Degeneracy, degree, and heavy tails in quantum annealing

Andrew D. King, 1,* Emile Hoskinson, 1 Trevor Lanting, 1 Evgeny Andriyash, 1 and Mohammad H. Amin 1,2

1 D-Wave Systems Inc., Burnaby B.C.
2 Department of Physics, Simon Fraser University, Burnaby B.C.

(Dated: April 27, 2016)

Both simulated quantum annealing and physical quantum annealing have shown the emergence of “heavy tails” in their performance as optimizers: The total time needed to solve a set of random input instances is dominated by a small number of very hard instances. Classical simulated annealing, in contrast, does not show such heavy tails. Here we explore the origin of these heavy tails, which appear for inputs with high local degeneracy—large isoenenergetic clusters of states in Hamming space. This category includes the low-precision Chimera-structured problems studied in recent benchmarking work comparing the D-Wave Two quantum annealing processor with simulated annealing. On similar inputs designed to suppress local degeneracy, performance of a quantum annealing processor on hard instances improves by orders of magnitude at the 512-qubit scale, while classical performance remains relatively unchanged. Simulations indicate that perturbative crossings are the primary factor contributing to these heavy tails, while sensitivity to Hamiltonian misspecification error plays a less significant role in this particular setting.

I. INTRODUCTION

The recent development of quantum annealing (QA) processors has spurred research into how to construct input instances—Ising spin instances—that might confirm or refute any purported computational advantage conferred by such hardware [1–10]. These recent works have proposed various important considerations such as error sensitivity, thermal hardness [10], ground state degeneracy, consistency of energy scale across a random ensemble [6, 7], potential barrier shape [4, 5], and the existence of classical phase transitions [8]. In parallel to this line of research is the consideration of suitable performance metrics for exact and approximate optimization [9, 11, 12].

Several previous efforts to quantify the performance of D-Wave hardware relative to physically motivated software solvers have focused on the median case performance for a given input class [2, 6]. However, Ref. [12] cautions that the median case does not give a complete picture of a given solver’s performance. Simulated quantum annealing (SQA) [13, 14], which has been advanced as an effective model for D-Wave hardware, shows dramatically varying performance over particular random ensembles of low precision problems: the total time required to solve all instances is dominated by a few instances at the tail of the distribution. In contrast, simulated annealing (SA) does not show nearly as strong of a variation in performance when run on the very same instances. The presence of so-called heavy tails in the distribution of the time to solution, as seen in both SQA simulations [12] and D-Wave hardware results [2, 14], points to an intrinsic physical mechanism that could explain the difference in performance between classical-physics-based approaches such as SA and the quantum-physics-based approaches such as SQA and D-Wave hardware.

In this paper we examine the hypothesis that local degeneracy—the abundance of large isoenenergetic clusters of spin states—plays a crucial role in the appearance of heavy tails. We compare the performance of a D-Wave 2X quantum annealing processor (DW2X) [11] with SA on random ±1 Ising spin glass instances with high local degeneracy and low local degeneracy, where local degeneracy is controlled via parity of qubit connectivity rather than selection of coupling strengths [9, 15]. Our results indicate that heavy tails are a consequence of degeneracy in the final target Hamiltonian and a particular implementation of (S)QA wherein all qubits follow an identical annealing schedule. Under these circumstances, small gap perturbative anticrossings appear in the system eigenspectrum late in the anneal, at which point tunneling is suppressed. These circumstances arise when random low-precision Ising spin glass instances are imposed on the full DW2X qubit connectivity graph. However, they do not arise when the connectivity of that graph is varied as described herein, and they become less common when more realistic [16–19] or combinatorially interesting problem classes are subject to (S)QA [3, 20].

II. QUANTUM ANNEALING PROCESSOR

A. Hamiltonians in the Ising model

Quantum annealing in the Ising model aims to find low-energy states in a system of n interacting spins via evolution of the time-dependent Hamiltonian

\[
\mathcal{H}_S(s) = \frac{\mathcal{E}(s)}{2} \mathcal{H}_P - \frac{\Delta(s)}{2} \sum_i \sigma_i^z
\]

\[
\mathcal{H}_P = \sum_{i<j} J_{ij} \sigma_i^x \sigma_j^x + \sum_i h_i \sigma_i^z.
\]

(1)

Here \( s \in [0, 1] \) is the annealing parameter indicating progress through the anneal; \( \sigma_i^x \) and \( \sigma_i^z \) are Pauli ma-
trices acting on spin $i$; $\Delta(s)/2$ and $\varepsilon(s)/2$ are the energy scales corresponding to the transverse and longitudinal fields, respectively, and satisfy $\Delta(0) \gg \varepsilon(0)$ and $\Delta(1) \ll \varepsilon(1)$; $\mathcal{H}_P$ is the Ising problem Hamiltonian [21]. An isolated qubit with a bias $h = 1$ has an energy gap of $\Delta(s) + \varepsilon(s)$ between the two eigenvalues of its Hamiltonian $\mathcal{H}_S(s)$, where $\Delta(s)$ is the contribution due to tunneling and $\varepsilon(s)$ is the contribution due to the Ising Hamiltonian.

B. Floppy qubits and degeneracy in $J = \pm 1$
Chimera-structured Ising instances

Qubits in a D-Wave processor are connected as a Chimera graph [22] (see Fig. 1), in which most qubits are coupled to six others, provided that only a few qubits and couplers are inoperable. For a classical $n$-qubit state $\vec{s} = \{s_1, \ldots, s_n\} \in \{-1,1\}^n$, each qubit $s_i$ experiences an effective field

$$h_i^{\text{eff}}(\vec{s}) = h_i + \sum_{j > i} J_{ij}s_j.$$  \hspace{1cm} (2)

In the case $h_i^{\text{eff}}(\vec{s}) = 0$, we say that qubit $s_i$ is floppy in $\vec{s}$; equivalently, changing $s_i$ does not change the energy of the state. Floppy qubits and the resulting degeneracy have a well-understood contribution to perturbative crossings [23, 24], and have been studied from the perspective of differentiating heuristics that operate over time-dependent quantum versus thermal annealing potential [25, 26]. In short, floppy qubits lead to wide, degenerate valleys in the state space that are favored in the time-dependent quantum potential early in the anneal. We now illustrate the effect of floppy qubits on degeneracy in the random Ising instances often used to study D-Wave processors [1, 2, 11, 12, 14].

Let $(h,J)$ be a random Ising spin glass instance with $h = 0$ and with each operable coupler $J_{ij}$ taking a value in $\{-1,1\}$ uniformly at random. A qubit $s_i$ coupled to $d$ others (i.e. of degree $d$), has $\Pr[h_i^{\text{eff}}(\vec{s}) = 0] = \binom{d}{d/2}/2^d$ when $d$ is even and $\Pr[h_i^{\text{eff}}(\vec{s}) = 0] = 0$ when $d$ is odd (Fig. 2). Qubits of degree 2, 4, and 6 have respective probabilities 50%, 38%, and 31% of being floppy in a random spin configuration. This probability approaches zero for large degree. To put these probabilities in perspective, in a 512-qubit Chimera-structured instance there are 384 degree 6 qubits, of which we expect 120 to be floppy in a random state. This results in large isoenergetic clusters: due to the bipartite structure of Chimera, we expect every qubit in such a state to be in an isoenergetic hypercube of size $2^{60}$ [27]. Degeneracy will be less severe in local minima, where effective fields are not randomly distributed—for highly degenerate instances studied, roughly 5–10% of qubits in low-lying excited states were floppy.

For positive integer $k$ we denote by $U_k$ the class of instances in which each coupler is set to a value in $\{-1, \pm 2, \ldots, \pm k\}$ chosen uniformly at random. Refs. [9, 15] avoided the local degeneracy seen in $U_k$ instances by drawing coupling strengths from Sidon sets; for example $S_{28}$ instances draw couplings from $\{\pm28, \pm19, \pm13, \pm8\}$. Although this construction is effective in reducing local degeneracy, it brings a practical difficulty: Input to D-Wave processors must be rescaled so that entries of $h$ and $J$ have absolute value at most 1, so $S_{28}$ instances must be rescaled by a factor of 1/28. Compared with $U_1$ instances, $S_{28}$ instances are therefore subject to a 28-fold inflation of relative noise, control error, and thermal effects relative to the final (classical) gap when run on DW2. This effect increases for random spin glasses as the classical gap shrinks [7, 15], and decreases as excitations from ground state are accepted as viable solutions [9]. This paper demonstrates that even a 5-fold reduction of energy scale significantly degrades DW2X performance on low-degeneracy $U_1$ instances (see Fig. 6).

In this work we employed an alternative construction that suppresses or enhances local degeneracy without shrinking the classical gap through normalization. First we simply disabled, by setting to zero, a subset of cou-
mplers that heuristically minimizes or maximizes the number of qubits with even degree. We then imposed a $U_1$ instance on the remaining couplers. More specifically, we chose a target degree $d \in \{3, 4, 5, 6\}$ and attempted to make as many qubits as possible with this degree, while retaining the underlying gridlike structure of the graph by leaving inter-cell couplers mostly intact [28]. We denote by $U^d_k$ the class of instances in which a $U_k$ instance is imposed on a subgraph with target degree $d$; note that $U_k = U^6_k$. Ideal subgraph examples are shown in Fig. 1. The study of DW2X and SA performance on random instances with controlled degree (and consequently foppiness) illuminates the role of local degeneracy in the emergence of heavy tails.

### III. RESULTS

The testbed consisted of random $U_1$ and $U_4$ Ising instances on Chimera subgraphs of varying size, from $3 \times 3$ unit cells ($C_3$) to $10 \times 10$ unit cells ($C_{10}$). These problems contain between 72 and 765 operable qubits. For each problem size, range $k$, and target degree $d$, we constructed 1000 $U^d_k$ instances, each on a newly generated subgraph with target degree $d$ [29]. We used the same processor and qubit configuration as Ref. [11]. Further details are given in the appendix.

#### A. The smoking gun: Heavy tails for $U^3_3$ instances, but not $U^3_1$ instances

Ref. [12] ran SQA and SA on the same 200-qubit ($C_5$) $U^6_1$ instances, each containing 120 degree 6 (and therefore potentially floppy) qubits, and investigated the ground state probability $p$ of each solver on a given instance. While they consider $\tau = 1/p$, we consider the number of repetitions $R$ needed to achieve 99% probability of finding a ground state [14]:

$$R = \log(1 - 0.01) / \log(1 - p). \quad (3)$$

The two solvers exhibit similar success probability in the median case, but for the hardest instances SA has success probability orders of magnitude larger than SQA [12]. If these “heavy tails” appear for QA across broad selections of input classes, it would call into question the potential utility of the general QA implementation given in Eq. 1. Fig. 3 shows that on our inputs, DW2X shows a heavy tail on $U^3_3$ instances but not on $U^3_1$ instances. This difference is not observed in SA results, which indicates that degree and degeneracy play a different role in QA than in SA.

The $U^3_3$ instances here have many degree 5 vertices, which may explain why they show less of a heavy tail than $U^3_1$ instances. As for why $U^3_5$ instances show more of a heavy tail than $U^3_3$ instances, the cause is unknown but we provide one possible explanation: $U^3_5$ instances contain more active couplers inside each unit cell of the Chimera graph, which may lead to the development of strongly coupled domains of qubits in unit cells; increased density of internal coupling has the effect of suppressing the effective transverse field on these domains, reducing their tunneling rate in the vicinity of a late anticlassing.

#### B. Performance scaling on bimodal spin glasses

Fig. 4 shows performance of DW2X and SA at the $C_4$ and $C_5$ scales, which were the respective sizes of of D-Wave One and D-Wave Two processors [2, 14]. Again, performance of DW2X relative to SA improves enormously in high percentiles for $U^3_3$ instances, where removal of couplers minimizes degeneracy, versus $U^3_1$ instances. Fig. 13 shows performance scaling on $U^3_3$, $U^3_1$, $U^3_4$, and $U^3_5$ instances. These results suggest that several threads of research that probed for computational advantage in $U^6_1$ problems [2, 9, 10] might be more successfully directed towards $U^3_3$ problems. Neither SA nor DW2X performance responds monotonically to increas-
Computing $L_s(\bar{\theta})$ for a range of $s$ and for $\bar{\theta}$ representing many classical ground and excited states gives a heuristic idea of when instantaneous and final ground states coincide according to the mean-field Hamiltonian. We define the mean-field crossing time $s_{SV}^*$ as the maximum value of the annealing parameter $s$ for which no $\bar{\theta}$ with $L_s(\bar{\theta})$ minimize $H_{SV}(s)$ corresponds to a classical ground state. We use $s_{SV}^*$ as a mean-field approximation of the location of the minimum gap of the quantum Hamiltonian $H_S$. This approach assumes that the problem has a single small gap resulting from an avoided level crossing, for example between a large metastable valley and a valley containing a classical ground state.

Figure 5 shows distributions of $s_{SV}^*$ for instances of different degrees. There is a strong correlation between a late crossing and low DW2X success probability. Typically, a very late $s_{SV}^*$ implies a crossing between a semiclassical superposition of many low-energy excited states and a relatively small number of ground states [24]. Here $U_1^3$ and $U_1^5$ instances show very similar tail shapes, in contrast to DW2X output. This may simply be due to the tunneling dynamics in a quantum annealing processor.

D. Tails and energy scale

SA, DW2X, and other physically-motivated solvers are sensitive to multiplication of the problem Hamiltonian

rotors on the unit circle, while $\sigma_x$, $\sigma_y$, and $\sigma_z$ are replaced with $\sin \theta$, 0, and $\cos \theta$ respectively:

$$H_{SV}(s) = \frac{\varepsilon(s)}{2} \left( \sum_{i<j} J_{ij} \cos \theta_i \cos \theta_j + \sum_i h_i \cos \theta_i \right)$$

Ref. [38] showed that for certain inputs, the minimum gap location given by $H_{SV}(s)$ closely matches the location given by exact diagonalization. Here we aim to associate ground and excited states of the classical Hamiltonian with global and local minima of the time-dependent semiclassical Hamiltonian $H_{SV}(s)$, estimating the gap location as the last moment at which the global minimum of $H_{SV}(s)$ corresponds to a classical excited state.

For any annealing parameter $s$ and rotor state $\bar{\theta}$, we can map $\bar{\theta}$ to a local minimum of $H_{SV}(s)$, denoted $L_s(\bar{\theta})$, using a form of gradient descent (details are given in the appendix). If we take $\bar{\theta}$ as a rotor state in $\{0, \pi\}^n$ that naturally represents a classical state, then early in the anneal, $L_s(\bar{\theta})$ be an instantaneous ground state in the direction of the transverse field. Late in the anneal, the transverse field will have little effect, so $L_s(\bar{\theta})$ will only be an instantaneous ground state of $H_{SV}(s)$ if $\bar{\theta}$ represents a classical ground state.

C. Mean-field spectra of $U_1$ instances

Heavy tails in problem sets with high local degeneracy might be explained by perturbative crossings late in the anneal for hard instances [23, 24], when a superposition of many excited classical states ceases to be an instantaneous quantum ground state [30–33]. This crossing results in a minimum eigengap late in the anneal, when the tunneling rate is small. As a result the annealer tends to get stuck in a local minimum, leading to low success probability.

Here we illustrate the relationship between gap position and local degeneracy. For 127-qubit $U_1$ instances we cannot determine the gap position using exact or even approximate diagonalization, as the instances have too many low-energy states. As an alternative we resort to the mean-field spin vector Hamiltonian [34, 35] derived in Ref. [36] as the semiclassical limit of the spin-coherent states path integral [37, 38]. Qubits are represented by
$\mathcal{H}_P$ by a scaling factor $\alpha$. Reducing $\alpha$ in SA is equivalent to increasing temperature, but in DW2X there are additional considerations of noise, error, and transverse field.

When multi-qubit tunneling is the dominant mechanism of solution for DW2X (likely not the case for $U_k^d$ instances [8, 36, 37]), performance is best when $\alpha$ is maximized and relative temperature is minimized [4, 5]. In contrast, Fig. 6 shows that heavy tails become less prevalent in DW2X results for $U_4^3$ input as we reduce $\alpha$. This may be due to thermal excitation prior to diabatic Landau-Zener transition [39] through a perturbative crossing, as hypothesized in Ref. [24]. Reducing $\alpha$ degrades DW2X performance on easy $U_4^1$ and all $U_3^1$ instances, which agrees with the evidence in Fig. 5 that these instances are not governed by late perturbative crossings.

Performance of DW2X on $U_3^3$ and $U_4^1$ instances is almost identical: in both regimes a qubit is floppy in a random state with probability only 7%, far lower than in $U_4^1$ instances. Thus these instances are less degenerate, but are also run with couplings normalized by a factor of $\frac{1}{4}$ to within the range $[-1, 1]$. Even at $\alpha = 0.2$, $U_4^1$ results shows a markedly more prominent tail as compared to $U_3^3$ results; the same is not true for $U_3^3$ and $U_4^3$ run at $\alpha = 0.8$, where the classical gap is the same as $U_1^3$ instances at $\alpha = 0.2$.

Simulated annealing, by contrast, sees relatively consistent performance at $\alpha = 1$ across the input sets studied, and its performance degrades consistently as $\alpha$ is lowered relative to a fixed temperature schedule with $\beta_{\text{final}} = 5$ relative to $J \in [-1, 1]$.

### E. Error sensitivity

Heavy tails in QA are most severe when energy scale is highest, i.e. when relative noise and error should be minimal. The existence of heavy tails for SQA with no Hamiltonian misspecification [12] indicates that this phenomenon—and more generally, poor QA results on certain low-precision instances—cannot be solely attributed to control error. To explore the contribution of classical control error to heavy tails, we ran our 127-qubit $U_4^1$ and $U_4^3$ instances 100 times each using 10 rep-
from Boltzmann sampling, matrix condition number, or mixing time of a thermal process as studied in Ref. [10]. This perspective is further justified by heavy tails in SQA results [12] where the Hamiltonian is not prone to error, and by the fact that error sensitivity as measured by the spread of quantiles does not increase monotonically as energy scale decreases, increasing relative control error (see appendix).

IV. DISCUSSION

A. Does local degeneracy challenge the viability of the QA platform?

How important is local degeneracy in interesting problem classes? Low-precision Chimera-structured problems have been used to benchmark D-Wave processors as a matter of simplicity and supposed resistance to control error. The extreme local degeneracy in these instances arises from the fact that the Chimera qubit interaction graph is dominated by qubits of degree six, which is even. When we disable couplers to suppress even-degree qubits, the tails associated with local degeneracy disappear. This cannot be attributed to reduction of classical hardness: Figs. 4 and 13 show that for SA, $U_1^3$ instances are easier than $U_3^4$ and $U_3^5$ instances, while for DW2X the opposite is true. This further indicates that the number of couplings has limited value as a measure of hardness in a random Ising problem.

Heavy tails are not present to the same extent in higher-precision instances, as we see in DW2X $U_4$ performance results. However, we cannot dismiss the practical importance of low-precision problems, as there are hard low-precision combinatorial problems in the Ising model that are of more general interest than Chimera-structured $U_1$ instances. Fully connected $±1$ Sherrington-Kirkpatrick instances [3, 41] are well understood and of interest beyond quantum annealing, but for a large instance each qubit is coupled to many others, and is therefore unlikely to be floppy regardless of the parity of its degree since $(d^2/d) / 2^d$ scales as $d^{-1/2}$ (see Section II B). Any given small set of qubits is also unlikely to be floppy. More recently there has been interest in locked constraint satisfaction problems [20, 42, 43], which are hard by design and have low-precision expressions in the Ising model [44], for example 2-in-4-SAT and 3-in-6-SAT. Locked constraint satisfaction problems explicitly avoid local degeneracy, and ground states are expected to be pairwise distant in Hamming space. These constraint satisfaction problems tell us that hard, interesting problems in the Ising model do not need to have high precision or high degeneracy.

FIG. 7. (Top) Varying DW2X performance on $U_1^4$ and $U_1^3 \mathcal{C}_4$ instances; each instance is run using 10 spin reversal transformations, each of which is repeated 10 times for a total of 100 experiments per instance. The experiment quantiles are then sorted by hardness as measured by $R$. The middle line for example indicates the median of 100 experiments for each of 1000 instances. $U_1^4$ results show a heavy tail at each experiment quantile, in contrast with $U_1^3$ results. (Bottom) Results for each quantile are divided by median performance $R_{median}$, and plotted against $R_{median}$. Bottom plots for $U_1^2$ and $U_1^4$ are restricted to the mutual domain. Percentiles increase bottom to top.
B. Probing for speedup in highly degenerate instances

Previous work has advocated searching for quantum speedup in hard $U^6_1$ problems [2, 10, 12, 25]. The relationship between degree and degeneracy provides two limitations of this choice, in addition to the fact that the classical potential of $U_1$ instances presents little challenge to simulated thermal solvers [8]. First, $U^6_1$ problems have enormous local degeneracy, resulting in perturbative crossings late in the anneal for a subset—a heavy tail—of instances. Those instances in the 99th percentile for QA time to solution will be exactly those instances that have virtually impossible to avoid altogether as we relax the definitions of local and degeneracy. If high-precision instances do not have floppy qubits, they will at least have nearly floppy qubits, and it is unreasonable to expect that the time-dependent QA eigenspectrum will accurately reflect classical energy levels throughout the anneal. In this case, can we deal with degeneracy and near-degeneracy? Refs. [23] and [45] propose two methods to mitigate perturbative crossings by manipulating the relative degeneracy of ground and excited states via addition of ancillary qubits. These perturbative crossings, which arise in the presence of degenerate clusters of excited classical states [31, 32], might also be circumvented by algorithmic [46] or random [47] choice of the initial Hamiltonian and qubit annealing trajectories.

V. CONCLUSIONS

For a broad range of problems, degeneracy is a significant challenge to simulated thermal solvers [8]. First, $U^6_1$ instances normally studied in the literature to $U^6_1$ instances yields enormous improvements in QA performance relative to SA.

VI. ACKNOWLEDGMENTS

The authors are grateful to Richard Harris for helpful advice regarding this manuscript, and to Damian Steiger for useful discussions about degeneracy and heavy tails. We thank Quntao Zhuang and Helmut Katzgraber for discussions on an earlier version of this paper.

[1] C. McGeoch and C. Wang, in *Proceedings of the ACM International Conference on Computing Frontiers* (ACM, 2013) p. 23.
[2] T. Rønnow, Z. Wang, J. Job, S. Boixo, S. Isakov, D. Wecker, J. Martinis, D. Lidar, and M. Troyer, Science 345, 420 (2014), http://www.sciencemag.org/content/345/6195/420.full.pdf.
[3] D. Venturelli, S. Mandrà, S. Knysz, B. O’Gorman, R. Biswas, and V. Smelyanskiy, Phys. Rev. X 5, 031040 (2015).
[4] S. Boixo, V. N. Smelyanskiy, A. Shabani, S. V. Isakov, M. Dykman, V. S. Denchev, M. Amin, A. Smirnov, M. Mohseni, and H. Neven, arXiv preprint arXiv:1502.05754 (2015).
[5] V. S. Denchev, S. Boixo, S. V. Isakov, N. Ding, R. Bab-bush, V. Smelyanskiy, J. Martinis, and H. Neven, arXiv preprint arXiv:1512.02206 (2015).
[6] I. Hen, T. Albash, J. Job, T. Rønnow, M. Troyer, and D. Lidar, “Probing for quantum speedup in spin glass problems with planted solutions,” (2015), arXiv preprint arXiv:1502.01663v2.
[7] A. D. King, T. Lanting, and R. Harris, “Performance of a quantum annealer on range-limited constraint satisfaction problems,” (2015), arXiv preprint arXiv:1502.02098v2, arXiv:1502.02098v2.
[8] H. G. Katzgraber, F. Hamze, and R. S. Andrist, Phys. Rev. X 4, 021008 (2014).
[9] H. G. Katzgraber, F. Hamze, Z. Zhu, A. J. Ochoa, and H. Munoz-Bauza, Phys. Rev. X 5, 031026 (2015).
[10] V. Martin-Mayor and I. Hen, Nat. Sci. Rep. 5 (2015).
[11] J. King, S. Yarkoni, M. M. Nevisi, J. P. Hilton, and E. G. Rieffel, D. Venturelli, B. O’Gorman, M. B. Do, Z. Zhu, A. J. Ochoa, S. Schnabel, F. Hamze, and H. G. Katzgraber, “Best-case performance of quantum annealers on native spin-glass benchmarks: How chaos can affect success probabilities,” (2015), arXiv:1505.02278v1.
[12] E. G. Rieffel, D. Venturelli, B. O’Gorman, M. B. Do, E. M. Prystay, and V. N. Smelyanskiy, Quantum Information Processing 14, 1 (2015).
[13] D. Venturelli, D. J. Marchand, and G. Rojo, “Quantum annealing implementation of job-shop scheduling,” (2015), arXiv:1506.08479v1.
[14] R. Babush, A. Perdomo-Ortiz, B. O’Gorman, W. Macready, and A. Aspuru-Guzik, “Construction of energy functions for lattice heteropolymer models: Efficient encodings for constraint satisfaction programming and quantum annealing,” in Advances in Chemical Physics: Volume 155 (John Wiley & Sons, Inc., 2014) pp. 201–244.
[15] A. Perdomo-Ortiz, J. Fluegemann, S. Narasimhan, R. Biswas, and V. N. Smelyanskiy, Eur. Phys. J. Spec. Top. 224, 131 (2015).
[16] L. Zdeborová and M. Mézard, Phys. Rev. Lett. 101, 078702 (2008).
[17] M. Johnson, M. Amin, S. Gildert, T. Lanting, F. Hamze, N. Dickson, R. Harris, A. Berkley, J. Johansson, P. Bunyk, et al., Nature 473, 194 (2011).
[18] P. Bunyk, E. Hoskinson, M. Johnson, E. Tolkacheva, F. Altomare, A. Berkley, R. Harris, J. Hilton, T. Lanting, A. Przybylsz, et al., IEEE Trans. Appl. Supercond. (2014).
[19] N. G. Dickson, New J. Phys. 13, 073011 (2011).
[20] N. Dickson, et al., Nat. Commun. 4, 1903 (2013).
[21] S. Boixo, T. Albash, F. M. Spedalieri, N. Chancellor, and D. A. Lidar, Nat. Commun. 4 (2013).
[22] T. Albash, W. Vinci, A. Mishra, P. A. Warburton, and D. A. Lidar, Phys. Rev. A 91, 042314 (2015).
[23] Given 120 floppy qubits in a bipartite graph such as Chimera, there is a stable set $S$ of at least 60 qubits with no couplings between them; the set of states reached by flipping any subset of $S$ is an isoeenergetic hypercube of size $2^{|S|}$ in Hamming space.
[24] We avoid degree 2 qubits because in $U_1$ problems they can be reduced, as subdivisions of smaller problems [48], and therefore only add complexity in the form of degenerate subspaces.
[25] Instances were generated on distinct random graphs to minimize the influence of structural anomalies that might arise in particular graphs.
[26] M. H. S. Amin, Phys. Rev. Lett. 100, 130503 (2008).
[27] M. H. S. Amin and V. Choi, Phys. Rev. A 80, 062326 (2009).
[28] B. Altshuler, H. Krovi, and J. Roland, Proc. Nat. Acad. Sci. USA 107, 12446 (2010).
[29] A. P. Young, S. Knysy, and V. N. Smelyanskiy, Phys. Rev. Lett. 104, 020502 (2010).
[30] N. Dickson, T. F. Rønnow, M. Troyer, and D. A. Lidar, Eur. Phys. J. Spec. Top. 224, 111 (2015).
[31] S. W. Shin, G. Smith, J. A. Smolin, and U. Vazirani, arXiv preprint arXiv:1406.7087 (2014).
[32] S. Muthukrishnan, T. Albash, and D. A. Lidar, “Tunneling and speedup in quantum optimization for permutation-symmetric problems,” (2015), arXiv:1511.03910v1.
[33] C. Zener, in Proc. R. Soc. A, Vol. 137 (The Royal Society, 1932) pp. 696–702.
[34] A. D. King and C. C. McGeoch, arXiv preprint arXiv:1410.2628 (2014).
[35] D. Sherrington and S. Kirkpatrick, Phys. Rev. Lett. 35, 1792 (1975).
[36] L. Zdeborová and M. Mézard, J. Stat. Mech. Theor. Exp. 2008, P12004 (2008).
[37] L. Zdeborová and F. Krzakala, SIAM Journal on Discrete Mathematics 25, 750 (2011).
[38] A. Douglass, A. D. King, and J. Raymond, in Theory and Applications of Satisfiability Testing–SAT 2015 (Springer, 2015).
[39] Q. Zhuang, Phys. Rev. A 90, 052317 (2014).
[40] N. G. Dickson and M. H. Amin, Phys. Rev. A 85, 032303 (2012).
[41] E. Farhi, J. Goldstone, D. Gosset, S. Gutmann, H. B. Meyer, and P. Shor, Quantum Information and Computation 11, 181 (2011).
[42] R. Diestel, Graph Theory, 4th Edition, Graduate texts in mathematics, Vol. 173 (Springer, 2012).
[43] S. V. Isakov, I. N. Zintchenko, T. F. Rønnow, and M. Troyer, Computer Physics Communications 192, 265 (2015).
[44] A. Selby, arXiv preprint arXiv:1409.3934v1 (2014).

Appendix A: Experimental details

1. Quantum annealing processor

The quantum annealing processor used in this work, and in Refs. [11, 44], was a D-Wave Two X processor operating at 12mK. Fig. 8 shows the annealing schedule of the transverse and Ising Hamiltonians (see Eq. 1). All experiments were run using an anneal time of 20μs, and samples were drawn in batches of 1000 using random spin reversal transformations except where otherwise specified. Experiments were terminated after finding 100 ground state samples or drawing $10^5$ samples, except where indicated otherwise (e.g. Fig. 7).
2. Simulated annealing

Simulated annealing was run with a linear schedule in \( \beta \) running from \( \beta_0 = 0.01 \) to \( \beta_f = 5 \), with the input Hamiltonian \( J \) scaled such that \( \max_{ij} |J_{ij}| = 1 \). For performance scaling experiments shown in Fig. 13 performance was measured at optimal anneal length (number of Monte Carlo sweeps) chosen from the set \( \{2^3, 2^4, \ldots \} \), similar to Ref. [7]. In order to estimate the time required in Fig. 13 we used the single-thread speed of 6.65 spin updates per nanosecond claimed in Ref. [49]. Samples were drawn in batches of 100, and experiments were terminated after finding 100 ground states or drawing \( 10^4 \) samples.

Appendix B: Mean-field crossing time

In the context of \( H_{SV} \) (Eq. 4) we consider the effective field \( h_i^{\text{eff}}(\vec{\theta}) \) on qubit \( i \) in the state \( \vec{\theta} \) to be

\[
h_i^{\text{eff}}(\vec{\theta}) = h_i + \sum_{j \neq i} J_{ij} \cos \theta_j. \tag{B1}\]

For a given annealing parameter \( s \), consider the locally optimal value of \( \theta_i \) minimizing \( H_{SV} \) subject to \( (\theta_j \mid j \neq i) \), i.e.

\[
\theta_i^* = \arg \min_{\theta} \left( \frac{\varepsilon(s)}{2} h_i^{\text{eff}}(\vec{\theta}) \cos \theta - \frac{\Delta(s)}{2} \sin \theta \right)
= \arccot \left( \frac{-\varepsilon(s) h_i^{\text{eff}}(\vec{\theta})}{\Delta(s)} \right) \tag{B2}\]

given that \( \Delta(s) \) and \( \varepsilon(s) \) are positive. Since modifying \( \theta_i \) can alter some \( \theta_i^* \), we calculate the associated local minimum \( L_s(\vec{\theta}) \) using an iterative method that replaces \( \theta_i \) with \( (\theta_i + \theta_i^*)/2 \) for every spin \( \theta_i \) in turn, and iterates until convergence.

Appendix C: Testbed construction

Every \( U_k^d \) is constructed over a subgraph \( G' \) of the available qubit connectivity graph \( G \). For each instance, we construct \( G' \) iteratively at random by removing one edge at a time. An edge \( uv \) can be chosen for deletion if:

1. \( u \) has maximum degree in \( G' \)
2. There is no edge \( u'v' \) in \( G' \) such that \( d(u') \geq d(u) \) and \( d(v') > d(v) \).
3. If \( u \) and \( v \) are in different unit cells (see [22]) then \( d(u) = d(v) \) and there is no edge \( u'v' \) for which both \( d(u') = d(v') = d(u) \) and \( u' \) and \( v' \) are in the same unit cell.
4. \( d(u) \geq d(v) \geq 3 \).
TABLE I. Degree distributions of instances with target degree 3 to 6. Proportions of qubits with each degree are shown for instance sizes $C_4$, $C_6$, $C_8$ and $C_{10}$, which have 128, 288, 512 and 800 qubits respectively (some are inoperable). Remaining qubits have degree 0.

| $C_d$ | degree 2 | degree 3 | degree 4 | degree 5 | degree 6 |
|-------|----------|----------|----------|----------|----------|
| $U_3^7$ | 7% | 92% | – | – | – |
| $U_4^7$ | – | 4% | 95% | – | – |
| $U_5^7$ | – | – | 4% | 95% | – |
| $U_6^7$ | – | – | – | 53% | 46% |

| $C_6$ | degree 2 | degree 3 | degree 4 | degree 5 | degree 6 |
|-------|----------|----------|----------|----------|----------|
| $U_3^1$ | 10% | 86% | – | – | – |
| $U_4^1$ | – | 10% | 86% | – | – |
| $U_5^1$ | – | 2% | 14% | 81% | – |
| $U_6^1$ | – | 1% | 7% | 35% | 53% |

| $C_8$ | degree 2 | degree 3 | degree 4 | degree 5 | degree 6 |
|-------|----------|----------|----------|----------|----------|
| $U_3^1$ | 9% | 88% | – | – | – |
| $U_4^1$ | – | 8% | 89% | – | – |
| $U_5^1$ | – | 1% | 10% | 87% | – |
| $U_6^1$ | – | 1% | 2% | 34% | 61% |

| $C_{10}$ | degree 2 | degree 3 | degree 4 | degree 5 | degree 6 |
|----------|----------|----------|----------|----------|----------|
| $U_3^1$ | 10% | 85% | – | – | – |
| $U_4^1$ | – | 10% | 85% | – | – |
| $U_5^1$ | – | 2% | 10% | 83% | – |
| $U_6^1$ | – | 2% | 5% | 28% | 61% |

5. Removing $uv$ does not disconnect the graph.

Appendix D: Error sensitivity and hardness

Fig. 10 shows error sensitivity data corresponding to Fig. 7 when problems are run at energy scale $\alpha \in \{0.2, 0.6\}$ rather than full energy scale $\alpha = 1.0$. For each energy scale, $U_3^1$ and $U_5^1$ instances of similar hardness $R_{\text{median}}$ show very similar error sensitivity ($R/R_{\text{median}}$).

Fig. 11 shows analogous data from experiments using noisy SA as the solver rather than DW2X. For each experiment consisting of 1000 simulated annealing samples drawn using $2^{10}$ Monte Carlo sweeps, each entry of $h$ and $J$ corresponding to a physically existing coupler or qubit is perturbed by an independent Gaussian error with standard deviation 0.03. This error is relative to full energy scale, so experiments run at energy scale 0.2 will experience Hamiltonian misspecification errors five times as large as those experienced by experiments run at full energy scale. Both hardness and error sensitivity increase as energy scale decreases, which is both expected due to inflation of relative error, and contrary to what is seen in DW2X data.

As further evidence of the relation between DW2X error sensitivity and DW2X hardness, Fig. 12 plots the spread of relative performance between 75th and 25th percentiles for DW2X and noisy SA, versus hardness as experienced by each solver. The results indicate that median DW2X performance is a good indicator of DW2X error sensitivity, and that median SA performance is a good indicator of SA error sensitivity, but that median performance of one solver is not as good an indicator of error sensitivity in the other solver.

Appendix E: Performance scaling data

Fig. 13 shows performance scaling of SA and DW2X on $U_d^4$ instances of varying size for $d \in \{3, 4, 5, 6\}$. $C_L$ instances use up to $8L^2$ qubits in an $L \times L$ subgrid of unit cells in the Chimera graph [11]. At the $C_8$ scale, which is the size of the previous generation of D-Wave Two processor, DW2X solves $U_4^4$ problems at the 95th percentile around 1000 times faster than $U_5^4$ problems at the 95th percentile. As in Ref. [2], SA anneal time is optimized, while for DW2X, the minimum available anneal length of 20µs is used; these curves should not be used to characterize the underlying problem-solving dynamics of the DW2X processor.
FIG. 10. (Top) Varying DW2X performance on $U_3^3$ and $U_4^4$ instances; each instance is run using 10 spin reversal transformations, each of which is repeated 10 times for a total of 100 experiments per instance. (Bottom) Results for each quantile are divided by median performance $R_{\text{median}}$, and plotted against $R_{\text{median}}$. Bottom plots for $U_3^3$ and $U_4^4$ are restricted to the mutual domain.
FIG. 11. (Top) Varying noisy SA (noise $\sigma = 0.03$) performance on $U^3_3$, $U^3_4$, $U^4_3$, and $U^4_4$ instances; each instance is run using 10 spin reversal transformations, each of which is repeated 10 times for a total of 100 experiments per instance. The experiment quantiles are then sorted by hardness as measured by $R$. (Bottom) Results for each quantile are divided by the median quantile performance $R_{\text{median}}$, and plotted against $R_{\text{median}}$. Bottom plots for $U^3_1$ and $U^4_1$ are restricted to the mutual domain.
FIG. 12. Hardness, as seen by QA (DW2X) and noisy SA, versus error sensitivity, as seen by QA and noisy SA. As in Figs. 7 and 10, we run 100 experiments on each of 1000 $U_3^1$ instances and 1000 $U_4^1$ instances at the $C_4$ scale. We then sort each of the 100 quantiles by hardness, use the entrywise ratio of the 75th and 25th percentiles (the spread of quantiles) as a measure of error sensitivity, and plot against the median. Close agreement between $U_3^1$ and $U_4^1$ data indicates a close relationship between error sensitivity as seen by a particular solver, and hardness as seen by a particular solver, regardless of input parameters. Here we see very close agreement between DW2X error sensitivity and DW2X hardness, and agreement between noisy SA error sensitivity and noisy SA hardness.
FIG. 13. Time to solution for SA (top) and DW2X (bottom) on $U_3^j$ instances with sizes from $C_3$ to $C_{10}$. SA uses an optimized number of Monte Carlo sweeps. DW2X performance on high percentiles indicates the relative absence of heavy tails in $U_3^3$ instances and to a lesser extent $U_3^5$ instances. SA performance indicates that this difference cannot be explained by $U_3^3$ instances being fundamentally easier.