Minimal constraints in the parity formulation of optimization problems

Martin Lanthaler\textsuperscript{1,∗} and Wolfgang Lechner\textsuperscript{1,2}

\textsuperset{1} Institute for Theoretical Physics, University of Innsbruck, A-6020 Innsbruck, Austria
\textsuperset{2} Parity Quantum Computing GmbH, A-6020 Innsbruck, Austria

E-mail: Martin.Lanthaler@uibk.ac.at and Wolfgang.Lechner@uibk.ac.at

Abstract

As a means to solve optimization problems using quantum computers, the problem is typically recast into an Ising spin model whose ground-state is the solution of the optimization problem. An alternative to the Ising formulation is the Lechner–Hauke–Zoller model, which has the form of a lattice gauge model with nearest neighbor four-body constraints. Here we introduce a method to find the minimal strength of the constraints which are required to conserve the correct ground-state. Based on this, we derive upper and lower bounds for the minimal constraints strengths. We find that, depending on the problem class, the exponent ranges from constant $\alpha = 1$ to quadratic $\alpha = 2$ scaling with the number of logical qubits.

1. Introduction

Combinatorial optimization problems are ubiquitous in a wide range of scientific fields. Most of these problems can be reformulated as an Ising-spin glass problem\cite{1}, which is the starting point for digital (e.g. quantum approximate optimization algorithm (QAOA))\cite{2} and analog quantum optimization algorithms (e.g. adiabatic quantum computation, LHZ architecture, Ising formulation of optimization problems)\cite{3–6}. The efficiency of QAOA, in particular the claim of substantial speedup is currently under debate (for review see\cite{7} or\cite{8}). It is not ruled out that highly coherent AQO may be more efficient than classical algorithms, at least for some classes of problems\cite{9–15}. In QAOA, low energy states are found via a variational procedure which is considered a promising route for near term quantum optimization\cite{16–18}. However, in both, adiabatic and digital algorithms, the problem Hamiltonian contains long-range interactions which requires either embedding schemes\cite{19–22} or large numbers of SWAP operations.

An alternative to the spin glass paradigm has been recently introduced\cite{22–25}. By a conceptual division of logical qubits, defining the optimization problem and the physical qubits available in the laboratory one maps the logical Ising Hamiltonian $H_{\text{logic}} = \sum_{\langle ij \rangle} J_{ij} \sigma_z^{(i)} \sigma_z^{(j)}$ to the physical Hamiltonian $H_{\text{phys}} = H_1 + H_c$

$$H_{\text{phys}} = \sum_{\langle ij \rangle} J_{ij} \sigma_z^{(i)} \sigma_z^{(j)} - \frac{1}{2} \sum_{\langle ij \rangle} c_j \sigma_z^{(i)} \sigma_z^{(j)} \sigma_z^{(i+1,j)} \sigma_z^{(i+1,j)}.$$  

(1)

This mapping is done by introducing a physical qubit for each pair of logical qubits, where the $z$-component corresponds to the relative orientation of two logical qubits i.e. $\sigma_z^{(i,j)} := \sigma_z^i \sigma_z^j$. The overhead in qubits for all-to-all pair interactions is quadratic and this increased number of degrees of freedom is compensated by constraints. Arranging the physical spins on a 2D lattice, allows to construct the constraints from four-local interaction on individual plaquettes consisting of 4 neighbouring spins. Figure 1 sketches the layout and the labeling of the physical qubits $(i,j)$ and plaquettes $[i,j]$, where the labels only run over pairs with $i < j$. The four-body constraints ensure that the low energy sub-space of the physical system $H_{\text{phys}}$ is exactly the spectra of the logical Hamiltonian $H_{\text{logic}}$ by adjusting the constraint strengths $c_j \in \mathbb{R}$. The
Following the parity architecture [22], the logical spectra can be found (up to a global shift) as a subset of the parity Hamiltonians spectra. To ensure a separation, local four-body terms are introduced via penalty terms $P_{i,j} = \sigma_i \sigma_j + \sigma_{i+1} \sigma_{j+1} / 2$, as denoted in equation (1). If the corresponding strengths $c_{ij}$ are chosen large enough, the original spectra is well separated from all the other eigenenergies ((b), left). The required strengths can be lowered by allowing states orthogonal to the logical subspace down to some energy threshold $e$. To gain a maximal fidelity w.r.t. some annealing schedule, this threshold is most likely to be between the logical ground state energy $l_0$ and the first exited one $l_0 + \Delta l$. For $c_{ij} = c$, finding the minimal strength $c(e)$ involves minimizing over subspaces with defined number of violated constraints and then taking the largest $c$. A systematic scheme to construct violating states is shown in panels (c) and (d). Starting from a state with no parity constraint violated and flipping the spins in the blue shaded region, one can construct all states with certain parity constraints violated.

constraints $c_{ij}$ have to be chosen large enough to separate the allowed logical subspace, i.e. states which do have a translation back into the logical picture, from states which do not have a counterpart in the logical model. In favour to reduce the magnitude of the constraint strengths we drop the requirement for a full separation between logical sub-spectra and the other eigenvalues. For a perfect (infinite slow) annealer, the optimal choice would be to penalize all states which disobey at least one constrain s.t. they are above the logical ground state energy. But a real annealing device operates on a finite time scale, which introduces diabatic transitions to higher levels. Setting the penalty as discussed, the final Hamiltonian has a degenerated ground space spanned by the logical ground state and at least one constraint-violating state. Increasing the penalties above the ideal annealer value, opens up the minimal gap, while, on the other hand increases the spectral radius of the final Hamiltonian. By the adiabatic theorem, an increasing spectral gap is favorable in terms of a faster running time, however a larger spectral radius requires a slower sweep in order to obtain the same fidelity (see the review paper reference [7]). We argue in appendix A, that for interesting classes of optimization problems, a good balance is obtained by setting the constraint strength s.t. the physical and the logical Hamiltonian do have the ground state energy and the first exited energy in common. This is only validated numerically for small system sizes, and it is unclear what's the situation for large sizes. If this scales up, this would makes our definition comparable to works like reference [26] which measured success probability, not the probability that the space spanned by the ground and first exited state were preserved. For the rest of the paper we adopt our energy-threshold definition of minimal constraints, where we explore two interesting energy thresholds: the logical ground state energy and the first exited one. This allows us to examine the final Hamiltonian only, instead of analyzing the whole annealing schedule. This simplified picture makes it possible to do reasonable statistics up to logical system sizes of $n = 25$ ($n_{\text{phys}} = 30$), when we restrict us to single-violator-states only i.e. states where exactly one of the parity constraints are violated.

In this paper, we determine the minimum constraint strengths $\hat{c}_{ij}$ that satisfy either only the lowest and or in addition the first exited state of the problem Hamiltonian. We show that finding the minimal constraints can be rewritten as a linear program. In the homogeneous setting $c_{ij} = c$, we derive a series of upper and lower bounds to the optimal constraint strength $c$ allowing to approximate the optimal values.
Different classes of optimization problems are modelled by considering $J_{ij}$ as independent and identically distributed (i.i.d.) random variables with probability density function (pdf) $f(\mu, \sigma^2)$. In the case $\mu/\sigma \rightarrow \pm \infty$ we derive analytic solutions to the minimal constraint problem, which are naturally related to the problem of solving MaxCut on the complete graph $K_n$ or the total ferromagnetic problem respectively. By a simple argument the authors in [26] concluded, that in the antiferromagnetic case the constraints should grow at least linearly with the system size. We show, that in this case the constraints even have to grow quadratic in system size. Also for random $J_0 \in \{-1, 1\}$ the authors of [26] expect the constraints to scale linearly with the size of the problem. We find, for the case of $\mu/\sigma$ finite, the large size scaling of the expected optimal constraint strength is mainly determined by the sign of the expectation value $\mu$.

If $\mu$ is negative, and we aim to preserve the first exited logical state, the large size scaling is linear. But this scaling is due to the choice of the energy threshold to be the first exited one. It stems from the fact that Hamiltonians as the total-ferromagnetic one has an ordered and unfrustrated ground state and a (in system size) linear growing gap (cf figure 7 (upper panel)). As the ferromagnetic example shows, in that regime the optimal constraint strengths are closer to the optimal-annealer case. Hence, preserving the logical ground state only would be the natural definition in that regime—which leads to constant constraint strengths. On the contrary, in the case $\mu$ positive the logical gap is neglectable s.t. it does no matter whether we define the threshold to be the logical ground state or the first exited one. In both cases we observe a quadratic scaling of the minimal constraints. The point $\mu = 0$ is interesting for symmetry reasons. By relying on results from extreme value theory we argue, that for standard Gaussian couplings, choosing the constraint strengths of order $\sqrt{n} \log(n)$ could be enough to ensure that the physical ground state faithfully represents the logical ground state.

2. Constraints

The constraints separate the subspace of allowed configurations from the unphysical subspace. The strength of the constraints have to be large compared to the energy of the local field energies in the system. In the following we derive both, upper and lower bounds for the constraint strength.

2.1. Minimal constraint problem

Our goal is to find the minimal strength of the constraints, such that lowest and first excited states of the physical and logical Hamiltonian coincide w.r.t. the parity translation. Intuitively, this can be understood as follows: in the extreme case, where the constraints are set to zero, each spin would point in the direction of the local field acting on the spin in order to minimize the energy of the system. On the contrary, if the constraints are infinitely large, these states are generally forbidden and the condition of an even number of spins up per plaquette is enforced. We consider the case of finite constraints where the local field term and the constraint energies are competing. In this case it might be energetically favorable to violate a constraint for rearranging the spins with respect to their local fields. Our goal is to find the lowest constraint energy such that this case can be ruled out. The minimal energy and the scaling w.r.t. the number of qubits depends on the statistics of the local fields which in turn is associated with classes of optimization problems. Therefore, we derive the minimal constraints for different classes of optimization problems.

We consider $n$ logical spins $[n] := \{1, \ldots, n\}$ with all-to-all connectivity. Thus, the system contains $m := (n(n-1))/2$ interactions that are mapped to $m$ spins in the parity scheme. The interaction strengths can be viewed as weights on the edges of a complete graph $K_n = (V_n, E_n)$ with $V_n := [n]$ and $E_n := \{(i,j) \in [n]^2, i < j\}$. We label the physical spins with elements of $E_n$ (cf figure 1(c)). Plaquettes are labeled by elements from $E_{n-1}$ and to distinguish them from sites, we replace the curly brackets () with square ones [:]. Furthermore, we denote the sample mean of a random variable $X$ by $\overline{X}$.

The space of physical states $\{-1, 1\}^n$ can be decomposed into a family of subspaces $(S_{\omega})_{\omega \subseteq E_{n-1}}$—according to which plaquettes they obey/violate the constraints. In the homogeneous case $c_0 = c$ we have that states with a different number of violated constraints belong to different energy sectors of $H$, and hence get different penalties from $H$, i.e. the number of ‘defects’ times $c$. In that sense $S_{\omega} := S_{\omega}^c$ should denote the logical subspace where the local constraints on every plaquette are satisfied. More general, given a tuple $\omega \subseteq E_{n-1}$ of plaquettes, the subspaces $S_{\omega}$ are defined as states being simultaneous eigenstates to the stabilizers $P_{(i,j)} = -\sigma_z^{(i)} \sigma_z^{(i+1)} \sigma_z^{(i+1)} \sigma_z^{(i+2)}$ with corresponding eigenvalues $1/2$ if $(i,j) \in \omega$ and $-1/2$ otherwise. Note, on the lowest row of plaquettes indexed with $j = i + 1$, the stabilizers $P_{(i,j)}$ are given by three-local terms $\sigma_z^{(i+1)} \sigma_z^{(i+2)} \sigma_z^{(i+1+j)} / 2$. In figure 1(c) we show an example for a state satisfying all constraints beside the one corresponding to the loop 23–34–42. That state belongs to the subspace $S_{[2,3]}$. Likewise figure 1(d) shows another state with two unsatisfied constraints i.e. being element of $S_{[2,3],[3,5]}$.  

3
The constraint strengths $c_{ij}$ can be chosen either to be all identical (homogeneous case) or we can individually set them to the optimal constraint strength for each plaquette. In the latter case, the objective cost function depends on all individual constraints $\text{cost}(c_{12}, \ldots, c_{n-1,n})$, which in the linear case is the sum of the constraint strengths. We define the value $a_k$ as the lowest eigenenergy of $H_f$ restricted to states belonging to the subspace $S_{\omega}$. With this definition, the problem of minimizing the constraints strengths can be written as a linear program: to this end, the cost-function has to be minimized under the restrictions

$$\sum_{(i,j) \in \omega} c_{ij} \geq -a_\omega + e, \quad \forall \omega \subseteq E_{n-1}, \omega \neq \emptyset,$$

where the value $e$ denotes the imposed lower energy threshold to state which disobeys the constraints.

In the homogeneous case $c_{ij} = c$, the linear program equation (2) reduces to

$$c = \max \left\{ e - a_1, \frac{1}{2}(e - a_2), \ldots, \frac{1}{q}(e - a_q) \right\},$$

were $a_k := \min\{a_\omega : \omega \subseteq E_{n-1}, |\omega| = k\}$ and $q$ denotes the number of plaquettes. More compact this can be written as

$$c = c(e) = \max_{|\omega|} \langle \psi | e - H_f | \psi \rangle,$$

Here, the penalty Hamiltonian $H_f$ does not discriminate between states with the same number of parity constraints violated. Hence, the denominator $\langle \psi | H_f | \psi \rangle$ counts the number of violated constraints regarding the state $|\psi\rangle$. Since $a_1$ is the lowest eigenenergy of $H_f$ w.r.t. the subspace $S_1 := S_{[1,2]} \cup \ldots \cup S_{[n-2,n-1]}$, where a single parity condition is unsatisfied, the corresponding state gets a penalty of $c$. This penalty has to be chosen large enough to bridge the gap between $a_1$ and $e$, which explains the first term in equation (3).

Similar to $S_1$ we define $S_k$ as the subspace of states with two unsatisfied parity constraints. $a_2$ is then given as the lowest eigenvalue of $H_f$ restricted to $S_2$. Since all these states will penalised twice by $H_f$, the strength of $c$ has to be at least half the difference of $a_2$ and $e$. This explains the second term in equation (3). Finally, other cases with $k > 2$ follow by including states with more than two unsatisfied parity constraints.

In general, every term appearing in equation (3) is of the form $c_{-k} := (e - a_k)/k$ and can be seen as a lower bound for the optimal constraint strength. To get upper bounds, we consider the fact that the spectrum of $H_f$ is contained in the interval $[p_0, -p_0]$ with boundaries $p_0 := -\sum_{|\omega|} |f_\omega|$. With the definition

$$c_1 := \max \left\{ c_{-1}, c_{-2}, \ldots, c_{-n}, \frac{1}{n+1}(e - p_0) \right\},$$

a series of upper bounds can be derived according to

$$c \leq c_0 \leq c_{-1} \leq \ldots \leq c_1 \leq c_0 \leq 2|p_0|. $$

Note, that we included the trivial bound $2|p_0|$ and defined $c_0 := e - p_0$ in equation (6). Two important energy thresholds $e$ are given by the energy of the logical ground state $b_0$ and the first excited logical state $b_0 + \Delta t$. These define via equation (4) constraints strengths: (a) $C_0 := |c(b_0)|$ strong enough to lift every state which does not belong to the logical subspace above the logical ground state and (b) $C_1 := |c(b_0 + \Delta t)|$ defined to penalize them further s.t. in addition the physical gap equals the logical gap.

### 2.2. Single violator approximation

In order to make the problem numerically more accessible, we focus on the first lower bound $c_{-1}$ rather than $c$. Thus, only states with one parity constraints violated are considered. We like to call them the single violators states (cf figure 1(c)). This is numerically well justified since for all models studied in this manuscript we observe the ordering

$$c \approx c_{-1} \geq c_{-2} \geq \ldots,$$

(cf figures 3 and 4).

It is easy to see, if the $f_\omega$ are $m$ i.i.d. random variables, the expected minimal constraint strength $\overline{c}$ cannot grow faster than quadratic in $n$, since by the central limit theorem we have

$$2|p_0| \rightarrow N(2m\mu_{abs}, \sigma_{abs}^2 m)$$

for $n \rightarrow \infty$, where $N(\mu, \sigma)$ denotes the normal distribution and $\mu_{abs}$ and $\sigma_{abs}^2$ are the mean and variance of the positive random variables $|f_\omega|$. Therefore, the trivial upper bound scales quadratic $2|p_0| = \Theta(n^2)$, and with $\overline{c} \leq 2|p_0|$ one further concludes that the minimal constraint strength cannot grow faster than quadratic
in $n$ i.e. $\mathcal{O} = \mathcal{O}(n^2)$ (note: a function $f(n) \in \mathcal{O}(n^2)$ means the function $f(n)$ grows asymptotically no faster than $n^2$. In contrast, in the big-$\Theta$ notation, $f \in \Theta(n^2)$ means $f(n)$ grows asymptotically as fast as $n^2$).

Furthermore, if $J_{ij}$ are i.i.d. random variables, with pdf $f_{\mu/\sigma}(x)$, the problem of determining the scaling of $\mathcal{T}$ does only depends on the ratio $\mu/\sigma$. This can be seen by noting that a rescaling of the pdf $f(x) \mapsto f(k^{-1}x)$ is equivalent to multiplying the random variables by a constant factor $J_{ij} \mapsto k J_{ij}$. Hence, the strengths of the optimal constraints are multiplied by an overall factor of $k$ whereas the functional dependency on the size, i.e. the scaling of the optimal constraints, is not affected. On the other hand, for each random variable it is true that $\langle k J_{ij} \rangle = k \langle J_{ij} \rangle$ and $\text{var}(k J_{ij}) = k^2 \text{var}(J_{ij})$ i.e. rescaling of $J_{ij}$ does not alter $\mu/\sigma$. In conclusion, the scaling of the optimal constraints can only depend on the ratio $\mu/\sigma$.

3. Results

Using the bounds equation (5) we evaluate the optimal constraints, for general ensembles of systems with different specific connectivity, bias and variance. In particular, the scaling of the average optimal constraint strength $\mathcal{T}$ with the system size for classes of problems. Let us first introduce two examples for typical optimization problems.

Let $G = (E, V)$ denote a simple graph. Then, the MaxCut problem asks for two disjoint sets of vertices $V_1$ and $V_2$ with $V_1 \cup V_2 = V$, such that the number of cutting edges is maximal i.e. $(e_1, e_2)$ with $e_1 \in V_1$ and $e_2 \in V_2$. As second example, the MinBisection (or graph-bipartitioning) problem for a graph with even number of nodes, requires to minimize the number of cutting edges while balancing the size of the two subsets $|V_1| = |V_2| = |V|/2$.

These graph partitioning problems can be easily mapped onto an Ising problem by introducing one spin per node. The MaxCut problem can be reformulated as an antiferromagnetic Ising model, i.e. $J_{ij} = 1$ for all $(i,j) \in E$, where the ground state corresponds to the solution of the optimization problem. If $b_i$ denotes the smallest eigenvalue of

$$H_{\text{MaxCut}} = \sum_{(i,j) \in E} \sigma_i^z \sigma_j^z,$$

then the maximal cut is given by $\text{cut}_{\text{max}} = (-b_i + |E|)/2$.

Similarly, the MinBisection problem can be encoded into an ferromagnetic Ising model with magnetization fixed to zero, i.e. $J_{ij} = -1$ for all $(i,j) \in E$, with $\sum \sigma_i^z = 0$. The corresponding Hamiltonian reads as

$$H_{\text{MinBisection}} = -\sum_{(i,j) \in E} \sigma_i^z \sigma_j^z + u \left(\sum \sigma_i^z\right)^2,$$

where the second term of equation (13) guarantees, that the magnetization of the ground state is zero, given the energy penalty $u$ is larger than $\text{min}(4d_{\text{max}}, n)/4$, with $d_{\text{max}}$ the maximal degree of $G$ [1].

3.1. Numerical results

The general case of randomly distributed $J_{ij}$ values with a given bias $\mu$ and standard deviation $\sigma$ can be treated numerically. In the following we investigate and compare three different distributions.

(a) Normal distribution $N(\mu, \sigma)$ with mean $\mu$ and variance $\sigma^2$.

(b) Uniform distribution on the interval $[a, b]$ with $\mu/\sigma = \sqrt{3}(a + b)/|a - b|$.

(c) A bimodal distribution with two possible assignments $\{c_1, c_2\}$, which are i.i.d with the expectation value of $J_{ij}$.

For our numerical results we sample from random instances of particular optimization problems, calculate their ground state and lowest single violator energies and interpolate the sample mean with a powerlaw fit. Finding $c_1$ involves minimisation over subspaces with a single parity defect. These single violator states we enumerate them by flipping spins starting from a state from the logical subspace (cf figure 1(c)). This allows us to do reasonable statistics up to sizes of $n = 25$. In the parity picture this corresponds to $m = 300$ physical spins and $q = 276$ plaquettes.

Assuming i.i.d. random variables $J_{ij} = J$, homogeneous constraint strengths $c_{ij} = c$ and single violator approximation $\mathcal{T} \approx \mathcal{T}_n$, the numerics in figure 2 suggests that the scaling of the minimal constraint strength mainly depends on $\mu$, the expectation value of $J$. The limits $\mu/\sigma \to \pm \infty$ are analytically well understood, showing a linear and a quadratic scaling respectively. Furthermore, at $\mu/\sigma = 0$ all three analyzed distributions show a linear scaling, including the SK-SpinGlass model where $J_{ij}$ are standard normal distributed random variables (cf figure 3). Note, that this behaviour may only occur in small systems, as we will further elaborate in section 3.2(c).
Figure 2. Scaling of the minimal constraint strengths $c$ for the particular cases when the coupling strengths $J_{ij}$ are chosen normally, uniformly or according to a bi-modal discrete distribution with expectation value $\mu$ and variance $\sigma^2$. The plot shows the exponent $\alpha$ obtained from the power-law fit $n \rightarrow \beta n^\alpha + \gamma$ onto the lower bound $c-1$ (single violator states) for simulations up to system sizes of $n = 25$. Solid lines show the obtained scaling if we define the minimal constraint to be s.t. the parity mapping preserves the first exited state energy, while the dashed line considers the relaxed version where we only care about the ground state energy. In that case, we only plot the exponent for negative $\mu$ since outside the gray shaded region they do not differ significantly. Notice that we did not assign an error bars because we found that the errors provided by the fitting routine is very small. It would be misleading to include some really small errors unless it is clear how finite size effect play a role in scaling, especially knowing the large-size scaling exponents should be bounded by two.

Figure 3. Numerical simulation for random instances for SK-SpinGlass problems i.e. $J_{ij}$ i.i.d. $\sim N(0, 1)$ setting the energy threshold $e = l_0 + \Delta l$ to be the first excited logical energy. The log–log-plot spans the logical system size from $n = 4$ to $n = 25$. The corresponding number of instances drop from $10^5$ for $n = 4$ to $64$ for $n = 25$. Shown are mean and variance for upper bounds (left) and for the lower bounds (right). This series of bounds are obtained by including states with one, two and three violated constraints respectively. The dash–dotted curve is the result from the fitting model $n \rightarrow \beta n^\alpha + \gamma$ applied on $c-1$.

As paradigmatic examples of combinatorial optimization problems—satisfying the assumption of independent random variables—we consider now two graph partitioning problems on random Erdős–Rényi graphs i.e. graphs where each edge has a fixed probability $p$ of being present.

(a) MaxCut: solving MaxCut for random graphs, corresponds to independently choosing $J_{ij}$ with probability $p$ to be either 1 or 0, respectively. Since all strengths fulfill $J_{ij} \geq 0$, it follows that $\mu/\sigma \geq 0$. 


Figure 4. (a) Numerical simulations of MaxCut instances based on random Erdős–Rényi graphs with $p = 0.4$. The energy threshold $\epsilon = l_0 + \Delta l$ is defined to be the first excited logical energy. Shown are variance and mean of upper bounds (left) and lower bounds (right) to the minimal constraint strengths. Furthermore, building on semidefinite programming, an efficient calculable lower bound is provided. (b) Numerical results for MinBisection instances on random graphs with $p = 0.5$ for upper and lower bound respectively.

Figure 5. (Left) Fully ferromagnetic and anti-ferromagnetic ground states and single violator states. The ground state of the ferromagnetic Hamiltonian, corresponds to collecting all vertices in the same set. Missing a single one gives the first exited state, which hurts $n−1$ of the interactions. Flipping the spin $(1,n)$ introduces a defect, which comes with an energy cost of $2$. (Right) For $n$ even, a solution to MaxCut split the set of vertices into two subsets of equal size. The difference between cutting edges (white circles) and in-set edges (black circles) grow linear $l_0 = −n/2$. The first exited state is obtained by splitting into sets with $n/2 − 1$ and $n/2 + 1$ elements respectively, which gives an additional linear energy shift of $2(n−1)$. The ground state single violator states are obtained by collecting the vertices into three sets, i.e. for $n = 6$ into $\{1,2\}, \{3,4\}$ and $\{5,6\}$. The pictographs show the large size pattern of the corresponding states. The numbers indicate the energy of the relevant states.

Figure 4 shows our numerical results for system sizes up to $n = 25$, where we find a sub-quadratic increase of $c_T$ with the system size. As we will see in the following analysis, this sub-quadratic scaling becomes quadratic for larger problem sizes. To this end we derive an efficiently calculable lower bound by utilizing a semidefinite program relaxation of finding the ground state configuration of Hamiltonian (9) with $\sigma_i^z \in \{-1,1\}$. If $\text{opt}_{\text{sdp}}$ denotes the optimal value of the semidefinite program

$$\max \sum_{(i,j) \in E} \frac{1 - X_{ij}}{2}, \quad X_{ii} = 1 \forall i \in [n], X \succeq 0$$

then $\text{cut}_{\text{max}} \leq \text{opt}_{\text{sdp}}[27]$.

To be computationally efficient, we upper bound the minimal single-violator contribution $a_1 \leq a_1^*$ by sampling from quadratically many single-defect states. To achieve this, the set of vertices $V = [n]$ is partitioned into three disjoint sets containing consecutive nodes $A = \{1,2,\ldots,k\}$, $B = \{k+1,\ldots,j\}$ and $C = \{j+1,\ldots,n\}$, where all sets contain at least two elements. The lower right part of figure 5, demonstrates what such a partition of $n = 6$ nodes into sets $A = \{1,2\}$, $B = \{3,4\}$ and $C = \{5,6\}$
looks like in the parity picture. Note, that this particular state has a single unsatisfied constraint at
plaque [2, 4]. With \( b_0 = 2 \leq l_2 \) and \( \text{cut}_{\max} = (-l_0 + |E|)/2 \) we can derive a lower bound on
\( c = l_2 - a_1 \geq c_{-1, \text{sdp}} \) by defining

\[
c_{-1, \text{sdp}} := -2 \cdot \text{opt}_{\text{sdp}} + |E| + 2 - a_1^+.
\]

Figure 4 includes this lower bound \( c_{-1, \text{sdp}} \) for the class of random graphs with edge probability \( p = 0.4 \)
up to system sizes of \( n = 100 \). We find, that after a sub-quadratic increase, the growth rate of \( \tau \) becomes
quadratic for large sizes \( n \).

(b) MinBisection: it turns out, that the special case of MaxCut problems on the complete graph \( K_n \), is key
to understand the behaviour of the minimal constraints in the MinBisection problem for large system
sizes \( n \). For fixed \( p, n \) scales at least linearly in \( n \). Hence, the inverse \( u^{-1} \) faster approaches zero than \( n^{-1} \).
This means, if \( n \) is large enough

\[
\frac{H_{\text{MinBisection}}}{u} \propto \sum_{(ij) \in E} \sigma_i^z \sigma_j^z - \frac{1}{u} \sum_{(ij) \in E} \sigma_i^+ \sigma_j^+,
\]

is well approximated by the first term \( \sum_{(ij) \in E} \sigma_i^z \sigma_j^z \), i.e. equals the Hamiltonian for MaxCut problems,
when restricted to the class of complete graphs \( K_n \). Therefore, for large \( n \), the expected value of \( c_{-1} \)
grows quadratic, independent of the choice of \( p \), as we will further elaborate in the subsequent section
(cf figure 4(b)).

3.2. Analytical results
Now we investigate two limiting cases \( (\mu, \sigma \to \pm \infty) \) where the minimal constraint problem can be solved
analytically. Furthermore, we discuss the case of normally distributed interaction strengths with \( \mu = 0 \) and
\( \sigma \neq 0 \).

(a) Antiferromagnetic limit: for \( \mu, \sigma \to \infty \) the minimal constraint problem can be connected to the
MaxCut problem on the complete graph \( K_n \). In this case all interactions are antiferromagnetic, and can
be set to \( J_{ij} = 1 \) due to the rescaling property mentioned above. Assuming \( n = |V| \) to be an even
number of vertices, then the maximal cut is given by \( (n/2)^2 \) and thus, the lowest eigenvalue of the
corresponding Ising Hamiltonian is given by \( b_0 = -n/2 \) and has a gap of 2 i.e. independent of \( n \). For
the instructive example of \( n = 3k \ (k \in \mathbb{N}) \), the minimal single violator energy is given by
\( a_1 = -\frac{2}{3} \left(1 + \frac{2}{a} \right) \). Then the lower bounds \( c_{-1} \) can be analytically found where the largest is given by

\[
c_{-1} = \frac{n^2}{6} + \begin{cases} 2, & \text{if } n = 3k \text{ for some } k \in \mathbb{N} \vspace{3pt} \\ 4, & \text{else} \end{cases}.
\]

The scaling of \( c = c_{-1} \) is therefore \( \Theta(n^2) \) i.e. quadratic in \( n \). For a graphical representation of the
relevant states see figure 5.

(b) Ferromagnetic limit: the limit \( \mu, \sigma \to -\infty \) is reached when setting \( J_{ij} = -1 \) i.e. all pair interactions are
ferromagnetic. In this case, the lowest eigenvalue is given by \( b_0 = -n(n-1)/2 \) and shows a gap of
\( 2(n-1). \) In the parity picture, violating a single constraint can be done at minimal cost of a single
spinflip. Starting from the ground state in \( S_\mathcal{F} \), flipping the spin \((1, n) \) results in a state from \( S_{[1,n-1]} \) with
energy \( b_0 + 2 \) and thus \( c = c_{-1} = 2n - 4. \) The scaling of \( c \) is therefore \( \text{linear} \) in \( n \). As we mention in
appendix A and especially in the discussion around figure 7 (upper panel), this scaling strongly
depends whether we choose our minimal constraints to be \( C_0 \) or \( C_1 \). Since we numerically observe a
maximum in success probability around \( c \approx C_0, \) we only need constant constraint strength to be
consistent with the definition given in [26].

(c) SK-SpinGlass: of special interest is the case \( \mu, \sigma = 0 \). The numerical results for \( \text{cut}_{-1} \) as function of \( n \)
is shown in figure 2. The following arguments suggest that in the Gaussian case the scaling of \( \text{cut}_{-1} \) goes as
\[ \sqrt{n \log(n)} \]
To begin with, we note, that for every fixed spin configuration \( \vec{\sigma} \in \{-1, 1\}^n \) the eigenvalues of
\( H_{\text{logic}}(\vec{\sigma}) \) can be seen as a random variable w.r.t. the distribution determining the interaction strengths.
If this distribution is Gaussian, then—according to the central limit theorem—the eigenvalues are also
normally distributed with variance \( \sigma^2 = n(n-1)/2 \).

However, the \( 2^n \) different eigenvalues are clearly not independent. They are strongly correlated with
covariance matrix [28]
\[
E(H_{\text{logic}}(\vec{\sigma}_2)H_{\text{logic}}(\vec{\tau}_2)) = \frac{1}{2} \left( \sum_{i=1}^{n} \sigma_i^j \tau_i^j \right)^2 - \frac{n}{2}.
\]  

(15)

Note, that finding the minimum \(\min_{\vec{\tau}} H_{\text{logic}}(\vec{\sigma}_2)\) is challenging due to the presents of these correlations.

In the case of independent random variables, the limiting order statistics can be classified into one of three universally classes according to extreme value theory [29]. Moreover, for \(m\) Gaussian variables \(\{G_1, \ldots, G_m\}\) the limiting distribution of their minimum \(M = \min(G_1, \ldots, G_m)\) is given by the Gumbel distribution \(\text{Gumbel}(\alpha, \beta)\) with parameters

\[
\alpha = F^{-1} \left( 1 - \frac{1}{m} \right), \quad \beta = F^{-1} \left( 1 - \frac{1}{em} \right) - \alpha.
\]

(16)

Here \(e\) is the Euler constant and \(F^{-1}\) denotes the quantile function, which in the case of standard normal variables is given by the probit \(\sqrt{2}\text{erf}^{-1}(2p - 1)\). By fixing the distribution, the parameters \(\alpha\) and \(\beta\) only depend on the number of variables \(m\). We want to emphasise, that setting \(m = 2^n\) models all eigenvalues as independent. Since the factor \(\sigma^{-1}\) normalizes each eigenvalue, \(\overline{M}_{\text{ind}} := -\sigma(\alpha + \Gamma\beta)\) is the expected minimal energy in the independent case (\(\Gamma\) is the Euler–Mascheroni constant).

Numerically we see, that choosing the eigenvalues to be independent, results (on average) in too small energies \(\overline{I}_0 > \overline{M}_{\text{ind}}(m = 2^n)\). However, we observe that by introducing a free parameter \(\delta < 1\) in order to decrease the number of realizations \(m = 2^{\delta n}\) gives a reasonably good approximation \(\overline{I}_0 \approx \overline{M}_{\text{ind}}(m = 2^{\delta n})\) for \(\delta \approx 0.798\) 158(4) (cf. figure 8) [30]. If one allows \(\delta\) to be a function of \(n\) then the statement is trivial, but interestingly a constant \(\delta\) gives rise to a good approximation. To which extend our analysis captures the large \(n\) behaviour is outside the scope of the present work. However, as we show in the appendix \(\overline{M}_{\text{ind}} \approx -\sqrt{\delta} \log(2)n^2\) for large \(n\). Since \(\sqrt{\delta} \log(2) \approx 0.743\) this is in accordance with Parisi’s result [31, 32]

\[
\overline{I}_0 = (-0.763167 \ldots + o(1))n^2.
\]

(17)

On the other hand, finding \(a_{ij}\) corresponds to finding the groundstate of the physical Hamiltonian \(H_{\text{phys}}\) restricted to the subspace of single violator states. Finding the single violator state w.r.t. the plaquette indexed by \([k, l]\) is equivalent to finding the ground state of \(H_{[k, l]} := \sum J_{ij}\sigma_i^j\sigma_l^j\) (cf. figure 1(c)), with

\[
J_{ij}^k := \begin{cases} -J_{ij} & \text{if } i \leq k \text{ and } j > l \\ J_{ij} & \text{else} \end{cases}.
\]

(18)

Hence, minimizing over all plaquettes gives the minimal single-violator energy \(a_1\). Due to symmetry, \(-J_{ij}\) is again a standard normal Gaussian variable and therefore the eigenenergies associated to single violators are distributed according to \(N(0, \sigma^2)\). Similar to equation (15) these eigenstates are highly correlated, but nevertheless we find a numerically well justified approximation \(\overline{r} \approx \overline{M}_{\text{ind}}(m = 2^{\delta n}(n + 1)/12)\), with \(\delta\) as above. Here, the quadratic terms count the average number of spins that have to be flipped to induce a parity defect according to figure 1(c). As we show in appendix B, the energy difference between single violator ground states and logical ground state \(\overline{I}_0 - a_1\) scales to leading order in \(n\) as

\[
f_1(n) := \frac{1}{2\sqrt{\delta} \log(2)} \sqrt{n} \log \left[ \frac{n(n+1)}{12} \right].
\]

(19)

Since the gap \(\Delta_1\) can be neglected for large \(n\), the scaling of \(\overline{r} = \overline{I}_0 - a_1\) (cf. figure 8).

Similar arguments can be used for the scaling analysis of the remaining lower bounds \(\overline{I}_0 - a_k\) (cf. figure 8).

Here, the subspace \(S_k\) spanned by states with \(k\) parity constraints violated has \(O(n^k)\) elements. Setting \(p_k(n)\) to a polynomial of order \(2k\) and modelling \(\overline{r}\) as a \(2^n p_k(n)\) i.i.d. Gaussian variables results in a scaling of \(c_{-k}\) as \(\sqrt{\delta} \log(2^{n}k) \sim \sqrt{n} \log(n)\) analog to equation (19).

(d) General case: finally, let us discuss the case when \(J_{ij}\) are neither centered distributed random variables nor close to the limits \(\mu/\sigma = \pm \infty\). In general the bound \(c_{-1}\) depends on the lowest eigenvalue \(b_0\), the gap \(\Delta_1\) and the smallest single violator energy \(a_1\).

If \(\mu \neq 0\), the distributions of eigenenergies of the logical Hamiltonian are shifted in contrast to the previously considered case of \(\mu = 0\). As an example we consider in the following Gaussian distributed couplings \(J_{ij} \sim N(\mu, 1)\). Our argument builds on the fact that two normal variables with parameters \((\mu_1, \sigma_1^2)\) and \((\mu_2, \sigma_2^2)\) add up to a single normal variable \(N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)\). Thus, the eigenvalue corresponding to the all-ones logical state \((1, \ldots, 1)\) is distributed via \(N(\mu m, m)\), where \(m\) denotes the
number of physical spins. Likewise, the distribution of a state with equally many $-1$ and $+1$ is centered around zero $\sim N(0, m)$. More generally, for fixed $k$, there are $2 \binom{n}{k}$ combinations of eigenvalues, where every combination is distributed according to a Gaussian with mean

$$\mu_k = \mu \frac{(n-2k)^2-n}{2}$$

and variance $m$. This ‘splitting’ is the reason for the different behaviour of the smallest eigenvalue in the two cases $\mu > 0$ and $\mu < 0$.

In the case $\mu > 0$, the probability that the lowest energies originate from one of the $2 \binom{n}{k}$ states centered around zero is largest. Since these are exponentially many states, we assume the expectation value for the ground state energy to be similar to the $\mu = 0$ case. The single violator ground state for the limit $\mu/\sigma \to \infty$, denoted by $\phi_{\infty}$ (cf figure 5 (bottom)), is Gaussian distributed w.r.t. $f_l$ with mean $-\mu n^2 \left( \frac{1}{n} + O(n^{-1}) \right)$ and variance $m$. Therefore, the single violator ground-state energy lies on average quadratically deeper than the logical ground state energy.

On the contrary, for $\mu < 0$, the larger $|\mu|$ is, the more likely it is, that the smallest eigenvalue is one of the ferromagnetic states $(1, \ldots, 1)$ or $(-1, \ldots, -1)$. As shown, in the total ferromagnetic case, the gap $\Delta_k$ grows linearly with $n$. Since $b_0 - a_1 = 2$ this linear increase is the main contribution to $c_1$. In the large-$n$ limit we expect that the gap scales linearly. Since our numerical simulation is limited up to $n = 25$, we see this linear behaviour only if $|\mu|$ is large enough. As one can further see in figure 2 as $\mu$ gets smaller, the scaling exponent drops due to the fact that the expected difference $b_0 - a_1$ tends to get smaller for more negative $\mu$. However, by further increasing $|\mu|$ the linear scaling of the gap becomes apparent.

4. Discussion

We have shown numerically and analytically the scaling of the minimal constraint strengths in the parity based encoding for various classes of optimization problems. In the parity scheme, the optimization problem is encoded in the local fields only, and the addition of a penalty term favours the logical state too. For the models we studied, we noted, the choice of definition matters if the class of Hamiltonians has ordered ground states and where the gap grows linearly with system size. This is true for the extreme case of ferromagnetic interactions and for a range of random models with i.i.d. couplings with presumable negative couplings. Here, the optimal constraints are close to $C_0$.

On the other hand figure 2 suggests that the logical gap is neglectable for the large size scaling in the studied i.i.d. models with $\mu/\sigma > -0.5$. Here, the ground states are expected to be highly frustrated. We analyzed the Sherrington–Kirkpatrick-model and concluded (at least for small system size) the optimal constraint strength in terms of an optimal fidelity lies around $C_1$. Same holds for the anti-ferromagnetic model on the full graph (cf figure 7). From the simulations although it not clear how these findings scale up. At least for the SK-model it seems the best strengths lie in-between $C_0$ and $C_1$. If the maximal fidelity is size-independent best for constraint strengths around $C_1$, the scaling laws reported in this paper would be translated into scaling law for optimal strengths when optimality is measured in terms of success probability. In the region, where the logical gap is neglectable compared the quadratically deep lying single-violator states, obviously, both definitions via $C_0$ and $C_1$ lead to the same big-size scaling behaviour.

Hence, it does not matter which definition is used. Although, it could be that with growing system size, the best fidelity could be obtained for constraint strengths larger than $C_1$ but our simulation on small systems don’t support this possibility. Nevertheless, for problem classes where this holds our method would still provides a lower bound on the exact behaviour. When considering real optimization problems the problems tend to be frustrated ones and we think the same arguments are applicable.

For the i.i.d. models that we have examined, we have found that the large size scaling of the optimal constraints is mainly determined by the sign of the bias $\mu$. The observed differences stem from two distinct effects: (a) linear growth of the gap and (b) quadratic deep lying single violator states. The scaling of the gap is the important factor for problems with predominantly negative couplings. We found here, preserving the gap may be counterproductive in terms of optimal fidelity and optimal schedules may be archived with constant constraint strengths. On the other side, properties of the single violator states govern the regime...
with predominately positive couplings. We find the optimal constraints tend to grow quadratic in system size as depicted in figure 2.

Acknowledgments

The authors thank Kilian Ender and Clemens Dlaska for the useful discussions and comments on the paper. Work was supported by the Austrian Science Fund (FWF) through an START Grant under Project No. Y1067-N27 and the SFB BeyondC Project No. F7108-N38, the Hauser-Raspe foundation, and the European Union’s Horizon 2020 research and innovation program under Grant Agreement No. 817482. This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No. HR001120C0068. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of DARPA.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Appendix A. Annealing and fidelity

In the model of adiabatic quantum computation, a quantum system prepared in the ground state of the initial Hamiltonian \( H_{\text{init}} \) is drawn into the ground state of a final Hamiltonian \( H_{\text{probl}} \) through adiabatic evolution. The final Hamiltonian or problem Hamiltonian itself encodes the optimization problem \([3, 7]\). For the computation, the time-dependent Hamiltonian interpolates between \( H_{\text{init}} \) and \( H_{\text{probl}} \) via

\[
H(t) = (1 - s(t))H_{\text{init}} + s(t)H_{\text{probl}}, \tag{A1}
\]

where \( s(t) \) changes from 0 to 1 as \( t \) runs from \( t_0 \) to \( t_0 + T \). Here, \( T \) defines the run-time of the computation. If this process is performed slowly enough i.e. \( T \) is big enough, the quantum adiabatic theorem will ensure the system stays in its instantaneous ground state \([7]\). At the end of the sweep, the ground state of \( H_{\text{probl}} \) is successfully prepared and the optimal solution of the optimization can be obtained by measuring the final state. In the case of quantum annealing \([4]\) the initial Hamiltonian is often referred to as the driver Hamiltonian and is of the form \( \sum_i \sigma_x \), where \( \sigma_x \) denotes the Pauli operator acting on qubit \( i \). The ground state of this driver Hamiltonian can be easily prepared since it’s a product state \( |+\rangle \), with \( |+\rangle = \frac{1}{\sqrt{2}} (|0\rangle + |1\rangle) \) the plus-state. The problem Hamiltonian \( H_{\text{probl}} \) is usually a two-local Ising Hamiltonian encoding some QUBO problem and in our case coincides with the physical Hamiltonian \( H_{\text{phys}} \) implementing the LHZ parity encoded version of the optimization problem. In contrast to a perfect annealer with infinite running time \( T = \infty \), a real annealer operates on a finite time scale. This induces diabatic transitions to higher states s.t. the annealing process does not deterministically find the ground state of the problem Hamiltonian. If we neglect any source of noise the evolution under the Hamiltonian equation (A1) is governed by Schrödinger’s equation

\[
|\psi(t,c)\rangle = T \exp \left( -\frac{i}{\hbar} \int_0^T H(t')dt' \right) |+\rangle^\otimes n_{\text{phys}}. \tag{A2}
\]

We measure the success of the annealing schedule by the fidelity i.e. the overlap between the final state under Schrödinger’s evolution and the ground state of the physical Hamiltonian \( \langle p_0 | \psi(t,c) \rangle \) by

\[
\mathcal{F}(c) := |\langle p_0 | \psi(t,c) \rangle|^2. \tag{A3}
\]

In the relevant regime when \( c \geq C_0 \), the lowest physical ground state coincides with the logical ground state and is independent of \( c \). Consequently, \( \mathcal{F}(c) \) measures the success of solving the logical optimization problem. Another parameter of interest is the minimal gap \( \Delta_{\text{min}} \) i.e. the smallest energy difference between instantaneous ground state and corresponding first exited one.

(a). **SK-model** figure 6 shows numerical simulation for small system sizes of five logical qubits. The corresponding parity encoding requires ten physical qubits. We simulate the dynamics given by equation (A2) and calculate fidelity and minimal gap for random SK-problems i.e. the \( J_{ij} \) are i.i.d. standard Gaussian distributed. For different annealing times \( T \in \{5, 10, 50\} \) and different constraint strengths \( c \) we sample trough 1024 instances. Of particular interest are two cases: \( c = C_0 \) and \( c = C_1 \). In the first case we penalize all unwanted state s.t. they are higher than the logical ground state. In the
latter case, we increase the penalty s.t. in addition the first exited energy in the parity picture corresponds to the first exited logical energy. Note, the computation of $C_0$ and $C_1$ can be done by equation \( (A4) \) and is exponential in the system size $n$. By definition, as the constraint strength $c$ approaches $C_0$ from above, the minimal gap $\Delta_{\text{min}}$ has to go to zero. Numerically, we see: increasing $c$ s.t. all unwanted states are gaped away from the logical ground state, opens up the minimal gap. Furthermore, we conclude: on average, the minimal gap peaks around $C_1$ and continues to decrease for penalties above $C_1$ as depicted in figure 6 (middle) where mean, median (i.e. 0.5-quantile) and 0.25- and 0.75-quantiles for the minimal gap is shown.

As a measure of computational success, the fidelity is more relevant than the minimal gap. Both quantities are tightly connected, but the fidelity does not only depend on the minimal gap but also on the annealing time and more concrete on the whole annealing schedule. Going with different speed through a linear annealing schedule we observe in figure 6 (left) a maximal fidelity around $c \approx C_1$ and a decrease towards $C_0$. This can be explained due to ‘leakage’ into unwanted states, which are in energy nearby the logical ground state if the constraint strength is too close to $C_0$. On contrast, a big penalty strength results into a big spectral radius of the final Hamiltonian, which leads to a smaller minimal gap.

Finally, we investigate the effect of the system size in figure 6 (right). At fixed annealing time we vary the system size from three to seven logical qubits. We note a slow shift of the best strength from initially around $C_1$ for $n_{\text{logi}} = 3$ towards $C_0$ for $n_{\text{logi}} = 7$. But, due to the high computational costs a fully quantum mechanical simulation has, we dropped the number of instances from 1024 for $n_{\text{phys}} = 3$ to 32 in the case of 21 physical spins. Hence, it is hard to tell from the data if the trend continues for bigger sizes. If one fixes the anneal-time and alters the system size the overall performance drops. In order to get a consistent probability of success the anneal time has to increase with growing system size. A longer sweep of $T = 50$ on $n_{\text{logi}} = 6$ qubits has a nearly identically penalty-fidelity-graph as it is the case with a faster sweep of $T = 10$ at system size $n_{\text{logi}} = 4$ (not depicted here).

As interesting and important edge cases, we investigated in figure 7 the ferromagnetic and anti-ferromagnetic problem on $K_5$ i.e. the full Graph on five vertices.

(b) **Anti-ferromagnetic:** as the fully anti-ferromagnetic problem equals the MAX-CUT problem on $K_5$, the ground states are highly frustrated. Moreover, the ground state is not unique since every partition into $3 \pm 2$ vertices is a valid solution. Computing the fidelity in equation \( (A3) \) we have to take this degeneracy into account and sum over all relevant states. Figure 7 (lower panel) shows the success probability to end up in one of the ground states in dependency of the penalty strength for different annealing times $T \in \{5, 10, 50\}$. As in the case of the SK-model we observe a maximal fidelity for constraints around $c \approx C_1$. 

**Figure 6.** Annealing performance for random SK-instances. (Left and middle) System size of five logical spins with $2^{20}$ random SK-instances. The middle panel shows the minimal gap during a linear annealing sweep in dependence of the constraint strength. Shown are mean, median, 0.25- and 0.75-quantil of the obtained distributions. Important landmark strengths are $C_0 := c(l_0)\) (in equation \( (4) \) —when every unwanted state is penalized at least to the eigenenergy of the logical ground state and $C_1 := c(l_0 + \Delta)$—the required strength to push them above the first exited logical energy. (Left) The fidelity calculated via equation \( (A3) \) for three different annealing times $T \in \{5, 10, 50\}$ when performing a linear annealing schedule. (Right) The annealing time $T = 10$ is fixed to show the effect of growing sizes $n_{\text{phys}} \in \{3, 6, 10, 15, 21\}$. For the smallest sizes 1024 random instances where drawn and for the biggest two 256 and 32 respectively.
Figure 7. Fidelity in dependence of the parity constraint strength $c$ for the ferromagnetic and the anti-ferromagnetic problem on $K_5$. Horizontal axis labeling is the same as in figure 6. Curves for three different annealing times $T \in \{5, 10, 50\}$ are shown.

(c) Ferromagnetic: the fully ferromagnetic model has two ordered ground states $|0,...,0\rangle$ and $|1,...,1\rangle$. The parity encoding does not discriminate between them and maps both onto the same state. As depicted in figure 5 (left panel) the corresponding ground state in the parity model is given by the all-spins-up state $|0\rangle_{\text{phys}} \otimes |1\rangle_{\text{phys}}$. This is simultaneously the ground state of the parity Hamiltonian equation (1), even in the case of zero penalty $c = 0$. The lowest state with one violated constraint can be obtained at the cost of one single spin flip. All other states with violated constraints lie higher, s.t. even negative penalty strength $c$ can be allowed and still satisfying the condition of not having unwanted states below the logical ground state. In the ferromagnetic case, where the ground state is ordered and has no frustration, the success fidelity is best when $c \approx C_0$ and drops for higher constraint strengths (cf figure 7 (upper panel)).

Appendix B. SK-model details

The expected minimum of $m$ Gaussian variables is given as the mean of the corresponding Gumbel distribution via $M_{\text{ind}} = -\sigma(\alpha + \Gamma \beta)$ with parameters as in equation (16). Including the known quantile-function for Gaussian variables one has

$$M_{\text{ind}} = -\sqrt{m(m-1)} \left[ (1-\Gamma) \text{erf}^{-1} \left( 2 \left( 1 - \frac{1}{m} \right) - 1 \right) + \Gamma \text{erf}^{-1} \left( 2 \left( 1 - \frac{1}{em} \right) - 1 \right) \right].$$  \hspace{1cm} (B1)

As described in the main text, we incorporate the correlations between the eigenvalues of the SK-model by introducing a factor $\delta$ and observing $l_0 \approx M_{\text{ind}}(m = 2^m)$ for a fixed size independent $\delta$ (cf figure 8).

In order to motivate this simplified idea, we want to derive the large $n$ scaling and compare it to a well known result of Parisi [31]. For $x$ close to one, the inverse error function can be approximated by $\text{erf}^{-1}(x) \approx \sqrt{-\log(1-x^2)}$ [33]. With $x = 2 \left( 1 - 2^{-2m} \right) - 1$ we have $1 - x^2 = (2^{2m} - 1) 2^{-2mn}$ and after decomposing the log of the product as the sum, we neglect the $-1$ over the $2^{2m}$ and arrive at

$$\alpha \approx \sqrt{(\delta n - 2) \log(2)}. \hspace{1cm} (B2)$$

Doing the same for the second inverse error function term in equation (B1) results into

$$\beta + \alpha \approx \sqrt{(\delta n - 2) \log(2) + 1}, \hspace{1cm} (B3)$$
where the additional 1 comes from $\log(e) = 1$. Neglecting this contribution gives $l_0 \approx -\sigma \alpha$ i.e.

$$l_0 \approx -\sqrt{n(n-1)} \sqrt{2n-2} \log(2). \quad \text{(B4)}$$

After expanding both square root expressions to leading order it holds

$$\overline{l_0} \approx -\frac{2}{3} \log(2) n^2 \approx -0.7436 \cdot n^2. \quad \text{(B5)}$$

This motivates our approach when we compare it to asymptotic behaviour obtained by Parisi in equation (17).

As discussed in the main text, finding the minimal single violator energy $a_1$ can be done by solving a family of ground state problems with Hamiltonians $H_{k,l}$, $k < l$ defined via equation (18). If we denote by $q = (n-2)(n-1)/2$ the number of plaquettes, there are $q$ possible ways to violate a single constraint. Furthermore, the standard Gaussian distribution is symmetrically centered around zero, hence every eigenvalue of the Hamiltonians $H_{k,l}$ is again distributed according $N(0, \sigma^2)$. One could make two assumptions: (i) the minimum of $2^{n/2}$ independent variables corresponds to the expected lowest eigenvalue of the SK-model and (ii) the eigenvalues of different $H_{k,l}$ are independent of each other. As conclusion, one would expect: a number of $2^{n/2}q$ independent random variables distributed according $N(0, \sigma^2)$ can give us the expected smallest single violator energy. But, assumption (ii) leads to way to low estimates for $\overline{l_0}$ (the $H_{k,l}$ are clearly not independent of each other). Following this approach an upper bound for $l_0 - a_1$ is obtained. As described in the main text we found, choosing $m = 2^{nh} p(n)$ Gaussian’s $\sim N(0, \sigma^2)$ with the quadratic polynomial $p(n) = n(n+1)/12$ is well suited for modeling the expected single violator ground state energy (cf figure 8 for a comparison with the numerical data). Actually, the concrete form of $p(n)$ does not even play a role. Each (quadratic) polynomial leads to the same functional large $n$ behaviour for $l_0 - a_1$.

Setting $m = 2^{nh} p(n)$ in equation (B1) and using the same approximations as before, one derives

$$\overline{a_1} \approx -\sqrt{n(n-1)} \sqrt{2n-2} \log(2) + \epsilon, \quad \text{(B6)}$$

with $\epsilon := \log(p(n))$. Using $\sqrt{1+x} \approx 1 + x/2$ for small $x$ to approximate the square root expressions in equations (B4) and (B6) leads to

$$\overline{l_0} - \overline{a_1} \approx \sqrt{n(n-1)} \frac{\epsilon}{2 \sqrt{2n \log(2)}}. \quad \text{(B7)}$$

Another Taylor expansion of $\sqrt{n(n-1)}$ reveals the leading order behaviour of $\Theta(\sqrt{n} \log(n))$ equation (19).
References

[1] Lucas A 2014 Front. Phys. 2 5
[2] Farhi E, Goldstone J and Gutmann S 2014 A quantum approximate optimization algorithm (arXiv:1411.4028[quant-ph])
[3] Farhi E, Goldstone J, Gutmann S and Sipser M 2000 Quantum computation by adiabatic evolution (arXiv:quant-ph/0001106)
[4] Susa Y, Yamashiro Y, Yamamoto M, Hen I, Lidar D A and Nishimori H 2018 Phys. Rev. A 98 042326
[5] Kadowaki T and Nishimori H 1998 Phys. Rev. E 58 5355–63
[6] Kadowaki T 2002 Study of optimization problems by quantum annealing (arXiv:0203020[quant-ph])
[7] Albash T and Lidar D A 2018 Rev. Mod. Phys. 90 015002
[8] Hauke P, Katzgraber H G, Lechner W, Nishimori H and Oliver W D 2020 Rep. Prog. Phys. 83 054401
[9] Katzgraber H G, Hanne F, Zhai Z, Ochao A J and Munoz-Bauza H 2015 Phys. Rev. X 5 031026
[10] Lidar D A et al 2014 Science 345 420–24
[11] Mbang G B, Privitera L, Arceci L and Santoro G E 2019 Phys. Rev. B 99 064201
[12] Santoro G E 2002 Science 295 2427–30
[13] Dickson N G and Amin M H S 2011 Phys. Rev. Lett. 106 050502
[14] Hartmann A and Lechner W 2019 New J. Phys. 21 043025
[15] Hartmann A and Lechner W 2019 Phys. Rev. A 100 032110
[16] Zhou L, Wang S-T, Choi S, Pichler H and Lukin M D 2020 Phys. Rev. X 10 021067
[17] Farhi E and Harrow A W 2016 Quantum supremacy through the quantum approximate optimization algorithm (arXiv:1602.07674[quant-ph])
[18] Wecker D, Hastings M B and Troyer M 2016 Phys. Rev. A 94 022309
[19] Choi V 2008 Quant. Inf. Process. 7 193–209
[20] Choi V 2010 Quant. Inf. Process. 10 343–53
[21] Bunyk P I et al 2014 IEEE Trans. Appl. Supercond. 24 1–10
[22] Lechner W, Hauke P and Zoller P 2015 Sci. Adv. 1 e1500838
[23] Lechner W 2018 Quantum approximate optimization with parallelizable gates (arXiv:1802.01157[quant-ph])
[24] Glaetzle A W, van Bijnen R M W, Zoller P and Lechner W 2017 Nat. Commun. 8 15813
[25] Leib M, Zoller P and Lechner W 2016 Quantum Sci. Technol. 1 015008
[26] Albash T, Vinci W and Lidar D A 2016 Phys. Rev. A 94 022327
[27] Goemans M X and Williamson D P 1995 J. ACM 42 1115
[28] Panchenko D 2012 J. Stat. Phys. 149 362–83
[29] Kotz S and Nadarajah S 2000 Extreme Value Distributions (Singapore: World Scientific)
[30] By curiosity, we note that value is really close 2/π = 0.797884... The value 2/π can be related to a normal distribution: If X is a standard normal variable, then |X| defines another random variable with mean 2/π.
[31] Parisi G 1980 J. Phys. A: Math. Gen. 13 115–21
[32] Talagrand M 2006 Ann. Math. 163 221
[33] Streleck A J 1968 Math. Comput. 22 144