemble excited state computation seen in in Diabatic Quantum Annealing [6].

The optimal annealing curve needs an annealing region for these results to hold, but [14] showed that annealing regions are common in optimal curves. In practical terms, this means that QAOA can safely be bootstrapped up, using the results of lower circuit depth optimization to produce a good guess for the variational parameters in a higher circuit depth setting. It also means that QAOA parameters can be used to kickstart the processes of finding the optimal annealing schedule, which in general has better performance than QAOA. Furthermore, our results contradict the common logic that annealing paths should be monotonic. Monotonicty is a holdover from the infinite time adiabatic limit [1], and a monotonic schedule will always improve the state over doing nothing [16]. However, our results show that adding an oscillation to the annealing schedule, with a frequency dependent on the spectral gap, improve performance. The amplitudes of these oscillations vanish in the asymptotic limit.

We begin in Section II by defining all the relevant algorithms and providing background material on them. Section III provides the original numerical motivation for this work, representing the QAOA asymptotic curves of [7, 8] and the numeric connection between these curves and the optimal annealing schedules of [14]. In order to prove this connection, our numerics are broken up into two thrusts. The first analytic push in Section IV relates to the optimal annealing curves themselves, showing how the oscillatory behavior arises. This section explores the properties of the initial and final bangs which serve to spread the initial ground state out and then bring it back together with the intermediate annealing region providing a nearly-adiabatic procedure. The oscillations result from properties similar to counter-diabatic driving terms from shortcuts to adiabaticity [17–19]. The second analytic thrust in Section V presents work involving product formula expansions. We show that a product formula that matches up with macroscopic features of a curve, such as an oscillatory pattern, can have a massive reduction in the resulting error. This reduction combined with incoherent effects and additional optimization explains the step size of QAOA and how it relates back to the optimal annealing curves. In **??** we review the implications and caveats of our work, providing possible directions for future study and developement. Finally in Section VI we summarize and conclude.

II. THE ALGORITHMS

All of the analog quantum algorithms considered here fit within a framework described by the Hamiltonian

$$\hat{H}(t) = u(t)\hat{B} + (1 - u(t))\hat{C}. \tag{1}$$

The Hamiltonian, $\hat{B}$, is often described as the "mixer" and encodes quantum mixing (e.g. a uniform transverse field on qubits). $\hat{C}$ is known as the "problem" Hamiltonian and encodes the optimization task that is trying to be solved (e.g. a diagonal Hamiltonian with the target cost function along the diagonal). In all examples, the initial state of the system is taken to be the ground state of $\hat{B}$, and the target state of the system is taken to be the ground state of $\hat{C}$. The control function, $u(t) \in [0, 1]$, is what changes between the various algorithms, with each algorithm providing a different ansatz for how to design this control function. We will use t_f to denote the total runtime of each algorithm.

A. Adiabatic Quantum Computing

Adiabatic Quantum Computing was original proposed to solve combinatorial optimization problems [1]. The function $u(t)$ is taken to be a monotonic function, starting at $u(0) = 1$ and ending at $u(t_f) = 0$. If the change in $u(t)$ is slow enough, the quantum adiabatic theorem [5] guarantees adiabaticity which means that the system will stay in the same relative eigenstate throughout the evolution. Notably, this is usually employed to ensure that a system starting in the ground state of $\hat{B}$ at $t = 0$ will evolve into the ground state of $\hat{C}$ at $t = t_f$.

This necessitates that the Hamiltonian, $\hat{H}(t)$ maintains a non-zero spectral gap throughout, with some exceptions (e.g. ground state subspaces or ground state degeneracy at $t = t_f$). A commonly cited condition to ensure adiabaticity is that

$$t_f \gg \frac{||\frac{d\,\hat{H}(t)}{d\,(t/t_f)}||}{\min_t \Delta(t)^2}, \tag{2}$$

where $\Delta(t)$ is the spectral gap of the Hamiltonian at time t. This naive condition often works in practice, and the formal version, while more complicated [5], depends roughly on the same parameters, potentially with worse exponents.

Therefore, a large part of the analysis of adiabatic quantum algorithms involves spectral theory to determine the size of $\Delta(t)$. In many hard optimization problems, this spectral gap can be exponentially small during what are known as avoided level crossings.

Often the monotonic annealing schedule, $u(t)$, is taken to be a linear ramp (or some other hardware-determined shape), but the ramp can be optimized to slow down when the gap is small and speed up when the gap is large. This optimization is necessary to recover the Grover square-root speed-up for unstructured search [23], and there is good evidence that optimization of the schedule in general can lead to a similar square root speed-up over unoptimized schedules [27]. One generalization of the original unstructured search technique from Roland and Cerf is the quantum adiabatic brachistochrone [24]. It has been noted that such optimized schedules might

be unachievable realistically due to the level of precision needed in the analog setting [28].

B. Quantum Annealing

Quantum Annealing was originally proposed [2] before Adiabatic Quantum Computing and was justified not by the adiabatic theorem, but instead by comparison to classical thermal annealing. In practice, the setup of quantum annealing is roughly the same as Adiabatic with $u(0) = 1$ and $u(t_f) = 0$ and some smooth, usually monotonic, ramp between.

The relative definitions of Quantum Annealing and Adiabatic Quantum Computation are slightly ambiguous and vary throughout the field. In this paper, we will use one of the more common definitions of Quantum Annealing as a generalization of AQC, with AQC being a subclass of Annealing. Whereas AQC requires adiabaticity, meaning the state of the system most always track the ground state, Quantum Annealing allows for non-adiabatic effects as well as adhering to adiabaticity. These non-adiabatic effects might be due to thermal noise or simply going too fast; although, in this paper, we will consider only unitary dynamics. These non-idealities could mean that the final state is an excited state that is deemed good-enough for practical purposes.

It is also possible to utilize the sped up behavior and engineer the dynamics to depopulate the ground state and then repopulate it, utilizing the power of higher excited states for intermediate computational steps. This is the basis of Diabatic Quantum Annealing [6]. While diabatic algorithms show promise, it is currently unclear how to efficiently engineer the desired effects. This paper could be interpreted as addressing this question, and we pointed interested readers to Appendix D where we describe a practical algorithm for potentially engineering a useful diabatic evolution.

C. Quantum Approximate Optimization Algorithm

While QAOA is sometimes described in the digital quantum circuit framework, it is ultimately an analog quantum algorithm. The control function is no longer a smooth function but instead takes on a pulsed bang-bang form where $u(t)$ can only equal 0 or 1, meaning we are only applying either $\hat{B}$ or $\hat{C}$ but not linear combinations of them. The original [3] formulation of QAOA is best described using unitaries and the state of the system $|x(t)\rangle$

$$|x(t_f)\rangle = \left[\prod_{i=1}^{p} e^{-i\beta_i \hat{B}} e^{-i\gamma_i \hat{C}}\right] |x(0)\rangle \,. \tag{3}$$

The angles or times $\vec{\gamma}$ and $\vec{\beta}$ describe how long to apply each bang, and the total runtime for this algorithm is $t_f = \sum_{i=1}^{p} (\gamma_i + \beta_i)$. The number of layers in QAOA, p, is usually fixed with the γ_i and β_i allowed to vary freely to whatever values they want.

QAOA is a variational algorithm with the γ_i and β_i acting as the variational parameters. This hybrid algorithm uses a classical optimizer to optimize these angles, using a quantum computer to sample and estimate the final energy $\langle E\rangle = \langle x(t_f)| \hat{C} |x(t_f)\rangle$. The goal is to find the γ_i and β_i that minimize $\langle E\rangle$ and that therefore prepare a state that is hopefully close to the target state.

As it was originally proposed, QAOA was conceived as a generalized discretization of Quantum Annealing. Indeed, a Suzuki-Trotter expansion of an Annealing or Adiabatic annealing schedule would result in a bang-bang pattern similar to QAOA. Numerics results [7, 8] have shown that the optimal angles do not approach a Trotterization with the ideal bang-lengths remaining roughly constant as p is increased; whereas, a Trotterization would expect these angles to become vanishingly small as $p \to \infty$. A key goal of the current paper is to explain this phenomenon and describe this large p behavior.

D. Optimal Schedules

In a previous study [14], so of the current authors examined the analog quantum algorithm problem through the lens of optimal control theory. We asked what properties an optimal $u(t)$ must have in order to produce the lowest possible $\langle E\rangle$ within a given t_f.

The resulting procedure takes on a bang-anneal-bang form (more accurately termed bang-*-bang) with a finite-length $u = 0$ bang at the beginning and a finite-length $u = 1$ bang at the end. Our analytics suggested multiple possibilities in the middle, but in all numerics tested (mostly focusing on the Ising model with some Heisenberg model numerics), the middle region was dominated by a smooth non-monotonic annealing region. The form of this anneal was not studied extensively, and the exact shape of this region as well as a heuristic picture of the evolution is one of the main contributions of the current paper.

The initial and final bangs in a bang-anneal-bang procedure are guaranteed to decrease as t_f increases and should vanish in the limit $t_f \to \infty$. As we show later in this paper, this actually corresponds to recovering the adiabatic limit, and these initial and final bangs can be thought of as exciting the system into a diabatic annealing regime.

Given the looseness of the definition of Quantum Annealing, these optimal bang-anneal-bang protocols can rightly be classified as the optimal quantum annealing schedules. In that sense, short-time versions of these optimal protocols really are just a form of diabatic quantum annealing.

1. Alternative Formulation

In the original presentation of bang-anneal-bang protocols [14], the protocols were found via gradient descent using the gradient $\Phi(t) = \frac{\delta \langle E \rangle}{\delta u(t)}$. In that original paper, this gradient was formulated in optimal control language, including a fictitious Lagrange multiplier state that required detailed knowledge of the optimal control formulation to calculate.

A more physically meaningful representation of the gradient $\Phi(t)$ is possible in the Heisenberg picture of quantum mechanics. For completeness we present this simplified form of the gradient here:

$$\Phi(t) = i \langle x(0)| \left[\hat{C}(t_f), \left(\hat{C}(t) - \hat{B}(t) \right) \right] |x(0)\rangle \quad (4)$$

This form does not depend on the Lagrange multiplier and only cares about physically meaningful parameters, making it potentially more useful experimentally.

III. NUMERICALLY COMPARING OPTIMAL CURVES AND QAOA

Our main results are inspired by two separate pieces of numeric evidence. The first is the asymptotic large-p structure of QAOA as has already been presented in [7, 8]. The second is asymptotic large-t_f behavior of the optimal bang-anneal-bang curves. This behavior of the bang-anneal-bang curves was explored partially in the Appendices of our previous paper [14], but here we formalize those results and connect them to the behavior of QAOA.

A. QAOA Curves

One of the primary sources of excitement with QAOA is the ability to predict the γ_i and β_i in a given situation from other situations. It has been observed numerically [29] that QAOA angles for a given instance can be used as a good initial guess for the angles for similar instances. More relevant for our purposes, for a single problem instance, the QAOA angles for one value of p, the number of layers, are similar to the values of the angles for another p.

Specifically, suppose that the optimal QAOA angles for a given p are given by γ_i and β_i, then there are continuous functions $\gamma_p(s)$ and $\beta_p(s)$ for $s \in [0, 1]$ such that $\gamma_p\left(\frac{i-1}{p-1}\right) = \gamma_i$ and $\beta_p\left(\frac{i-1}{p-1}\right) = \beta_i$. As p increases, it has been noted numerically [7, 8] that these functions $\gamma_p(s)$ and $\beta_p(s)$ converge to the same asymptotic functions $\gamma(s)$ and $\beta(s)$.

These asymptotic curves should not be confused with a simple Suzuki-Trotter expansion of some underlying annealing curve. A Suzuki-Trotter expansion requires small time steps in order for its errors to be manageable, so if this were interpreted just as a Suzuki-Trotter expansion with no additional bells and whistles, the errors would be large. That said, Zhou *et al.* [7] did look at what would happen if these curves were used as annealing curves. Specifically, they constructed $u(s) = \frac{\beta(s)}{\beta(s)+\gamma(s)}$ as an annealing curve which works in part because it always seems to be the case that $\beta(s)$ is dominant at the beginning and $\gamma(s)$ is dominant at the end (the reason why will be explained in a later section of this paper). The resulting annealing curve captured a well known effect from diabatic quantum annealing, diabatic cascades, providing an empirical link between QAOA and Diabatic Quantum Annealing.

An example of these asymptotic QAOA curves is given in Fig. 1

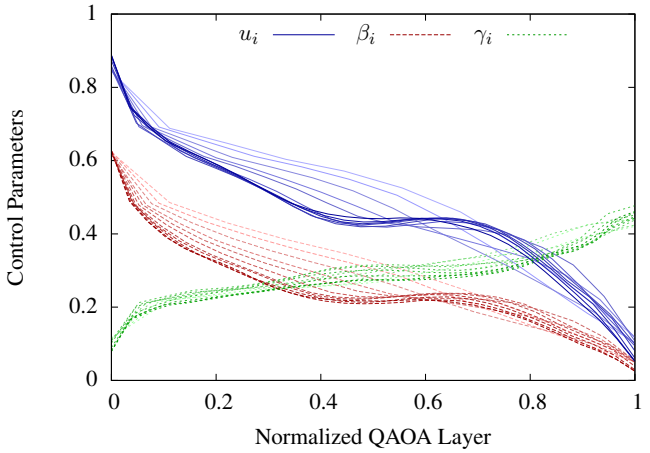


FIG. 1: This plot shows the QAOA variational parameters for a given problem instance at several different values of p. Plotted are the γ_i, β_i, and $u_i = \frac{\beta_i}{\beta_i+\gamma_i}$. The x-axis is the normalized QAOA layer $\frac{i-1}{p-1}$. The lighter curves are for lower p (starting at $p = 10$), and the darker curves are for higher p (ending at $p = 30$). These curves do vary slightly, but especially at higher p, they settle into some smooth asymptotic curve. This data was gathered for $\hat{B}$ being a transverse field and $\hat{C}$ being a randomized Ising model on $n = 12$ qubits.

B. Bang-Anneal-Bang Oscillations

The new numerics that inspired this current study involve the behavior of bang-anneal-bang curves when compared to QAOA. QAOA procedures are not usually thought of in terms of their total runtimes but instead in terms of the number of layers p. However, it is possible to look at the runtime of a QAOA algorithm as $t_f = \sum_{i=1}^{p} (\gamma_i + \beta_i)$. A natural question to ask then is what the optimal bang-anneal-bang curve is for that length of time. The numeric answer is exemplified in Fig. 7.

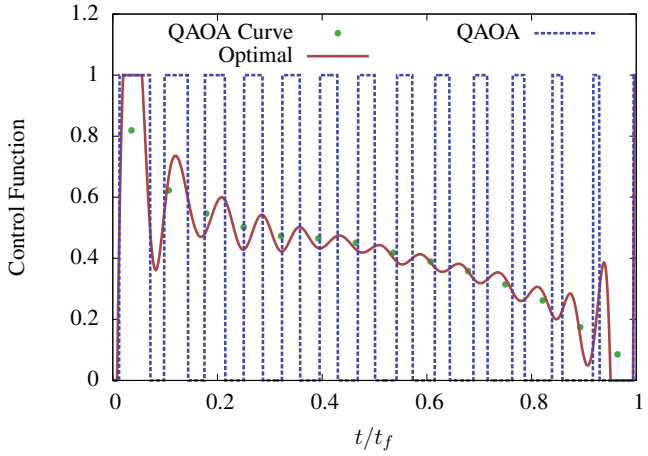


FIG. 2: This plot numerically demonstrates some of the key points of this paper, showing a $p = 14$ QAOA protocol and the optimal bang-anneal-bang protocol that takes the same length of time. Notice first that the optimal bang-anneal-bang protocol oscillates in such a way that it fits 14 oscillations into this time frame. Also, in green, we plot the QAOA curve, defined as $u_i = \frac{\beta_i}{\beta_i+\gamma_i}$ which tracks the underlying annealing portion of the bang-anneal-bang curve. These properties have been seen numerically in every Ising model we have studied.

In Fig. 7, we plot the QAOA curve for a given problem instance (of a randomized Ising model) alongside the optimal bang-anneal-bang curve that takes the same length of time. For ease of optimization, the QAOA instance here used the same length for each layer (with that length also being treated as a variational parameter), but all the qualitative properties apply in the normal QAOA setting as well. Also plotted on this plot is the QAOA curve defined by $u_i = \frac{\beta_i}{\beta_i+\gamma_i}$ with these points located on the x-axis in the middle of that QAOA layer.

There are two key qualitative points to be made here. First the optimal bang-anneal-bang curve oscillates about some base curve. The period of these oscillations matches up with the length of the QAOA layers, with there being $p = 14$ QAOA layers and 14 oscillations of the bang-anneal-bang curve. Second the underlying curve that is being oscillated about matches up with the QAOA curve. These are very general properties and are seen in every numerical instance we looked at.

This behavior suggests a deep connection between QAOA and bang-anneal-bang procedures. Furthermore, the bang-anneal-bang procedure approaches an adiabatic procedure for long time scales with the initial and final bangs becoming vanishingly small. The rest of the paper will be devoted to explaining these connections.

In Section IV, we explain where these oscillations come from, employing an asymptotic near-adiabatic perturbative analysis. In the long t_f limit, the period of oscillations turns out to be inversely proportional to the instantaneous spectral gap. Numerically, the examples we can access are not in this asymptotic regime, but the same analytic mechanism can explain the origin of these oscillations even if the time scales are not long enough for the spectral gap to govern the period timescales.

Then in Section V, we explain the connection between these oscillations and QAOA by interpreting QAOA as a large time-step product formula (a.k.a. Trotterization) of the underlying bang-anneal-bang curve. Due to the large time scales involved, QAOA cannot ordinarily be interpreted as a Product Formula without incurring untenable errors. However, we show first that a product formula aligning with an underlying oscillation incurs less error overall and second that careful coherent effects across multiple layers could account for the remaining error.

IV. DERIVING THE OSCILLATIONS

First we consider how to characterize the optimal annealing curve, specifically the oscillatory pattern in it. These optimal curves were studied extensively in [14] where it was found that these optimal annealing curves provably always start with an initial bang and end with a final bang, with the lengths of those bangs decreasing to zero as the runtime increases. The interior region mostly had a smooth annealing form which was quite obvious for transverse field Ising models but appeared to a greater or lesser degree in other models.

In the large runtime limit, these optimal curves approach a pure annealing region, and the oscillations noted above become smaller and smaller in amplitude. This is consistent with the adiabatic theorem [5] which in the long time regime guarantees that a monotonically decreasing function will do the ground state to ground state transfer we want. Interestingly, there are conjectures that the adiabatic path which transfers all eigenstates in the initial Hamiltonian to the equivalent eigenstates in the final Hamiltonian is the shortest path through the control landscape (in the Lie algebra generated by $\hat{B}$ and $\hat{C}$) that transfers the ground state to ground state [22]. Their conjecture is proven in the adiabatic and near-adiabatic limit but is harder to prove far away from this limit. Also note that their result only talks about transfer to the exact final ground state, meaning that our setup, where we do not have enough resources to enact the full transfer, is slightly different, albeit approaching their case for long times. Therefore, we expect the optimized annealing schedules for long times to approach an optimized adiabatic schedule, similar to what was derived by Roland and Cerf [23] for the unstructured search problem or in the quantum adiabatic brachistochrone [24].

It is possible to emulate an adiabatic protocol in a shorter period of time, using shortcuts to adiabaticity and counter-diabatic (CD) protocols [**?**], and most notably, it is even possible to emulate the effects of a CD addition to the Hamiltonian using only the original Hamiltonians with a fast oscillation of the control function [25]. This method involves setting points in the evolu-

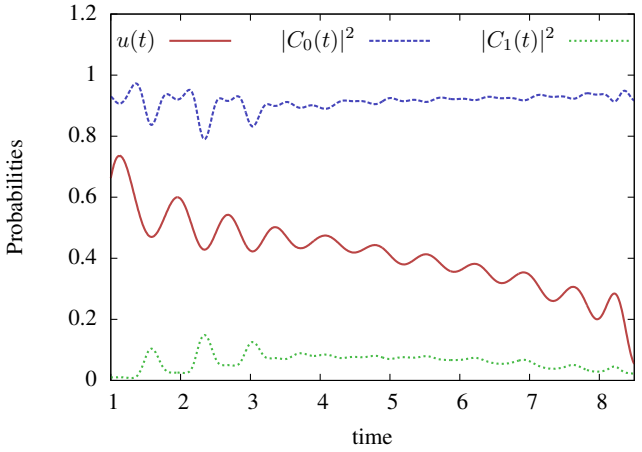


FIG. 3: Here we zoom in on the bang-anneal-bang curve from Fig. 7 during the annealing region. In addition, we plot the probabilities of being in the instantaneous ground state, $|C_0(t)|^2$, and first excited state, $|C_1(t)|^2$. This plot shows that this annealing region is transferring the states adiabatically with the populations roughly maintained from the beginning to end of the anneal. There is variation in these amplitudes but they roughly return to themselves after a full oscillation of the annealing curve.

tion where you want the instantaneous eigen-distribution to match the original eigen-distribution. Then, a Magnus expansion is used on the unitary generated by CD driving and the unitary generated by evolution using the original Hamiltonians with a new control function. By matching the control function up to the CD unitary, the effects of the CD can be emulated to greater precision, matching up specifically at the points in the evolution you chose to match. Their method relies on a user-defined period but results in oscillatory functions with periods that evenly divide that user defined frequency. Their method also relies on full knowledge of the counter-diabatic driving term which we lack and which is difficult to find for large system sizes.

Before we proceed, we need to comment on whether we actually should expect an adiabatic evolution, or at least one that keeps the instantaneous eigen-distribution constant, potentially only at certain points (at least constant between the beginning and end of the annealing region). Numerically, we do see this in the optimized annealing or bang-anneal-bang curves. For long times, as stated previously, the anneal is just an optimized adiabatic schedule with very small oscillation amplitudes. Whereas for shorter times, the oscillations are quite pronounced, and an examination of the eigen-distribution, such as in Fig. 3 reveals that the instantaneous eigen-distribution does indeed transfer over, matching up at the beginning and end of the anneal as well at points roughly in line with the periods of the oscillations.

A. Near Adiabatic Approximation

Therefore, it seems natural that the annealing region of the optimal curves is emulating a fast-forwarded adiabatic protocol. Unfortunately, the Magnus expansion method of [25] uses the frequency of the period of their oscillations as a fit parameter with their method relying on a perturbative approach as this period becomes small.

Therefore, we consider a different approach here to getting such an oscillatory counter-diabatic procedure. To demonstrate this approach, we will first restrict down to the setting where only the ground state and first excited state are relevant. This is true in the near-adiabatic limit, and our numerics in Fig. 3 back this up.

The methods we use here are similar to those used in adiabatic boundary cancellation methods [**?**], and we specifically follow results from [26].

Consider a case where we have some control function given by

$$u(t) = u_0(t) + c(t) \tag{5}$$

so that the Hamiltonian is given by

$$\begin{aligned} \hat{H}(t) &= \hat{H}_0(t) + \hat{H}_c(t) \\ &= \left(u_0(t)\hat{B} + (1-u_0(t))\hat{C}\right) + \left(c(t)(\hat{B}-\hat{C})\right) \end{aligned} \tag{6}$$

The function $u_0(t)$ is a given function determined by the adiabatic nature of the problem, such as an optimized adiabatic schedule coming from a Roland and Cerf [23] style analysis optimized exactly and not just to satisfy the naive adiabatic condition. The function $c(t)$ represents our control freedom, and we can choose it so that the adiabatic passage described by $u_0(t)$ is followed as precisely as possible.

We will furthermore label the instantaneous eigenstates of $\hat{H}_0(t)$ by $|n_0(t)\rangle$, and at the moment we will ignore any degeneracy (or just work in the subspace defined by the symmetries of our Hamiltonian and ground state). Throughout this section (unless otherwise noted), we use the $*_0$ subscript to indicate that these quantities are relative to the eigenframe determined by $u_0(t)$ rather than the full eigenframe determined by $u(t)$.

Now, we can express our current state in terms of these eigenstates by

$$|\psi(t)\rangle = \sum_i C_i(t)\,|i_0(t)\rangle\,. \tag{7}$$

We want to make the assumption that $|C_0(t)|$ and $|C_1(t)|$ are much larger than all other probability amplitudes. This assumption comes in two pieces: first, it assumes that $\dot{u}_0$ is small so that the system is evolving approximately adiabatically. Second, and more importantly for us, it assumes that $c(t)$ is small so that even the small deviations from the base curve are not enough to break our approximation to a two level system. These assumptions are both in place to ensure an effective two-level system.

Let's apply the Schrödinger equation to our state. In that case we get

$$\sum_j i\left(\frac{d\,C_j(t)}{d\,t}\left|j_0(t)\right\rangle + C_j(t)\frac{d}{d\,t}\left|j_0(t)\right\rangle\right) = \sum_j \hat{H}(t)C_j(t)\left|j_0(t)\right\rangle \tag{8}$$

What we want to do is use the orthonormality of the eigenstates $|n_0(t)\rangle$ to pull out components. For instance, the left-hand side of this equation has a component on the eigenstate $|k_0(t)\rangle$ of

$$i\left(\frac{d\,C_k(t)}{d\,t} + \sum_{j\neq k} C_j(t)\left\langle k_0(t)\right|\frac{d}{d\,t}\left|j_0(t)\right\rangle\right) \tag{9}$$

where we have set the phases of the eigenstates by requiring that $\langle k_0(t)|\frac{d}{dt}|k_0(t)\rangle = 0$. Notably in stoquastic Hamiltonians, this choice of phase means that the instantaneous ground state will maintain the same phase throughout the evolution, which we take to be real and positive. Similarly, the stoquastic first excited state can be taken to be real throughout.

It is now time to look at the eigenvalues. Specifically, we will say that $\hat{H}_0(t)\,|k_0(t)\rangle = \lambda_k(t)$ and will choose $\lambda_0(t) = 0$ at all times without loss of generality. Now looking at the time derivative of the eigen-equation

$$\frac{d}{d\,t}(\hat{H}_0(t)\left|j_0(t)\right\rangle) = \frac{d}{d\,t}(\lambda_j(t)\left|j_0(t)\right\rangle). \tag{10}$$

Carrying out the time derivatives yields

$$\frac{d\,\hat{H}_0(t)}{d\,t}\left|j_0(t)\right\rangle + \hat{H}_0(t)\frac{d}{d\,t}\left|j_0(t)\right\rangle = \frac{d\,\lambda_j(t)}{d\,t}\left|j_0(t)\right\rangle + \lambda_j(t)\frac{d}{d\,t}\left|j_0(t)\right\rangle. \tag{11}$$

Now consider the inner product of this equation with another eigenstate $|k_0(t)\rangle$ such that $k \neq j$:

$$\left\langle k_0(t)\right|\frac{d\,\hat{H}_0(t)}{d\,t}\left|j_0(t)\right\rangle) + \left\langle k_0(t)\right|\hat{H}_0(t)\frac{d}{d\,t}\left|j_0(t)\right\rangle = \frac{d\,\lambda_j(t)}{d\,t}\left\langle k_0(t)|j_0(t)\right\rangle + \lambda_j(t)\left\langle k_0(t)\right|\frac{d}{d\,t}\left|j_0(t)\right\rangle. \tag{12}$$

We can act on the bra states with the Hamiltonian and eliminate one element through orthogonality of eigenstates to get

$$\frac{\langle k_0(t)|\frac{d\,\hat{H}_0(t)}{d\,t}\left|j_0(t)\right\rangle)}{(\lambda_j(t)-\lambda_k(t))} = \left\langle k_0(t)\right|\frac{d}{d\,t}\left|j_0(t)\right\rangle. \tag{13}$$

This expression will let us deal with the time derivative of the eigenstates just fine. Furthermore, in Eq. (9) this time derivative of eigenstates is multiplied by the amplitudes $C_j(t)$. By our assumptions, only $C_0(t)$ and $C_1(t)$ will be relevant, and we can discard cases where $j \neq 0, 1$.

Now, we need to look at the right-hand side of Eq. (8) which is significantly harder to deal with. If we were in the actual adiabatic reference frame, the Hamiltonian would just give its eigenvalues, but now looking at the overlap with the $|k_0(t)\rangle$ state, we get

$$\left\langle k_0(t)\right|\hat{H}(t)\sum_j C_j(t)\left|j_0(t)\right\rangle = C_k(t)\lambda_k(t) + \sum_j C_j(t)\left\langle k_0(t)\right|\hat{H}_c(t)\left|j_0(t)\right\rangle \tag{14}$$

Let's define some terminology

$$\begin{aligned}\gamma(t) &\equiv \left\langle 0_0(t)\right|(\hat{B}-\hat{C})\left|1_0(t)\right\rangle)\\ \Delta(t) &\equiv \lambda_1(t)\\ \kappa_i(t) &\equiv \left\langle i_0(t)\right|(\hat{B}-\hat{C})\left|i_0(t)\right\rangle\end{aligned}$$

where we are using the fact that $\lambda_0(t) = 0$. In the stoquastic setting all of these quantities are real, and we will treat them as such going forward. For a non-stoquastic setting, γ and κ_i could be complex.

Then pulling everything together, the Schrödinger equation for the ground state and first excited state amplitudes give

$$i\left(\frac{d\,C_0(t)}{d\,t} + C_1(t)\frac{\gamma(t)\dot{u}_0(t)}{\Delta(t)}\right) = c(t)\left(C_0(t)\kappa_0(t) + C_1(t)\gamma(t)\right) \tag{15}$$

$$i\left(\frac{d\,C_1(t)}{d\,t} - C_0(t)\frac{\gamma(t)\dot{u}_0(t)}{\Delta(t)}\right) = \Delta(t)C_1(t) + c(t)\left(C_0(t)\gamma(t) + C_1(t)\kappa_1(t)\right) \tag{16}$$

With these equations, we can separate out the amplitudes and phases so that

$$C_i(t) = A_i(t)e^{i\varphi_i(t)}. \tag{17}$$

It is simple algebra to separate out the real and imaginary parts of the differential equations which results in equations in terms of $A_0(t)$, $A_1(t)$, and $\varphi(t) \equiv \varphi_0(t) - \varphi_1(t)$. The resulting differential equations are (suppressing all functional dependencies for space purposes)

$$\dot{\varphi} = \Delta + c(\kappa_0 - \kappa_1) + \frac{A_0^2 - A_1^2}{A_0A_1}\left(c\gamma\cos(\varphi) - \frac{\gamma\dot{u}_0}{\Delta}\sin(\varphi)\right) \tag{18}$$

$$\dot{A}_0 = -\left(c\gamma\sin(\varphi) + \frac{\gamma\dot{u}_0}{\Delta}\cos(\varphi)\right)A_1 \tag{19}$$

$$\dot{A}_1 = \left(c\gamma\sin(\varphi) + \frac{\gamma\dot{u}_0}{\Delta}\cos(\varphi)\right)A_0 \tag{20}$$

It should be noted at this point that only two assumptions have been made. First we are assuming that $\dot{u}_0$ and $c(t)$ are small enough to ensure that we are acting in an effective two level system. The second assumption is

that we have stoquasticity which ensures that the γ and κ_i functions are real. If we did not have stoquasticity, these functions could be complex which would have just made the algebra above to separate our amplitudes and phases slightly more complicated without fundamentally changing the results.

The A equations can be integrated to give

$$A_0(t) = a\cos\left(\int_0^t dt'\left(c\gamma\sin(\varphi) + \frac{\gamma\dot{u}_0}{\Delta}\cos(\varphi)\right) + \vartheta\right) \tag{21}$$

$$A_1(t) = a\sin\left(\int_0^t dt'\left(c\gamma\sin(\varphi) + \frac{\gamma\dot{u}_0}{\Delta}\cos(\varphi)\right) + \vartheta\right), \tag{22}$$

where a and ϑ are integration constants that can be set such that $a\cos\vartheta$ is the initial population of the ground state and $a\sin\vartheta$ is the initial population of the first excited state. The signs here don't matter quite so much because any sign could just be absorbed into the φ phases.

It is obvious that maintaining the same populations from beginning to end here translates to the trig argument at time t_f,

$$\Theta_0[u(t)] = \int_0^{t_f} dt\left(c\gamma\sin(\varphi) + \frac{\gamma\dot{u}_0}{\Delta}\cos(\varphi)\right), \tag{23}$$

being close to a multiple of π. In practice a non-zero multiple of π would correspond to swapping the populations back and forth during the anneal which is inconsistent with the assumptions we made about being near-adiabatic with low leakage. Therefore, we want Θ_0 to be as close to zero as possible, meaning that the problem has simplified to finding the $c(t)$ that ensures $\Theta_0 \approx 0$.

In most of the examples shown, the oscillations fit neatly into the time allowed, giving a whole number of oscillations. This is largely because we look at cases where the time for the optimal procedure is the same as the time that QAOA takes. If we look at other times, the oscillations are not regular, destroying the arguments discussed in Sec. V as expected since QAOA should only get the enhancement at discrete time steps. The point here is that we only expect $c(t)$ to have a nice, simple oscillatory structure when t_f neatly divides into periods of the oscillations.

In Appendix A 1, we work in the perturbative limit of $\dot{u}_0$ small what form $c(t)$ can take to ensure that Θ is zero:

$$c(t) = \frac{4\pi\frac{d\,\Delta}{d\,u_0(t)}}{(\Delta(u_0(t)))^3}\dot{u}_0(t)^2\sin(\Delta(u_0(t))t) + \mathcal{O}(\dot{u}_0^3) \tag{24}$$

So long as t_f is a multiple of the period of oscillations $\tau = \frac{2\pi}{\Delta(u_0(t))}$, then these oscillations will ensure that the amplitudes follow the eigenbasis associated with $u_0(t)$

Behind the scenes, this perturbative expression relies on having $c(t)$ be out of phase with the oscillations due to the phase difference $\varphi(t)$. Specifically, we need $c(t)$ and $\cos(\varphi(t))$ to be out of phase to ensure cancellation. Up to the first order in $\dot{u}_0$, the phase difference scales like $\varphi(t) = \Delta(u_0(t))t$, so we wind up with oscillations with period inversely proportional to the spectral gap.

Unfortunately, the numerics shown in this paper for bang-anneal-bang curves do not fall into the perturbative regime described above. In these numerics $\dot{u}_0$ is large enough that $\dot{\varphi}$ is no longer dominated by the spectral gap and begins oscillating at a higher frequency. As stated in the numerics section, determining the bang-anneal-bang curves for larger t_f becomes unfeasible due to the quality of the solution and the difficulty of determining the gradient when so many solutions are good up to numerical precision.

While it is not possible to solve for this frequency perturbatively any more, based on the perturbative analysis in Appendix A 1, as well as numerical simulations of Eqs. (18-20) outside of the perturbative regime of $\dot{u}_0$, the best way to follow adiabaticity along $u_0(t)$ still relies on the phase $\varphi(t)$. In this nonperturbative regime, it becomes easier to deal with the full adiabatic reference frame that follows the eigenstates of the full $u(t)$.

In Appendix A 2, we derive the near-adiabatic differential equations again following the full control function $u(t)$ instead of $u_0(t)$. In this setting, you need to impose boundary conditions that $u(0) = u_a$ at the start of the relevant region and that $u(t_f) = u_b$ at the end of the region. The resulting equations are similar, and the key quantity is still given by an argument that is functionally similar to Θ_0:

$$\Theta[u(t)] \equiv \int_0^{t_f} dt\frac{\gamma(u(t))\dot{u}}{\Delta(u(t))}\cos(\varphi(t)) \tag{25}$$

Unfortunately, this form does not lend itself to a perturbative approach anymore because the small quantity is now $u_a - u_b$. Fortunately, this makes it even more clear how to ensure that this quantity should be close to zero. Namely, by roughly having $\frac{\gamma(u(t))\dot{u}}{\Delta(u(t))} \propto \sin(\varphi(t))$ we can effectively oscillate more quickly and cancel out more of the effects from the overall drift in $u(t)$ needed to transition from u_a to u_b. This intuition coincides perfectly with the numerics in the problem.

To see this in practice, in Fig. 4 we plot $u(t)$, $\cos(\varphi(t))$, and $\frac{\gamma(u(t))\dot{u}}{\Delta(u(t))}$ as determined numerically for an optimal annealing curve resulting from a model using a randomized Ising model. This plot was constructed so that the time allotted for the annealing curve corresponds with the amount of time that $p = 10$ QAOA needed; though, note that the beginning and ending portions of the graph are cutoff to avoid unaesthetic and distracting discontinuities that occur during the initial and final bangs.

As can be seen from Fig. 4, the frequency of the phase oscillations do indeed mesh up with the frequency of the optimal annealing curve. In fact looking at these oscillations are exactly out of phase with cosine of the amplitude phase difference, as expect from our analytics. Note that this is still for relatively low t_f where the annealing curve has relatively large amplitude oscillations,

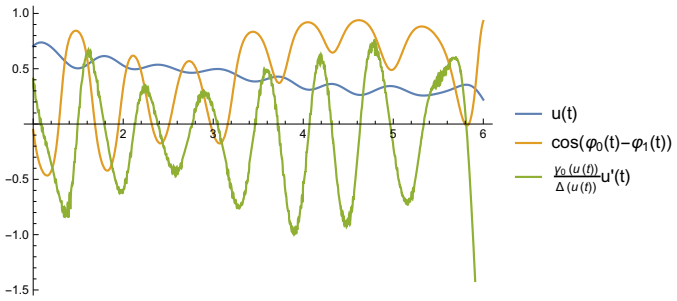


FIG. 4: This plot shows the optimal annealing curve, $u(t)$, as well as two of the quantities that go into Eq. (25). This plot is for a randomized $n = 8$ Ising model, and the initial and final bits of time have been cut off to focus on just the annealing region of interest. Notably from this curve, we can see that the phase difference between the ground and first excited state matches up exactly with the oscillatory pattern of $u(t)$ and is out of phase with $\dot{u}(t)$ as we expect from the analytic arguments surround Eq. 25. The jaggedness in this plot is due to numerical differentiation.

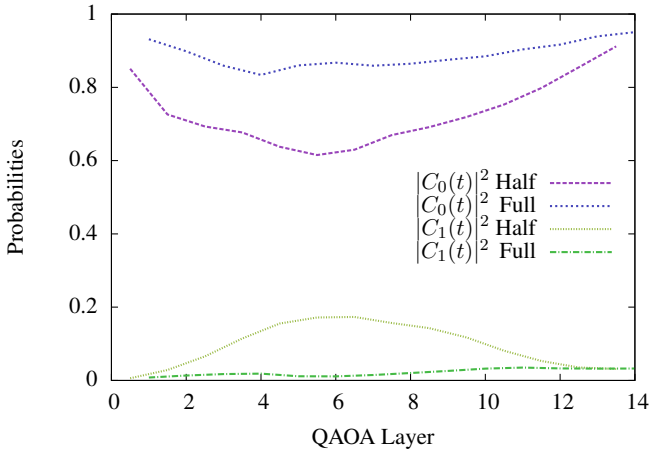


FIG. 5: The instantaneous ground state and first excited state probabilities are plotted versus the QAOA layer. The instantaneous eigenbasis is defined based of $u_i = \frac{\beta_i}{\beta_i+\gamma_i}$ at the end of the ith layer. "Half" means that this is measured after half the QAOA layer (after just the $\hat{C}$ bang), and "Full" means the probabilities are measured after the full QAOA layer (both the $\hat{C}$ and the $\hat{B}$ bangs). QAOA seems to roughly follow an adiabatic-like procedure with the ground state population mostly being preserved. The problem instance displayed here is the same as in Figs. 7 & 3

meaning the the resulting oscillations are not exactly sinusoidal in shape and the period does not mesh up with the asymptotic expectation of the spectral gap.

It is possible to get some data for larger t_f, but we run into numerical limitations. These procedures are already getting fantastically close to the ground state energy, and with higher t_f, the amount of benefit from the optimal annealing procedure versus a rather good procedure can often be less than the numerical inaccuracies in our integration techniques over such long time periods. What limited results we can glean from higher t_f agree with the results presented here and indicate decreasing amplitudes for the oscillations in $c(t)$ which is consistent with the perturbative limit as well.

V. PRODUCT FORMULA ERROR

Based on the numerics shown in Section III, it seems that QAOA is emulating the behavior of the optimal bang-anneal-bang curve that takes the same amount of time. This section will seek to elucidate how QAOA can emulate the annealing-like curve despite the step sizes being large enough to throw off the usual error analysis of Product Formulas, also known as Trotterization.

To further see how QAOA is emulating the bang-anneal-bang curve, compare Figs. 3 & 5. These show the probabilities of being in the ground state and first excited state of the instantaneous evolutions for the bang-anneal-bang curve and QAOA. For QAOA, the instantaneous eigenbasis is determined by u_i the ratio of $\hat{B}$ bang for that layer. Both procedures roughly track the ground state with some leakage, mostly into the first excited state. QAOA as a rougher procedure has more leakage.

Let's suppose that there is some optimal procedure $u(t)$ such that the evolution governed by

$$\hat{H}(t) = u(t)\hat{B} + (1 - u(t))\hat{C} \tag{26}$$

gets us as close to the target state as we can get in time t_f. The control function is then further given by

$$u(t) = u_0(t/t_f) + c(t, t_f), \tag{27}$$

where $c(t)$ is some oscillatory function. For concreteness, we take

$$c(t, t_f) = -c_0(t_f) \sin\left(\frac{2\pi}{\tau}t + \phi\right), \tag{28}$$

where c_0 is some amplitude, τ some period, and ϕ some phase. We have included the negative sign and specified down to sine since this will make some interpretation easier later on. In essence, we have oscillations whose pattern of up and down match our preexisting pattern of up and down in QAOA protocols and also in the optimal bang-anneal-bang protocols.

For the purposes of this section, we will zoom in on a small region of the annealing curve where $\dot{u}_0$ is small and approximately constant. Then we want to ask the question how good a product series approximation of that evolution would be.

The actual evolution will be governed by the unitary time evolution operator

$$\hat{U}(t_f, 0) = \exp_{\mathcal{T}}\left(-i\int_0^{t_f} dt\, \hat{H}(t)\right) \tag{29}$$

where $\exp_{\mathcal{T}}$ denotes the time ordered exponential. Our goal will be to ask what happens if we use a Product Formula to break this up into a QAOA-like format such as the one layer approximation

$$\hat{U}_{PF}(t_f, 0) = \prod_{k=0}^{p-1} \hat{U}_1 \left(k\Delta t + \Delta t, k\Delta t\right), \tag{30}$$

where Δt is the length of the QAOA layer. In this section, we assume that every QAOA layer uses the same Δt, and as we will see, this corresponds to the frequency of oscillations in the optimal annealing curve being constant with time (again assuming that $\dot{u}_0$ is small enough). Therefore, we simply set $\Delta t = \frac{t_f}{p}$. It is appropriate to interpret this section as looking at a small region of the optimal annealing curve in the adiabatic limit where the oscillations occur on much shorter timescales than the gross changes in the curve. When we specify down to our specific control problem, we get

$$\begin{aligned}\hat{U}_1 \left(t_0 + \Delta t, t_0\right) = &\exp_{\mathcal{T}} \left(-i \int_{t_0}^{t_0+\Delta t} dt\, \hat{H}_0(t)\right) \\ &\times \exp_{\mathcal{T}} \left(-i \int_{t_0}^{t_0+\Delta t} dt\, \hat{H}_1(t)\right)\end{aligned} \tag{31}$$

In our setting $\hat{H}_0(t) = u(t)\hat{B}$ and $\hat{H}_1(t) = (1-u(t))\hat{C}$ which will simplify down this expression greatly to

$$\begin{aligned}\hat{U}_1 \left(t_0 + \Delta t, t_0\right) = &\exp \left(-i\hat{B} \int_{t_0}^{t_0+\Delta t} dt\, u(t)\right) \\ &\times \exp \left(-i\hat{C} \int_{t_0}^{t_0+\Delta t} dt\, (1-u(t))\right),\end{aligned} \tag{32}$$

where we have dropped the time ordering since $\hat{H}_0(t)$ and $\hat{H}_1(t)$ commute with themselves at all t (but not with each other).

Our core result is that taking $\Delta t = \tau$, the size of the Trotter slice equal to the period of the annealing oscillation, while keeping the ratio of the bang lengths proportional to $u_0(t)$, leads to a lower bound on the Trotterization error than if the oscillations were not considered. In this way it becomes advantageous for QAOA to match its layer length to the period of the optimal annealing oscillations and its ratio of $\hat{B}$ bang lengths and $\hat{C}$ bang lengths to the value of the base annealing function $u_0(t)$.

We show this enhancement in two different settings. First in Appendix C 1 we show this enhancement in the context of the standard operator error for product formulas. To cite our conclusion from the appendix, we find that

$$\begin{aligned}&||\hat{U}(t_f, 0) - \hat{U}_{PF}(t_f, 0)|| \\ &\leq \left|\left|\left[\hat{B}, \hat{C}\right]\right|\right| \frac{\Delta t^2 p}{2} \left(1 - \frac{c_0^2}{\pi}\right),\end{aligned} \tag{33}$$

specifically when $\Delta t = \tau$ and $\phi = 0$. In the case of $c_0 = 0$, this is equal to the standard error bound for product formulas. This enhancement washes out when $\Delta t \neq \tau$, so the enhancement is specifically dependent on matching the size of the Trotter steps to the period of the annealing oscillations.

Unfortunately, this bound on the error from unitaries specifically scales linearly with p, the number of QAOA layers. This is unfortunate because it means that rather than getting tighter with QAOA layer's the bound gets looser.

Our second setting, is specific to an adiabatic anneal where the goal is to maintain the populations of eigenstates, specifically the ground state in our setting. The overall Trotter error bound in this setting was recently tightened significantly by Yi and Crosson [30]. The same oscillatory enhancement found in the case of operator errors appears in this setting as well, but the method requires a perturbative limit which we do not have. We rederive the previous results and modify them for our setting in Appendix C 2. The key result of this method is that [30] reduce the scaling of the Trotter error down to $\mathcal{O}(\frac{1}{p}) + \mathcal{O}(\frac{\Delta t}{p})$ when trying to simulate an adiabatic evolution.

To address the breakdown of this method in our case, this method requires $\Delta t \in \mathcal{O}(n^{-1})$ which is inconsistent with our goal of $\Delta t = \tau$ since τ numerically seems to be a constant in the QAOA setting. The fact that Δt is small is necessary to employ the Magnus expansion which is a perturbative technique essential to the method of [30]. The enhancement still exists for cases when $\Delta t = \tau/M$ for an integer M, so it is possible to take $M \in \mathcal{O}(n)$ and get an enhancement. So even a product formula in this perturbative small Δt regime, there is a benefit to matching up with the annealing oscillations.

It is not possible to fully apply the Yi and Crosson method to our setting because of the perturbative Δt issues. Our product formula enhancement works partially in this setting where the error on the product formula for an adiabatic evolution scales as $\mathcal{O}(p^{-1})$ instead of the $\mathcal{O}(p)$ we see in the operator setting.

All of the results so far were based on just doing a simplistic product formula approximation of the underlying bang-anneal-bang curve. Of course, QAOA has more freedom than this and can modify the parameters to do a smarter approximation than just a product formula. It is allowed to modify the angles away from what a product formula would do in order to achieve more enhancement. Specifically, it could be possible to coherently match up the leakage between multiple QAOA layers. All the upper bounds described above assume a worst case scenarios that assume the errors from adjacent QAOA layers add coherently, but it is possible to design things so that the errors subtract coherently to some degree. We explore one way of doing this in Appendix I. There, we assume that the bang-bang unitaries are a coarse-graining and approximation of the underlying annealing unitaries and so introduce additional error, but only between low-lying

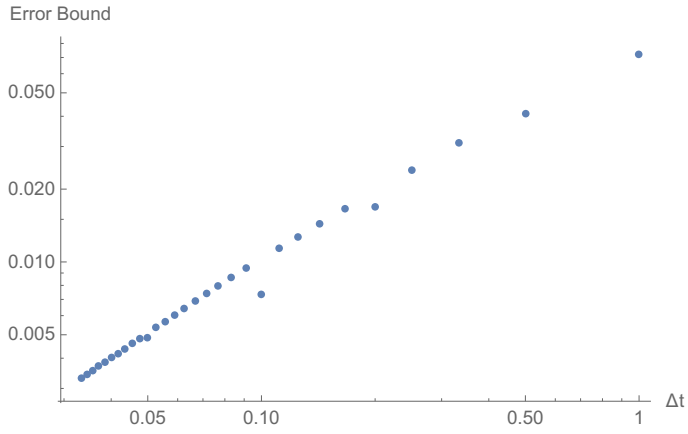


FIG. 6: This plot shows the upper bound on the Trotter error in the unitaries from Eq. 33 for a fixed t_f changing p. The oscillation period in this case was taken to be $\tau = 0.1$. Eq. 33 is specifically for the case when $\Delta t = \tau$ with this plot being more general. As can be seen, the proportional amount of enhancement is greatest when $\Delta t = \tau$, but there is a lesser enhancement when there is an integer multiple difference between these two quantities.

eigenstates. Therefore, we are able to implement a very similar strategy to coherently cancel this additional error, namely by introducing an oscillation in the bang ratios that is inversely proportional to the energy gap plus the error from the bang-bang approximation.

We should note that the results in this section should all be taken as analytic evidence supporting the numeric evidence from III. None of this is a proof that QAOA is Trotterizing the underlying bang-anneal-bang curve. The bounds described in this section do exhibit an enhancement when we match up Δt and the oscillatory period, τ, but the bounds are not tight enough to describe the exact setting we see in the numerics. We leave it up to future work to tighten these bounds even further to the setting of QAOA.

VI. CONCLUSION

The optimal bang-anneal-bang protocol is by construction the most efficient way to operate a quantum annealer or analog quantum computer. This protocol demonstrates structure including the initial and final bangs explored in [14] that vanish in the long t_f limit.

This work explored the structure of the annealing region in more detail. Do to previous results about the optimality of the adiabatic path [22], we expect and indeed see that in the long t_f limit, the annealing region looks like an optimized adiabatic schedule, similar to what was derived by [23]. Furthermore, that optimal annealing section has an oscillatory pattern superposed on top of it. In the adiabatic limit, the amplitude of these oscillations should vanish to recover a monotonic annealing ramp. However, in the near-adiabatic limit, these oscillations are helpful in managing the leakage between the ground state and first excited state. We derive the near-adiabatic form of these oscillations and describe their dependence on the phase difference between the ground state and first excited state amplitudes outside of this perturbative limit.

This analysis of the near-adiabatic limit of the annealing curve should be of interest in itself since it can be used to potentially enhance adiabatic protocols with little additional a priori information.

We, furthermore, explore the connections between QAOA and this oscillatory structure of the bang-anneal-bang curves. Numerically, we see that QAOA matches up exactly with the underlying bang-anneal-bang curve. The length of the QAOA layers matches up with the oscillation period of the annealing curve, and the ratio of the bang lengths within the QAOA layer matches up with the average value of the annealing curve within that period.

This behavior provides an explanation for the QAOA asymptotic curve behavior seen in [7, 8]. The behavior of the optimal bang-anneal-bang curve can be understood asymptotically where it approaches an optimized adiabatic procedure with a fixed curve form. If QAOA is emulating this optimal curve, then QAOA should also be approaching a fixed asymptotic form.

We sought to provide analytic evidence for this matching up between the QAOA curve and the optimal procedure. Our results do show that there is an decrease in the error of a product formula if the product formula step size matches the oscillations in an annealing curve being Trotterized. Furthermore, this enhancement requires that the ratio of the bangs follows the annealing curve, just as we see in the numerics. Unfortunately, this error bound scales unfavorably with p, the number of QAOA layers, failing to match up with the scaling in practice. Based on other methods, we provide further arguments for how this additional scaling behavior could occur, but it remains an open question how to tighten this analysis to match the exact scaling seen in QAOA in practice.

As a result of this analytic and numeric work, we not only achieve an explanation for the asymptotic large p behavior of QAOA, but we also better understand the optimal bang-anneal-bang procedure. One of the main difficulties with the optimal bang-anneal-bang curve is that it is not feasible to construct this protocol on real hardware. The protocol requires too much information about the intermediate quantum state and requires treating an entire smooth curve as a variational parameter. To address these issues, in Appendix D, we construct a new algorithm that uses the results of this paper to create an ansatz with very few variational parameters that outperforms quantum annealing and QAOA.

This algorithm uses a QAOA procedure to find the form of the annealing region of the bang-anneal-bang procedure and then uses this to create an ansatz. This ansatz then treats the lengths of the initial bang, final bang, and

some basic properties of the oscillatory pattern to create an ansatz with fewer variational parameters than QAOA. In practice, this algorithm outperforms QAOA and quantum annealing but falls slightly short of the full optimal bang-anneal-bang protocol.

Acknowledgments

The research of L.T.B. was partially supported by a National Institute of Standards and Technology (NIST) National Research Council (NRC) Research Postdoctoral Associateship Award in the Information Technology Lab (ITL). The research was supported by the U.S. Department of Energy Award No. DE-SC0019449 (insert short phrase describing Yaroslav's contribution), AFOSR (insert short phrase describing Przemek's contribution), and ARO MURI (insert short phrase describing Ani's contribution).

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, under the Quantum Computing Application Teams program. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

[1] E. Farhi, J. Goldstone, S. Gutmann, M. Sipser, arXiv:quant-ph/0001106 (2000).
[2] T. Kadowaki, H. Nishimori, *Phys. Rev. E* **58**, 5355 (1998).
[3] E. Farhi, J. Goldstone, S. Gutmann, arXiv:1411.4028 (2014).
[4] J. Preskill, *Quantum* **2**, 79 (2018).
[5] S. Jansen, M. Ruskai, R. Seiler, "Bounds for the adiabatic approximation with applications to quantum computation," *J. Math. Phys.* **48**, 102111 (2007).
[6] E. J. Crosson, D. A. Lidar, arXiv:2008.09913 (2020).
[7] L. Zhou, S. Wang, S. Choi, H. Pichler, M. D. Lukin, *Phys. Rev. X* **10**, 021067 (2020).
[8] G. Pagano, A. Bapat, P. Becker, K. S. Collins, A. De, P. W. Hess, H. B. Kaplan, A. Kyprianidis, W. L. Tan, C. Baldwin, L. T. Brady, A. Deshpande, F. Liu, S. Jordan, A. V. Gorshkov, C. Monroe, arXiv:1906.02700 (2019).
[9] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, E. F. Mishchenko, *The Mathematical Theory of Optimal Processes*, (Interscience Publishers, New York, 1962).
[10] Z. C. Yang, A. Rahmani, A. Shabani, H. Neven, C. Chamon, *Phys. Rev. X*, 7(2):1–8 (2017).
[11] A. Bapat, S. Jordan, arXiv:1812.02746 (2018).
[12] G. B. Mbeng, R. Fazio, G. Santoro, arXiv:1906.08948 (2019).
[13] C. Lin, Y. Wang, G. Kolesov, U. Kalabić, *Phys. Rev. A* **100**, 022327 (2019).
[14] L. T. Brady, C. L. Baldwin, A. Bapat, Y. Kharkov, A. V. Gorshkov, arXiv:2003.08952 (2020).
[15] J. Roland, N. J. Cerf, *Phys. Rev. A* **65**, 042308 (2002).
[16] A. Callison, M. Festenstein, J. Chen, L. Nita, V. Kendon, N. Chancellor, arXiv:2007.11599 (2020).
[17] M. Demirplak, S. A. Rice, *J. of Phys. Chem. A* **107**, 9937 (2003).
[18] M. V. Berry, *J. of Phys. A* **42**, 365303 (2009).
[19] D. Guéry-Odelin, A. Ruschhaupt, A. Kiely, E. Torrontegui, S. Martínez-Garaot, J. G. Muga, *Rev. Mod. Phys.* **91**, 045001 (2019).
[20] J. Huyghebaert, H De Raedt, *J. Phys. A: Math. Gen.* **23** 5777 (1990).
[21] D. Poulin, A. Qarry, R. D. Somma, F. Verstraete, *Phys. Rev. Lett.* **106**, 170501 (2011).
[22] M. Bukov, D. Sels, A. Polkovnikov, *Phys. Rev. X* **9**, 011034 (2019).
[23] J. Roland, N. J. Cerf, *Phys. Rev. A* **65**, 042308 (2002)
[24] A. T. Rezakhani, W. J. Kuo, A. Hamma, D. A. Lidar, P. Zanardi, *Phys. Rev. Lett.* **103**, 080502 (2009).
[25] F. Petiziol, B. Dive, F. Mintert, S. Wimberger, *Phys. Rev. A.* **98**, 043436 (2018).
[26] L. T. Brady, W. van Dam, arXiv:1801.04349 (2018).
[27] M. Jarret, B. Lackey, A. Liu, K. Wan, "Quantum adiabatic optimization without heuristics," arXiv:1810.04686 (2018).
[28] I. Hen, "How Quantum is the Speedup in Adiabatic Unstructured Search?", *Quant. Inf. Proc.* **18**, 162 (2019).
[29] F. G. S. L. Brandao, M. Broughton, E. Farhi, S. Gutmann, H. Neven, "For Fixed Control Parameters the Quantum Approximate Optimization Algorithm's Objective Function Value Concentrates for Typical Instances," arXiv:1812.04170 (2018).
[30] C. Yi, E. Crosson, "Spectral Analysis of Product Formulas for Quantum Simulation," arXiv:2102.12655 (2021).
[31] M. C. Tran, Y. Su, D. Carney, J. M. Taylor, "Faster Digital Quantum Simulation by Symmetry Protection," arXiv:2006.16248 (2020).

Appendix A: Near-Adiabatic

In this section of the appendix, we explore additional features of the near-adiabatic limit. In section A 1, we carry out a perturbative analysis of the near-adiabatic equations to find the form of the oscillations, $c(t)$, in the limit where the base ramp is changing slowly $\dot{u}_0 \ll 1$.

In section A 4, we examine the near-adiabatic equations in the limit where bang-bang control is enforced and explore what form the bang-bang control would need to be to get close to the results derived for the oscillations. This section should be taken with a grain of salt because the bang-bang control makes it harder to ensure that the

near-adiabatic assumptions are satisfied. This section is exploratory and helps explain features of our work but should not be taken as rigorously.

Finally, section A 3 derives the optimal control equations in the near-adiabatic limit. These differential equations, if solvable would give not only the oscillatory portion $c(t)$ but the entire function $u(t) = u_0(t) + c(t)$, enforcing $u(0)$ and $u(t_f)$. These equations might be of interest to experts or numericists but no long lend themselves to a perturbative analysis, making them less useful within the current context.

1. Perturbation and Ansatz for $c(t)$

Now, we want to zoom in one one small region of the optimal annealing curve. In this small region (we will call the start of this region $t = 0$ for convenience), the base curve $u_0(t)$ is varying very slowly such that $\dot{u}_0$ is a small quantity, and the region we have zoomed in on is such that $u_0(t)$ is approximately linear so that $u_0(t) \approx u_0^{(0)} + \dot{u}_0 t$ where $\dot{u}_0$ is approximately constant and small and $u_0^{(0)} \equiv u_0(0)$.

Next, we will Taylor expand $\Delta(u_0(t))$ and $\gamma(u_0(t))$ about $u_0(0) = u_0^{(0)}$ so that

$$\Delta(u_0(t)) \approx \Delta(u_0^{(0)}) + \left.\frac{d\,\Delta(u_0(t))}{d\,u_0(t)}\right|_{u_0(t)\to u_0^{(0)}} \dot{u}_0 t \tag{A1}$$

$$\gamma(u_0(t)) \approx \gamma(u_0^{(0)}) + \left.\frac{d\,\gamma(u_0(t))}{d\,u_0(t)}\right|_{u_0(t)\to u_0^{(0)}} \dot{u}_0 t \tag{A2}$$

For convenience, we shorten the notation here so that

$$\Delta(u_0(t)) \approx \Delta^{(0)} + \Delta^{(1)}\dot{u}_0 t \tag{A3}$$

$$\gamma(u_0(t)) \approx \gamma^{(0)} + \gamma^{(1)}\dot{u}_0 t \tag{A4}$$

Our strategy here will be to look at all of these equations perturbatively in the limit of small $c(t)$ and small $\dot{u}_0$. Our goal is to have the $A_0(t)$ and $A_1(t)$ equations return to themselves after a set period of time. To match up with our numeric results we will require them to return only at the end of one period of oscillation for $\varphi(t)$. Ultimately, this means that the integrals in Eqs. (21) & (22) should be zero after a full period, perturbatively in our small parameters. It is fairly obvious to see that the integral is zero to zeroeth order, and some simple calculations show that $c(t) = 0$ is the solution to first order in $\dot{u}_0$, so we will likely need to go to at least second order in $\dot{u}_0$. This incidentally means that we expect $c \in \mathcal{O}(\dot{u}_0^2)$, so we will roughly equate the size of $c(t)$ and $\dot{u}_0^2$.

We start with the $\dot{\varphi}$ equation, Eq. (18), and look for the solution to this equation to first order in $\dot{u}_0$ (which means that $c(t)$ will be negligibly small). Firstly, since the A_i terms only appear attached to small parameters, we can approximate them by their zeroeth order constants, and the same goes for Δ and γ. The $\sin\varphi$ and $\cos\varphi$ terms are troublesome, but they are already multiplied by small parameters, so we can approximate them by the zeroeth order solution to this equation $\varphi(t) \approx \Delta^{(0)}t$ (using as a boundary condition that $\varphi(0) = 0$). With these approximations and iterations, the first order solution to the equation is

$$\begin{aligned}\varphi(t) \approx{}& \Delta^{(0)}t \\ &+ \left(\frac{1}{2}\Delta^{(1)}t^2 + \frac{\gamma^{(0)}(\cos(\Delta^{(0)}t) - 1)}{(\Delta^{(0)})^2\tan(2\vartheta)}\right)\dot{u}_0\end{aligned} \tag{A5}$$

We next go to the integral of interest and expand all our parameters to second order in $\dot{u}_0$:

$$\begin{aligned}0 \approx{}& \int_0^{t_f} dt\; c(t)\gamma^{(0)}\sin(\Delta^{(0)}t) \\ &+ \int_0^{t_f} dt\frac{\gamma^{(0)}}{\Delta_0^{(0)}}\dot{u}_0\cos(\Delta^{(0)}t) \\ &+ \int_0^{t_f} dt\frac{\gamma^{(0)}}{\Delta_0^{(0)}}\left(\frac{\gamma^{(1)}}{\gamma^{(0)}} - \frac{\Delta^{(1)}}{\Delta^{(0)}}\right)\dot{u}_0^2 t\cos(\Delta^{(0)}t) \\ &+ \int_0^{t_f} dt\frac{\gamma^{(0)}}{\Delta_0^{(0)}}\left(\frac{1}{2}\Delta^{(1)}t^2\right)\dot{u}_0^2\sin(\Delta^{(0)}t) \\ &+ \int_0^{t_f} dt\frac{\gamma^{(0)}}{\Delta_0^{(0)}}\left(\frac{\gamma^{(0)}(\cos(\Delta^{(0)}t) - 1)}{(\Delta^{(0)})^2\tan(2\vartheta)}\right)\dot{u}_0^2\sin(\Delta^{(0)}t)\end{aligned} \tag{A6}$$

With the exception of the first line, every quantity here is known. Furthermore, in order to capture just one period of the oscillations, we will look at these integrals over the range such that t_f encompasses one full oscillation for most of these terms that are $\mathcal{O}(\dot{u}_0^2)$ this means that $t_f = \frac{2\pi}{\Delta^{(0)}}$ but for the second equation, that is only $\mathcal{O}(\dot{u}_0)$, we need to look at the period to first order which gives $t_f = \frac{2\pi}{\Delta^{(0)}} - \frac{2\pi^2\Delta^{(1)}\dot{u}_0}{(\Delta^{(0)})^3}$. The relevant integrals are relatively easy to carry out with all but the first, second, and fourth lines vanishing (with the second and fourth lines giving the same contribution up to the relevant order):

$$\begin{aligned}0 \approx{}& \int_0^{2\pi/\Delta^{(0)}} dt\; c(t)\gamma^{(0)}\sin(\Delta^{(0)}t) \\ &- 4\frac{\pi^2\gamma^{(0)}\Delta^{(1)}}{(\Delta^{(0)})^4}\dot{u}_0^2\end{aligned} \tag{A7}$$

Now, without any ansatz for $c(t)$ this is as far as we can go. We will take the relatively simple ansatz (inspired by our numerics) that

$$c(t) = c_0\sin(\omega t + \theta) \tag{A8}$$

with c_0 being the quantity we expect to be in $\mathcal{O}(\dot{u}_0^2)$. Furthermore, we would like this perturbation to match up between adjacent oscillations at the endpoints, so up to the appropriate order, we take $\omega = m\Delta^{(0)}$ for $m \in \mathbb{Z}^+$. Using this anstaz, we can integrate the remaining integral in our condition to get

$$0 \approx \frac{\pi\gamma^{(0)}c_0\cos\theta}{\Delta^{(0)}}\delta_{m,1} - 4\frac{\pi^2\gamma^{(0)}\Delta^{(1)}}{(\Delta^{(0)})^4}\dot{u}_0^2 \tag{A9}$$

where $\delta_{m,1}$ is 1 if $m = 1$ and 0 for all other $m \in \mathbb{Z}^+$. So the only solution is if $\omega = \Delta^{(0)}$, meaning that we are on resonance with the natural oscillations in the system. It is simple to solve for c_0 which yields (reinstating the $+\mathcal{O}(\dot{u}_0^3)$ that has been implicit throughout):

$$c_0 = \frac{4\pi\Delta^{(1)}}{(\Delta^{(0)})^3 \cos\theta}\dot{u}_0^2 + \mathcal{O}(\dot{u}_0^3) \tag{A10}$$

Our goal is to get the smallest perturbation possible that still accomplishes the desired task, so taking $\theta = 0$ is appropriate. Therefore, we settle upon the result that in this limit, an oscillation described by

$$c(t) = \frac{4\pi\Delta^{(1)}}{(\Delta^{(0)})^3}\dot{u}_0^2 \sin(\Delta^{(0)}t) + \mathcal{O}(\dot{u}_0^3) \tag{A11}$$

will cancel out the non-adiabatic effects and allow for more precise adiabatic passage.

Since we had focused in on a region where all of this is linear, we further expect that over a larger swath of time, the oscillations will follow the generalized pattern

$$c(t) = \frac{4\pi\frac{d\,\Delta}{d\,u_0(t)}}{(\Delta(u_0(t)))^3}\dot{u}_0(t)^2 \sin(\Delta(u_0(t))t) + \mathcal{O}(\dot{u}_0^3) \tag{A12}$$

2. Adiabatic Frame

In the main text, we derived the near-adiabatic limit for the case of the instantaneous eigenframe evolving alongside the base curve $u_0(t)$ so that the full control function was given by $u(t) = u_0(t) + c(t)$. In this setting $c(t)$ was our actual free function with $u_0(t)$ fixed and $c(0) = c(t_f) = 0$. It is possible to treat this entire problem a control problem and seek out the $u(t)$ that maintains populations the best between $t = 0$ and $t = t_f$ subject to the constraint that $u(0) = u_a$ and $u(t_f) = u_b$.

The full version of this problem would just result in bang-anneal-bang curves in general, but here we are interested in just the smooth annealing region. One feature of this smooth annealing region is that the bangs have already excited up some of the state into the first excited state, making the near-adiabatic approximation even more relevant. For the purposes of this section, we will zoom in one one small region of the annealing curve and still implicitly assume that u_a and u_b are not that far apart.

We consider the probability amplitudes, $C_i(t)$ of being in the instantaneous eigenstates of a system, $|j(u(t))\rangle$, with instantaneous eigenenergies, $\lambda_j(u(t))$. So our state can be written as

$$|\psi(t)\rangle = \sum_j C_j(t)\,|j(u(t))\rangle\,, \tag{A13}$$

Applying the Schrödinger equation to this state yields a set of coupled differential equations

$$i\left(\frac{d\,C_k(t)}{d\,t} + \sum_j C_j(t)\,\langle k(u(t))|\,\frac{d}{d\,t}\,|j(u(t))\rangle\right) = \lambda_k(u(t))C_k(t). \tag{A14}$$

Now, a few assumptions will be made, the first being that the ground and first excited states are non-degenerate. This could in general be satisfied by going into a symmetric subspace and looking at the relevant probability amplitudes within that symmetric subspace (for instance with the transverse field Ising model, we will consider the subspace defined by the usual Ising rotational symmetry). The second and more relevant assumption is that $|C_0| \gg |C_1| \gg |C_2| \gg \ldots$ which is just a statement that we are in the near-adiabatic limit of evolution. For our purposes, we will assume that the amplitudes for the second excited state and above are small enough throughout the evolution to be negligible. Furthermore, we will later consider $|C_1|$ to be a small quantity relative to $|C_0|$ for approximation purposes. The last requirement is that we set $\lambda_0(t) = 0$ which can be done without loss of generality.

Applying these assumptions and following the calculations of [26], we derive the equations:

$$\frac{d\,C_0(t)}{d\,t} = -\frac{\gamma(u(t))\dot{u}(t)}{\Delta(u(t))}C_1(t) \tag{A15}$$

$$\frac{d\,C_1(t)}{d\,t} = \frac{\gamma(u(t))\dot{u}(t)}{\Delta(u(t))}C_0(t) - i\Delta(u(t))C_1(t), \tag{A16}$$

where the new functions represent

$$\gamma(u) \equiv \langle\varphi_0(u)|\,(\hat{B} - \hat{C})\,|\varphi_1(u)\rangle \tag{A17}$$

$$\Delta(u) \equiv \lambda_1(u) - \lambda_0(u) \tag{A18}$$

This leaves us with two coupled, complex differential equations. To make things more explicit, we now split the C variables into real amplitudes and phases such that

$$C_0(t) = e^{i\varphi_0(t)}A_0(t)$$
$$C_1(t) = e^{i\varphi_1(t)}A_1(t)$$

These can be inserted into the differential equations. After some algebra, including separating out real and imaginary components, the differential equations reduce to the real equations

$$\dot{\varphi} = \Delta - \frac{A_0^2 - A_1^2}{A_0 A_1}\frac{\gamma\dot{u}}{\Delta}\sin(\varphi) \tag{A19}$$

$$\dot{A}_0 = -\frac{\gamma\dot{u}}{\Delta}\cos(\varphi)A_1 \tag{A20}$$

$$\dot{A}_1 = \frac{\gamma\dot{u}}{\Delta}\cos(\varphi)A_0 \tag{A21}$$

where $\varphi(t) \equiv \varphi_0(t) - \varphi_1(t)$

The A equations can be integrated to give

$$A_0(t) = a\cos\left(\int_0^t dt' \frac{\gamma\dot{u}}{\Delta}\cos(\varphi) + \vartheta\right) \tag{A22}$$

$$A_1(t) = a\sin\left(\int_0^t dt' \frac{\gamma\dot{u}}{\Delta}\cos(\varphi) + \vartheta\right). \tag{A23}$$

These equations are similar to what was seen in the main text, and once again we are left with the conclusion that at the end of the evolution region we are considering, we want

$$\Theta[u(t)] \equiv \int_0^{t_f} dt \frac{\gamma\dot{u}}{\Delta}\cos(\varphi) \tag{A24}$$

to be close to a multiple of π. Though, we again have the caveat that having $\Theta[u(t)]$ equal to any multiple of π other than zero would violate the assumptions of near-adiabaticity.

3. Optimal Control

Our setup is to take a procedure that goes from time 0 to time t_f moving from $u(0) = u_1$ at the beginning to $u(t_f) = u_2$ at the end. We want to ensure that the instantaneous eigenstate populations are maintained during that evolution, at least from the beginning to the end (but not necessarily in the middle), so we want to minimize

$$\begin{aligned} J = &|C_0^*(t_f)C_0(t_f) - C_0^*(0)C_0(0)| \\ &+ |C_1^*(t_f)C_1(t_f) - C_1^*(0)C_1(0)|. \end{aligned} \tag{A25}$$

The actual form of whether we are looking at the L^1 or L^2 norm of the difference between the probabilities is largely irrelevant, and another choice could be made with little consequence. We have also written out the probabilities explicitly as $|C|^2 = C^*C$ which will be helpful shortly.

Now, we will treat this as an optimal control problem, seeking to find the conditions on $u(t)$ such that J is minimized. In order to enforce Eqs. A15 & A16, we introduce Lagrange Multipliers $D_0(t)$ and $D_1(t)$ so that

$$\begin{aligned} J = &|C_0^*(t_f)C_0(t_f) - C_0^*(0)C_0(0)| \\ &+ |C_1^*(t_f)C_1(t_f) - C_1^*(0)C_1(0)| \\ &+ \int_0^{t_f} dt \left[D_0(t)\left(\dot{C}_0(t) + \frac{\gamma(u(t))\dot{u}(t)}{\Delta(u(t))}C_1(t)\right)\right. \\ &\left. + D_1(t)\left(\dot{C}_1(t) - \frac{\gamma(u(t))\dot{u}(t)}{\Delta(u(t))}C_0(t) + i\Delta(u(t))C_1(t)\right)\right] \\ &+ c.c. \end{aligned} \tag{A26}$$

where the final *c.c.* indicates that we need to complex conjugates of the third and forth lines, just to treat the variables and their complex conjugates equally (remember that $u(t)$ is purely real).

Now, we just perform a Calculus of Variations analysis of this using $C_0(t)$, $C_1(t)$, $C_0^*(t)$, $C_1^*(t)$, and $u(t)$ as the variational parameters. In doing this procedure, it is important to remember that the C variables are fixed at $t = 0$ but not at t_f and that $u(t)$ is fixed at both end points.

Note that under a full optimal control theory analysis, such as [14], there would be restrictions on $u(t)$ such as $u(t) \in [0, 1]$. In this setting, we will ignore this restriction for ease of analysis, and this ignoring is justified by the fact that we are interested specifically at how this system behaves in an annealing region. We are using this analysis explicitly to look at the annealing rather than bang-bang portions of the control function, and any of our results here should be taken explicitly within that context. Also note that any restrictions on the C variables is already taken care of by the fact that Eqs. A15 & A16 are being enforced by the Lagrange multipliers. These equations came from the Schrödinger equation, so the Cs will obey all necessary properties of probability amplitudes.

The resulting end point equations yield the boundary conditions for the D variables

$$\begin{aligned} D_0(t_f) &= -\text{sgn}(|C_0(t_f)|^2 - |C_0(0)|^2)C_0^*(t_s) \\ D_1(t_f) &= -\text{sgn}(|C_1(t_f)|^2 - |C_1(0)|^2)C_1^*(t_s). \end{aligned}$$

Any changes to using the L^1 or L^2 norm originally would have shown up here and would have just resulted in slightly different boundary conditions.

The variational procedure for the D Lagrange multipliers just results in Eqs. A15 & A16 again as expected, and the variational procedure for the C variables results in

$$\dot{D}_0(t) = -\frac{\gamma(u(t))\dot{u}(t)}{\Delta(u(t))}D_1(t) \tag{A27}$$

$$\dot{D}_1(t) = \frac{\gamma(u(t))\dot{u}(t)}{\Delta(u(t))}D_0(t) - i\Delta(u(t))D_1(t). \tag{A28}$$

These are essentially following their own Schrödinger evolution. Also note that based on the boundary conditions for the Ds, we have $D_1(t)$ roughly the same size as $C_1(t)$ and $D_0(t)$ roughly the same size as $C_0(t)$. Hence we can use the same hierarchy of $C_0 \gg C_1$ with these new variables.

The last equation, resulting from the variations of $u(t)$ is the one that is actually important here. Assuming the gap is nonzero, the resulting condition can be written as (suppressing functional dependencies for space reasons)

$$\gamma\left(D_0\dot{C}_1 + \dot{D}_0C_1 - C_0\dot{D}_1 - \dot{C}_0D_1\right) = iC_1D_1\Delta\Delta' \tag{A29}$$

where $\Delta' = \frac{\partial\Delta(u(t)}{\partial u(t)}$. The natural next step is to use Eqs. A15, A16, A27, & A28 to eliminate the time derivatives of the C and D variables:

$$\begin{aligned} &(C_1(t)D_0(t) + D_0(t)C_1(t))\,\gamma(u(t)) \\ =& -C_1(t)D_1(t)\Delta'(u(t)) \end{aligned} \tag{A30}$$

This gives us the full set of optimal control equations that are necessary for an optimal procedure. Unfortunately, this formalism does not lend itself to the perturbative analysis discussed in the previous sections. These results are presented for completeness.

4. Requisite Bang-Bang

Let's go back to the near-adiabatic frame that is following $u_0(t)$ but instead of having $u(t) = u_0(t) + c(t)$, let's just impose that $u(t) = b(t)$ where $b(t)$ is a bang-bang function. If we did this, then the near-adiabatic differential equations would become

$$i\left(\frac{d\,C_0(t)}{d\,t} + C_1(t)\frac{\gamma(t)\dot{u}_0(t)}{\Delta(t)}\right) = C_0(t)\upsilon_{00}(t,b(t)) + C_1(t)\upsilon_{01}(t,b(t)) \tag{A31}$$

$$i\left(\frac{d\,C_1(t)}{d\,t} - C_0(t)\frac{\gamma(t)\dot{u}_0(t)}{\Delta(t)}\right) = C_0(t)\upsilon_{10}(t,b(t)) + C_1(t)\upsilon_{11}(t,b(t)) \tag{A32}$$

where

$$\upsilon_{ij}(t,b) = \delta_{b1}\upsilon_{ij1}(t) + \delta_{b0}\upsilon_{ij0}(t) \tag{A33}$$

$$= \delta_{b1}\left\langle i_0(t)\right|\hat{B}\left|j_0(t)\right\rangle + \delta_{b0}\left\langle i_0(t)\right|\hat{C}\left|j_0(t)\right\rangle. \tag{A34}$$

For stoquastic Hamiltonians in this restricted subspace, υ_{ij} is symmetric and real. Unfortunately, this basis is not easily related to either the initial or final time bases, but we can use that the instantaneous basis to relate them via

$$u_0(t)\upsilon_{ij1}(t) + (1-u_0(t))\upsilon_{ij0}(t) = \delta_{j1}\delta_{i1}\Delta(t). \tag{A35}$$

For convenience, we will rewrite all equations in terms of υ_{ij1} variables. Then, the differential equations become

$$i\left(\frac{d\,C_0(t)}{d\,t} + C_1(t)\frac{\gamma(t)\dot{u}_0(t)}{\Delta(t)}\right) = C_0(t)\upsilon_{001}(t) + C_1(t)\upsilon_{011}(t) \tag{A36}$$

$$i\left(\frac{d\,C_1(t)}{d\,t} - C_0(t)\frac{\gamma(t)\dot{u}_0(t)}{\Delta(t)}\right) = C_0(t)\upsilon_{011}(t) + C_1(t)\upsilon_{111}(t) \tag{A37}$$

when $b(t) = 1$ and

$$i\left(\frac{d\,C_0(t)}{d\,t} + C_1(t)\frac{\gamma(t)\dot{u}_0(t)}{\Delta(t)}\right) = -\frac{u_0(t)}{1-u_0(t)}\left(C_0(t)\upsilon_{001}(t) + C_1(t)\upsilon_{011}(t)\right) \tag{A38}$$

$$i\left(\frac{d\,C_1(t)}{d\,t} - C_0(t)\frac{\gamma(t)\dot{u}_0(t)}{\Delta(t)}\right) = \frac{\Delta(t)C_1(t)}{1-u_0(t)} - \frac{u_0(t)}{1-u_0(t)}\left(C_0(t)\upsilon_{011}(t) + C_1(t)\upsilon_{111}(t)\right) \tag{A39}$$

when $b(t) = 0$

These can be separated into phase and amplitude just as in the main text in the continuous case. The phase difference is determined by two differential equations depending on whether $b(t) = 1$,

$$\varphi' = \upsilon_{111} - \upsilon_{001} + \frac{A_0^2 - A_1^2}{A_0 A_1}\left(\upsilon_{011}\cos(\varphi) - \frac{\gamma\dot{u}_0}{\Delta}\sin(\varphi)\right), \tag{A40}$$

or $b(t) = 0$,

$$\varphi' = \frac{\Delta}{1-u_0} - \frac{u_0}{1-u_0}(\upsilon_{111} - \upsilon_{001}) + \frac{A_0^2 - A_1^2}{A_0 A_1}\left(-\frac{u_0}{1-u_0}\upsilon_{011}\cos(\varphi) - \frac{\gamma\dot{u}_0}{\Delta}\sin(\varphi)\right). \tag{A41}$$

These would then need to be matched at the boundaries when $b(t)$ switches

The $A_0(t)$ and $A_1(t)$ differential equations can also be written down, but it is possible to solve those differential equations as before. Using boundary matching, it is possible to find that

$$A_0(t) = a\cos(\Theta_{bb}[b(t)] + \vartheta) \tag{A42}$$

$$A_1(t) = a\sin(\Theta_{bb}[b(t)] + \vartheta) \tag{A43}$$

exactly as before, but now the Θ function takes on a new form:

$$\Theta_{bb}[b(t)] = \int_0^{t_f} dt\frac{\gamma\dot{u}}{\Delta}\cos(\varphi) + \int_{b=0} dt\sin(\varphi)\upsilon_{011} - \int_{b=1} dt\sin(\varphi)\upsilon_{011}\frac{u_0}{1-u_0}, \tag{A44}$$

where on the second line, the first integral is evaluated over all times when $b(t) = 0$ and the second integral is evaluated over all time when $b(t) = 1$.

As always, out goal is to ensure that Θ_{bb} is close to zero at the end of the procedure, and in this case our control is whether $b(t) = 0, 1$.

A perturbative expansion of these equations in terms of $\dot{u}$ still results in a set of equations that is not solvable analytically.

Appendix B

Recall the Schrödinger equation, Eq. (A38) and (A39), for the ground state and first excited state amplitudes,

$$i\left(\frac{d\,C_0(t)}{d\,t} + C_1(t)\frac{\gamma(t)\dot{u}_0(t)}{\Delta(t)}\right) = c(t)\left(C_0(t)\kappa_0(t) + C_1(t)\gamma(t)\right) \tag{B1}$$

$$i\left(\frac{d\,C_1(t)}{d\,t} - C_0(t)\frac{\gamma(t)\dot{u}_0(t)}{\Delta(t)}\right) \tag{B2}$$
$$= \Delta(t)C_1(t) + c(t)\left(C_0(t)\gamma(t) + C_1(t)\kappa_1(t)\right)$$

we assume γ_0 =const, κ_j =const, $c(t) = c_0\sin(i\omega t + \theta)$, $\Delta(t) =: \Delta_0 + \Delta_1 t$, $\Omega = \gamma_0\dot{u}_0/\Delta$. In the new notation, we have two-level Hamiltonian that we will analyse within time-dependent perturbation theory with

$$H_0 = \begin{pmatrix} 0 & 0 \\ 0 & \Delta_0 \end{pmatrix} \tag{B3}$$

and perturbation

$$V = \begin{pmatrix} c_0\kappa_0\sin(\theta + t\omega) & c_0\gamma_0\sin(\theta + t\omega) - i\Omega \\ c_0\gamma_0\sin(\theta + t\omega) + i\Omega & c_0\kappa_1\sin(\theta + t\omega) + \Delta_0 + \Delta_1 t \end{pmatrix}$$

Up to second order,

$$|\psi_I(t)\rangle = \left[1 - i\int_0^t d\tau\mathcal{V}(t) - \int_0^t d\tau\mathcal{V}(\tau)\int_0^\tau d\tau_2\mathcal{V}(\tau_2)\right]|\psi(0)\rangle$$

with $\mathcal{V}(t) = W(t)^\dagger V(t)W(t)$ where

$$W(t) = \begin{pmatrix} 1 & 0 \\ 0 & e^{-i\Delta_0 t} \end{pmatrix} \tag{B4}$$

We perform these integrals, and expand order by order in $\dot{u}$ with $t = t_f = 2\pi/\Delta_0 + \delta t_f$, $\delta t_f = \mathcal{O}(\dot{u}), \Delta_1 = \mathcal{O}(\dot{u})$, $c_0 = \mathcal{O}(\dot{u}^2)$, $\omega - \Delta = \mathcal{O}(\dot{u})$.

To come back from interaction picture $_I$, we use that $|\psi(t)\rangle = W(t)\,|\psi_I(t)\rangle$.

First order is given by

$$U^{(1)} = \begin{pmatrix} 0 & 0 \\ 0 & -\frac{2i\pi^2\Delta_1}{\Delta_0^2} \end{pmatrix} \tag{B5}$$

which for small perturbation leads to a rotation (phase pick up) of the excited state due to the i in front.

Second order $U^{(2)}$ equals

$$\begin{pmatrix} \frac{2i\pi\Omega^2}{\Delta_0^2} & U_{12}^{(2)} \\ U_{12}^{(2)*} & -\frac{2\pi\left(i\Delta_0^2\Omega^2 + \pi^3\Delta_1^2 + i\Delta_1\Delta_0^3\delta t_f\right)}{\Delta_0^4} \end{pmatrix} \tag{B6}$$

where diagonal terms correspond to the phase imprint. And we focus on the off-diagonal term

$$U_{12}^{(2)} = -\frac{\pi\gamma_0 c_0(\cos(\theta) + i\sin(\theta))}{\Delta_0} - \frac{2\pi(\pi - i)\Delta_1\Omega}{\Delta_0^3} - \Omega\delta t_f$$

Since all parameters are real, we need $\theta \neq 0$ to cancel imaginary term $\propto \Omega$. If we chose δt_f such that it cancels the real part of the off-diagonal expression, then $\theta = \pi/2$ and we can cancel the imaginary part by the smallest possible c_0.

This way we get

$$\delta t_f = -\frac{\left(2\pi^2\right)\Delta_1}{\Delta_0^3} \tag{B7}$$

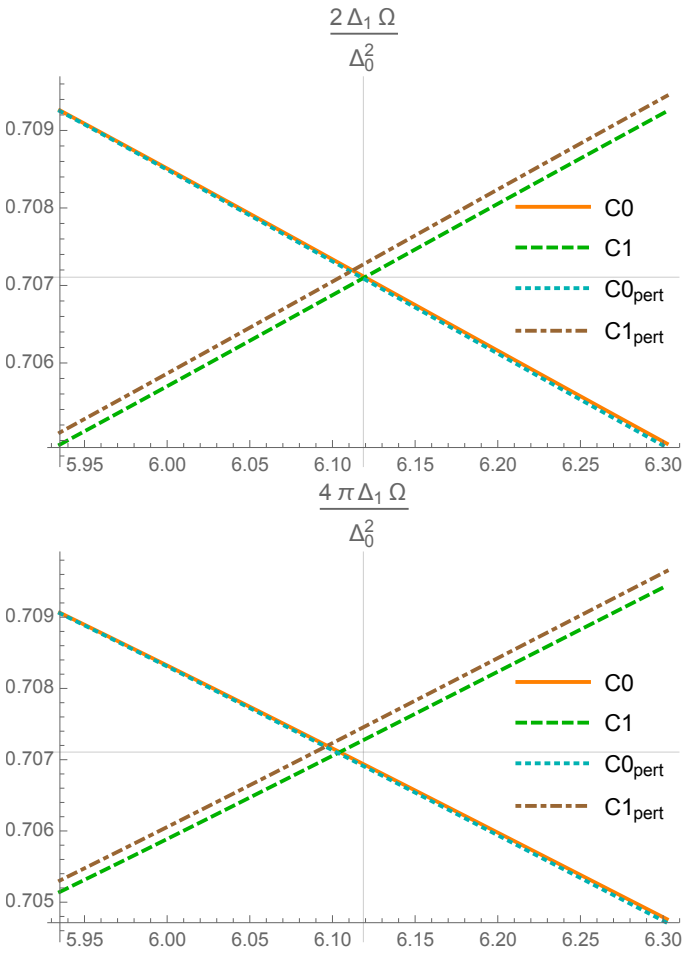


FIG. 7: Comparison of the exact numerics with the perturbative expressions up to the second order. Vertical gray thin line indicate the position of $t = t_f + \delta t_f$. We start from equal superposition in 0 and 1, and use two different expressions for c_0: (left) Przemek's, (right) LucasB's. However both has some systematic deviations between the exact theory and the perturbative expressions–maybe there is a mistake on my side.

which agrees with LucasB, and

$$c_0 = \frac{2\Delta_1\Omega}{\gamma_0\Delta_0^2} \tag{B8}$$

which is smaller by factor 2π compared to the LucasB's expression.

From numerics, we get that that Przemek's expressions agree better than LucasB's. I've tried to figure out from Lucas' calculation the 2π difference but couldn't find a mistake.

Appendix C: Trotterization Error

1. Standard Operator Error Scaling

In this section, our goal will be to find the bound on the matrix norm error between unitaries given in Eqs. 29 & 30 under the assumption that the evolution is governed by the oscillatory function given in Eq. 27

This sequence of arguments will initially follow the appendix of [20]. For another good reference on this, try [21].

Finding the error between these two unitaries is fairly straightforward and is laid out well in [20]. We rederive this result in Appendix F. To cite the result

$$||\hat{U}(t_f,0)-\hat{U}_{PF}(t_f,0)|| \leq \sum_{k=0}^{p-1}\int_{k\Delta t}^{(k+1)\Delta t} ds \int_{k\Delta t}^{s} dr \left|\left|\left[\hat{H}_0(r),\hat{H}_1(s)\right]\right|\right|. \tag{C1}$$

The matrix norm used in this proof was the standard operator norm. Also, notably, this result does not rely on perturbative methods like the Baker-Campbell-Hausdorff equation or the Magnus expansion.

Now, it is fairly straightforward to specify down to the form we are using in which case

$$||\hat{U}(t_f,0)-\hat{U}_{PF}(t_f,0)|| \leq \left|\left|\left[\hat{B},\hat{C}\right]\right|\right| \sum_{k=0}^{p-1}\int_{k\Delta t}^{(k+1)\Delta t} ds \int_{k\Delta t}^{s} dr\, u(r)(1-u(s)). \tag{C2}$$

Next, the form in Eq. (C2) is a little unruly to work with. It actually is much easier to go over to Fourier space where

$$u(t) = \int_{-\infty}^{\infty} d\xi\, \tilde{u}(\xi) e^{2\pi i t \xi}. \tag{C3}$$

Of course if we take the Fourier transform of Eq. (27), we would get

$$\tilde{u}(\xi) = \tilde{u}_0(\xi) + \frac{c_0}{2i}\left(e^{i\phi}\delta\left(\xi+\frac{1}{\tau}\right) - e^{-i\phi}\delta\left(\xi-\frac{1}{\tau}\right)\right). \tag{C4}$$

Putting this Fourier transformed version in allows us to easily do the integrals over s and r, resulting in

$$||\hat{U}(t_f,0)-\hat{U}_{PF}(t_f,0)|| \leq \left|\left|\left[\hat{B},\hat{C}\right]\right|\right| \sum_{k=0}^{p-1}\int_{-\infty}^{\infty} d\xi \int_{-\infty}^{\infty} d\eta \tilde{u}(\xi)\left(\delta(\eta)-\tilde{u}(\eta)\right) \times \left[\frac{\left(e^{2i\pi\Delta t\eta}\left(1-e^{2i\pi\Delta t\xi}\right)\eta - \left(1-e^{2i\pi\Delta t\eta}\right)\xi\right)e^{2i\pi\Delta t k(\eta+\xi)}}{4\pi^2\eta\xi(\eta+\xi)}\right]. \tag{C5}$$

Notice that all the k dependence is in the last line, so we can carry out the k sum fully to get

$$||\hat{U}(t_f,0)-\hat{U}_{PF}(t_f,0)|| \leq \left|\left|\left[\hat{B},\hat{C}\right]\right|\right| \int_{-\infty}^{\infty} d\xi \int_{-\infty}^{\infty} d\eta \tilde{u}(\xi)\left(\delta(\eta)-\tilde{u}(\eta)\right) \times \frac{\left(\eta\left(-1+e^{2i\pi\Delta t\xi}\right)+\xi\left(1-e^{2i\pi\Delta t\eta}\right)\right)\left(1-e^{2i\pi\Delta t p(\eta+\xi)}\right)}{4\pi^2\eta\xi(\eta+\xi)\left(e^{2i\pi\Delta t\xi}-e^{-2i\pi\Delta t\eta}\right)}. \tag{C6}$$

From this point, putting in Eq. (C4) and simplifying down is quite possible; although, the fully general expression is a bit messy and not horribly informative. One possible simplification that is quite informative is the case where $\tau \to \Delta t$ in which case manipulation can simplify all of this nicely down to

$$||\hat{U}(t_f,0)-\hat{U}_{PF}(t_f,0)|| \leq \left|\left|\left[\hat{B},\hat{C}\right]\right|\right| \int_{-\infty}^{\infty} d\xi \int_{-\infty}^{\infty} d\eta \tilde{u}_0(\xi)\left(\delta(\eta)-\tilde{u}_0(\eta)\right) \times \frac{\left(\eta\left(-1+e^{2i\pi\Delta t\xi}\right)+\xi\left(1-e^{2i\pi\Delta t\eta}\right)\right)\left(1-e^{2i\pi\Delta t p(\eta+\xi)}\right)}{4\pi^2\eta\xi(\eta+\xi)\left(e^{2i\pi\Delta t\xi}-e^{-2i\pi\Delta t\eta}\right)} - \left|\left|\left[\hat{B},\hat{C}\right]\right|\right| c_0^2\frac{\Delta t^2 p\cos(\phi)}{2\pi}. \tag{C7}$$

If we undo the Fourier transform, this further reduces to

$$||\hat{U}(t_f,0)-\hat{U}_{PF}(t_f,0)|| \leq \left|\left|\left[\hat{B},\hat{C}\right]\right|\right| \sum_{k=0}^{p-1}\int_{k\Delta t}^{(k+1)\Delta t} ds \int_{k\Delta t}^{s} dr\, u_0(r)(1-u_0(s)) - \left|\left|\left[\hat{B},\hat{C}\right]\right|\right| c_0^2\frac{\Delta t^2 p\cos(\phi)}{2\pi}. \tag{C8}$$

In other words, the sine function we added onto the control function essentially becomes decoupled from the rest of the error in the case that its period is the same as the Trotter slice size. Furthermore, the first line of the error bound will always be positive (remember that $u_0 \in [0,1]$), but the second line can be negative, effectively reducing the error in the Trotterization. As promised, choosing $\phi = 0$ results in the maximum improvement.

The improvement in the error is proportional to $\Delta t^2 p$ which is coincidentally the same rough scaling as the term above, so this term will actually be competitive and could contribute greatly to the error bound. To see this more precisely, note that the first term in the bound can be upper bounded quite easily by $\frac{1}{2}\Delta t^2 p$ so that

$$||\hat{U}(t_f,0)-\hat{U}_{PF}(t_f,0)|| \leq \left|\left|\left[\hat{B},\hat{C}\right]\right|\right| \frac{\Delta t^2 p}{2}\left(1-\frac{c_0^2}{\pi}\cos(\phi)\right). \tag{C9}$$

Note that $c_0 \leq 0.5$ at the very worst to ensure that $u(t) \in [0,1]$, so it is not possible for this bound to be below zero.

2. Adiabatic Trotter Error

In this subsection our goal will be to bound the Trotter error by looking at the error on the ground state fidelity

directly. It should be noted that our analysis indicates that the underlying annealing curve adiabatically transfers not just the ground state but also higher excited states, with this reducing down to just ground state adiabaticity in the limit of $t_f \to \infty$. The results in this section focus on just the ground state, but similar results can be derived for any excited states, and those results can be simultaneously applicable.

The methods in this section closely follow the results of Yi and Crosson [30] who themselves draw inspiration from [5] and [31]. Specifically, this result can be thought of as a modification of their Proposition 1 (proven in their Appendix F) to the setting where the underlying annealing curve has an oscillatory structure. In practice, this modification is exactly the same as the modification to the usual Trotter operator error formula, meaning we can recover the oscillatory enhancement and still have the improved scaling analysis of Yi and Crosson.

As a reminder, the control function is

$$u(t) = u_0(t/t_f) + c(t, t_f), \tag{C10}$$

where $u_0(s)$ is a smooth monotonically decreasing function, and

$$c(t, t_f) = -c_0(t_f) \sin\left(\frac{2\pi}{\tau} t + \phi\right). \tag{C11}$$

We discretize our adiabatic evolution over time t_f into M steps with "short" timestep $\Delta t = t_f/M$:

$$\begin{aligned} &\hat{U}_1\left(t + \Delta t, t\right) \\ &= \exp\left(-i \int_t^{t+\Delta t} \mathrm{d}t' (u(t')\hat{B} + (1 - u(t'))\hat{C})\right) \end{aligned} \tag{C12}$$

We want to evaluate the integral

$$\int_t^{t+\Delta t} \mathrm{d}t' u(t') = \int_t^{t+\Delta t} \mathrm{d}t' \left(u_0(\frac{t}{t_f}) + c(t, t_f)\right) \tag{C13}$$

For our purposes, we will assume that $u_0(t/t_f)$ is approximately constant over this interval, which is valid for large t_f. In other words, we assumed here that $u_0'(\frac{t}{t_f})$ is approximately constant over this small interval. We will also now introduce $s \equiv \frac{t}{t_f}$ as a normalized time. In this case

$$\begin{aligned} &\int_t^{t+\Delta t} \mathrm{d}t' u(t') \\ &= \Delta t u_0(s) + c_0 \frac{\tau}{2\pi} (\cos(\frac{2\pi}{\tau}(t + \Delta t) + \phi) - \cos(\frac{2\pi}{\tau} t + \phi)) \end{aligned} \tag{C14}$$

We can use trig identities to reduce this further to

$$\begin{aligned} \int_t^{t+\Delta t} \mathrm{d}t' u(t') &= \Delta t u_0(s) \\ &\quad - c_0 \frac{\tau}{\pi} \sin(\frac{\pi \Delta t}{\tau}) \sin(\frac{2\pi}{\tau}(t + \frac{\Delta t}{2}) + \phi) \end{aligned} \tag{C15}$$

We will now define $\sigma \equiv \tau/t_f$ and $\Delta s \equiv \Delta t/t_f$ so that

$$\int_t^{t+\Delta t} \mathrm{d}t' u(t') = \Delta t U(s) \tag{C16}$$

where for convenience, we have defined

$$U(s) \equiv u_0(s) - c_0 \frac{\sigma}{\Delta s \pi} \sin\left(\frac{\pi \Delta s}{\sigma}\right) \sin\left(\frac{2\pi}{\sigma}(s + \frac{\Delta s}{2}) + \phi\right) \tag{C17}$$

In the limit of $\Delta s \to 0$, this just reduces to $U(s) \to u(s)$.

To recover the original results of [30], take $u_0 = 1 - s$ for a linear ramp and $c_0 = 0$ for no oscillations. This is part of the discretization error that we will later assume is smaller than the adiabatic error.

It is also important to note that $c_0 \ll u_0'(s)$ (specifically elsewhere we found that $c_0 \in \mathcal{O}(\dot{u}_0^2)$). This means that the sinusoids can be counted as a correction to the $u_0(s)$ terms rather than their own term. This will be useful when bounding quantities since we can then treat these two as a whole rather than seperate quantities to bound

Each such discrete unitary is then Trotterized to first order:

$$\begin{aligned} \hat{U'}_1\left(t + \Delta t, t\right) = &\exp\left(-i \Delta t U(s) \hat{B}\right) \\ &\times \exp\left(-i \Delta t (1 - U(s)) \hat{C}\right), \end{aligned} \tag{C18}$$

we define the effective Hamiltonian for this Trotterized evolution by

$$\tilde{H}(t) = i \log\left(\hat{U'}_1\left(t + \Delta t, t\right)\right) / \Delta t. \tag{C19}$$

In the limit of $\Delta t \to 0$, the discretization step size, there is a continuous Hamiltonian defined by this. This effective Hamiltonian has the nice property that $\tilde{H}(0) = \hat{B}$ and $\tilde{H}(t_f) = \hat{C}$, so evolution under $\tilde{H}(t)$ for slow t_f can be described as an adiabatic process. The bang-anneal-bang curves approach an adiabatic procedure with vanishing initial and final bangs in the large t_f limit, so this is appropriate in our setting for large t_f (corresponding to large p for QAOA).

The core idea of this method then is to bound the error on the evolution, not using operator errors but using the adiabatic theorem directly. This will result in tighter scaling in terms of the number of Trotter or QAOA slices p but will introduce scaling with the p independent spectral gap of $\tilde{H}(t)$. This technique will use the adiabatic theorem of Jansen, Ruskai, and Seiler [5].

We split the effective Hamiltonian into two parts such that

$$e^{-i\Delta t \tilde{H}(s)} = \left(e^{-i\Delta t U(s)\hat{B}} e^{i\Delta t U(s)\hat{C}}\right) e^{-i\Delta t \hat{C}} \tag{C20}$$

$$\equiv e^{-i\Delta t \hat{G}(U(s))} e^{-i\Delta t \hat{C}}. \tag{C21}$$

To reiterate our ultimate goal, we would want to show that there is an enhancement when $\Delta t = \tau$. As we will

see, this goal is not consistent with the assumptions of this method, which we will discuss later in this section.

In order to utilize the Adiabatic condition bounds of [5], it is necessary to compute the matrix norms of derivatives of $\tilde{H}(s)$ with respect to s. Using Magnus expansion techniques, [30] bounds the norms of these derivatives by

$$||\frac{d}{d\,s}\tilde{H}|| \leq ||\frac{d}{d\,s}\hat{G}||\ \mathcal{F}_1(2\Delta t||\hat{C}|| + 2\Delta t||\hat{G}||) \tag{C22}$$

$$||\frac{d^2}{d\,s^2}\tilde{H}|| \leq ||\frac{d^2}{d\,s^2}\hat{G}||\ \mathcal{F}_1(2\Delta t||\hat{C}|| + 2\Delta t||\hat{G}||) \tag{C23}$$
$$+ 2\Delta t||\frac{d}{d\,s}\hat{G}||^2\ \mathcal{F}_0(2\Delta t||\hat{C}|| + 2\Delta t||\hat{G}||)$$

Here the $\mathcal{F}$ functions are defined to be

$$\mathcal{F}_0(x) = \sum_{j=0}^{\infty} x^j = \frac{1}{1-x} \tag{C24}$$

$$\mathcal{F}_1(x) = \sum_{j=1}^{\infty} \frac{x^{j-1}}{j} = -\ln(1-x)/x \tag{C25}$$

$$\mathcal{F}_2(x) = \sum_{j=2}^{\infty} \frac{x^j}{j^2} = -\int_0^x dx' \ln(1-x'). \tag{C26}$$

We present both norm derivatives here, but since our results only effect the constant prefactors and not overall scaling, it will be enough to keep track of $||\frac{d}{d\,s}\tilde{H}||$ because our results do not alter the dominant term in [30] which depends on $||\frac{d}{d\,s}\tilde{H}||$.

The next step in this process is to bound the norms of the derivatives of $\hat{G}$. This is possible by looking at

$$\frac{d}{d\,s}e^{-i\Delta t\hat{G}(U(s))} = \tilde{E}(s,\Delta t)e^{-i\Delta t\hat{G}(U(s))} \tag{C27}$$

where

$$\tilde{E}(s,\Delta t) = i\Delta t\frac{d\,U}{d\,s}\,e^{i\Delta t\,U(s)\hat{B}}(\hat{B}-\hat{C})e^{-i\Delta t\,U(s)\hat{C}} \tag{C28}$$

The $\frac{d\,U}{d\,s}$ is the only new portion of our results compared to [30], and to recover their results exactly, you would need to set $\frac{d\,U}{d\,s} \to 1$. In our setting, this derivative is just

$$\frac{d\,U}{d\,s} = u_0'(s) - \frac{2c_0}{\Delta s}\cos\left(\frac{2\pi}{\sigma}(s+\frac{\Delta s}{2})+\phi\right)\sin\left(\frac{\pi\Delta s}{\sigma}\right). \tag{C29}$$

Again in the limit of small Δs, this reduces to just

$$\frac{d\,U}{d\,s} \to \frac{d\,u}{d\,s} = u_0'(s) - \frac{2\pi c_0}{\sigma}\cos\left(\frac{2\pi}{\sigma}s+\phi\right). \tag{C30}$$

A nice feature of Eq. C27 is that it has an exact solution in the form of the Magnus Expansion. The terms in the Magnus expansion can then be bounded as in [30], and we follow a similar bounding but now keeping track of $\frac{d\,U}{d\,s}$. For instance, we can work out that the first two derivatives of $\hat{G}$ can have their norms bounded by

$$||\hat{G}|| \leq U(s)D_- + \frac{1}{2\Delta t}\mathcal{F}_2\left(2\Delta t D_- U(s)\right) \tag{C31}$$

$$||\hat{G}'|| \leq U'(s)D_- + \Delta t D_1 U(s)U'(s)\mathcal{F}_1\left(2\Delta t D_- U(s)\right) \tag{C32}$$

Here we define $D_- \equiv ||\hat{B}-\hat{C}||$, $D_0 \equiv ||\hat{C}||$, and $D_1 \equiv ||\left[\hat{B},\hat{C}\right]||$. Now, we can finally get back to the bound on $||\tilde{H}'||$ from Eq. C22.

$$\begin{aligned}||\tilde{H}'|| &\leq ||\hat{G}'||\ \mathcal{F}_1(2\Delta t||\hat{C}|| + 2\Delta t||\hat{G}||) \\ &\leq \left[U'(s)D_- + \Delta t D_1 U(s)U'(s)\mathcal{F}_1\left(2\Delta t D_- U(s)\right)\right] \\ &\quad \times \mathcal{F}_1\left(2\Delta t D_0 + 2\Delta t U(s)D_- + \mathcal{F}_2\left(2\Delta t D_- U(s)\right)\right)\end{aligned} \tag{C33}$$

These functions can be bounded if their arguments are $x < 1/2$

$$\mathcal{F}_1(x) \leq 1 + x \tag{C34}$$

$$\mathcal{F}_2(x) \leq \frac{x^2}{2}(1+x) \tag{C35}$$

This allows us to bound

$$\begin{aligned}||\tilde{H}'|| \leq\ & \left[U'(s)D_- + \Delta t D_1 U(s)U'(s)\left(1 + 2\Delta t D_- U(s)\right)\right] \\ & \times \Bigg[1 + 2\Delta t D_0 + 2\Delta t U(s)D_- \\ & + \frac{\left(2\Delta t D_- U(s)\right)^2}{2}\left(1 + 2\Delta t D_- U(s)\right)\Bigg]\end{aligned} \tag{C36}$$

We make the same assumption as Yi and Crosson that D_0, $D_- \in \mathcal{O}(n)$, $D_1 \in \mathcal{O}(n^2)$. The other essential assumption here is that $\Delta t \in \mathcal{O}(n^{-1})$ in order to make the arguments of the $\mathcal{F}$ functions small. Unfortunately, this assumption is extremely hard to justify in our QAOA setting since we observe that $\Delta t \approx \tau$. As discussed elsewhere in this paper, τ is proportional to the spectral gap of the problem, and the spectral gap is $\mathcal{O}(1)$ in most reasonable problems with it dipping down into polynomially small and exponentially small values during phase transitions. At the moment, we will still assume that $\Delta t \in \mathcal{O}(n^{-1})$, but this is the point in the argument where this method breaks down in our setting.

With these assumptions, the second term is proportional to $\mathcal{O}(1)$ and so we drop it. On the other hand, these assumptions mean that the first term is $\mathcal{O}(n)$.

With these assumptions, $2\Delta t D_- U(s) \leq 1/2$ and so

$$||\tilde{H}'|| \leq \left[U'(s)D_- + \frac{3}{2}\Delta t D_1 U(s)U'(s)\right]. \tag{C37}$$

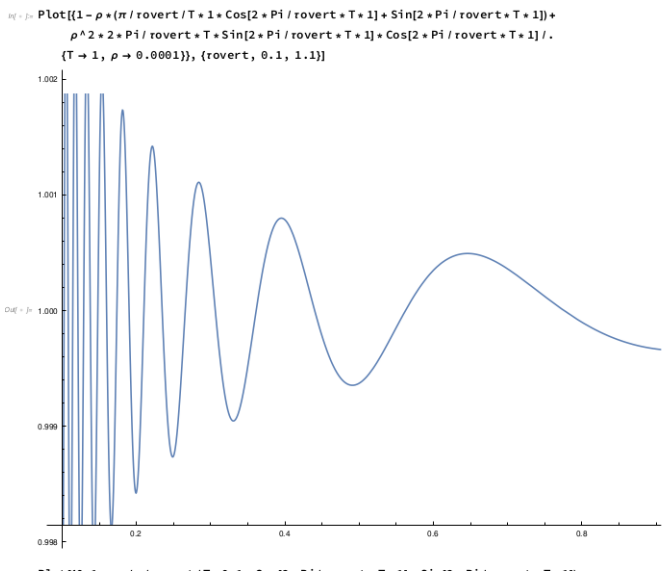


FIG. 8:

Finally,

$$\begin{aligned}
\epsilon_{\text{tro}} \leq & \epsilon'_{\text{adb}} + \epsilon'_{\text{tot}} && \text{(C38)}\\
\leq & \epsilon_{\text{adb}} + \epsilon_{\text{tot}} && \text{(C39)}\\
= & \mathcal{O}(\epsilon_{\text{tot}}) && \text{(C40)}\\
= & \mathcal{O}(\mathcal{G}(t_f, \tilde{H})) && \text{(C41)}\\
\leq & \mathcal{O}\left(\frac{1}{t_f \lambda^3} ||\tilde{H}'(s)||^2\right) && \text{(C42)}\\
= & \mathcal{O}\left(\frac{U'(s) D_-}{t_f \lambda^3}\right) + \mathcal{O}\left(\frac{(\Delta t D_1 U(s) U'(s))^2}{t_f \lambda^3}\right) && \text{(C43)}\\
= & \mathcal{O}\left(\frac{U'(s) D_-}{t_f \lambda^3}\right) + \mathcal{O}\left(\frac{t_f D_1^2 U(s)^2 U'(s)^2}{p^2 \lambda^3}\right), && \text{(C44)}
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{G}(t_f, \tilde{H}) = & \frac{1}{t_f}\left(\frac{\|\tilde{H}'(0)\|}{\tilde{\lambda}(0)^2} + \frac{\|\tilde{H}'(1)\|}{\tilde{\lambda}(1)^2}\right) && \text{(C45)}\\
& + \frac{1}{t_f}\int_0^1 \mathrm{d}s \left(\frac{\|\tilde{H}''(s)\|}{\tilde{\lambda}(s)^2} + 7\frac{\|\tilde{H}'(s)\|^2}{\tilde{\lambda}(s)^3}\right)
\end{aligned}$$

and in the second line we used

$$\begin{aligned}
\epsilon_{\text{disc}} \equiv & |\epsilon'_{\text{adb}} - \epsilon_{\text{adb}}| \ll \epsilon_{\text{adb}}, && \text{(C46)}\\
\tilde{\epsilon}_{\text{disc}} \equiv & |\epsilon'_{\text{tot}} - \epsilon_{\text{tot}}| \ll \epsilon_{\text{tot}}. && \text{(C47)}
\end{aligned}$$

ϵ'_{adb} is the error from a finite time implementation of an adiabatic process, ϵ_{adb}, plus discretization error, ϵ_{disc}. ϵ'_{tot} is the error from a finite time implementation of an adiabatic process plus discretization error plus Trotter error. ϵ_{tot} is the error in doing the discrete and Trotterized evolution adiabatically. $\tilde{\epsilon}_{\text{disc}}$ is the error in discretizing and Trotterizing the adiabatic discretized and Trotterized process.

Plotting this second term produces minima in error at $\Delta t = \tau/M$ for $M \in \mathcal{Z}$, as plotted in Figure 8. This is very similar to what we found in the usual Trotter error in Figure 6.

This means that there are enhancements in the error bound when the Trotter slices evenly divides up the period of oscillation into an integer number of slices. Ideally based on our QAOA numerics, we expect this integer to be one, but as already statement, that violates the perturbative assumptions in this method. These assumptions go back to the use of the perturbative Magnus expansion. This method would allow for $\Delta t = \frac{\tau}{\mathcal{O}(n)}$ to achieve the enhancement, but that is not consistent with numerics for QAOA, at least at the system sizes currently accessible by classical simulators and current quantum devices.

Appendix D: Bang-Anneal-Bang Ansatz Algorithm

One of the leading problems with the bang-anneal-bang curves is how to actually construct them efficiently. These optimal curves always seem to have the same qualitative structure, but working out the exact shape and length of various features is key to implementing these schedules effectively. Formally, these schedules can be found via a gradient descent procedure, using the analytically constructed gradient $\Phi(t) = \frac{\delta\langle E(t_f)\rangle}{\delta u(t)}$. This gradient can be analytically constructed but requires information from experimentally inaccessible intermediate times, and numerically estimating this gradient can prove cumbersome for an entire continuous function.

To address these issues, we here present a variational algorithm that produces a good approximation of the bang-anneal-bang optimal path using. This algorithm will not produce the exact optimal procedure but will approximate it, and in our numerical trials it produces better results, given fixed time, than either QAOA or a simple, linear annealing schedule. The number of variational parameters can be adjusted depending on the available resources.

This algorithm is based off the insight that the asymptotic curve derived from QAOA angles is one-in-the-same with the base curve in the annealing region of the bang-anneal-bang curves. Specifically, if QAOA is parameterized in terms p layers with β_i, the angles for mixer $\hat{B}$ bangs, and γ_i, the angles for problem $\hat{C}$ bangs, then the asymptotic QAOA can be found in the large p limit by

$$v\left(\frac{i-1}{p-1}\right) = \frac{\beta_i}{\beta_i + \gamma_i} \qquad \text{(D1)}$$

This behavior was noted numerically in [7, 8]. The current work provides justification for the existence of these asymptotic curves and links them to the bang-anneal-bang protocols. Specifically [7] interpreted this $v(s)$ as an annealing curve which resulted in a good annealing procedure that actually captured well known effects from

diabatic quantum annealing. Our current algorithm is an improvement on this that captures even more of the structure and power of the optimal procedure.

In bang-anneal-bang protocols, this $v(s)$ has roughly the same functional form as the base curve $u_0(s)$ that determines the shape of the annealing region, up to a superposed oscillatory pattern. Furthermore, the period of that oscillatory pattern coincides with the length of the QAOA layers.

Therefore, it should be possible to use an existing QAOA procedure to get a good guess as to what the bang-anneal-bang procedure should look like. The initial and final bangs are not well captured by QAOA but can be inserted in later as variational parameters. Therefore, we propose the following hybrid variational algorithm for approximating the bang-anneal-bang curves.

1. Perform a normal QAOA procedure for large enough p to be able to identify the shape of $v\left(\frac{i-1}{p-1}\right)$. In practice, we have found that $p \sim 5$ can start to see the pattern with $p \sim 10-20$ clearly identifying the pattern.

2. Interpret the QAOA curve $v\left(\frac{i-1}{p-1}\right)$ as a smooth annealing region.

3. Create an ansatz for the bang-anneal-bang curve that is one $u = 0$ bang at the beginning, the annealing curve, defined by $v\left(\frac{i-1}{p-1}\right)$ in the middle, and one $u = 1$ bang at the end. Furthermore, superpose an oscillatory curve $c(t) = A(t)\sin(\omega(t)t + \phi)$ onto the annealing region so that this region is described by $v(t) + c(t)$.

   - The lengths of the initial, $\tilde{\gamma}$, and final, $\tilde{\beta}$, bangs are variational parameters.
   - There are multiple ways to parameterize the interior anneal:
     - The length of the anneal can be fixed to be the same as the time the QAOA procedure took minus the bang lengths, $T_{QAOA} - \tilde{\gamma} - \tilde{\beta}$; the frequency of oscillations can be chosen to be $\omega(t) = 2\pi p/T_{QAOA}$; and the amplitude of the oscillations $A(t) = A$ is taken to be a variational parameter.
     - The length of the anneal can be fixed to be $T_{QAOA} - \tilde{\gamma} - \tilde{\beta}$; the frequency of oscillations $\omega(t) = \omega$; and amplitude of oscillations $A(t) = A$ are taken to be static variational parameters.
     - The length of the anneal can be fixed to be $T_{QAOA} - \tilde{\gamma} - \tilde{\beta}$; the frequency of oscillations $\omega(t)$ is chosen to be a variable function so that the period of a given oscillation matches the length of the corresponding QAOA layer. The amplitudes of oscillation can either be fixed to be the same or treated as seperate variational parameters in each oscillation.
     - The length of the anneal, T_A, can be treated as a variational parameter[**?**]. The frequency can be taken as fixed $\omega(t) = 2\pi p/T_{QAOA}$ or allowed to vary as a free fitted parameter as in previous versions. The amplitude of oscillation is a single variational parameter or can be binned into different regions with the amplitude in each region being treated as a variational parameter.
     - Adjust this ansatz as suits the system at hand and how many variational parameters the specific setting is capable of handling.
   - Based on analytics, the optimal ϕ should be 0, but for optimization purposes it might be beneficial to treat this phase as a variational parameter as well.
   - In the end, this procedure will result in an ansatz with at least three ($\tilde{\beta}$, $\tilde{\gamma}$, and A), but possibly more, variational parameters.

4. Using the constructed anstaz, run a variational algorithm to determine the optimal values of the selected variational parameters, attempting to optimize with respect to the final energy of the state with respect to $\hat{C}$.

This procedure will always produce a better protocol than just interpreting $u(s) = v(s)$, and the number of variational parameters can be small. The most intensive part of this from a variational standpoint is the initial QAOA procedure to discover the shape of $v(s)$. Given the asymptotic nature of this curve, it is possible to find $v(s)$ for a given p (corresponding to a QAOA procedure that takes time t_f) and then scale it up into a bang-anneal-bang ansatz for a larger t_f.

Since the base annealing curve is related asymptotically to an optimized adiabatic schedule, it could be possible to use insight from the adiabatic limit to bypass the QAOA step entirely and create an ansatz for $v(s)$ a priori. For instance, in the unstructured search problem, it could be possible to use Roland & Cerf's [23] optimized adiabatic annealing schedule as a guess for the $u_0(t)$ base curve. If knowledge of the spectral gap is available, similar curves could be constructed for other problem instances.

Below we present some of the results for this algorithm, simulated on a classical computer, using direct evaluation of the Schrödinger equation. Three different levels of the above algorithm are used. The first listed as "Basic Interpolation" is the form used by [7] where the QAOA derived asymptotic curve $v(s)$ is just interpreted as an annealing curve. The second, "Sine Interpolation," superposes on

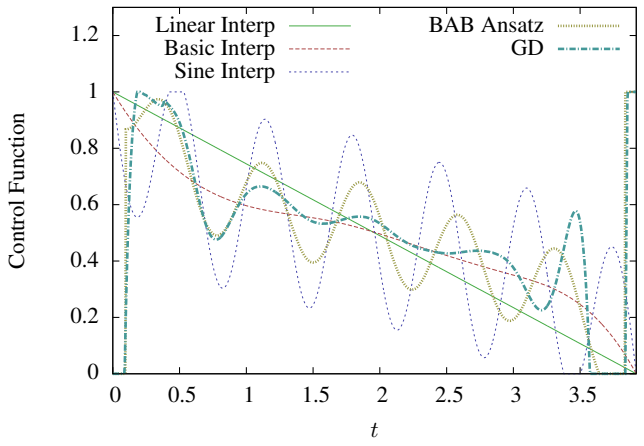


FIG. 9: Various versions of our algorithm, with "BAB Ansatz" being the most advanced in terms of number of parameters. The BAB Ansatz vaguely follows the optimal procedure found using gradient descent, "GD." The energies resulting from each procedure can be found in Fig. 10

Model	Energy
Linear	-7.743
Basic	-7.970
Sine	-8.178
QAOA	-8.322
BAB	-8.366
GD	-8.417
Ground	-8.836

FIG. 10: Energies related to the procedures shown in Fig. 9. The more information is used in constructing the ansatz, the closer we get to the optimal energy possible for this time. Notably, our full BAB Ansatz is required to perform better than the underlying QAOA procedure used to kickstart this procedure. From our experience, this ordering is fairly representative of the relative qualities of the algorithms with BAB outperforming QAOA but not quite reaching the quality of the Optimal procedure.

top of this a sine curve whose period is equal to the average duration of a QAOA layer. Finally, "BAB Ansatz" is our full Bang-Anneal-Bang Ansatz, here treating $\tilde{\gamma}$ (the initial bang), $\tilde{\beta}$ (the final bang), ω (the frequency of the oscillations), A (the amplitude of the oscillations), and ϕ (the phase of the oscillations) to all be variational parameters. For all procedures, the time allotted is t_f, the same as the time that the QAOA procedure took.

Fig. 9 shows the first example. The kickstarting QAOA procedure used $p = 6$, and also shown are a basic linear ramp, and the exact optimal bang-anneal-bang procedure found via a gradient descent ("GD") procedure. The resulting energies of all the relevant procedures are summarized in the accompanying table, Fig. 10

Note that we have attempted to run this algorithm without the initial QAOA procedure to find an estimate for $v(s)$ and instead using just a linear ramp for $v(s)$. The resulting procedure performs poorly compared to any of our procedures employing the QAOA derived asymptotic curve.

Appendix E: Adding in Energy Scaling

The quantity we actually care about bounding is

$$\begin{aligned}\Delta E(t_f,p) &= \langle\psi(0)|\,\hat{U}_{Tr}^\dagger(t_f,0)\,\hat{C}\,\hat{U}_{Tr}(t_f,0)\,|\psi(0)\rangle \\ &\quad - \langle\psi(0)|\,\hat{U}^\dagger(t_f,0)\,\hat{C}\,\hat{U}(t_f,0)\,|\psi(0)\rangle \qquad \text{(E1)}\\ \implies |\Delta E| &= \|\hat{U}_{Tr}^\dagger\hat{C}\hat{U}_{Tr} - \hat{U}^\dagger\hat{C}\hat{U}\|\,|\langle\Psi|\Psi\rangle| \qquad \text{(E2)}\end{aligned}$$

where $\|\cdot\|$ denotes the operator norm ($\sup\{\|\hat{U}\,|\psi\rangle\,\| : |\langle\psi|\psi\rangle| = 1\}$) and $|\langle\psi|\psi\rangle| = 1$.

$$\begin{aligned}&\|\hat{U}_{Tr}^\dagger\hat{C}\hat{U}_{Tr} - \hat{U}^\dagger\hat{C}\hat{U}\\ =&\frac{1}{2}\|(\hat{U}_{Tr}^\dagger + \hat{U}^\dagger)\hat{C}(\hat{U}_{Tr} - \hat{U}) + (\hat{U}_{Tr}^\dagger - \hat{U})\hat{C}(\hat{U}_{Tr} + \hat{U})\|\\ \le&\frac{1}{2}\Big[(\|\hat{U}_{Tr}^\dagger\| + \|\hat{U}^\dagger\|)\|\hat{C}\|\|\hat{U}_{Tr} - \hat{U}\| \qquad \text{(E3)}\\ &+ \|\hat{U}_{Tr}^\dagger - \hat{U}^\dagger\|\|C\|(\|\hat{U}_{Tr}\| + \|U\|)\Big]\\ \le&\Big[(1+1)\|C\|\|\hat{U}_{Tr} - \hat{U}\| + \|\hat{U}_{Tr} - \hat{U}\|\|C\|(1+1)\Big]\\ =&2\|C\|\|\hat{U}_{Tr} - \hat{U}\|.\end{aligned}$$

Therefore,

$$|\Delta E| \le 2\|C\|\|\hat{U}_{Tr} - \hat{U}\|. \qquad \text{(E4)}$$

In Eq. E3 we used the identity

$$\begin{aligned}&(\hat{U}_{Tr}^\dagger + \hat{U}^\dagger)\hat{C}(\hat{U}_{Tr} - \hat{U}) + (\hat{U}_{Tr}^\dagger - \hat{U})\hat{C}(\hat{U}_{Tr} + \hat{U})\\ =&\hat{U}_{Tr}^\dagger\hat{C}\hat{U}_{Tr} - \hat{U}^\dagger\hat{C}\hat{U} + \hat{U}^\dagger\hat{C}\hat{U}_{Tr} \qquad \text{(E5)}\\ &- \hat{U}_{Tr}^\dagger\hat{C}\hat{U} + \hat{U}_{Tr}\hat{C}\hat{U}_{Tr} - \hat{U}^\dagger\hat{C}\hat{U} - \hat{U}^\dagger\hat{C}\hat{U}_{Tr} + \hat{U}_{Tr}^\dagger\hat{C}\hat{U}\\ =&2(\hat{U}_{Tr}\hat{C}\hat{U}_{Tr} - \hat{U}^\dagger\hat{C}\hat{U}).\end{aligned}$$

In Eq. E3 we used the properties that $\|\hat{U}\| \le 1$ for $\hat{U}$ unitary and $\|\hat{A}^\dagger\| = \|\hat{A}\|$.

Appendix F: Error from Non-Optimized Trotterization

Let $F(t_k, t_{k-1}) = U_1^{(1)}(t_k, t_{k-1})U_+(t_k, t_{k-1})$.

$$\frac{\partial F}{\partial t_k} = \frac{\partial U_-^{(+)}}{\partial t_k}U_+ + U_-^{(1)}\frac{\partial U_+}{\partial t_k}. \qquad \text{(F1)}$$

$$\frac{\partial U_-^{(+)}}{\partial t_k} \tag{F2}$$
$$= iB(t_k)\exp_-\left(i\int_{t_{k-1}}^{t_k} B(s)\mathrm{d}s\right)\exp_-\left(i\int_{t_{k-1}}^{t_k} A(s)\mathrm{d}s\right)$$
$$+ i\exp_-\left(i\int_{t_{k-1}}^{t_k} B(s)\mathrm{d}s\right)A(t_k)\exp_-\left(i\int_{t_{k-1}}^{t_k} A(s)\mathrm{d}s\right).$$

$$\frac{\partial U_+}{\partial t_k} = -i(A(t_k)+B(t_k))U_+. \tag{F3}$$

$$\frac{\partial F}{\partial t_k} = \cancel{iB(t_k)U_-^{(1)}U_+} + \cancel{iA(t_k)U_-^{(1)}U_+} + \left[\exp_-\left(i\int_{t_{k-1}}^{t_k} B(s)\mathrm{d}s\right), iA(t_k)\right]\exp_-\left(i\int_{t_{k-1}}^{t_k} A(s)\mathrm{d}s\right) \tag{F4}$$
$$- \cancel{i(A(t_k)+B(t_k))U_-^{(1)}U_+} - i[U_-^{(1)}, A(t_k)+B(t_k)]U_+$$
$$= \cancel{i\left[\exp_-\left(i\int_{t_{k-1}}^{t_k} B(s)\mathrm{d}s\right), A(t_k)\right]\exp_-\left(i\int_{t_{k-1}}^{t_k} A(s)\mathrm{d}s\right)U_+} \tag{F5}$$
$$- i\exp_-\left(i\int_{t_{k-1}}^{t_k} B(s)\mathrm{d}s\right)\left[\exp_-\left(i\int_{t_{k-1}}^{t_k} A(s)\mathrm{d}s\right), \cancel{A(t_k)} + B(t_k)\right]U_+$$
$$- i\left[\exp_-\left(i\int_{t_{k-1}}^{t_k} B(s)\mathrm{d}s\right), \cancel{A(t_k)} + \cancel{B(t_k)}\right]\exp_-\left(i\int_{t_{k-1}}^{t_k} A(s)\mathrm{d}s\right)U_+$$
$$= i\exp_-\left(i\int_{t_{k-1}}^{t_k} B(s)\mathrm{d}s\right)\int_{t_{k-1}}^{t_k}\mathrm{d}u\exp_-\left(i\int_{t_{k-1}}^{u} A(s)\mathrm{d}s\right)[B(t_k), iA(u)]\exp_+\left(-i\int_{t_k}^{u} A(s)\mathrm{d}s\right)U_+^{(1)}U_-^{(1)}U_+ \tag{F6}$$
$$= -\exp_-\left(i\int_{t_{k-1}}^{t_k} B(s)\mathrm{d}s\right)\int_{t_{k-1}}^{t_k}\mathrm{d}u\exp_-\left(i\int_{t_{k-1}}^{u} A(s)\mathrm{d}s\right)[-A(u), B(t_k)]\exp_+\left(-i\int_{t_k}^{u} A(s)\mathrm{d}s\right) \tag{F7}$$
$$\times\exp_+\left(-i\int_{t_{k-1}}^{t_k} A(s)\mathrm{d}s\right)\exp_+\left(-i\int_{t_{k_1}}^{t_k} B(s)\mathrm{d}s\right)F$$
$$= \exp_-\left(i\int_{t_{k-1}}^{t_k} B(s)\mathrm{d}s\right)\int_{t_{k-1}}^{t_k}\mathrm{d}u\exp_-\left(i\int_{t_{k-1}}^{u} A(s)\mathrm{d}s\right)[A(u), B(t_k)]\exp_+\left(-i\int_{t_{k-1}}^{u} A(s)\mathrm{d}s\right) \tag{F8}$$
$$\times\exp_+\left(-i\int_{t_{k-1}}^{t_k} B(s)\mathrm{d}s\right)F$$

where we used the identity

$$[e^A, B] = -\int_0^1 \mathrm{d}s e^{(1-s)A}[B, A]e^{sA}. \tag{F9}$$

Integrating the last equation and using $F(t_{k-1}, t_{k-1}) = 1$ yields $F(t_k, t_{k-1}) = \int_{t_{k-1}}^{t_k} C(v, t_{k-1})F(v, t_{k-1})\mathrm{d}v$ which produces

$$\|U_+(t_k, t_{k-1}) - U_+^{(1)}(t_k, t_{k-1})\| \tag{F10}$$
$$\leq \int_{t_{k-1}}^{t_k}\mathrm{d}v\|C(v, t_{k-1})\| \tag{F11}$$
$$\leq \int_{t_{k-1}}^{t_k}\mathrm{d}v\int_{t_{k-1}}^{v}\mathrm{d}u\|[A(u), B(v)]\|. \tag{F12}$$

Appendix G: Phase Cancellation Method

We start at time t with $1-c \approx 1$ of our population in the ground state and $c \ll 1$ in the first excited state.

Assuming that we choose δt small enough such that $A(t+\delta t)+B(t+\delta t)-A(t)-B(t)$ is small, it follows from perturbation theory,

$$\begin{aligned} |0(t+\delta t)\rangle &= (1-\mathcal{O}(\delta t))\,|0\rangle_t + \mathcal{O}(\delta t)e^{i\phi_0\delta_t}\,|1\rangle_t \\ &\quad +\mathcal{O}(\delta t^2), \end{aligned} \tag{G1}$$

and

$$\begin{aligned} |1(t+\delta t)\rangle &= c\mathcal{O}(\delta t)\,|0\rangle_t + (c-c\mathcal{O}(\delta t))e^{i\phi_0\delta_t}\,|1\rangle_t \\ &\quad +\mathcal{O}(\delta t^2). \end{aligned} \tag{G2}$$

These are all written in the eigenbasis at time t (hence the subscripts t on the kets).

We relate the new eigenstates at time $t+\delta t$ to these eigenstates at time t by making the ansatz that nearest energy neighbors dominate coupling:

$$\begin{aligned} |0(t)\rangle_t &= (1-\mathcal{O}'(\delta t)\,|0\rangle_{t+\delta t} + \mathcal{O}'(\delta t)e^{i\alpha_0(t)\delta t}\,|1\rangle_{t+\delta t}\,, \\ &\quad +\mathcal{O}'(\delta t^2). \end{aligned} \tag{G3}$$

and

$$\begin{aligned} |1(t)\rangle_t &= \mathcal{O}'(\delta t)\,|0\rangle_{t+\delta t} + (1-\mathcal{O}'(\delta t))e^{-\alpha_1(t)\delta t}\,|1\rangle_{t+\delta t} \\ &\quad +\mathcal{O}'(\delta t)e^{i\alpha_1'(t)\delta t}\,|2\rangle_{t+\delta t} + \mathcal{O}'(\delta t^2). \end{aligned} \tag{G4}$$

Note that this is a purely spectral ansatz. It has no information about the population of the states.

Combining these two sets of equations together allows us to calculate the leakage from the ground state to the first excited state:

$$\begin{aligned} &\langle 1|_{t+\delta t}\,U(t,t+\delta t)\,|0\rangle_t \\ &= \langle 1|_{t+\delta t}\,|0(t+\delta t)\rangle \\ &= \mathcal{O}'(\delta t)e^{i\alpha_0(t)\delta t} + \mathcal{O}(\delta t)e^{i\alpha_1(t)\delta t + i\phi_0(t)\delta t} + \mathcal{O}(\delta t^2). \end{aligned} \tag{G5}$$

The same technique permits us to calculate the leakage from the first excited state to the second excited state:

$$\begin{aligned} &\langle 2|_{t+\delta t}\,U(t,t+\delta t)\,|1\rangle_t \\ &= \langle 2|_{t+\delta t}\,|1(t+\delta t)\rangle_t \\ &= c\mathcal{O}'(\delta t)e^{i\phi_1(t)\delta t + \alpha_1'(t)\delta t} + c\mathcal{O}(\delta t)e^{i\phi_1'(t)\delta t + i\alpha_2(t)\delta t} + \mathcal{O}(\delta t^2). \end{aligned} \tag{G6}$$

We now make the assumption that we are near-adiabatic during the full evolution from the mixer Hamiltonian A at time 0 to the problem Hamiltonian B at time τ. This means we only have to ever consider leakage to nearest-neighbor energies throughout the evolution.

For $0 \le \Delta t \le \tau$, this means that the probability of leaking from the ground state to the first excited state is,

$$P_{0\to1}(\Delta t) = \frac{1}{2}\left(1+\cos(\alpha_0\Delta t + \phi_0\Delta t - \alpha_1\Delta t\right), \tag{G7}$$

and the probability of leaking from the first to the second excited state is,

$$P_{1\to2}(\Delta t) = \frac{c}{2}\left(1+\cos(\phi_1\Delta t + \alpha_1'\Delta t - \phi_1'\Delta t - \alpha_2)\right). \tag{G8}$$

(I've assumed here that all these phases are constant with time for brevity's sake. More formally, they need to be written as time integrals over the 0 to Δt time interval.)

We know from perturbation theory that $\phi_0(t) = -\phi_1(t) = E_1(t) - E_0(t)$ and $\phi_1'(t) = E_0(t) - E_2(t)$. On the other hand, the alphas are related to how fast the *eigenstates* change during the evolution from 0 to τ. Since we are near the adiabatic regime, this rate of change is considerably smaller than $E_1 - E_0$ or $E_2 - E_1$: $|\alpha_0|, |\alpha_1|, |\alpha_1'| \ll |\phi_0|, |\phi_1|, |\phi_1'|$.

As a result, we can simplify the equations above to

$$P_{0\to1}(\Delta t) = \frac{1}{2}\left(1+\cos((E_1-E_0))\Delta t)\right), \tag{G9}$$

and

$$P_{1\to2}(\Delta t) = \frac{c}{2}\left(1+\cos\left((E_2-E_1)\Delta t\right)\right). \tag{G10}$$

To minimize leakage you want to follow the adiabatic curve when $P_{0\to1}(\Delta t) + P_{1\to2}(\Delta t) \approx 0$ and oscillate away from the adiabatic base curve when $P_{0\to1}(\Delta t) + P_{1\to2}(\Delta t) \gg 0$.

Since $P_{1\to2}(\Delta t)$ is linear in $c \ll 1$, it follows that the rate of this oscillation back and forth from the adiabatic base curve will be dominated by $P_{0\to1}(\Delta t)$. Thus, the frequency of this oscillation will be determined by the energy gap between the ground state and the first excited state.

Appendix H: Product Formula Perturbations

One important question is whether $\Delta_k = \tau_k$ is a true minimum or just an enhancement. In other words, is it beneficial to wiggle slightly away from this. Furthermore, if it is a true minimum, how much can we wiggle away before messing up the fact that we are in a minimum well.

The results in Sections V and **??** look only at the case where the product formula step size Δt_k matches up with the periods of the oscillations τ_k. What happens if we consider small perturbations such that

$$\Delta t_k = \tau_k + \epsilon \tag{H1}$$

for a small ϵ.

For the purpose of this section, we break the error, $||\hat{U}(0,\Delta t_k) - U_{PF}(0,\Delta t_k)||$, up into three portions. The first is just the base error due to u_0 alone without the superposed oscillations. This section will not consider this portion of the error because we are only concerned with the enhancement due to the oscillations. The second portion is the cross terms between $u_0(t)$ and the oscillations,

which we write as $\mathbb{E}_{CT}$. In the main portion of the paper, we found that $\mathbb{E}_{CT} = 0$. The last portion, due to just the behavior of the oscillatory portions, is $\mathbb{E}_{Os}$ which is what is responsible for the enhancement we see in the actual results.

Luckily, for both the single oscillation and the aggregate case, simple algebra shows that the cross term error, $\mathbb{E}_{CT}$, cancel out at the first order in ϵ as well. Therefore, those terms are consistent with being in an extremum. For the single oscillation case, the error due to the ϵ shift is (setting $\phi = 0$)

$$\mathbb{E}^{(1)}_{CT} = \left|\left|\left[\hat{B}, \hat{C}\right]\right|\right| \int_{-\infty}^{\infty} d\lambda \frac{iC_k\epsilon^2\left(1 - e^{2i\pi\lambda\tau_k}\right)\tilde{u}_0(\lambda)}{2\lambda\tau_k} + \mathcal{O}(\epsilon^3) \tag{H2}$$

When we look at the error in the cross terms for the entire procedure, it becomes

$$\begin{aligned}\mathbb{E}_{CT} = \left|\left|\left[\hat{B}, \hat{C}\right]\right|\right| \int_{-\infty}^{\infty} d\lambda \Bigg(& \frac{iC_0\epsilon^2\tilde{u}_0(t)}{2\lambda\tau(1 - e^{2\pi i\lambda\tau})} \\ & \times (e^{2i\pi\lambda\tau} + (2p-1)e^{2i\pi\lambda(p+1)\tau} - (2p+1)e^{2i\pi\lambda p\tau} + 1)\Bigg) \\ & + \mathcal{O}(\epsilon^3)\end{aligned} \tag{H3}$$

When we consider the case of a single oscillation, the correction from the oscillation, $\mathbb{E}_{Os}$, has no term that is linear in ϵ, meaning that $\Delta t_k = \tau_k$ is an exact minimum. To be exact this term comes out to be (again setting $\phi = 0$)

$$\mathbb{E}^{(1)}_{Os} \leq \left|\left|\left[\hat{B}, \hat{C}\right]\right|\right| \left(-\frac{C_k\tau_k^2}{2\pi} - \frac{\pi C_k\epsilon^3}{3\tau_k} + \mathcal{O}\left(\epsilon^4\right)\right). \tag{H4}$$

This result is mildly problematic because it means that this enhancement is not a minimum here but a higher order critical point. This problem gets fixed when we consider more oscillations and is a result just of the fact that we are considering a single oscillation here.

To understand how to fix this, we go instead to the aggregate case in the main text where all the oscillations are considered together (under the approximation that the frequency of the oscillations is constant). There, the expansion results in

$$\begin{aligned}\mathbb{E}_{Os} = \left|\left|\left[\hat{B}, \hat{C}\right]\right|\right| \Bigg(& -\frac{C_0 p\tau^2}{2\pi} \\ & + \frac{1}{6}\pi C_0 p\left(2p^2 - 3p + 1\right)\epsilon^2\Bigg) + \mathcal{O}(\epsilon^3)\end{aligned} \tag{H5}$$

Note that this is all before incoherent errors or any additional optimizations have been accounted for.

Appendix I: Bang-Bang Coherent Error Cancellation

We make the ansatz that for every pair of alternate bangs for total time Δt, the ground state and first excited state are still mostly themselves like in time-dependent perturbation theory:

$$\begin{aligned}|0(t+\Delta t)\rangle_t &= (1-\Theta)|0(t)\rangle_t + \Theta e^{i\phi_0(t)\Delta t}|1(t)\rangle_t \\ &\quad + \ldots. && \text{(I1)} \\ |1(t+\Delta t)\rangle_t &= \Theta e^{i\phi_1(t)\Delta t}|0(t)\rangle_t + (1-\Theta)|1(t)\rangle_t \\ &\quad + \Theta e^{i\phi_1'(t)\Delta t}|2(t)\rangle_t + \ldots. && \text{(I2)}\end{aligned}$$

We make the additional ansatz that the eigenstates before the Δt evolution are the same as those after but with additional nearest-neighbor overlap:

$$\begin{aligned}|0(t)\rangle_t &= (1-\Theta')|0(t)\rangle_{t+\Delta t} + \Theta' e^{i\alpha_0(t)\Delta t}|1(t)\rangle_{t+\Delta t} \\ &\quad + \ldots. && \text{(I3)} \\ |1(t)\rangle_t &= \Theta' e^{i\alpha_1(t)\Delta t}|0(t)\rangle_{t+\Delta t} + (1-\Theta')|1(t)\rangle_{t+\Delta t} \\ &\quad + \Theta' e^{i\alpha_1'(t)\Delta t}|2(t)\rangle_{t+\Delta t} + \ldots. && \text{(I4)}\end{aligned}$$

Substituting Eq. I3 & I4 into Eq. I1

$$\begin{aligned}&|0(t+\Delta t)\rangle_t && \text{(I5)} \\ &= (1-\Theta)[(1-\Theta')|0(t)\rangle_{t+\Delta t} + \Theta' e^{i\alpha_0(t)\Delta t}|1(t)\rangle_{t+\delta t}] \\ &\quad + \Theta e^{i\phi_0(t)\Delta t}[\Theta'|0(t)\rangle_{t+\Delta t} + (1-\Theta')e^{i\alpha_1\Delta t}|1(t)\rangle_{t+\Delta t}].\end{aligned}$$

This implies that

$$\begin{aligned}&\langle 1(t)|_{t+\Delta t} U(t, t+\Delta t)|0(t)\rangle_t \\ &= \langle 1(t)|_{t+\Delta t}|0(t+\Delta t)\rangle_t && \text{(I6)} \\ &= \Theta' e^{i\alpha_0\Delta t} + \Theta e^{i\alpha_1(t)\Delta t + i\phi_0(t)\Delta t}, && \text{(I7)}\end{aligned}$$

where I only kept the terms up to first order in Θ and Θ'. It then follows that

$$\begin{aligned}&P(t+\Delta t)_{0\to 1} \\ &= |\langle 1(t)|_{t+\Delta t} U(t, t+\Delta t)|0(t)\rangle_t|^2 && \text{(I8)} \\ &= \Theta^2 + \Theta'^2 && \text{(I9)} \\ &\quad + 2\Theta\Theta'\cos(\alpha_0(t)\Delta t + \phi_0(t)\Delta t - \alpha_1(t)\Delta t).\end{aligned}$$

Similarly, substituting Eq. I3 & I4 into Eq. I2

$$\begin{aligned}|1(t+\Delta t)\rangle_t &= \Theta e^{i\phi_1(t)\Delta t}[(1-\Theta')|0(t)\rangle_{t+\Delta t} && \text{(I10)} \\ &\quad + \Theta' e^{i\alpha_0(t)\Delta t}|1(t)\rangle_{t+\Delta t}] \\ &\quad + (1-\Theta)[\Theta' e^{i\alpha_1(t)\Delta t}|0(t)\rangle_{t+\Delta t} \\ &\quad + (1-\Theta')|1(t)\rangle_{t+\Delta t}].\end{aligned}$$

This implies that

$$\begin{aligned}&\langle 0(t)|_{t+\Delta t} U(t, t+\Delta t)|1(t)\rangle_t \\ &= \langle 0(t)|_{t+\Delta t}|1(t+\Delta t)\rangle_t && \text{(I11)} \\ &= \Theta e^{i\phi_1(t)\Delta t} + \Theta' e^{i\alpha_1(t)\Delta t}, && \text{(I12)}\end{aligned}$$

where again I only kept the terms up to first order in Θ and Θ'. It then follows that

$$\begin{aligned}&P(t+\Delta t)_{1\to 0} \\ &= |\langle 0(t)|_{t+\Delta t} U(t, t+\Delta t)|1(t)\rangle_t|^2 && \text{(I13)} \\ &= \Theta^2 + \Theta'^2 && \text{(I14)} \\ &\quad + 2\Theta\Theta'\cos(\alpha_1(t)\Delta t + \phi_1(t)\Delta t).\end{aligned}$$

In the case when time-dependent perturbation theory holds, $\phi_0(t) = -\phi_1(t) = E_1(t) - E_0(t)$. Here we expect some error $\delta E(t)$ such that $\phi_0(t) = -\phi_1(t) = E_1(t) - E_0(t) + \delta E(t)$.

$\alpha_0(t)\Delta t$ and $\alpha_1(t)\Delta t$ are some adiabatic energy phases which we expect to vary very slowly in time and so we can consider them to be near zero for some time interval.

As a result, Eqs. I9 and I13 for $P(t + \Delta t)_{0\to1}$ and $P(t + \Delta t)_{1\to0}$ are cosines that oscillate at a frequency of $E_1(t) - E_0(t) + \delta E(t)$. When the difference between $P(t + \Delta t)_{0\to1}$ and $P(t+\Delta t)_{1\to0}$ is maximally positive, you want to slow down how much you're varying the unitary with time and become more adiabatic so that you minimize leakage from the ground state to the first excited state. When the difference is maximally negative, you can get away with speeding up the rate of varying the unitary as this will mostly increase leakage back from the first excited state to the ground state.

The discrete bang-bang situation differs from the continuous case in possessing a different oscillation period because of the added error, $E_1(t) - E_0(t) + \delta E(t)$, and the unitaries corresponding to the pairs of bangs are constrained to be fixed for Δt, which means we can only evaluate Eqs. I9 & I13 for integer multiples m of Δt:

$$\begin{aligned} & P(t + m\Delta t)_{0\to1} \\ = \; & |\, \langle 1(t)|_{t+m\Delta t}\, U(t, t + m\Delta t)\, |0(t)\rangle_t \,|^2 \qquad \text{(I15)} \\ = \; & \Theta^2 + \Theta'^2 + 2\Theta\Theta' \cos((\alpha_0(t) + \phi_0(t) - \alpha_1(t))m\Delta t), \end{aligned}$$

and

$$\begin{aligned} & P(t + m\Delta t)_{1\to0} \\ = \; & |\, \langle 0(t)|_{t+m\Delta t}\, U(t, t + m\Delta t)\, |1(t)\rangle_t \,|^2 \qquad \text{(I16)} \\ = \; & \Theta^2 + \Theta'^2 + 2\Theta\Theta' \cos((\alpha_1(t) + \phi_1(t))m\Delta t). \end{aligned}$$

We know that $\Delta t \approx 2\pi(E_1 - E_0)^{-1}$ since we found that the pairs of bangs minimize discretization error (in a particular sense) when they have the same total time as one annealing oscillation period.

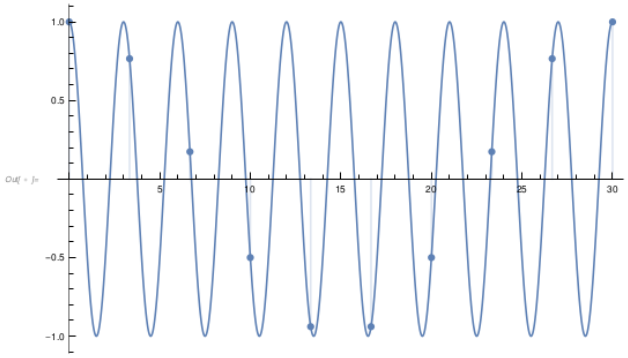
Hence, the period of oscillation of Eqs. I15 & I16 can be determined by finding the minimal integer m such that the difference in cosines has undergone one oscillation:

$$\begin{aligned} & min_{m\in\mathbb{Z}} |\phi_1(t) m\Delta t - 2\pi| \\ = \; & min_{m\in\mathbb{Z}} \left| m \frac{E_1(t) - E_0(t) + \delta E(t)}{E_1(t) - E_0(t)} - 1 \right|. \qquad \text{(I17)} \end{aligned}$$

The period of oscilation will then be $m_{\min}\Delta t$.

For instance, in the following example we set $E_1 - E_0 + \delta E = 2\pi/3$, $\delta E = 0.1 \times 2\pi/3$, $\alpha_0 = \alpha_1 = 0.001$ (a small value; we actually neglect them here but throw them in for the last two figures) and so $m_{\min} = 9$. In other words, it takes 9 $\Delta t = 2\pi/(E_1 - E_0 + \delta E)$ timesteps to see one oscillation from cosines oscillating with a period of $E_1 - E_0$:

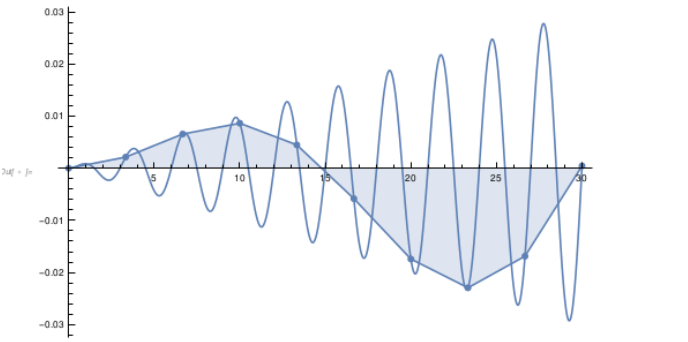
```
In[ ]:= Show[Plot[Cos[2 * Pi / 3 * t], {t, 0, 3 * 10}], (*E1-E0+δE=2*Pi/3*)
          DiscretePlot[Cos[2 * Pi / 3 * t], {t, 0, 3 * 10, 2 * Pi / 0.9 * 3 / 2 / Pi}]
          (*let δE=0.1*(E1-E0+δE) <=> Δt=2*Pi*1/(0.9*(E1-E0+δE))*)]
```



The underlying continuous anneal curve has actually oscillated 10 times during that one coarse bang-bang oscillation; $9\Delta t$ is the optimal period of oscillation for the bang-bang around the underlying anneal values.

In other words, if you examine the difference of the cosines at this sampling frequency of Δt, you will see it repeat every $9\Delta t$.

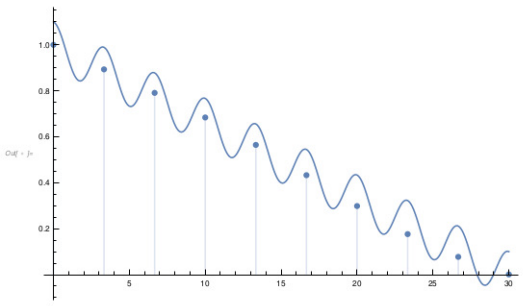
```
In[ ]:= Show[Plot[{Cos[α1 * t - α0 * t + ΔE * t] - Cos[α1 * t + ΔE * t]} /.
            {α1 → 0.001, α0 → 0.001, ΔE → 2 * Pi / 3}, {t, 0, 3 * 10}], (*ΔE=E1-E0+δE=2*Pi/3*)
          DiscretePlot[Cos[α1 * t - α0 * t + (ΔE) * t] - Cos[α1 * t + (ΔE) * t] /.
            {α1 → 0.001, α0 → 0.001, ΔE → 2 * Pi / 3, δE → 0.1 * 2 * Pi / 3},
           {t, 0, 3 * 10, 2 * Pi / 0.9 * 3 / 2 / Pi}, PlotMarkers → {Automatic}, Joined → True]]
```



Again, you want to slow down the bang-bang curve when the difference is maximally positive and speed it up when the difference is maximally negative.

This action on the curve is illustrated in the left plot below for the bang-bang approximation to a simple underlying "cartoon" $u(t)$ anneal oscillating curve. Since the overall anneal curve is decreasing, while the difference in cosines is increasing at first, you can effectively slow the anneal curve down and speed it up by adding the difference in the cosines to it to construct the optimal bang-bang curve:

```
Show[Plot[
  (1 - t / (3 * 10)) + 0.1 * Cos[(ΔE - δE) * t] /. {ΔE → 2 * Pi / 3, δE → 0.1 * 2 * Pi / 3}, {t, 0, 3 * 10}],
 DiscretePlot [(1 - t / (3 * 10)) + 0.1 * Cos[(ΔE - δE) * t + Pi / 2] +
   2 * (Cos[α1 * t - α0 * t + (ΔE) * t] - Cos[α1 * t + (ΔE) * t]) /. {α1 → 0.001, α0 → 0.001,
   ΔE → 2 * Pi / 3, δE → 0.1 * 2 * Pi / 3}, {t, 0, 3 * 10, 2 * Pi / 0.9 * 3 / 2 / Pi}]]
```



After this transformation, the bang-bang curve is above the anneal curve about 1/3rd of the way and below it about 2/3rd of the way. Squinting at the numerics in Fig. 7 (where we know the period of oscillating leakage from bang-bang to the first excited state is t_f), maybe this might also be happening.

You will always find solutions that look like this if $\delta E \ll E_1(t) - E_0(t)$ and $\alpha_0(t), \alpha_1(t) \approx 0$ over long enough time intervals.