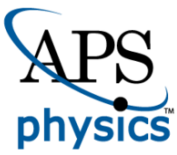

Optimal Protocols in Quantum Annealing and Quantum Approximate Optimization Algorithm Problems

Lucas T. Brady, Christopher L. Baldwin, Aniruddha Bapat, Yaroslav Kharkov, and Alexey V. Gorshkov

Phys. Rev. Lett. **126**, 070505 — Published 19 February 2021

DOI: 10.1103/PhysRevLett.126.070505

Optimal Protocols in Quantum Annealing and QAOA Problems

Lucas T. Brady,[1,2,*] Christopher L. Baldwin,[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*

(Dated: December 21, 2020)

Quantum Annealing (QA) and the Quantum Approximate Optimization Algorithm (QAOA) are two special cases of the following control problem: apply a combination of two Hamiltonians to minimize the energy of a quantum state. Which is more effective has remained unclear. Here we analytically apply the framework of optimal control theory to show that generically, given a fixed amount of time, the optimal procedure has the pulsed (or "bang-bang") structure of QAOA at the beginning and end but can have a smooth annealing structure in between. This is in contrast to previous works which have suggested that bang-bang (i.e., QAOA) protocols are ideal. To support this theoretical work, we carry out simulations of various transverse field Ising models, demonstrating that bang-anneal-bang protocols are more common. The general features identified here provide guideposts for the nascent experimental implementations of quantum optimization algorithms.

Introduction. The ongoing development of Noisy Intermediate Scale Quantum devices is guided by the question of how to leverage limited resources to best prepare the desired state of a system. Both Quantum Annealing (QA) [1, 2] and the Quantum Approximate Optimization Algorithm (QAOA) [3] use a combination of two Hamiltonians to try to prepare the ground state of one of the Hamiltonians. QA smoothly interpolates between the two Hamiltonians, whereas QAOA applies one or the other in sequence. It has previously been unclear which method, if either, is the most efficient. This question informs not only algorithm design but also hardware design, since near-term devices will be restricted in their capabilities.

Previous works [4–7] have applied the formalism of optimal control theory, in particular Pontryagin's Maximum/Minimum Principle [8], to this problem. It has been suggested on the basis of Pontryagin's principle that a "bang-bang" protocol, as in QAOA, is optimal [4]. Yet as we demonstrate, some of the assumptions behind this result are not true in general. We show that hybrid protocols consisting of both bang-bang and annealing segments are often best, backing up our analytic results with numerics for Ising models. Ref. [7] found an optimal bang-singular-bang [48] protocol for the unstructured search problem, an effective two-level system. That work can be thought of as a special case of our results, which apply for general Hamiltonians and generically see the singular region take on a smooth annealing structure.

QA is closely related to Adiabatic Quantum Computing [1], in which the Hamiltonian interpolates from a simple "mixer" to the desired "problem" Hamiltonian. The adiabatic theorem guarantees that the system, if initially in the ground state of the mixer Hamiltonian and deformed sufficiently slowly, will remain in the ground state throughout. QA generalizes this to allow for non-adiabatic protocols. Even in adiabatic regimes, optimization of the annealing schedule can potentially give polynomial speedups over both classical algorithms and unoptimized quantum schedules [10, 11], showing that the shape of the schedule is important for quantum advantage.

QAOA, on the other hand, applies the mixer and problem Hamiltonians alternately, using the timings of these pulses ("bangs") as variational parameters to be optimized over [3]. There is evidence that restricted forms of QAOA are more powerful than *adiabatic* protocols with the same restrictions [12] (although both are known to be quantum universal [13, 14]), but this does not address whether a non-adiabatic annealing procedure can outperform QAOA.

Both protocols are candidate algorithms for near-term quantum devices. Determining any features of the optimal protocol a priori would be a significant advantage in designing both the algorithms and the hardware they can run on in this era of integrated hardware and software design.

Optimal control theory [8] is well-suited to address such questions. It has long been used in a variety of physics and chemistry fields [15–25]. Applications to QA/QAOA are more recent, beginning with Ref. [4] and continuing with Refs. [5–7]. The QA/QAOA problem is distinct from the majority of quantum optimal control problems in that it has only one control function to be applied in a large Hilbert space [49]. As a result, unlike in standard quantum optimal control, the desired state typically cannot be prepared exactly in finite time. Some work has been done to examine the standard theory in this limit [26], but the results in this direction remain sparse.

In what follows, we first carefully articulate the control problem under consideration, pulling in standard ideas from optimal control and applying them to the QAOA/Annealing problem. We then prove general state-

ments about where bang-bang vs. annealing forms are preferred and what forms those will take in practice and finally present numerical results that support the analytic results. Our results also provide both a novel recasting of QAOA within the Quantum Annealing framework and an analysis of Quantum Annealing in the low time regime where it deviates heavily from the adiabatic dynamics that inspired it.

Control Problem. The problem which both QA and QAOA seek to solve is as follows: given Hamiltonians $\hat{B}$ and $\hat{C}$, with the system in the ground state of $\hat{B}$ at time 0, find the protocol $u(t)$ which minimizes the energy

$$J \equiv \langle x(t_f)|\hat{C}|x(t_f)\rangle, \tag{1}$$

where the time evolution of $|x(t)\rangle$ is given by

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}|x(t)\rangle &= -i\hat{H}(t)|x(t)\rangle, \\ \hat{H}(t) &\equiv u(t)\hat{B} + (1-u(t))\hat{C}. \end{aligned} \tag{2}$$

To avoid extreme protocols, we require that

$$u(t) \in [0,1] \ \ \forall t. \tag{3}$$

Our analytic results apply for any two Hamiltonians where the initial state is the ground state of $\hat{B}$ and the target state is the ground state of $\hat{C}$[50]. Often in Annealing/QAOA problems $\hat{C}$ (the problem Hamiltonian) is diagonal in the computational basis, while $\hat{B}$ (the mixer Hamiltonian) is a transverse field on the qubits. As stated above, QA assumes a smooth $u(t)$ whereas QAOA assumes that $u(t)$ is bang-bang.

To apply optimal control theory to this problem, we interpret Eq. (2) as a constraint relating $|x(t)\rangle$ to $u(t)$ and account for it by introducing a Lagrange multiplier $|k(t)\rangle$. Thus the cost function is modified to

$$\begin{aligned} J = &\langle x(t_f)|\hat{C}|x(t_f)\rangle \\ &+ \int_0^{t_f} \mathrm{d}t \, \langle k(t)| \left[-\frac{\mathrm{d}}{\mathrm{d}t} - i\hat{H}(t)\right] |x(t)\rangle + \text{c.c.}. \end{aligned} \tag{4}$$

Optimal control theory [8, 34] then uses calculus-of-variations techniques to derive necessary conditions for a minimum of J (in addition to Eq. (2)):

$$\frac{\mathrm{d}}{\mathrm{d}t}|k(t)\rangle = -i\hat{H}(t)|k(t)\rangle, \tag{5}$$

$$|k(t_f)\rangle = \hat{C}|x(t_f)\rangle, \tag{6}$$

and finally, for all *allowed* variations $\delta u(t)$ of the protocol,

$$\begin{aligned} \frac{\delta J}{\delta u(t)}\delta u(t) &\equiv \ \Phi(t)\delta u(t) \geq 0, \\ \Phi(t) &= \left[i\langle k(t)|\big(\hat{C}-\hat{B}\big)|x(t)\rangle + \text{c.c.}\right]. \end{aligned} \tag{7}$$

Note that Eq. (7), which is a form of Pontryagin's Minimum Principle [8] applied to this problem, can be satisfied at any given time in one of three ways: i) $\Phi(t) = 0$; ii) $\Phi(t) > 0$ and $u(t) = 0$; iii) $\Phi(t) < 0$ and $u(t) = 1$. The first possibility, that the functional derivative is 0 at the minimum of $J(t)$, is natural from a calculus-of-variations perspective. The latter two are legitimate only because $u(t)$ is restricted to be between 0 and 1, and Eq. (7) needs to hold merely for all *allowed* $\delta u(t)$. However, situations in which $\Phi(t) = 0$ for an extended interval have historically been referred to as "singular" [51]. Previous works have argued that such situations are uncommon in practice and thus that the optimal protocol must be of bang-bang form [4]. One of our key results is that singular regions are in fact quite natural, meaning that the exceptions noted in Ref. [4] are often the rule [52] .

If an optimal protocol has a smooth annealing form in some interval, then $\Phi(t)$ must equal zero in that interval. Correspondingly, if $\Phi(t)$ is non-zero in some interval, then the protocol must be of bang-bang form in that interval.

Just as one transforms the Lagrangian of a dynamical system into a Hamiltonian by the Legendre transform, one can construct from the cost function J a "control Hamiltonian" $\mathbb{H}$ (not to be confused with the system Hamiltonian $\hat{H}$). For our problem, the control Hamiltonian [53] evaluates to

$$\mathbb{H}(t) = i \, \langle k(t)| \, \hat{H}(t) \, |x(t)\rangle + \text{c.c.}. \tag{8}$$

The derivation of $\mathbb{H}$, as well as a proof that it is a constant in time, is carried out in the Supplementary Material [43].

Time constraints. It is important that we restrict to a fixed runtime t_f. If protocols are allowed arbitrarily long times, then the problem becomes trivial: the adiabatic theorem guarantees that any sufficiently slow protocol will end in the desired ground state. Since adiabatic protocols are often prohibitively inefficient, we constrain ourselves to more feasible runtimes. One way to do so is to simply fix t_f ("hard constraint"), as we have done. Another would be to allow protocols of varying t_f but include a penalty term λt_f ("soft constraint") in the action [Eq. (4)].

There is a useful connection between these two means of enforcing the time constraint which shows that they ultimately yield the same protocol. Furthermore, this connection gives physical meaning to the control Hamiltonian $\mathbb{H}$. The value of $\mathbb{H}$ in a hard-constraint problem with given t_f equals the value of λ needed in a soft-constraint problem for the optimal protocol to have the same runtime t_f. We prove this equivalence in detail in the Supplementary Material [44].

As a final point, we will assume that t_f is small enough that the desired ground state cannot be reached exactly, which is often the case in the setting of variational state preparation. Given enough time, the control problem is under-constrained and has several optimal solutions that may in general be hard to characterize. This assumption

implies that $\mathbb{H}$ is strictly positive, since some amount of penalty is needed for t_f to be the optimal runtime.

Initial and final bangs. We now show that any optimal protocol for our control problem must both begin and end with a bang. For some finite time interval at the beginning, the protocol must have $u(t) = 0$, and for another finite time interval at the end, it must have $u(t) = 1$.

To prove this, write $\Phi(t) = \Phi_C(t) - \Phi_B(t)$, where

$$\Phi_X(t) \equiv i\langle k(t)|\hat{X}|x(t)\rangle + \text{c.c}, \tag{9}$$

for any operator $\hat{X}$. Note that Φ_B, Φ_C, and thus Φ are continuous functions of time, as is clear from the continuity of $|x\rangle$ and $|k\rangle$ [see Eqs. (2) and (5)]. Also, the control Hamiltonian can be written as

$$\mathbb{H} = u(t)\Phi_B(t) + (1 - u(t))\Phi_C(t). \tag{10}$$

Consider the final portion of the protocol first. Eq. (6) gives $\Phi_C(t_f) = \text{Re}[i\langle x|\hat{C}^2|x\rangle] = 0$. Eq. (10) then gives $\Phi_B(t_f) > 0$ (remember that $\mathbb{H} > 0$), and thus $\Phi(t_f) < 0$. The continuity of $\Phi(t)$ then implies that $\Phi(t) < 0$ for a finite interval ending at t_f. We thus have that $u(t) = 1$ for a finite interval at the end of the protocol.

The initial portion of the protocol can be treated similarly, albeit with one additional step. Note that by Eqs. (2) and (5), the time derivative of $\langle k(t)|x(t)\rangle$ is 0, thus $\langle k(0)|x(0)\rangle = \langle k(t_f)|x(t_f)\rangle$. Since the system is assumed to initially be in the ground state of $\hat{B}$,

$$\begin{aligned}\Phi_B(0) &= \text{Re}\big[i\langle k(0)|\hat{B}|x(0)\big] \\ &\propto \text{Re}\big[i\langle k(0)|x(0)\rangle\big] = \text{Re}\big[i\langle k(t_f)|x(t_f)\rangle\big].\end{aligned} \tag{11}$$

Eq. (6) thus gives $\Phi_B(0) = 0$. Identical reasoning to above then shows that $\Phi(t) > 0$ and $u(t) = 0$ for a finite interval at the beginning of the protocol.

These results make sense heuristically. At the beginning of the protocol, the system is in an eigenstate of $\hat{B}$ and thus application of $\hat{B}$ does nothing to the state. While the final state is not exactly an eigenstate of $\hat{C}$, it should be close to the target state which is the ground state of $\hat{C}$. Therefore, roughly we might expect that similar logic about the futility of $\hat{C}$ at the end might hold, as our results indeed show.

In the Supplementary Material [45], we discuss how the lengths of the initial and final bangs vary with the parameters of the problem. In particular, we find that they become small as t_f increases, with the procedure approaching the monotonic annealing schedule typical of adiabatic quantum computing. Our results can be interpreted as saying that optimal quantum annealing for short timescales deviates from the monotonic ramp characteristic of the adiabatic regime [54].

The condition for a singular region is that $\Phi(t)$ and all its time derivatives must be zero for a finite interval of time. This condition can be achieved in a variety of ways. It is known in classical systems [34] that uncontrollability can lead to singularities, but our system is guaranteed to be able to reach the desired state due to the adiabatic theorem since the initial and final states are the ground states of the two Hamiltonians. Even without reachability problems, there are other forms of singularities that exist and impose conditions on the optimal $u(t)$. In the Supplement [46], we explain these conditions. Numerous options are possible, largely based on the structure of the control space and how the derivatives of $\Phi(t)$ are set to zero. We put particular emphasis on the simplest to derive condition which is the form we have seen for all singular regions numerically and is given by:

$$u^*(t) = \frac{\Phi_{[[\hat{B},\hat{C}],\hat{C}]}(t)}{\Phi_{[[\hat{B},\hat{C}],\hat{B}]}(t) - \Phi_{[[\hat{B},\hat{C}],\hat{C}]}(t)}. \tag{12}$$

While this equation describes how a singular region behaves, it does not specify whether the critical point determined by such a singular condition will be the global minimum, for which we turn to numerics.

Numerical results. The basic form and structure of the optimal protocol has been discussed for general Hamiltonians. The variety of singularity types and bang-bang possibilities in the middle region turns out to be problem dependent. To shed light on these concepts in practice, we here support our analytic results with extensive numerical simulations for transverse-field Ising models; though, our analytic results are valid for other models. The mixer and problem Hamiltonians are

$$\hat{B} = -\sum_{i=1}^{N} \hat{\sigma}_i^x, \tag{13}$$

$$\hat{C} = \sum_{ij} J_{ij}\hat{\sigma}_i^z\hat{\sigma}_j^z. \tag{14}$$

We examine a variety of different couplings J_{ij}: long-range antiferromagnets $J_{ij} \propto |i-j|^{-\alpha}$ (inspired by current experimental apparatuses [27–29]), instances of all-to-all spin glasses having every J_{ij} chosen randomly from $[-1, 1]$, and instances of the MaxCut problem on random 4-regular graphs (J_{ij} being the adjacency matrix of the graph).

All models studied show the same qualitative behavior: the optimal protocol begins and ends with a bang and in between has extended annealing portions (possibly punctuated by additional bangs). We term such protocols "bang-anneal-bang," although a more precise name would be "bang-*-bang", where "*" indicates any combination of annealing and bangs. For concreteness, we shall present the results obtained for the MaxCut problem, with numerics for other models presented in the Supplementary Material [47].

To find the optimal protocol, we discretize the time evolution in Eqs. (2) and (5), and apply gradient de-

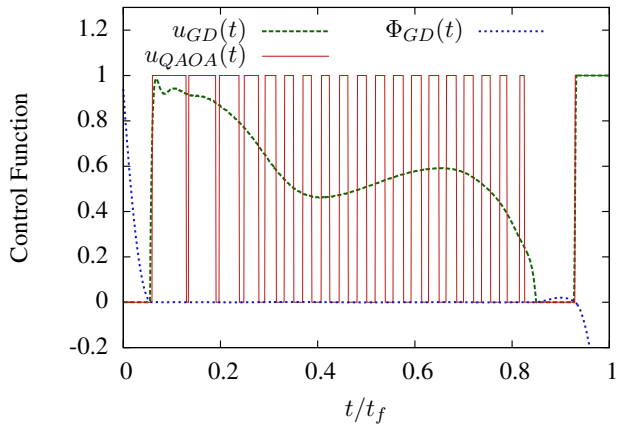

FIG. 1: Optimal control functions found through either gradient descent ($u_{GD}(t)$) or constrained-time QAOA ($u_{QAOA}(t)$) for a random instance of the MaxCut problem. Also shown is the gradient $\Phi_{GD}(t)$ for the gradient descent method. Parameters: $n = 8$ spins, total time $t_f = 2.0$, $2p = 40$ bangs for the QAOA method.

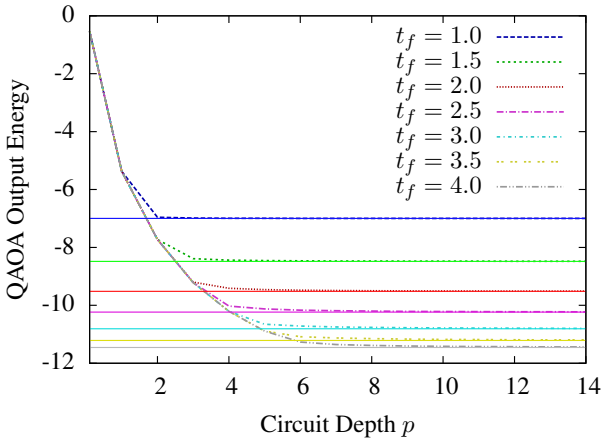

FIG. 2: QAOA output energy as a function of the circuit depth p (i.e., number of bangs), for an $n = 10$ MaxCut instance on a 4-regular graph. Dashed lines are the QAOA energies; solid horizontal lines are the energies obtained from gradient descent.

scent (specifically Nesterov's method [35]) to J as a functional of $u(t)$. Since $|x(0)\rangle$ is known, we first evolve forward in time to determine $|x(t)\rangle$, then compute $|k(t_f)\rangle$ through Eq. (6), then evolve *backwards* in time to determine $|k(t)\rangle$. The gradient descent could become trapped in a false minimum, so we perform multiple trials using different initial choices for $u(t)$. In practice, false minima appear to be rare, and most initial guesses found the optimum [55]. Fig. 1 shows a representative example of a protocol thus obtained, denoted $u_{GD}(t)$ (dashed green line), as well as the corresponding $\Phi_{GD}(t)$ (dashed blue line). As proven, it has bangs at the beginning and end. In the middle, either bangs or smooth anneals are possible based on our theoretical analysis. Numerically, we always find the middle region dominated by smooth anneals, possibly with an additional bang at the end as seen in Fig. 1. Also note the consistency between the behavior of $u_{GD}(t)$ and the sign of $\Phi_{GD}(t)$.

Fig. 1 also shows the result of QAOA for the same instance, using a fixed number of layers, $p = 20$, and optimizing over the length of each interval (with the sum constrained to be t_f). The bangs in the middle of the protocol, where gradient descent would produce an annealing segment, are significantly shorter than those at the beginning or end. This makes sense given the Suzuki-Trotter decomposition:

$$e^{-i(\beta\hat{B}+\gamma\hat{C})} = e^{-i\frac{\beta}{2p}\hat{B}}e^{-i\frac{\gamma}{p}\hat{C}} \cdot \left(e^{-i\frac{\beta}{p}\hat{B}}e^{-i\frac{\gamma}{p}\hat{C}}\right)^{p-1} e^{-i\frac{\beta}{2p}\hat{B}} + O\left(\frac{1}{p^2}\right). \quad (15)$$

A large number of short bangs serves as a reasonable approximation to an annealing segment. Fig. 1 suggests that QAOA is indeed attempting to approximate the bang-anneal-bang protocol found by gradient descent. Note that this behavior is only seen when QAOA is constrained to a fixed time but with increasing QAOA depth p. QAOA without the time constraint does not approach a Trotterization [29, 37].

For further evidence, Figs. 2 and 3 plot respectively the energy and "approximation quotient" of the QAOA output state as functions of the number of layers p. Here the approximation quotient is defined as $(E_{GD} - E_{QAOA})/E_{GD}$, where E_{GD} and E_{QAOA} are the output energies of gradient descent and QAOA. We see that the QAOA protocol performs worse than gradient descent, but with an error that decreases as p increases. Fitting the error to a power law $Cp^{-\nu}$, we find $\nu \approx 2.2$ for all t_f. This is reasonably consistent with the scaling expected from Eq. (15).

It should be noted that even though time-constrained QAOA approximates the bang-anneal-bang protocol, it may be the more effective approach in practice. QAOA has a much smaller parameter space to explore – the durations of the pulses as opposed to an entire function – and pulses may be simpler to implement experimentally than arbitrary combinations of Hamiltonians.

Conclusions. We have shown that for the control problem of minimizing the energy of a quantum state, the optimal protocol under time constraints is often of the bang-anneal-bang form. This shows that recent conjectures about the optimality of QAOA based on Pontryagin's principle are not as general as previously thought. Our results do not preclude bang-bang entirely, but they do provide the theoretical framework in which non-bang-bang protocols are possible. Furthermore, our numeric results indicate that these non-bang-bang protocols are extremely common for Ising models; though, there are

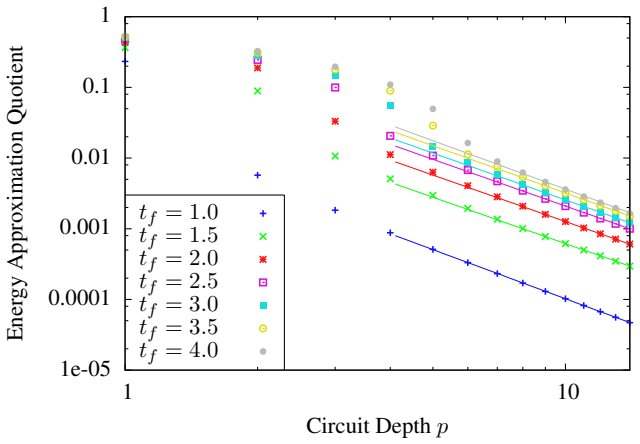

FIG. 3: Approximation quotient of QAOA output energy versus circuit depth p, for the same data as in Fig. 2. Solid lines are power-law fits, $Cp^{-\nu}$, to the last 5 data points in each curve. The fitted $\nu \approx 2.2$ for all t_f.

known non-Ising examples where bang-bang is still optimal [38–40]. Nonetheless, Pontryagin's principle and optimal control theory do serve as valuable tools. We have used them to prove that the optimal protocol must begin and end with a finite-length bang when not enough time is allowed for the desired state to be reached perfectly. Furthermore, optimal control theory provides guidance about the form of the protocol in the middle both analytically and numerically through the gradient, $\Phi(t)$.

Keep in mind that these results say nothing of the *practicality* of finding the optimal protocol. Since the algorithms require simulating the time evolution of the n-spin system, they are extremely expensive computationally on a classical computer. The main attraction of QAOA is that the time evolution required can be performed on a *quantum* computer, implemented experimentally on a real system [29, 41, 42]. The gradient descent method used in the present paper would be more difficult to implement on the same experimental setups that QAOA is typically implemented on; though other hardware and setups, such as [36] could utilize a similar gradient based method for optimization. It is obviously of great interest and utility to consider how one might better merge the tools of optimal control theory with current experimental capabilities.

Acknowledgments. We would like to thank P. Bienias and Z. C. Yang for helpful discussions. The numerics in this paper were done in part using code built upon QAOA code written by S. Jordan. The research of L.T.B. and C.L.B. was supported by National Institute of Standards and Technology (NIST) National Research Council (NRC) Research Postdoctoral Associateship Awards. The research was supported by the U.S. Department of Energy Award Number DE-SC0019449 (time-constrained optimization and analysis) and the U.S. Department of Energy Award Number DE-SC0019139 (quantum variational optimization).

* Electronic address: Lucas.Brady@nist.gov

[1] E. Farhi, J. Goldstone, S. Gutmann, M. Sipser, arXiv:0001106 (2000).
[2] T. Kadowaki, H. Nishimori, *Phys. Rev. E* **58**, 5355 (1998).
[3] E. Farhi, J. Goldstone, S. Gutman, arXiv:1411.4028 (2014).
[4] Z. C. Yang, A. Rahmani, A. Shabani, H. Neven, C. Chamon, *Phys. Rev. X*, **7**, 021027 (2017).
[5] A. Bapat, S. Jordan, arXiv:1812.02746 (2018).
[6] G. B. Mbeng, R. Fazio, G. Santoro, arXiv:1906.08948 (2019).
[7] C. Lin, Y. Wang, G. Kolesov, U. Kalabić, *Phys. Rev. A* **100**, 022327 (2019).
[8] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, E. F. Mishchenko, *The Mathematical Theory of Optimal Processes*, (Interscience Publishers, New York, 1962).
[9] S. Jansen, M. Ruskai, R. Seiler, *J. Math. Phys.* **48**, 102111 (2007).
[10] J. Roland, N. J. Cerf, *Phys. Rev. A* **65**, 042308 (2002).
[11] M. Jarret, B. Lackey, A. Liu, K. Wan, arXiv:1810.04686 (2018).
[12] E. Farhi, A. Harrow, arXiv:1602.07674 (2016).
[13] D. Aharonov, W. van Dam, J. Kempe, Z. Landau, S. Lloyd, O. Regev, Proc. 45th FOCS, p. 42-51 (2004).
[14] S. Lloyd, arXiv:1812.11075 (2018).
[15] H. Rabitz, R. de Vivie-Riedle, M. Motzkus, K. Kompa, *Science*, Vol. 288, 5467, pp. 824-828 (2000).
[16] J. Werschnik, E. K. U. Gross, *Journal of Phys. B*, Vol. 40, 18, pp. R175–R211 (2007).
[17] C. Brif, R. Chakrabarti, H. Rabitz, *New Journal of Phys.*, Vol. 12 (2010).
[18] J. P. Palao, R. Kosloff, *Phys. Rev. Lett.* **89**, 188301 (2002).
[19] J. P. Palao, R. Kosloff, *Phys. Rev. A* **68**, 062308 (2003).
[20] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrggen, S. J. Glaser, *Journal of Magnetic Resonance* 172, pp. 296-305 (2005).
[21] S. Montangero, T. Calarco, R. Fazio, *Phys. Rev. Lett.* **99**, 170501 (2007).
[22] A. V. Gorshkov, T. Calarco, M. D. Lukin, A. S. Sørensen, *Phys. Rev. A* **77**, 043806 (2008).
[23] M. Grace, C. Brif, H. Rabitz, I. A. Walmsley, R. L. Kosut, D. A. Lidar, *Journal of Physics B* **40**, 9 (2009).
[24] J. Cui, R. van Bijnen, T. Pohl, S. Montangero, T. Calarco, *Quantum Sci. Technol.* **2** 035006 (2017).
[25] A. Omran, H. Levine, A. Keesling, G. Semeghini, T. T. Wang, S. Ebadi, H. Bernien, A. S. Zibrov, H. Pichler, S. Choi, J. Cui, M. Rossignolo, P. Rembold, S. Montangero, T. Calarco, M. Endres, M. Greiner, V. Vuletić, M. D. Lukin, *Science*, Vol. 365, 6453, pp. 570-574 (2019).
[26] G. Riviello, K. M. Tibbetts, C. Brif, R. Long, R. Wu, T. Ho, H. Rabitz, *Phys. Rev. A* **91**, 043401 (2015).
[27] T. Koffel, M. Lewenstein, L. Tagliacozzo, *Phys. Rev. Lett.* **109**, 267203 (2012).
[28] J. Zhang, G. Pagano, P. W. Hess, A. Kyprianidis, P.

Becker, H. Kaplan, A. V. Gorshkov, Z. X. Gong, C. Monroe, *Nature* **551**, 601604 (2017).
[29] 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).
[30] O. L. Mangasarian, *J. SIAM Control*, Vol. 4, No. 1 (1966).
[31] R. Wu, R. Long, J. Dominy, T. Ho, H. Rabitz, *Phys. Rev. A* **86**, 013405 (2012).
[32] G. Riviello, C. Brif, R. Long, R. Wu, K. M. Tibbetts, T. Ho, H. Rabitz, *Phys. Rev. A* **90**, 013404 (2014).
[33] B. Russell, H. Rabitz, R. Wu, arXiv:1608.06198 (2016).
[34] D. E. Kirk, *Optimal Control Theory: An Introduction*, Dover Publications (1998).
[35] Y. Nesterov, *Introductory Lectures on Convex Optimization*, Springer (2004).
[36] I. Novikova, A. V. Gorshkov, D. F. Phillips, A. S. Sørensen, M. D. Lukin, R. L. Walsworth, *Phys. Rev. Lett.* **98**, 243602 (2007).
[37] L. Zhou, S. Wang, S. Choi, H. Pichler, M. D. Lukin, arXiv:1812.01041 (2018).
[38] A. Rahmani, T. Kitagawa, E. Demler, C. Chamon, *Phys. Rev. A* **87**, 043607 (2013).
[39] T. Karzig, A. Rahmani, F. von Oppen, G. Refael, *Phys. Rev. B* **91**, 201404(R) (2015).
[40] S. Bao, S. Kleer, R. Wang, A. Rahmani, *Phys. Rev. A* **97**, 062343 (2018).
[41] X. Qiang, X. Zhou, J. Wang, C. M. Wilkes, T. Loke, S. OGara, L. Kling, G. D. Marshall, R. Santagati, T. C. Ralph, J. B. Wang, J. L. OBrien, M. G. Thompson, J. C. F. Matthews, Nat. Photonics **12**, 534 (2018).
[42] J. S. Otterbach, R. Manenti, N. Alidoust, A. Bestwick, M. Block, B. Bloom, S. Caldwell, N. Didier, E. S. Fried, S. Hong, P. Karalekas, C. B. Osborn, A. Papageorge, E. C. Peterson, G. Prawiroatmodjo, N. Rubin, C. A. Ryan, D. Scarabelli, M. Scheer, E. A. Sete, P. Sivarajah, R. S. Smith, A. Staley, N. Tezak, W. J. Zeng, A. Hudson, B. R. Johnson, M. Reagor, M. P. da Silva, C. Rigetti, arXiv:1712.05771 (2017).
[43] Supplementary Material, Section S1
[44] Supplementary Material, Section S2
[45] Supplementary Material, Section S3
[46] Supplementary Material, Section S4
[47] Supplementary Material, Section S5
[48] The analysis specific to this system revealed that the middle region between two bangs should be a constant bang-like superposition of the two Hamiltonians.
[49] Even early work, e.g., Refs. [5, 7], on applying optimal control to QA/QAOA problems relied on small effective Hilbert spaces where the control analysis can take on a qualitatively different flavor.
[50] Oficially, the initial state only needs to be an eigenstate of $\hat{B}$, and the target state only needs to be an eigenstate of $\hat{C}$
[51] The terminology is due to the fact that, for problems linear in both the state $x(t)$ and control $u(t)$ without their product (unlike the Schrodinger equation), one can show that $\Phi(t)$ can only vanish in a finite-length interval if the control matrices are non-invertible.
[52] This point is obscured by the fact that some common classical systems cannot exhibit such singular regions and are always bang-bang [8, 34]. This is especially true of problems that are linear in both the control and the state and do not depend on their product.
[53] Eq. (7) is often expressed in terms of the control Hamiltonian, in which case singular regions are defined as those where $\delta\mathbb{H}/\delta u(t) = 0$. This is equivalent to the definition used here, since (see Eq. (10) below) $\delta\mathbb{H}/\delta u(t) = -\Phi(t)$.
[54] Based on prior work such as [10], we expect that even in the long-time regime, substantial optimization in the annealing schedule can lead to algorithmic improvements, and possibly, quantum advantage.
[55] all protocols shown were found starting from an initial guess of $u(t) = 0.5$