In this supplementary material we provide in depth derivations of the optimal control theory results used in the main text and explicate additional implications, details, and applications that would be of interest to experts and practitioners. In accordance with standard control theory, we present the derivation of the optimal control Hamiltonian from the more standard “Lagrangian” formulation in section S1, and in section S2 we show how the value of this Hamiltonian is constant and related to the type of time-constraint used in the main problem. Our main result from the text is that bang-anneal-bang protocols are common in QAOA/QA problems, and in section S3 we explore the lengths of the initial and final bangs, and these guaranteed bangs are shown to decrease in length as the runtime increases, and in section S4 various avenues for achieving the intermediate anneal from control singularities are derived and demonstrated for a test case. Finally in section S5 we present additional numeric verifications of our results from the main text.

S1. CONTROL HAMILTONIAN

In this section we will derive and explore the properties of the optimal control Hamiltonian $\mathbb{H}(t)$, presented in Eq. (9) of the main text. Much of this section will be the application of classical Control Theory [S1, S2] and the standard Calculus of Variations to the specific quantum problem at hand.

Consider a general quantity to be minimized (our $J$ or equivalently an action from Lagrangian Mechanics) given by

$$J = h(x(t_f), t_f) + \int_0^{t_f} dt g(x(t), \dot{x}(t), u(t), t)$$

where $x(t)$ is our state variable and $u(t)$ is our control variable, both possibly vectors. Our goal is to recast this in a traditional Lagrangian format and then perform a Legendre transform to get the corresponding Hamiltonian.

As a first step, note that the “action” can be recast as

$$J = \int_0^{t_f} dt \left[ g(x(t), \dot{x}(t), u(t), t) + \frac{d}{dt} h(x(t), t) \right] + h(x(0), 0),$$

where the final $h(x(0), 0)$ will be ignored from now on since our initial conditions are fixed and this term will not contribute to the minimization procedure:

$$J = \int_0^{t_f} dt \left[ g(x(t), \dot{x}(t), u(t), t) + \frac{d}{dt} h(x(t), t) \right]$$

Next, we calculate the generalized momenta via

$$p(t) = \frac{\partial L}{\partial \dot{x}},$$

and calculate the Hamiltonian

$$\mathbb{H} = p(t) \cdot \dot{x}(t) - L.$$

To restrict down to our case our state variables are given by $|x(t)\rangle$, $|k(t)\rangle$, $|x(t)\rangle$, $|k(t)\rangle$, and the control Lagrangian can be computed starting from the comparison of the cost function (with a Lagrange multiplier imposing the Schrödinger equation):

$$J = \langle x(t_f) | \hat{C} | x(t_f) \rangle$$

$$+ \int_0^{t_f} dt \left[ \langle k(t) \mid -i\hat{H}(t) \mid x(t) \rangle - \langle \dot{x}(t) \rangle \right] + c.c..$$

with (S1) and continuing through to Eq. (S3)

$$\mathbb{L} = \langle k(t) \mid -i\hat{H}(t) \mid x(t) \rangle - \langle \dot{x}(t) \rangle + c.c.,$$

The generalized momenta are given by

$$|p_x(t)\rangle = \frac{\partial \mathbb{L}}{\partial \langle \dot{x}(t) \rangle} = \hat{C} | x(t) \rangle - |k(t)\rangle,$$

$$|p_k(t)\rangle = \frac{\partial \mathbb{L}}{\partial \langle \dot{k}(t) \rangle} = 0,$$

$$p_u(t) = \frac{\partial \mathbb{L}}{\partial u(t)} = 0.$$

Therefore, calculation of the control Hamiltonian gives

$$\mathbb{H} = \langle p_x(t) | \dot{k}(t) \rangle + \langle \dot{x}(t) | p_x(t) \rangle - \mathbb{L}$$

$$= i \langle k(t) | \hat{H}(t) | x(t) \rangle + c.c..$$
Since \( \frac{dH}{dt} = 0 \) (remember that we are treating \( u(t) \) as a variable), the control Hamiltonian should be conserved for all time. To verify this we can take the full time derivative
\[
\frac{dH}{dt} = -i \langle \dot{x}(t) | \hat{H}(t) | k(t) \rangle + c.c. \quad \text{(S12)}
\]
\[
- i \langle x(t) | \hat{H}(t) | k(t) \rangle + c.c.
\]
\[
- i \langle x(t) | \dot{u}(t)(\hat{B} - \hat{C}) | k(t) \rangle + c.c.
\]
Using the Schrödinger equations,
\[
| \dot{x}^*(t) \rangle = -i \hat{H}^*(t) | x^*(t) \rangle, \\
| \dot{k}^*(t) \rangle = -i \hat{H}^*(t) | k^*(t) \rangle, \quad \text{(S13)}
\]
the first two lines exactly cancel with each other. The leaves the last line which can be rewritten using the notation of the main text as
\[
\frac{dH}{dt} = -i \langle x(t) | \dot{u}(t)(\hat{B} - \hat{C}) | k(t) \rangle + c.c. \quad \text{(S14)}
\]
\[
= -\dot{u}(t)\Phi(t).
\]
For optimal protocols we either are in a bang region where \( u(t) \) is constant implying \( \dot{u}(t) = 0 \) or we are in a singularity where \( \Phi(t) = 0 \). Therefore, \( \frac{dH}{dt} = 0 \) and the control Hamiltonian must be a conserved quantity.

Finally, we consider what modifications would need to be done to this picture by including a soft time constraint. Using the primes to refer to the setting where we have a soft time constraint and \( t_f \) is allowed to vary and unprimed quantities to refer to the original \( t_f \) fixed problem, we can relate
\[
J' = J + \lambda t_f \quad \text{(S15)}
\]
where \( \lambda \) gauges the strength of the soft constraint and makes additional time usage unfavorable. Going through the derivation, we get the following modifications
\[
L' = L + \lambda, \quad \text{(S16)}
\]
\[
\mathcal{H}' = \mathcal{H} - \lambda. \quad \text{(S17)}
\]
Obviously since \( \lambda \) is time independent, \( \mathcal{H}' \) is still a conserved quantity.

### S2. SOFT TIME CONSTRAINTS

This section explores the different forms of soft and hard time constraints, providing the details and derivations behind the explanation of “Time constraints” in the main text. This derivation also highlights the physical meaning of the control Hamiltonian, \( \mathcal{H}(t) \), discussed in the previous section S1.

We will consider two forms of time constraint: the first is a hard-constraint where we just give the system time \( t_f \) as considered in the main text, and the second imposes a linear cost for each additional amount of time taken. In terms of the hard constraint cost function (unprimed), the soft constraint cost function (primed) is
\[
J' = \langle x(t_f) | \hat{C} | x(t_f) \rangle + \lambda_t = J + \lambda_t. \quad \text{(S18)}
\]
For the most part, this does not modify the equations of motion except the condition due to variations in the final time:
\[
\left[ \langle x^*(t_f) | \hat{H}^*(t_f) | k^*(t_f) \rangle - \text{c.c.} \right] \delta t_f = 0, \quad \text{(S19)}
\]
which does not apply in the hard constraint case since \( \delta t_f = 0 \) and becomes
\[
\left( \langle x^*(t_f) | \hat{H}^*(t_f) | k^*(t_f) \rangle - \text{c.c.} \right) = i\lambda, \quad \text{(S20)}
\]
for the soft constraints which can be thought of as a condition solely on \( t_f \).

When \( \lambda = 0 \) and \( t_f \) is allowed to be free, Eq. (S19) combined with the form of \( \mathcal{H}(t_f) \), Eq. (S11), implies that \( \mathcal{H}(t) = 0 \), and we can verify this numerically in both the quantum adiabatic limit and for QAOA protocols that are constrained by \( p \) but not \( t_f \).

However, the introduction of \( \lambda \neq 0 \) modifies the \( \lambda = 0 \) control Hamiltonian \( \mathcal{H}(t) \) so that the soft-constrained version (primed) is
\[
\mathcal{H}'(t) = \mathcal{H}(t) - \lambda. \quad \text{(S21)}
\]
Since this problem has a finite final time, $\mathbb{H}(t) = 0$ (see Eq. (S20)), which means that the original control Hamiltonian for an equivalent hard constraint problem has $\mathbb{H}(t) = \lambda$.

Furthermore, if we look at a hard-constrained problem with a cutoff of $t_f$, the equations of motion are again the same, but the control Hamiltonian is equal to a non-zero constant, with the constant being dependent on $t_f$. Since the equations of motion are identical, this constant is just $\lambda$ again and just dictates how much soft-constraint would lead to the same $t_f$ as our hard-constrained solution. In practice, the dependence of $\lambda$ on $t_f$ (or vice-versa) is heavily influenced by the nature of the problem and how quickly the system approaches its ground state energy.

In Fig. (S1) we plot the value of the control Hamiltonian versus $t_f$ for $t_f$-constrained QAOA applied to a random MaxCut instance. To read off the equivalence between soft and hard constraints in this plot, draw a horizontal line from the $y$ axis at the value of $\lambda$ you have. Where that line intersects the desired $p$ curve is the time that a soft-constrained problem with $\lambda$ would prefer. Note that the $\lambda$ line could intersect at multiple points, in this case, take the right-most point. For instance, the intersections in the green shaded region will never be prefered by soft-constrained QAOA. Furthermore, any soft constraints, $\lambda$, above the black dashed line are so constricted that they prefer a $t_f = 0$ solution with the $p$ shown. As $p$ increases further, the curves slowly start encroaching upward into this region.

Since soft and hard constraints are theoretically equivalent through this factor $\lambda$, we focus on hard time constraints for most of the paper. Also note that $\mathbb{H}(t_f) = \lambda = i \langle x(t_f) | \left[ \hat{\mathcal{B}}, \hat{\mathcal{C}} \right] | x(t_f) \rangle$ can be measured and could be used as an estimate of the error rates in the system after a QAOA variational loop has been completed.

S3. NON-SINGULAR BANG LENGTH

One of our key results is that all optimal protocols must begin with a finite length bang (assuming we start in an eigenstate of $\hat{\mathcal{B}}$) and end with a finite length bang (assuming our cost function is an expectation value of $\hat{\mathcal{C}}$). One important question is whether we can estimate how long these respective bangs should be. Numerically, as seen in Fig. (1) from the main text, these initial and final bangs remain large on the timescale of the system.

While the initial and final bangs are undergoing Hamiltonian evolution under a static Hamiltonian, it is not possible to use simple a priori arguments to determine the exact lengths of the bangs. This is because of a lack of boundary information. For instance, in the initial bang, we know the starting point of $|x(0)\rangle$ but not the starting point of $|k(0)\rangle$, both of which we need to determine the first value of $t$ such that $\Phi(t) = 0$ which would herald the probable end of the bang. Similarly, at the end, we know that $|k(t_f)\rangle = \hat{\mathcal{C}}|x(t_f)\rangle$, but there is no a priori way of knowing what $|x(t_f)\rangle$ is, that being the goal of these algorithms.

For both the initial and final bang, the length of the bang is determined by the time it takes $|\Phi(t)|$ to transition from $\lambda$ to 0. Therefore, the value of $\lambda$ will determine the lion’s share of how long the bang takes. As described in the main text, $\lambda$ can be thought of as the penalty to the cost function for each extra amount of calculation. Therefore, the function $\lambda(t_f)$ will depend primarily on how the achieved QAOA energy scales with $t_f$. For instance, if the QAOA energy scales with $O(1/t_f)$ towards the true ground state, then the cost function $J'$ in Eq. (S18) will be the trade-off between $1/t_f$ scaling in the energy and $t_f$ scaling in the time cost, resulting in some balance that produces a preferred $t_f$ curve like that seen in Fig. S1. We expect $\lambda$ to decrease with increasing $t_f$, but the exact form of that decrease will be problem specific as discussed in the main text.

In addition to $\lambda$, one of the key factors determining the length of the bangs is how fast $\Phi(t)$ can change. To first order, we can approximate this by $\dot{\Phi}(t_f)$ and $\Phi(0)$. At the final time, it is easy to see that

$$\dot{\Phi}(t_f) = \langle x(t_f) | \left[ \hat{\mathcal{B}}, \hat{\mathcal{C}} \right] | x(t_f) \rangle.$$  \hspace{1cm} (S22)

For large enough $t_f$, this quantity should depend primarily on the nature of the ground state and a few excited states. Because of our boundary conditions, we cannot write $\dot{\Phi}(0)$ in terms of only $|x(0)\rangle$ and not $|k(0)\rangle$, but we similarly expect (and see numerically) that this quantity is roughly constant.

Therefore, up to multiplicative constants, we expect $\lambda$ to determine the scaling of the sizes of the bangs. Therefore, these bangs should become smaller and smaller as $t_f$ is increased. Eventually in the true $t_f \to \infty$ adiabatic limit, these bangs disappear recovering the standard form expected for quantum adiabatic computing.

S4. CLASSIFYING SINGULARITIES

In the main text we prove that all optimal protocols begin and end with non-singular bang regions. In this section, we find no evidence that singular regions (i.e. smooth annealing regions) cannot exist frequently in the middle region, and we derive the conditions for such singularities. In fact numerically as seen in Fig. (1) of the main text, we find that such singular regions are common in the true optimal protocol.

First, a singular region implies that $\Phi(t) = 0$ for an extended period of time, which by the constancy of the optimal control Hamiltonian [see Eq. (9) from the main text or Eq. (S11)] means that in a singular region with $u^*(t) \in (0, 1)$, $\Phi_C(t) = \lambda$ and $\Phi_B(t) = \lambda$.

In order for $\Phi_C(t)$ and $\Phi_B(t)$ to remain constant, all their time derivatives must be zero. Simple differentiation and application of the Shr"odinger equation show
that the first derivative condition reduces to

$$\Phi_{[[B,C]]} = 0,$$

(S23)

where $\Phi_X$ for any operator $X$ is defined by

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

(S24)

The second derivatives give the condition

$$0 = \Phi([B,C],B)u^*(t) + \Phi([B,C],C](1-u^*(t)).$$

(S25)

At this level, there are three possibilities. The first possibility is that

$$u^*(t) = \frac{\Phi([B,C],C]}{\Phi([B,C],B]}(t) - \Phi([B,C],C](t).$$

(S26)

A similar form of singular control was explored geometrically in [S3] where it was used to show a singular control for the Grover search problem that outperforms the traditional Grover problem (but has the same asymptotic scaling). The Grover problem in this setting was an effective two-level system, and the geometric form of the singular control found in [S3] was specific to that problem and its mapping onto the Bloch sphere.

The second possibility is that only $\Phi([B,C],B] = 0$, but this necessitates $u^*(t) = 0$, and while this is a singular protocol, it matches the form of a bang-bang procedure. There are known examples (for non-Ising models) of cases where such a bang-bang singularity is optimal [S9].

The last would be that $\Phi([B,C],B] = 0$ and $\Phi([B,C],C](t) = 0$ which would necessitate considering higher derivatives, and this eventuality is discussed later.

If we have the first case, that just guarantees the derivatives are zero at this level, so we would need to go up to higher levels and verify compliance there. That compliance could be through a condition on $u(t)$ such as seen in Eq. (S26), but as we go up the ladder to higher derivatives, the requirements become overconstrained with two equations to satisfy at the next level, four at the next, and so on, making this possibility unlikely.

Another possibility is that all the $\Phi_X(t) = 0$ due to the nature of the states $|x(t)\rangle$ and $|k(t)\rangle$ which is the only possible source of singularities in classical systems linear in $x(t)$ and $u(t)$ and then only if the control system lacks full controllability. This is not possible in our case by construction because our initial state, as a ground state of $\hat{B}$ must lie within the subspace defined by the shared symmetries of $B$ and $C$. Therefore, $|x(t)\rangle$ (and also $|k(t)\rangle$) must lie entirely within this same symmetry subspace. The Lie algebra generated by $\hat{B}$ and $\hat{C}$ can move us around fully within that subspace, with symmetries being what restricts down the controllability of the system, meaning that some operator within the Lie algebra generated by nested commutators of $\hat{B}$ and $\hat{C}$ must have a non-zero matrix element between $|x(t)\rangle$ and $|k(t)\rangle$.

For the reasons discussed above, we expect most singularities to be of the form shown in Eq. (S26) with other forms increasingly unlikely. The big question is whether we will see this form of singularity or a bang-bang form in general as the global, rather than local, optimum.

These singularities are similar to the forms seen in more traditional quantum optimal control literature for their version of singularities [S4, S5].

It should be noted that all the conditions listed here are necessary conditions for optimality, not sufficient. In order for these conditions to be sufficient, $J$ would need to be a convex function of $|x(t)\rangle$ [S6] which is not true in general for quantum systems. In quantum optimal control problems with no constraints on the control parameter and other conditions such as full controllability, false minima or traps that satisfy the optimal conditions but are not true optima are exceedingly rare [S5]. However, when control and time constraints apply (as in our case), traps become more likely.

Our case does have a large number of constraints, so the standard logic from quantum optimal control theory about the non-idealness of singular solutions [S4, S5] does not hold here. As was seen in the main text and discussed further in the next part of this section, in Ising models at least, the singularity from Eq. (S26) turns out to be the global minimum for the control landscape.
A. Forms of Singularities in Practice

While Fig. (1) in the main text was constructed using a gradient descent method, the singular regions must still follow the different forms of singularities possible for this problem as discussed in the previous section. This section discusses how the analytic results of the previous section apply to our numerics.

For instance, a large question is whether \( u(t) \) in Fig. (1) from the main text obeys the singularity condition in Eq. (S26) or if it obeys some other condition derived from a higher derivative as discussed in the previous section. The answer to this question is twofold. For the most part, the singular \( u(t) \) does obey Eq. (S26), especially in the middle of the smooth curve. However, it does not always follow this condition.

In Fig. S2, we reproduce Fig. (1) from the main text, highlighting the region where the singular control obeys Eq. (S26) to within a numerical tolerance. Notice that the highlighted region does not encompass the entire singular region. This implies that for portions of the evolution, \( u(t) \) is obeying some higher order singularity condition as discussed, but not derived, in the main text.

Note that our simulation requires a discretization of \( u(t) \) into 1001 \( \Delta t \) steps and that our gradient descent only achieves the optimal up to some precision based on how long we run it. Therefore, the \( u(t) \) shown in Fig. S2 could have numerical errors shortening the range of agreement with Eq. (S26), but it is possible that higher derivative conditions are relevant here.

S5. ADDITIONAL NUMERICs

In the main text we focused on a particular instance of the MaxCut problem to provide example numerics. In this section, we expand to examining other problems, most Ising but with also some Heisenberg models, all of which exhibit roughly the same qualitative behavior. We also examine more in depth the behavior of these optimal bang-anneal-bang curves as \( t_f \) and \( n \) change. There are two additional Ising models that we will be considering here, both conforming to the basic form

\[
\hat{C} = \sum_{i<j} J_{ij} \hat{\sigma}_i^z \hat{\sigma}_j^z. \tag{S27}
\]

The first model, Randomized Ising (abbreviated RI), will just involve a randomized all-to-all connected \( J_{ij} \) matrix where each coupling is chosen randomly from the range \([-1,1]\). The second model, Long-Range Ising (abbreviated LR), is motivated by our experimental collaboration [S10] and involves a 1D chain of qubits interacting via

\[
J_{ij} = \frac{J}{|i-j|^\alpha}. \tag{S28}
\]

where we have taken \( J = 1 \) in our dimensionless calculations.

First in Fig. S3 we show data from an instance of our RI model. In blue, we show the optimal \( u(t/t_f) \) procedure; in green we show a \( p = 20 \) constrained QAOA protocol (green dotted line) that takes the same amount of time. Red line highlights the regions where \( u(s) \) matches the singular condition from Eq. S26. Note that the singular condition does not match for the entire annealing region, and we claim that this deviation is due mostly to premature stopping of the gradient descent procedure. Each plot represents a different final \( t_f \) value.

In Fig. S4 we show results from the LR model for \( n = 5 \) and \( t_f = 2 \). Notice that the oscillations in the middle of the evolution have large enough amplitude that they reach the edge of the allowed domain and become bangs, at least for low \( \alpha \) plots. Our other work [S11] indicates that the oscillations decrease in amplitude as \( t_f \) increases, so for low \( t_f \), these chopped off amplitudes provide one route toward a mixed interior region with both bangs and anneals. It is also interesting that for \( \alpha \geq 1 \), the procedures are very similar in appearance. In general, we have noticed that similar problem instances often lead to qualitatively similar curves.

Finally to give a non-Ising example, we present Fig. S5 which shows the optimal protocol when the Hamiltonian
FIG. S4: The optimal protocols $u(t)$ for the Long Range Ising model (LR) for various values of the exponent $\alpha$ at $n = 5$ and $t_f = 2$. The curves still roughly follow a bang-anneal-bang pattern, but for some of the plots, the annealing region’s oscillations are large enough to flatten into bangs at the peaks.

The specific $J_{ij}$ matrix for this figure is the connectivity graph for an $n = 8$ 4-regular graph (the same one used in Fig. (1) of the main text). Most notably, this optimal protocol still maintains the bangs at the beginning and end and strongly requires them based off the sign and magnitude of $\Phi(t)$. In the interior region, both bangs and anneals are seen, as is allowed by the analytic analysis but is rarely seen in the smoother Ising model protocols. The $\Phi(t)$ here is rougher and indicates that further fine-tuning with gradient descent is possible. This tuning is much harder numerically for this model since the current $\hat{C}$ is non-diagonal and more complicated to simulate. Overall, this plot does match with the results of our analytic analysis.

FIG. S5: This plot shows the optimal protocol for the Heisenberg model where the problem Hamiltonian is given by $\hat{C} = \sum_{i<j} \frac{J_{ij}}{2} (\hat{\sigma}_i^x \hat{\sigma}_j^x + \hat{\sigma}_i^y \hat{\sigma}_j^y + \hat{\sigma}_i^z \hat{\sigma}_j^z)$. The specific $J_{ij}$ matrix here is the same one used in Fig. (1) of the main text. Note that the $\Phi(t)$ indicates that the gradient descent done to produce this figure was not as thorough. This is due largely to numerical limits and greater complexity of simulating the given $\hat{C}$ relative to the diagonal Ising model. Nevertheless, this figure demonstrates clear initial and final bangs with the interior region exhibiting both annealing and bang-bang behavior as our analytics predict.