Diagonal catalysts in quantum adiabatic optimization

Tameem Albash and Matthew Kowalsky

Diagonal Catalysts in Quantum Adiabatic Optimization

Tameem Albash[1, 2] and Matthew Kowalsky[3, 4]

[1] *Department of Electrical and Computer Engineering,
University of New Mexico, Albuquerque, New Mexico 87131, USA*
[2] *Department of Physics and Astronomy and Center for Quantum Information and Control,
CQuIC, University of New Mexico, Albuquerque, New Mexico 87131, USA*
[3] *Department of Physics and Astronomy, University of Southern California, Los Angeles, California 90089, USA*
[4] *Center for Quantum Information Science & Technology,
University of Southern California, Los Angeles, California 90089, USA*

We propose a protocol for quantum adiabatic optimization, whereby an intermediary Hamiltonian that is diagonal in the computational basis is turned on and off during the interpolation. This 'diagonal catalyst' serves to bias the energy landscape towards a given spin configuration, and we show how this can remove the first-order phase transition present in the standard protocol for the ferromagnetic p-spin and the Weak-Strong Cluster problems. The success of the protocol also makes clear how it can fail: biasing the energy landscape towards a state only helps in finding the ground state if the Hamming distance from the ground state and the energy of the biased state are correlated. We present examples where biasing towards low energy states that are nonetheless very far in Hamming distance from the ground state can severely worsen the efficiency of the algorithm compared to the standard protocol. Our results for the diagonal catalyst protocol are analogous to results exhibited by adiabatic reverse annealing, so our conclusions should apply to that protocol as well.

I. INTRODUCTION

The ubiquity of optimization problems continues to nurture the study and development of new algorithms to reduce the computational cost of solving them. Quantum adiabatic optimization [1–5] (QAO) is an approach that uses a sufficiently slow quantum evolution subject to an interpolating Hamiltonian $H(s), s \in [0, 1]$, to find the solution to the optimization problem encoded in the ground state of the problem Hamiltonian H_P at the end of the interpolation, $H(1) = H_P$. Recent progress in the development of large-scale qubit systems [6–9] has increased the impetus to better understand when we can expect QAO to provide a quantum advantage over classical algorithms for real-world optimization problems. To date, such advantages are only known in the oracular setting [10, 11][1].

In an analogous manner to its classical counterpart [13], QAO relies on the adiabatic theorem of quantum mechanics [14–18] to provide a guarantee that evolutions satisfying the adiabatic condition will have a high overlap with the ground state of H_P. The condition on how slowly the evolution must be performed is usually stated as a condition on the total evolution time t_f, whereby t_f must be much greater than some power of the inverse minimum energy gap $\Delta_{\min}^{-1}$ of $H(s)$ along the interpolation. This then gives a convenient way to express the computational time-cost of the algorithm, with the scaling of the minimum gap with system size n often reported as representing the efficiency of the QAO algorithm.

In the standard ('S') QAO setting, the interpolating Hamiltonian is written as a linear combination of a 'driver' Hamiltonian H_D and a 'problem' Hamiltonian H_P that is diagonal in the computational basis:

$$H_S(s) = (1-s)H_D + sH_P , \qquad (1)$$

where for simplicity we have taken a linear interpolation for the annealing schedule. We take the single qubit computational basis $\{|0\rangle, |1\rangle\}$ to be the eigenstates of the Pauli-z operator, $\sigma^z|0\rangle = |0\rangle$, $\sigma^z|1\rangle = -|1\rangle$, corresponding to the spin up and spin-down states respectively. The driver Hamiltonian is usually taken to be the uniform transverse field Hamiltonian $H_D = -\sum_i \sigma_i^x$.

Increased experimental control capabilities of quantum annealing system [19, 20] has lead to a resurgence of interest in different interpolation paths than those used in standard quantum annealing. Notably, there has been renewed interest in 'adiabatic reverse annealing' [21, 22] (ARA), or 'sombrero AQC' [23], whereby the interpolation starts from a diagonal local-field Hamiltonian $H_B = -\sum_i \varepsilon_i \sigma_i^z$ that encodes a classical spin configuration $\vec{\varepsilon} \in \{-1, 1\}^n$ as its ground state. The spin state $\vec{\varepsilon}$ will generically disagree with the ground state of H_P over some subset of the indices i. We denote the fraction of spins that agree with the ground state of H_P by c.

Studying the performance of ARA on the p-spin model, Ref. [21] found that above a critical value of c, the scaling of the minimum gap changes from exponentially closing to only polynomially closing, indicating an exponential improvement in the performance of the algorithm. These results have been taken as a positive indication that greater control of and choices for the interpolating Hamiltonian may result in dramatically improved performance for solving hard optimization problems.

Here we propose an alternative interpolating path

---

using the same terms of the ARA Hamiltonian, $\{H_\mathrm{D}, H_\mathrm{B}, H_\mathrm{P}\}$, but that is more akin to the standard interpolation in that the initial Hamiltonian is the driver Hamiltonian:

$$H_\mathrm{DC}(s) = (1-s)H_\mathrm{D} + \lambda s(1-s)H_\mathrm{B} + sH_\mathrm{P} . \qquad (2)$$

In the new interpolation, the Hamiltonian H_B is introduced as a 'catalyst' Hamiltonian [24–27] that is turned on and off during the interpolation. Because this catalyst is diagonal in the computational basis, we refer to it as a 'diagonal catalyst' (DC). This new approach reproduces the exponential improvement in the performance of solving the p-spin model over the standard QAO algorithm, with the added benefit that it makes clear the role of the catalyst H_B in providing the performance improvement: it works by biasing the energy landscape towards the target solution.

Our analysis also highlights two important limitations of this approach. First, as the value of p increases, the overall strength of the catalyst Hamiltonian must be made larger, and in the limit of $p \to \infty$, where the p-spin model becomes similar to the problem of unstructured search [10, 28], maintaining an exponential improvement requires the overall strength to increase linearly with system size.

Second, it becomes clear that biasing the energy landscape only works if we are biasing *towards* the target state, where the relevant distance measure is Hamming distance and not how close they are in energy. We highlight this by constructing instances where the low-lying energy states of H_P are far in Hamming distance from the ground state, and biasing towards these states makes the performance of the algorithm significantly worst than the standard protocol.

Our paper is structured as follows. In Sec. II, we study the ferromagnetic p-spin model, and show how a suitable choice of the diagonal catalyst can eliminate the first-order phase transition associated with this model. In Sec. III, we study the large p limit, corresponding to the problem of unstructured search and show how the diagonal catalyst fails to eliminate the first-order phase transition in this model unless the catalyst becomes infinitely strong. In Sec. IV, we study the Weak-Strong Cluster problem [29], which is another example where the first-order phase transition can be eliminated by a suitably chosen diagonal catalyst. In Sec. V, we present a failure mechanism of the protocol and show how the diagonal catalyst can exacerbate or introduce new bottlenecks to the standard protocol. In Sec. VI, we provide a discussion and concluding remarks about our protocol and its relationship to adiabatic reverse annealing.

II. FULLY-CONNECTED FERROMAGNETIC p-SPIN MODEL

We begin our analysis of the performance of our DC interpolation (Eq. (2)) by studying the fully-connected ferromagnetic p-spin model

$$H_\mathrm{P} = -n \left(\frac{1}{n} \sum_i \sigma_i^z \right)^p . \qquad (3)$$

The ground state of this problem Hamiltonian is the all-zero state, $|0\rangle^{\otimes n}$, corresponding to a magnetization density of 1. In the standard interpolation (Eq. (1)), the minimum gap along the interpolation closes exponentially with system size [28], and in the thermodynamic limit the closing of the gap is associated with a first-order phase transition.

A. Mean-field analysis

We first perform a mean-field analysis of our DC Hamiltonian in Eq. (2). The mean field analysis and the resulting free energy density in terms of the mean-field magnetization density provides a simple characterization of the energy landscape during the interpolation. Discontinuous jumps in the identity of the global minimum of the free energy are associated with a first-order phase transition, since the magnetization changes discontinuously when following the global minimum. Following the derivation of Ref. [21], the mean-field free energy in the zero-temperature limit can be written as

$$\begin{aligned} f(m) = {}&s(p-1)m^p \\ &-c\sqrt{s^2 \left(pm^{p-1} + \lambda(1-s)\right)^2 + (1-s)^2} \\ &-(1-c)\sqrt{s^2 \left(pm^{p-1} - \lambda(1-s)\right)^2 + (1-s)^2} , \end{aligned} \qquad (4)$$

where m denotes the mean-field magnetization density. For $\lambda = 0$, corresponding to the absence of the DC, the free energy density reverts to that of the standard interpolation. We identify the location of the first-order phase transition along the interpolation by solving for the global minimum of the free energy $f(m)$ and identifying the value of the interpolation parameter $s = s_*$ where a discontinuous jump in the global minimum occurs.

In the absence of a catalyst, $f(m)$ exhibits a degenerate double-well at some point along the interpolation. When the catalyst is turned on, the separation of the degenerate minima is reduced. Above a critical value c and when the catalyst is sufficiently strong, the two minima merge and the double well vanishes, and only a single minimum is realized during the entire anneal. This behavior is qualitatively similar to that observed for the ARA protocol [21]. Here, the role of the catalyst is to energetically bias the landscape towards $m = 1$, pushing the first minimum towards the second minimum until only one minimum remains. Further details are provided in Appendix A.

However, as the catalyst strength is further increased, the first-order phase transition reemerges because the catalyst biases the wrong configuration and hinders the system from reaching the fully-ferromagnetic ground
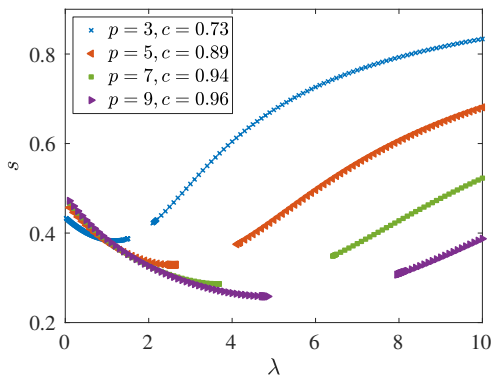
FIG. 1. Phase diagram from mean field theory showing first-order phase transition line for different p. Shown here are values of c where there is a range of λ values where the first-order phase transition is absent.

FIG. 2. Behavior of the catalyst strength λ^*, which maximizes the minimum gap, as a function of c for fixed p values. Inset: Behavior of the minimum gap evaluated at λ^*, $\Delta_{\min}^*$, as a function of c for the same p values.

state (for $c < 1$, there is always a subset of qubits that are biased in the 'wrong' direction). Therefore, for sufficiently high c, whose value depends on p, there is a range of λ for which the first-order phase transition can be avoided. We show snapshots of the phase diagram in Fig. 1.

B. Energy gap behavior

We supplement our mean-field analysis by calculating the minimum gap $\Delta_{\min}(\lambda) = \min_{s \in [0,1]} \Delta_\lambda(s)$ of the Hamiltonian (Eq. (2)) along the interpolation in the thermodynamic limit using the method developed in Ref. [30]. The method calculates the energy gap in the thermodynamic limit by considering fluctuations around the single global minimum of the semi-classical (large-spin) Hamiltonian density:

$$\mathcal{H}(s) = -(1-s)\left((1-c)m_2^x + cm_1^x\right)$$
$$-\lambda s(1-s)\left(cm_1^z - (1-c)m_2^z\right)$$
$$-s\left(cm_1^z + (1-c)m_2^z\right)^p ,\quad (5)$$

where $m_1^\alpha, m_2^\alpha \in [-1,1]$ correspond to the magnetization density in the $\alpha = x, z$ direction of the two clusters of spins. The method is suited to identify the position of the minimum gap in the parameter regime of (c, λ) where the first-order phase transition is absent. We show that the method accurately predicts the thermodynamic limit behavior in Appendix B. Our calculations of the minimum gap corroborate the conclusions from the mean-field analysis, in that we find that the minimum gap approaches a constant and does not close exponentially for sufficiently large c and λ values within the appropriate range.

However, we also observe a feature that can be expected from our mean-field analysis : the minimum gap exhibits a maximum value as a function of λ within the allowed range. The presence of a maximum can be expected because of the competing effects of increasing λ: while it biases the energy landscape towards the
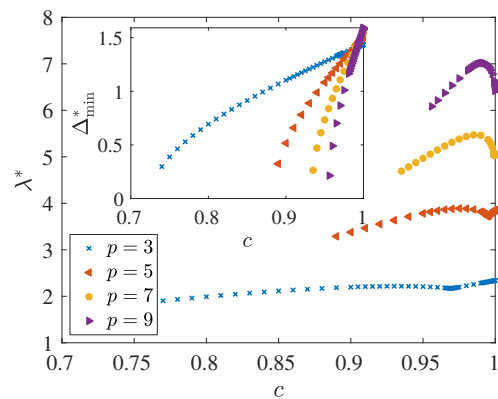
ferromangetic ground state, it ultimately does energetically favor different spin configurations. We define λ^* to be the value of λ that maximizes the minimum gap, $\lambda^* = \arg\max_\lambda \Delta_{\min}(\lambda)$, and the maximum minimum gap is then given by $\Delta_{\min}^* = \Delta_{\min}(\lambda^*) = \max_\lambda \Delta_{\min}(\lambda)$.

While $\Delta_{\min}^*$ grows monotonically as c approaches 1, the behavior of λ^* is not monotonic with c (see Fig. 2): it increases, decreases, and increases again (this last region becomes smaller as p increases). We provide an explanation for this behavior in Appendix C.

III. LARGE p LIMIT: UNSTRUCTURED SEARCH

So far, we have considered the case of fixed p while take the thermodynamic limit. We have so far observed that increasing p in this limit requires increasing c and λ to avoid the first-order transition (and exponentially closing gap). We therefore expect that as $p \to \infty$, $c \to 1$ and $\lambda \to \infty$.

In order to confirm this expectation, we write our p-spin Hamiltonian as:

$$H_P = -n\left(\mathbb{1} - \frac{1}{n}\sum_{i=1}^n (\mathbb{1} - \sigma_i^z)\right)^p = -n\left(\mathbb{1} - \frac{2\text{HW}}{n}\right)^p$$
$$= -n|0\rangle\langle 0| + n(-1)^p |2^n - 1\rangle\langle 2^n - 1|$$
$$-n\sum_{k=1}^{n-1}\left(1 - \frac{2k}{n}\right)^p \sum_{x:|x|=k} |x\rangle\langle x| ,\quad (6)$$

where $|x\rangle$ denotes the computational basis state with a bit-configuration representing the integer x, and HW $= \frac{1}{2}\sum_{i=1}^n (\mathbb{1} - \sigma_i^z)$ is the Hamming weight operator. If we now take the $p \gg 1$, we have that our p-spin Hamiltonian is approximately given by

$$H_P \approx -n|0\rangle\langle 0| - n(-1)^p |2^n - 1\rangle\langle 2^n - 1| .\quad (7)$$

This Hamiltonian is effectively the oracle Hamiltonian of unstructured search [10, 28], except for the factor of n and the additional penalty on the conjugate of the marked state. (Here $|0\rangle$ is the marked state, and $|2^n - 1\rangle$ is its conjugate.) The factor of n arises because we had made our p-spin Hamiltonian extensive (see Eq. (3)), and it leads to a gap between the ground state and the (degenerate) first excited state that grows with n.

We focus on the case of p odd. Even for $c = 1$, when the bias is entirely on the final ground state, for a fixed λ there is always a system size above which the gap scaling is exponential (see Fig. 3). Thus for fixed λ, our bias Hamiltonian is not able to eliminate the first-order phase transition typically associated with unstructured search. Above this size, the minimum gap scales as $\Delta_{\min} \sim \exp(n/(2\lambda))$, so it follows that in order to ensure a constant gap as the system size scales, λ must scale linearly with the system size. This finding is then consistent with our expectation from our fixed p analysis of the previous section. A similar result was also found in Ref. [31].
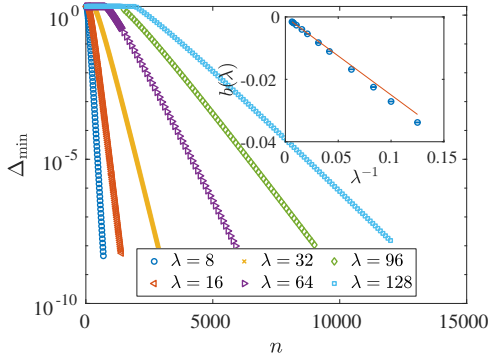

FIG. 3. Behavior of the minimum gap $\Delta_{\min}$ at $c = 1$ for the $p \gg 1$ problem Hamiltonian in Eq. (7). Inset: The fit of the exponential scaling parameter of the minimum gap, $\Delta_{\min} \sim \exp(b(\lambda)n)$, with $b(\lambda)$ approaching $1/(2\lambda)$ as λ becomes large, as indicated by the red line.

IV. WEAK-STRONG CLUSTER PROBLEM

Our next example is the prototypical large-spin tunneling problem studied in Ref. [27, 29, 30]. The problem is characterized by two fully-connected clusters of spins, one with a 'strong' local field in one direction, and the other with a 'weak' local field pointing in the opposite direction, depicted in Fig. 4. We will denote the former as the strong cluster and the latter as the weak cluster. In the standard QA protocol, the interpolation exhibits an exponentially closing gap, associated with the global minimum changing from the weak cluster being anti-aligned with the strong cluster to being aligned with the strong cluster. This event is then associated with the tunneling of $\mathcal{O}(n/2)$ spins.
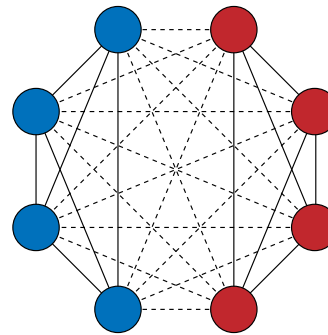

FIG. 4. Weak-Strong Cluster problem. Blue (left) and red (right) circles correspond to the two clusters of spins, one with a strong local field and the other with a weak local field in the opposite direction. Solid edges correspond to intra-cluster ferromagnetic couplings, and dashed edges correspond to inter-cluster ferromagnetic couplings.

The role of the diagonal catalyst can be readily understood in this case. We consider the following interpolating Hamiltonian:

$$H(s) = -2(1-s)\left(S_1^x + S_2^x\right) - s\left(2h_1 S_1^z - 2h_2 S_2^z\right.$$
$$\left. + \frac{4}{n}\left((S_1^z)^2 + (S_2^z)^2 + S_1^z S_2^z\right)\right) - \lambda s(1-s)\sum_{i=1}^{n}\epsilon_i \sigma_i^z, \quad (8)$$

where $S_1^\alpha = \frac{1}{2}\sum_{i=1}^{n/2}\sigma_i^\alpha$, $S_2^\alpha = \frac{1}{2}\sum_{i=n/2+1}^{n}\sigma_i^\alpha$. For simplicity, we restrict to the case where n is even. We take $h_1 = 1$ and $h_2 = 0.49$, such that the ground state at $s = 1$ has eigenvalues $(+1, +1)$ under $\frac{2}{n}S_1^z$ and $\frac{2}{n}S_2^z$ respectively. The terms associated with the local fields on the weak cluster (the terms proportional to h_2) are violated by the ground state at $s = 1$, but the instantaneous ground during the interpolation initially aligns with these local fields, which leads to the exponentially closing gap. A diagonal catalyst that reduces the alignment with the weak cluster local field during the interpolation can thus be expected to eliminate the exponentially closing gap. This intuitive argument already suggests that if we consider a catalyst of the form:

$$H_B = -\sum_{i=1}^{n/2}\sigma_i^z - \sum_{i=n/2+1}^{\frac{n}{2}(1+c)}\sigma_i^z + \sum_{i=\frac{n}{2}(1+c)+1}^{n}\sigma_i^z, \quad (9)$$

where we correctly bias the strong cluster and $\frac{n}{2}c$ qubits of the weak cluster and incorrectly bias $\frac{n}{2}(1-c)$ qubits of the weak cluster, then we should be able to eliminate the exponentially closing gap for $c > 1/2$.

In order to validate this intuition, we first consider the free energy density in the zero-temperature limit as
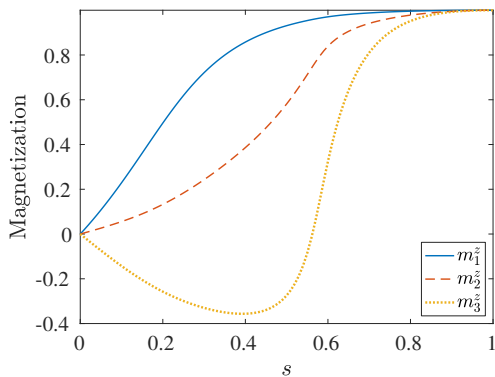
FIG. 5. Magnetization at the global minimum of the semi-classical Hamiltonian for the Weak-Strong Cluster problem (Eq. (11)) for the case of $c = 0.6$ and $\lambda = 1$ for which the first-order phase transition is avoided.

before:

$$f(m_1, m_2) = \frac{s}{4} \left(m_1^2 + m_2^2 + m_1 m_2 \right)$$
$$-\frac{1}{2} \sqrt{s^2 \left(m_1 + \frac{m_2}{2} + h_1 + \lambda(1-s) \right)^2 + (1-s)^2}$$
$$-\frac{c}{2} \sqrt{s^2 \left(m_2 + \frac{m_1}{2} - h_2 + \lambda(1-s) \right)^2 + (1-s)^2}$$
$$-\frac{1-c}{2} \sqrt{s^2 \left(m_2 + \frac{m_1}{2} - h_2 - \lambda(1-s) \right)^2 + (1-s)^2} . \tag{10}$$

For a sufficiently large c above $1/2$, we find we can avoid any discontinuous jump in the free energy for a sufficiently large λ. The mean field analysis can be corroborated by calculating the gap in the thermodynamic limit. The semi-classical Hamiltonian density is given by:

$$\mathcal{H}(s) = -\frac{1}{2}(1-s) \left(m_1^x + c m_2^x + (1-c) m_3^x \right)$$
$$-\frac{\lambda}{2} s(1-s) \left(m_1^z + c m_2^z - (1-c) m_3^z \right)$$
$$-s \left[\frac{h_1}{2} m_1^z - \frac{h_2}{2} \left(c m_2^z + (1-c) m_3^z \right) \right.$$
$$+\frac{1}{4}(m_1^z)^2 + \frac{1}{4} \left(c m_2^z + (1-c) m_3^z \right)^2$$
$$\left. +\frac{1}{4} m_1^z \left(c m_2^z + (1-c) m_3^z \right) \right] , \tag{11}$$

where m_1^α, $\alpha = x, y, z$, is the magnetization of the strong cluster, m_2 is the magnetization of the c-fraction of spins in the weak cluster with the correct bias, and m_3 is the magnetization of the $(1-c)$-fraction of spins in the weak cluster with the incorrect bias. We show in Fig. 5 how these three magnetizations behave in the absence of a closing gap, and the behavior is consistent with our mean field analysis.

V. BOTTLENECKS INTRODUCED BY DIAGONAL CATALYSTS

Our study of the p-spin and the Weak-Strong Cluster problems has revealed under what conditions we can expect a diagonal catalyst to enhance the performance of the adiabatic algorithm. Crucial to this success has been applying an energetic bias on states that are close in Hamming distance to the desired ground state. However, this implicitly assumes that we have a good reason (without knowing the ground state) to apply a bias on these states as opposed to other states. In the case of the p-spin model, states with low Hamming distance are also states with low energy, so biasing these states is a reasonable procedure.

However, it is easy to construct instances where this is not the case, meaning the low energy states are actually very far in Hamming distance from the true ground state. In this case, biasing towards low energy states can have a severely detrimental effect on the performance of the algorithm. We demonstrate this next with some striking cases. Our constructions are based on 'perturbative crossings' [32], a known bottleneck of QAO.

A. Perturbative Crossings

We first consider a one-dimensional Ising Hamiltonian on $n = 2k$ spins given by:

$$H_{\mathrm{P}} = \frac{1}{R} \left((R-1)\sigma_0^z - R\sigma_k^z - \sum_{i=0}^{n-1} J_{i,i+1} \sigma_i^z \sigma_{i+1}^z \right) \tag{12}$$

with $R \geq 4$, $\sigma_n^z \equiv \sigma_0^z$ (periodic boundary conditions), and

$$J_{i,i+1} = \begin{cases} R/2 & \text{if } i = k-1, k \\ R & \text{otherwise} \end{cases} \tag{13}$$

This system, depicted in Fig. 6, has anti-aligned local fields on opposite ends of the periodic chain with magnitudes 1 and $(1 - 1/R)$ respectively and Ising couplings that are uniform everywhere except at one end of the chain. The ground state is given by the all-zero bit string ($|\phi\rangle \equiv |0_{n-1} 0_{n-2} \ldots 0_0\rangle$), where the subscripts denote the qubit index as labeled in Fig. 6, with energy E_0, and the first excited state is double degenerate with states corresponding to the all-one state ($|\psi\rangle \equiv |1_{n-1} 1_{n-2} \ldots 1_0\rangle$) and the same state except with a 0 on the k-indexed qubit ($|\eta\rangle \equiv |1_{n-1} 1_{n-2} \ldots 1_{k+1} 0_k 1_{k-1} \ldots 1_0\rangle$) with energy E_1.

Let us now consider the annealing protocol with a diagonal catalyst, Eq. (2). We perform a perturbative analysis away from the point $s = 1$, with the perturbation parameter being $\Gamma = 1 - s$. The Hamiltonian up to first order is given by:

$$H^{(1)}(\Gamma) = H_{\mathrm{P}} + \Gamma V_1 = H_{\mathrm{P}} + \Gamma \left(H_{\mathrm{D}} + H_{\mathrm{I}} - H_{\mathrm{P}} \right) . \tag{14}$$
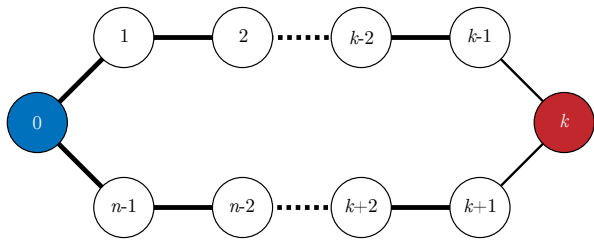
FIG. 6. Illustration of the Ising instance described by Eq. (12). The numbers inside the circles denote the labeling of the qubits. The links between circles denote ferromagnetic couplings (note the thickness denotes the strength of the coupling). The red circle (left-most) and blue circle (right most) denote qubits with non-zero Ising local fields in opposite directions.

At this order in perturbation theory, the first excited states $|\psi\rangle, |\eta\rangle$ and the ground state $|\phi\rangle$ are not coupled, i.e. $\langle\psi|V_1|\phi\rangle = \langle\eta|V_1|\phi\rangle = 0$. Thus, in order to determine how the first excited state degeneracy is broken, we can restrict ourselves to the degenerate subspace of $|\psi\rangle, |\eta\rangle$ at this order. In this subspace, we can represent V_1 as:

$$V_1 \to \begin{pmatrix} -E_1 + \varepsilon_1 - \lambda\epsilon_k & 1 \\ 1 & -E_1 + \varepsilon_1 + \lambda\epsilon_k \end{pmatrix}_{|\psi\rangle, |\eta\rangle}, \quad (15)$$

where $\varepsilon_1 = \lambda \sum_{i \neq k} \epsilon_i$. Specifically, the degeneracy is not only broken by the transverse field but also by the diagonal catalyst. Thus at first order in perturbation theory, the (non-degenerate) instantaneous first excited state energy is given by $E_1 - \Gamma(E_1 - \varepsilon_1 + \sqrt{1+\lambda^2})$. The instantaneous ground state energy is given by $E_0 - \Gamma(E_0 + \varepsilon_0)$, where $\varepsilon_0 = \lambda \sum_i \epsilon_i$. Therefore, the instantaneous gap at first order in perturbation theory is given by:

$$\Delta = \Delta_0 - \Gamma\left(\Delta_0 + \sqrt{1+\lambda^2} - \varepsilon_1 - \varepsilon_0\right), \quad (16)$$

where $\Delta_0 = E_1 - E_0$. Note that $\varepsilon_1 + \varepsilon_0 = 2\lambda \sum_{i \neq k} \epsilon_i + \lambda\epsilon_k$. The perturbative crossing is then expected to occur when this gap becomes zero, which occurs at a value of $\Gamma_{\mathrm{DC}}^* = \Delta_0 / \left(\Delta_0 + \sqrt{1+\lambda^2} - \varepsilon_1 - \varepsilon_0\right)$.

If we are to use a diagonal catalyst that energetically favors the first excited state $|\psi\rangle$, we would have $-\varepsilon_1 - \varepsilon_0 = (2n-1)\lambda$, which means that the perturbative crossing approaches $s = 1$ as the problem size grows, $\Gamma_{\mathrm{DC}}^* \sim 1/n$. Since the gap at the perturbative crossing scales as Γ_*^n, our results suggests a factorial scaling for the gap, $\Delta_{\min} \sim n^{-n}$.

This is to be contrasted with the standard protocol, where the instantaneous gap at first order would be given by:

$$\Delta_{\mathrm{S}} = \Delta_0 - \Gamma(\Delta_0 + 1), \quad (17)$$

and the perturbative crossing occurs at $\Gamma_{\mathrm{S}}^* = \Delta_0/(\Delta_0 + 1)$, which is independent of problem size. Thus we only expect an exponential scaling for the gap in this case, $\Delta_{\min} \sim c^{-n}$.
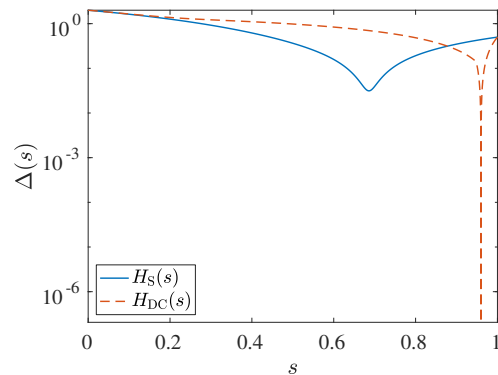

FIG. 7. Energy gap $\Delta(s)$ for $n = 6$ and $R = 4$ for the problem Hamiltonian in Eq. (12) using the standard protocol $H_{\mathrm{S}}(s)$ and the one with a diagonal catalyst $H_{\mathrm{DC}}(s)$ with $H_{\mathrm{I}} = \sum_{i=0}^{n-1} \sigma_i^z$ and $\lambda = 1$. This amounts to taking $\varepsilon_1 = -\lambda(n-1), \varepsilon_0 = -\lambda n$. For H_{DC}, first order perturbation theory predicts a level crossing at $s_{\mathrm{DC}}^* = 1 - \Gamma_{\mathrm{DC}}^* \approx 0.961283$, whereas the true avoided crossing calculated numerically occurs at $s_{\min} \approx 0.959768$. For H_{S}, first order perturbation theory predicts a level crossing at $s_{\mathrm{S}}^* = 2/3$, whereas the true avoided crossing calculated numerically occurs at $s_{\min} \approx 0.707325$.

To illustrate how this can affect performance, we show the energy gap of $H_{\mathrm{S}}(s)$ and $H_{\mathrm{DC}}(s)$ with $H_{\mathrm{B}} = \sum_{i=0}^{n-1} \sigma_i^x$ in Fig. 7, where we have chosen to bias towards the state $|\psi\rangle$. The minimum gap using the diagonal catalyst Hamiltonian is almost 5 orders of magnitude smaller than the standard Hamiltonian even for a 6 qubit problem.

B. Inducing a Perturbative Crossing

In the previous example, both $H_{\mathrm{S}}(s)$ and $H_{\mathrm{DC}}(s)$ exhibit a perturbative crossing. We now consider an example where only $H_{\mathrm{DC}}(s)$ exhibits a perturbative crossing. We consider a modification of our previous Hamiltonian such that it now takes the form

$$H_{\mathrm{P}} = \frac{1}{R}\left(R\sigma_0^z - (R-1)\sigma_k^z - \sum_{i=0}^{n-1} J_{i,i+1}\sigma_i^z\sigma_{i+1}^z\right). \quad (18)$$

For this Hamiltonian, the ground state is the all-one state $(|\psi\rangle)$, and the doubly degenerate first excited is now given by the all-zero state $(|\phi\rangle)$ and the state $|\eta\rangle)$.

In the standard annealing protocol, this problem does *not* exhibit a perturbative crossing. However, in the presence of a diagonal catalyst, a perturbative crossing can be induced. For example, if we take $H_{\mathrm{I}} = -\sum_i \sigma_i^z$, such that the catalyst biases the state $|\phi\rangle$, then in the first excited state subspace, the first order correction to the Hamiltonian in the perturbative parameter Γ is given by

$$V_1 \to \begin{pmatrix} -E_1 - n & 0 \\ 0 & -E_1 + n - 1 \end{pmatrix}_{|\phi\rangle, |\eta\rangle}. \quad (19)$$

The instantaneous energy gap at first order in perturbation theory is then given by:

$$\Delta_{\text{DC}}^{(1)} = \Delta_0 - \Gamma(\Delta_0 + 2n) \ , \qquad (20)$$

and the perturbative crossing is expected to occur at $\Gamma_{\text{DQ}}^* = \Delta_0/(\Delta_0 + 2n))$. We thus can expect the minimum gap to scale factorially in the system size.

We provide another example of such an induced perturbative crossing in Appendix D.

VI. DISCUSSION AND CONCLUSIONS

We have illustrated how introducing a catalyst Hamiltonian that is diagonal in the computational basis to the standard interpolating Hamiltonian of QAO can exhibit an exponential improvement in the efficiency of the algorithm for solving the ferromagnetic p-spin model and the weak-strong cluster problem. Our results are analogous to the exponential improvement observed for ARA [21], and the diagonal catalyst approach highlights the mechanism for this improvement: the Hamiltonian H_{B} biases the energy landscape towards the solution, and if the bias is sufficiently strong it can eliminate discontinuous jumps in the global minimum of the landscape, which are the bottleneck of the algorithm.

Our work also highlights the danger of over-selling this approach in the broader context of solving for the ground state of hard optimization problems using QAO. In the ferromagnetic p-spin models, the energy landscape is simple: the Hamming distance of a state from the ground state is directly correlated with its energy, so a bias Hamiltonian that progressively favors lower Hamming weight states will naturally help the algorithm reach the target ground state. However, in general, we expect there to be no correlation between a state's Hamming distance and its closeness in energy to the ground state, and in this more general case there is no evident prescription of how to pick the bias Hamiltonian. We highlighted this using examples where biasing low energy but high Hamming distance states impedes the algorithm, something that was already observed in the results of Ref. [23].

The operators appearing in the DC and ARA Hamiltonian are the same, and the only difference is how each term is turned on and off during the interpolation. Because the spectral gap is purely a property of the instantaneous Hamiltonian and not of the interpolation used to reach that Hamiltonian, we can expect both approaches to achieve the same minimum gap, assuming that both approaches are able to enact sufficiently rich interpolations. This suggests that the two approaches should achieve the same adiabatic algorithm efficiency. From this point of view, there does not appear to be a strong reason to prefer one approach over the other. However, there is one advantage of using a catalyst over ARA that we can identify. In ARA, the Hamiltonian H_{B} needs to energetically bias a single bit-string in order to ensure there is a unique initial ground state. This bit-string will have a subset of bits that disagree with the ground state of H_{P}. However, in the case of the catalyst, we need not place a bias on all qubits since the initial state of the algorithm is still the ground state of the driver Hamiltonian H_{D}. Thus with a catalyst, we can in principle avoid or mitigate biasing qubits incorrectly by not having H_{B} act on all qubits. While this may favor the catalyst approach, we believe that it is more important to identify which approach is more robust to experimental imperfections and noise. We hope to address this question in future work.

[1] A. B. Finnila, M. A. Gomez, C. Sebenik, C. Stenson, and J. D. Doll, "Quantum annealing: A new method for minimizing multidimensional functions," Chemical Physics Letters 219, 343–348 (1994).
[2] J. Brooke, D. Bitko, T. F., Rosenbaum, and G. Aeppli, "Quantum annealing of a disordered magnet," Science 284, 779–781 (1999).
[3] Tadashi Kadowaki and Hidetoshi Nishimori, "Quantum annealing in the transverse Ising model," Phys. Rev. E 58, 5355 (1998).
[4] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, Joshua Lapan, Andrew Lundgren, and Daniel Preda, "A quantum Adiabatic Evolution Algorithm Applied to Random Instances of an NP-Complete Problem," Science 292, 472–475 (2001).
[5] Giuseppe E. Santoro, Roman Martoňák, Erio Tosatti, and Roberto Car, "Theory of quantum annealing of an Ising spin glass," Science 295, 2427–2430 (2002).

[6] R. Harris, J. Johansson, A. J. Berkley, M. W. Johnson, T. Lanting, Siyuan Han, P. Bunyk, E. Ladizinsky, T. Oh, I. Perminov, E. Tolkacheva, S. Uchaikin, E. M. Chapple, C. Enderud, C. Rich, M. Thom, J. Wang, B. Wilson, and G. Rose, "Experimental demonstration of a robust and scalable flux qubit," Phys. Rev. B **81**, 134510 (2010).

[7] M W Johnson, P Bunyk, F Maibaum, E Tolkacheva, A J Berkley, E M Chapple, R Harris, J Johansson, T Lanting, I Perminov, E Ladizinsky, T Oh, and G Rose, "A scalable control system for a superconducting adiabatic quantum optimization processor," Superconductor Science and Technology **23**, 065004 (2010).

[8] A J Berkley, M W Johnson, P Bunyk, R Harris, J Johansson, T Lanting, E Ladizinsky, E Tolkacheva, M H S Amin, and G Rose, "A scalable readout system for a superconducting adiabatic quantum optimization system," Superconductor Science and Technology **23**, 105014 (2010).

[9] P. I Bunyk, E. M. Hoskinson, M. W. Johnson, E. Tolkacheva, F. Altomare, AJ. Berkley, R. Harris, J. P. Hilton, T. Lanting, AJ. Przybysz, and J. Whittaker, "Architectural considerations in the design of a superconducting quantum annealing processor," IEEE Transactions on Applied Superconductivity **24**, 1–10 (Aug. 2014).

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

[11] M. B. Hastings, "The Power of Adiabatic Quantum Computation with No Sign Problem," arXiv:2005.03791 (2020).

[12] Rolando D. Somma, Daniel Nagaj, and Mária Kieferová, "Quantum speedup by quantum annealing," Phys. Rev. Lett. **109**, 050501– (2012).

[13] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," Science **220**, 671–680 (1983).

[14] A. Einstein, "Beiträge z. quantentheorie," Verh. d. D. phys. Ges. **16**, 826 (1914).

[15] P. Ehrenfest, "Over adiabatische veranderingen van een stelsel in verband met de theorie der quanta," Verslagen Kon. Akad. Amsterdam **25**, 412 – 433 (1916).

[16] M. Born and V. Fock, "Beweis des adiabatensatzes," Zeit. f. Physik **51**, 165–196 (1928).

[17] T. Kato, "On the adiabatic theorem of quantum mechanics," J. Phys. Soc. Jap. **5**, 435 (1950).

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

[19] Andrew D. King, Juan Carrasquilla, Jack Raymond, Isil Ozfidan, Evgeny Andriyash, Andrew Berkley, Mauricio Reis, Trevor Lanting, Richard Harris, Fabio Altomare, Kelly Boothby, Paul I. Bunyk, Colin Enderud, Alexandre Fréchette, Emile Hoskinson, Nicolas Ladizinsky, Travis Oh, Gabriel Poulin-Lamarre, Christopher Rich, Yuki Sato, Anatoly Yu. Smirnov, Loren J. Swenson, Mark H. Volkmann, Jed Whittaker, Jason Yao, Eric Ladizinsky, Mark W. Johnson, Jeremy Hilton, and Mohammad H. Amin, "Observation of topological phenomena in a programmable lattice of 1,800 qubits," Nature **560**, 456–460 (2018).

[20] R. Harris, Y. Sato, A. J. Berkley, M. Reis, F. Altomare, M. H. Amin, K. Boothby, P. Bunyk, C. Deng, C. Enderud, S. Huang, E. Hoskinson, M. W. Johnson, E. Ladizinsky, N. Ladizinsky, T. Lanting, R. Li, T. Medina, R. Molavi, R. Neufeld, T. Oh, I. Pavlov, I. Perminov, G. Poulin-Lamarre, C. Rich, A. Smirnov, L. Swenson, N. Tsai, M. Volkmann, J. Whittaker, and J. Yao, "Phase transitions in a programmable quantum spin glass simulator," Science **361**, 162–165 (2018).

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

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

[23] Alejandro Perdomo-Ortiz, Salvador E. Venegas-Andraca, and Alán Aspuru-Guzik, "A study of heuristic guesses for adiabatic quantum computation," Quantum Information Processing **10**, 33–52 (2011).

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

[25] Yuya Seki and Hidetoshi Nishimori, "Quantum annealing with antiferromagnetic fluctuations," Phys. Rev. E **85**, 051112 (2012).

[26] Gabriel A. Durkin, "Quantum speedup at zero temperature via coherent catalysis," Phys. Rev. A **99**, 032315 (2019).

[27] Tameem Albash, "Role of nonstoquastic catalysts in quantum adiabatic optimization," Phys. Rev. A **99**, 042334 (2019).

[28] T. Jörg, F. Krzakala, J. Kurchan, A. C. Maggs, and J. Pujos, "Energy gaps in quantum first-order mean-field–like transitions: The problems that quantum annealing cannot solve," Europhys. Lett. **89**, 40004 (2010).

[29] Sergio Boixo, Vadim N Smelyanskiy, Alireza Shabani, Sergei V Isakov, Mark Dykman, Vasil S Denchev, Mohammad H Amin, Anatoly Yu Smirnov, Masoud Mohseni, and Hartmut Neven, "Computational multiqubit tunnelling in programmable quantum annealers," Nature Communications **7**, 10327 (2016).

[30] Kabuki Takada, Yu Yamashiro, and Hidetoshi Nishimori, "Mean-field solution of the weak-strong cluster problem for quantum annealing with stoquastic and nonstoquastic catalysts," Journal of the Physical Society of Japan **89**, 044001 (2020).

[31] Jacob Bringewatt, William Dorland, and Stephen P. Jordan, "Polynomial time algorithms for estimating spectra of adiabatic hamiltonians," Phys. Rev. A **100**, 032336 (2019).

[32] M. H. S. Amin and V. Choi, "First-order quantum phase transition in adiabatic quantum computation," Phys. Rev. A **80**, 062326 (2009).

[33] R. H. Dicke, "Coherence in spontaneous radiation processes," Phys. Rev. **93**, 99–110 (1954).

Appendix A: Deformation of the mean field free energy landscape in the presence of the catalyst

In the absence of a catalyst, f exhibits a degenerate double-well with one minimum at $m_1 = 0$ and the other at m_2, with $m_2 > m_1$. When the catalyst is turned on and increased in strength, the first minimum m_1 now occurs at a non-zero value, while m_2 remains at (approximately) the same position. There is therefore a reduction

in the separation between the two minima. Above a critical value c and when the catalyst is sufficiently strong, the two minima merge and the double well vanishes, and only a single minimum is realized during the entire anneal. We depict this behavior in Fig. 8.
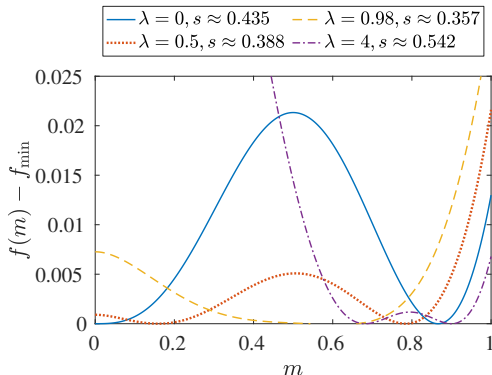

FIG. 8. The mean-field free energy (Eq. (4)) for the p-spin model with $p = 3$ and $c = 0.8$ and different λ values. The constant $f_{\min}$ is the value of the free energy at the minimum, and it is subtracted off so that all minima occur at zero. For $\lambda = 0$, and 0.5, the values of s is chosen to correspond to the point along the interpolation where the free energy exhibits doubly-degenerate minima. For $\lambda = 0.97$, there is only a single minimum along the entire interpolation.

Appendix B: Gap in the thermodynamic limit

We show how the the method of Ref. [21] accurately predicts the minimum gap in the thermodynamic limit for the models considered in Sec. II. We use the fact that the interpolating Hamiltonians are permutation symmetric for two subsets of the qubits. Since the initial ground state obeys these symmetries as well, we can focus on the symmetric subspace to which the ground state belongs. For the case of the ferromagnetic p-spin problem Hamiltonian, this subspace is spanned by the tensor product of Dicke states [33], $\{|j_1, m_1\rangle \otimes |j_2, m_2\rangle\}$, with $j_1 = cn/2, j_2 = (1 - c)n/2$ and $m_1 = -j_1, \ldots, j_1, m_2 = -j_2, \ldots, j_2$. This subspace is only $(2j_1 + 1) \times (2j_2 + 1)$ dimensional. In Fig. 9, we show how the minimum gap asymptotes to the thermodynamic limit value.

Appendix C: Non-monotonicity in λ^*

In order to understand the non-monotonic behavior of λ^* with c observed, it is useful to consider how the gap behaves during the interpolation. We show in Fig. 10 the gap behavior at $p = 3$ for representative c values. For $c \lesssim 0.93$, corresponding to the region where λ^* increases with c, the gap exhibits a single minimum along the interpolation, and the minimum grows in value and
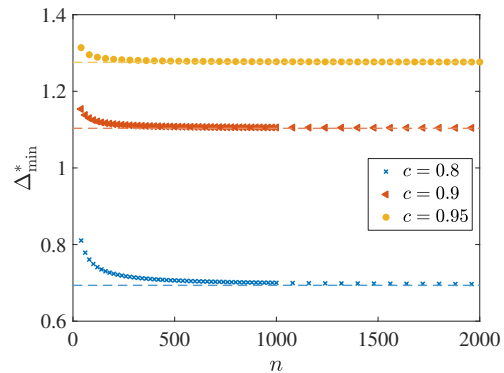

FIG. 9. Convergence of the finite n minimum gap values (data points) to the thermodynamic value (shown as dashed lines) for $p = 3$.

its location smoothly moves to larger s values as λ approaches λ^*.

For $c \gtrsim 0.93$, corresponding to the region where λ^* decreases and then increases with c, the gap exhibits two minima along the interpolation near λ^*. The first minimum (at the smaller s value) is initially lower in value but grows and passes the second minimum as λ approaches λ^* (the second minimum does not change significantly as λ approaches λ^*). Thus, this region of parameter space is characterized by the global minimum along the interpolation discontinuously jumping in s as λ approaches λ^*. The second minimum (and not the first) then determines the value of λ^*.

The second minimum occurs along the steepest portion of the rise of m_2^z from its negative to positive value (Fig. 11), so we can qualitatively associate the second minimum with the flipping of spins associated with the second cluster. Therefore, at sufficiently high c values, the optimum λ is associated with optimizing the hardness associated with flipping the misaligned small cluster.

Appendix D: Another example of inducing a perturbative crossing

We consider a problem Hamiltonian H_P of the form:

$$H_P = \frac{1}{R} \left(\sum_{i=0}^{n-1} h_i \sigma_i^z - \sum_{i=0}^{n-1} J_{i,i+1} \sigma_i^z \sigma_{i+1}^z \right) , \qquad (D1)$$

with $\sigma_n^z \equiv \sigma_0^z$ (periodic boundary conditions). We restrict ourselves to even length chains, but we distinguish between two types of chains. For $n = 4k, \ k = 2, 3, \ldots,$ we choose:

$$J_{i,i+1} = \begin{cases} \frac{R}{2} - 1 , & \text{if } i = \frac{n}{2}, n - 1 \\ R , & \text{otherwise} \end{cases} \qquad (D2a)$$

$$h_i = \begin{cases} -1 , & \text{if } i = \frac{n}{2} + 1, \frac{n}{2} + 2, \ldots, n - 1 \\ 1 , & \text{otherwise} \end{cases} , \qquad (D2b)$$
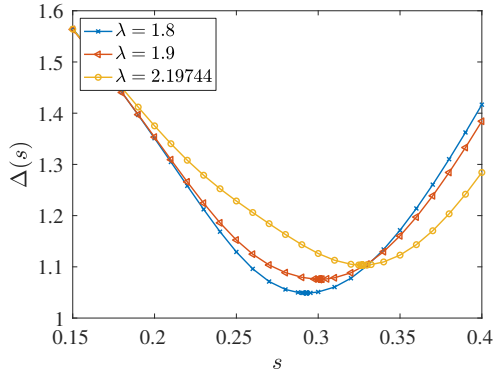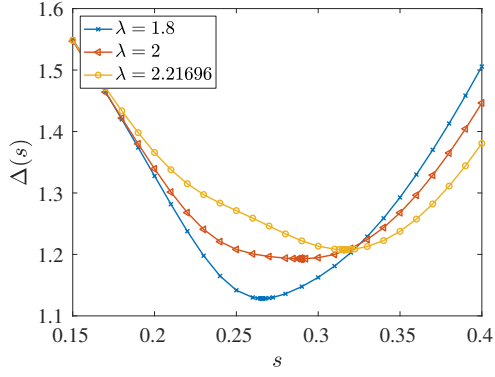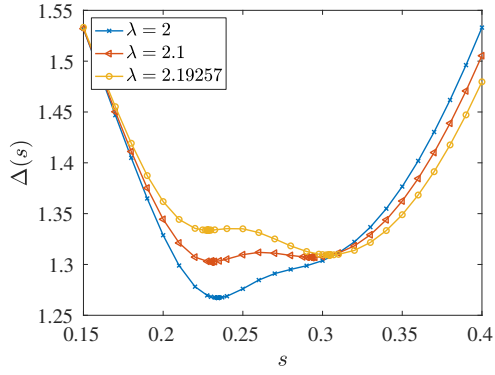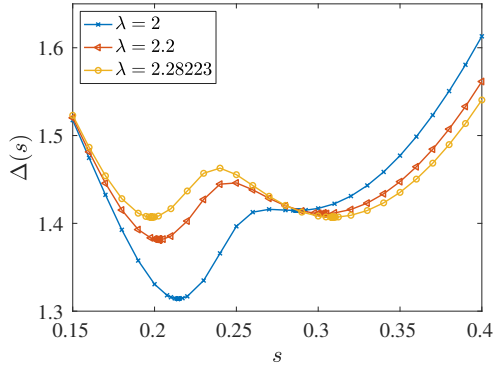
FIG. 10. Behavior of the gap $\Delta(s)$ for the $p = 3$ ferromagnetic p-spin model in the range of the interpolation where the the minimum gap occurs.
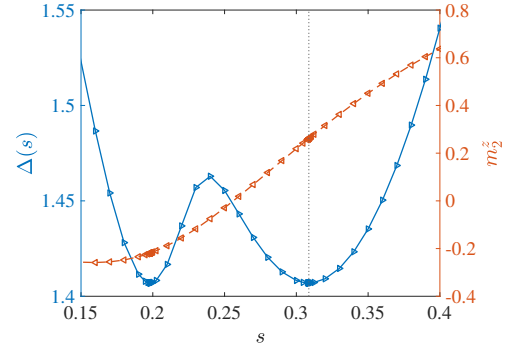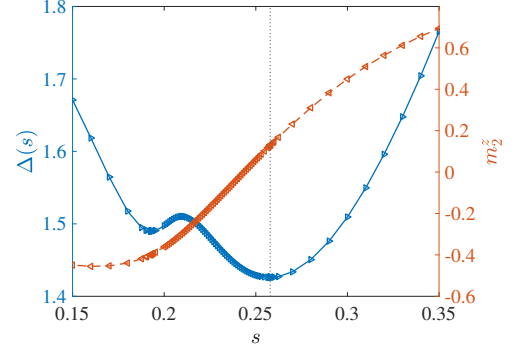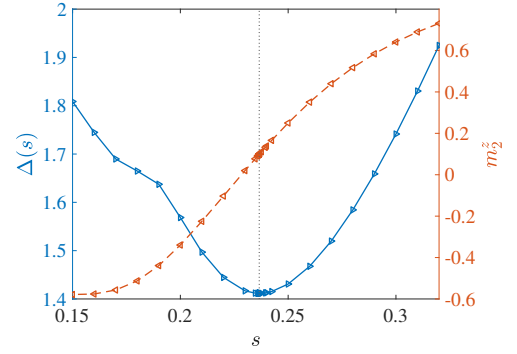

FIG. 11. Behavior of the gap $\Delta(s)$ and m_2^z for $c = 0.99$ at the optimal $\lambda = \lambda_*$ for the ferromagnetic p-spin model. Vertical dotted line corresponds to position of the minimum gap.

while for $n = 4k + 2$, $k = 1, 2, \ldots$, we choose:

$$J_{i,i+1} = \begin{cases} \frac{R-1}{2} \ , & \text{if } i = \frac{n}{2} - 1, n - 1 \\ R \ , & \text{otherwise} \end{cases} \tag{D3a}$$

$$h_i = \begin{cases} -1 \ , & \text{if } i = \frac{n}{2}, \frac{n}{2} + 1, \ldots, n - 1 \\ 1 \ , & \text{otherwise} \end{cases} \ . \tag{D3b}$$

We take $R = n/2$, which for our system sizes is always a positive integer.

The ground state of this Ising system is the state $|\phi\rangle \equiv |0_{n-1}0_{n-2}\ldots 0_{2k+1}1_{2k}1_{2k-1}\ldots 1_0\rangle$, where the subscript denotes the qubit index, with energy E_0. The doubly degenerate first excited states are the all-zero and

all-one bit strings ($|\eta\rangle = |0_{n-1}\ldots0_0\rangle, |\xi\rangle = |1_{n-1}\ldots1_0\rangle$) with energy E_1. For this problem Hamiltonian, the ground state is at least Hamming distance $n/2 - 1$ from the nearest first excited state, while the first excited states themselves are Hamming distance n apart. The Ising energy gap is given by $\Delta_0 = E_1 - E_0 = 2/R = 4/n$

Let us now consider the spectrum of the diagonal catalyst Hamiltonian near $s = 1$. The Hamiltonian at first order in the perturbative parameter $\Gamma = 1 - s$ is given by Eq. (14). At this order in perturbation theory, the ground state energy is modified to

$$E_0^{(1)}(\Gamma) = E_0 - \Gamma\left(E_0 + \varepsilon_n\right) , \qquad (D4)$$

where $\varepsilon_n = 0$ if $n = 4k + 2$ and $\varepsilon_n = 2$ if $n = 4k$. In the first excited state subspace, the first order perturbation is given by

$$V_1 = (H_D + H_B - H_P) \rightarrow \begin{pmatrix} -E_1 + n & 0 \\ 0 & -E_1 - n \end{pmatrix}_{|\eta\rangle,|\xi\rangle}, \qquad (D5)$$

where the state $|\xi\rangle$ is lowered in energy by H_B while the state $|\eta\rangle$ is raised in energy by H_B. Therefore the instantaneous first excited state energy is given by:

$$E_1^{(1)}(\Gamma) = E_1 - \Gamma\left(E_1 + n\right) . \qquad (D6)$$

The instantaneous energy gap at first order in the perturbation Γ is then given by:

$$\Delta_{DC}^{(1)}(\Gamma) = \Delta_0 - \Gamma(\Delta_0 + n - \varepsilon_n) . \qquad (D7)$$

Because the rate of decline of the instantaneous first excited state is faster than the instantaneous ground state, we predict a perturbative crossing to occur at $\Delta_{DC}^{(1)}(\Gamma_{DC}^*) = 0$:

$$\Gamma_{DC}^* = \Delta_0/\left(\Delta_0 + n - \varepsilon_n\right) . \qquad (D8)$$

This crossing gets closer and closer to $\Gamma = 0$ as n gets larger. Because the ground state and the first excited states are at least $n/2 - 1$ Hamming distance apart, we can expect the avoided level gap at this perturbative crossing to scale as $(\Gamma_{DC}^*)^{n/2} \sim n^{-n/2}$. This is in contrast to the standard forward anneal protocol, where the spectrum does *not* exhibit a perturbative crossing at first order in perturbation theory near $\Gamma = 0$.