

APPLIED PHYSICS

Experimental investigation of performance differences between coherent Ising machines and a quantum annealer

Ryan Hamerly^{1,2,*†}, Takahiro Inagaki^{3,*†}, Peter L. McMahon^{2,4,5,*†}, Davide Venturelli^{6,7}, Alireza Marandi^{4,8}, Tatsuhiro Onodera⁴, Edwin Ng⁴, Carsten Langrock⁴, Kensuke Inaba³, Toshimori Honjo³, Koji Enbutsu⁹, Takeshi Umeki⁹, Ryoichi Kasahara⁹, Shoko Utsunomiya², Satoshi Kako², Ken-ichi Kawarabayashi², Robert L. Byer⁴, Martin M. Fejer⁴, Hideo Mabuchi⁴, Dirk Englund¹, Eleanor Rieffel⁶, Hiroki Takesue³, Yoshihisa Yamamoto^{4,10}

Copyright © 2019
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S. Government
Works. Distributed
under a Creative
Commons Attribution
NonCommercial
License 4.0 (CC BY-NC).

Physical annealing systems provide heuristic approaches to solving combinatorial optimization problems. Here, we benchmark two types of annealing machines—a quantum annealer built by D-Wave Systems and measurement-feedback coherent Ising machines (CIMs) based on optical parametric oscillators—on two problem classes, the Sherrington-Kirkpatrick (SK) model and MAX-CUT. The D-Wave quantum annealer outperforms the CIMs on MAX-CUT on cubic graphs. On denser problems, however, we observe an exponential penalty for the quantum annealer [$\exp(-\alpha_{\text{DW}}N^2)$] relative to CIMs [$\exp(-\alpha_{\text{CIM}}N)$] for fixed anneal times, both on the SK model and on 50% edge density MAX-CUT. This leads to a several orders of magnitude time-to-solution difference for instances with over 50 vertices. An optimal-annealing time analysis is also consistent with a substantial projected performance difference. The difference in performance between the sparsely connected D-Wave machine and the fully-connected CIMs provides strong experimental support for efforts to increase the connectivity of quantum annealers.

INTRODUCTION

Optimization problems are ubiquitous in science, engineering, and business. Many important problems (especially combinatorial problems such as scheduling, resource allocation, route planning, or community detection) belong to the nondeterministic polynomial time (NP)-hard complexity class and, even for typical instances, require a computation time that scales exponentially with the problem size (1). Canonical examples such as Karp's 21 NP-complete problems (2) have attracted much attention from researchers seeking to devise new optimization methods because, by definition, any NP-complete problem can be reduced to any other problem in NP with only polynomial overhead. Many approximation algorithms and heuristics [e.g., relaxations to semidefinite programs (3), simulated annealing (4), and breakout local search (BLS) (5)] have been developed to search for good-quality approximate solutions and ground states for sufficiently small problem sizes. However, for many NP-hard optimization problems, even moderately sized problem instances can be time consuming to solve exactly or even approximately. Hence, there is a strong motivation to find alternative approaches that can consistently beat state-of-the-art algorithms.

Despite decades of Moore's law scaling, large NP-hard problems remain very costly even on modern microprocessors. Thus, there is a growing interest in special-purpose machines that implement a solver directly by mapping the optimization to the underlying physical dynamics. Examples include digital complementary metal-oxide semiconductor annealers (6, 7) and analog devices such as nanomagnet arrays (8), electronic oscillators (9, 10), and laser networks (11). Quantum adiabatic computation (12) and quantum annealing (13–16) are also prominent examples and may offer the possibility of quantum speedup (17–19) for certain NP-hard problems. However, all the nonphotonic analog optimization systems realized to date suffer from limited connectivity so that actual problems must, in general, first be embedded (20, 21) into the solver architecture native graph before they can be solved. This requirement adds an upfront computational cost (20, 22, 23) of finding the embedding (unless previously known) and, of most relevance in this study, in general results in the use of multiple physical pseudo-spins to encode each logical spin variable, which can lead to a degradation of time to solution.

Here, we perform the first direct comparison between the D-Wave 2000Q (DW2Q) quantum annealer and the coherent Ising machine (CIM) (24, 25). As we will see later, a crucial distinction between these systems is their intrinsic connectivity, which has a profound influence on their performance. Both systems are designed to solve the classical Ising problem, that is, to minimize the classical Hamiltonian

$$H = \frac{1}{2} \sum_{ij} J_{ij} \sigma_i \sigma_j + \sum_i h_i \sigma_i \quad (1)$$

where $\sigma_i = \pm 1$ are the Ising spins, J_{ij} are the entries of the spin-spin coupling matrix, and h_i are the Zeeman (bias) terms. The Ising problem is NP-hard for nonplanar couplings (26) and is one of the most widely studied problems in this complexity class. We focus on two canonical NP-hard Ising problems: unweighted MAX-CUT (2) and

¹Research Laboratory of Electronics, Massachusetts Institute of Technology, 50 Vassar Street, Cambridge, MA 02139, USA. ²National Institute of Informatics, Hitotsubashi 2-1-2, Chiyoda-ku, Tokyo 101-8403, Japan. ³NTT Basic Research Laboratories, NTT Corporation, 3-1 Morinosato Wakamiya, Atsugi, Kanagawa 243-0198, Japan. ⁴E. L. Ginzton Laboratory, Stanford University, Stanford, CA 94305, USA. ⁵School of Applied and Engineering Physics, Cornell University, Ithaca, NY 14853, USA. ⁶NASA Ames Research Center Quantum Artificial Intelligence Laboratory (QuAIL), Mail Stop 269-1, Moffett Field, CA 94035, USA. ⁷USRA Research Institute for Advanced Computer Science (RIACS), 615 National Avenue, Mountain View, CA 94035, USA. ⁸California Institute of Technology, Pasadena, CA 91125, USA. ⁹NTT Device Technology Laboratories, NTT Corporation, 3-1 Morinosato Wakamiya, Atsugi, Kanagawa 243-0198, Japan. ¹⁰ImPACT Program, Japan Science and Technology Agency, Gobancho 7, Chiyoda-ku, Tokyo 102-0076, Japan.

*These authors contributed equally to this work.

†Corresponding author. Email: rhamerly@mit.edu (R.H.); takahiro.inagaki.vn@hco.ntt.co.jp (T.I.); pmcmahon@stanford.edu (P.L.M.)

ground-state computation of the Sherrington-Kirkpatrick (SK) spin-glass model (27).

In the CIM, the spin network is represented by a network of degenerate optical parametric oscillators (OPOs). Each OPO is a nonlinear oscillator that converts pump light to its half-harmonic (28); it can oscillate in two identical phase states, which encode the value of the Ising spin (29, 30). Optical coherence is essential to the CIM, where the data are encoded in the phase of the light. As Fig. 1A shows, time multiplexing offers a straightforward way to generate many identical OPOs in a single cavity (30). A pulsed laser with repetition time T is used to pump an optical cavity with round-trip time $N \times T$. Parametric amplification is provided by the $\chi^{(2)}$ crystal; since this is an instantaneous nonlinearity, the circulating pulses in the cavity are identical and noninteracting. The approach is scalable using high-repetition rate lasers and long fiber cavities: OPO gain has been reported for up to $N = 10^6$ pulses, and stable operation achieved for $N = 50,000$ (31). Each circulating pulse represents an independent OPO with a single degree of freedom, a_i . Classically, a_i is a complex variable, which maps to the annihilation operator \hat{a}_i in quantum mechanics (32). A measurement-feedback apparatus is used to apply coupling between the pulses (24, 25). In each round trip, a small fraction of the light ($\sim 10\%$) is extracted from the cavity and homodyned against a reference pulse (the OPO pump is created from second harmonic generation of the reference laser, so there is good matching between the reference and the OPO signal light, which is at half the frequency of the pump). The homodyne result, in essence a measurement of a_i , is fed into an electronic circuit (consisting of an

analog-to-digital converter, a field-programmable gate array, and a digital-to-analog converter) that, for each pulse, computes a feedback signal that is proportional to the matrix-vector product $\sum_j J_{ij} a_j$. This signal is converted back to light using an optical modulator and a reference pulse and reinjected into the cavity. The measurement-feedback CIM has intrinsic all-to-all connectivity through its exploitation of memory in the electronic circuit [although the same effect can be obtained with optical delay-line memories in all-optical CIMs (29, 30)].

The OPO is a dissipative quantum system with a pitchfork bifurcation well adapted for modeling Ising spins: As the pump power is increased (Fig. 1B), the OPO state transitions from a below-threshold squeezed vacuum state (33–35) to an above-threshold coherent state (36). Because degenerate parametric amplification is phase sensitive, only two phase states are stable above threshold; thus, the OPO functions as a classical “spin” with states $\{|0\rangle, |\pi\rangle\}$ that can be mapped to the Ising states $\sigma_i = \{+1, -1\}$. The optimization process happens in the near-threshold regime where the dynamics are determined by a competition between the network loss and Ising coupling (which seek to minimize the product $\sum_{ij} J_{ij} a_i a_j$), as well as nonlinear parametric gain (which seeks to enforce the constraints $a_i \in \mathbb{R}$, $|a_i| = \text{const}$).

As an example, consider the Ising problem on the $N = 16$ Möbius ladder graph with antiferromagnetic couplings (37). Figure 1C shows a typical run of the CIM, resulting in a solution that minimizes the Ising energy [data from (24)]. The most obvious interpretation of the process is spontaneous symmetry breaking of a pitchfork bifurcation: Prepared in a squeezed vacuum state and driven by shot noise, the OPO state

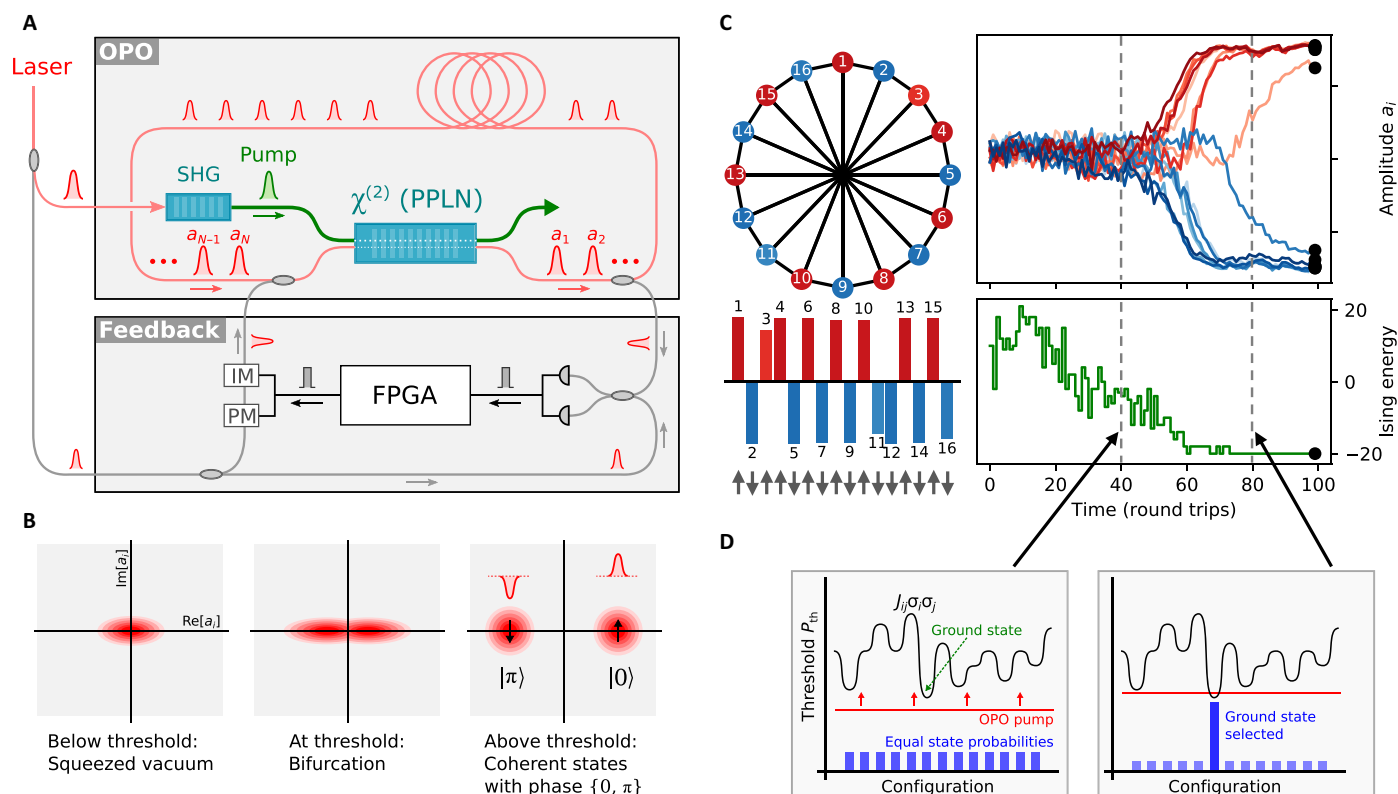


Fig. 1. Schematic and operation principle of the CIM. (A) CIM design consisting of time-multiplexed OPO and measurement-feedback apparatus. See (24, 25) for details. SHG, second harmonic generation; FPGA, field-programmable gate array; PPLN, periodically poled lithium niobate; IM, intensity modulator; PM, phase modulator. (B) OPO state during transition from below-threshold squeezed state to (bistable) above-threshold coherent state. (C) Solution of antiferromagnetic Ising problem on the Möbius ladder with the CIM, giving measured OPO amplitudes a_i and Ising energy H as a function of time in round trips. (D) Illustration of search-from-below principle of CIM operation.

bifurcates, during which its amplitudes a_i grow either in positive or negative value, and subsequently the system settles into the Ising ground state (or a low-lying excited state) (24, 38) [this is related to the Gaussian state model in (39)]. Another view derives from ground-state “search from below” (Fig. 1D). Here, the Ising energy is visualized as a complicated landscape of potential oscillation thresholds, each with its own spin configuration. If the OPO pump is far below the minimum threshold, all spin configurations will be excited with near-equal probability, but once the ground-state threshold is exceeded, its probability will grow exponentially at the expense of other configurations (30, 40). This ground-state selection process corresponds to the $40 \leq t \leq 60$ region in Fig. 1C.

The DW2Q quantum annealer used in this work is installed at NASA Ames Research Center in Mountain View, CA. The DW2Q has 2048 qubits, but its “Chimera” coupling graph (i.e., the graph whose edges define the nonzero J_{ij} terms in Eq. 1) is very sparse. Since most Ising problems are not defined on subgraphs of the Chimera, minor embedding is used to find a Chimera subgraph on which the corresponding Ising model has a ground state that corresponds to the classical ground state of the Ising model defined on the desired problem graph (20, 21). Native clique embeddings (Fig. 2A) (41) are pre-computed embeddings that can be used for fully connected problems or problems on dense graphs. Each logical qubit is associated with an L-shaped ferromagnetic chain of $\lceil N/\kappa \rceil + 1$ physical qubits, where 2κ is the number of qubits in each unit cell of the Chimera graph ($\kappa = 4$ in the DW2Q). Clique embeddings are desirable because all chain lengths are equal: This architecture simplifies the parameter setting procedure due to symmetry, and it is thought to prevent desynchronized freeze-out of chains during the calculation (42). However, the embedding introduces considerable overhead relative to the fully connected model: for N logical qubits, $N(\lceil N/\kappa \rceil + 1) \approx N^2/\kappa$ physical qubits are used. Because of the triad structure (22) of the embeddings (Fig. 2A), only approximately half of the annealer’s physical qubits are used, limiting the DW2Q to problems with $N \leq 64$ (the actual limit is $N \leq 61$ due to unusable qubits on the particular machine at NASA Ames).

RESULTS

SK spin-glass

As a first benchmarking problem, we consider the SK spin-glass model on a fully connected (i.e., maximally dense) graph, where the couplings $J_{ij} = \pm 1$ are randomly chosen with equal probability (27). Ground-state computation of the SK model is directly related to the graph partitioning problem, which is also NP-hard (43). For each problem size $2 \leq N \leq 61$, 20 randomly chosen instances were solved on the DW2Q. We consider as a performance metric the success probability P , defined as the fraction of runs on the same instance that return the ground-state energy (we will discuss the time to solution, which requires a more thorough analysis involving optimal annealing time, in a subsequent section).

To properly benchmark the DW2Q’s performance, it is important to optimize the embedding parameter J_c which sets the ratio of constraint couplings to logical couplings. The optimal J_c depends on the problem type and size and is found empirically. The DW2Q success probability is plotted as a function of J_c in Fig. 2B, from which we find that the optimal J_c scales roughly as $N^{1/2}$ (see Materials and Methods for details). This scaling is consistent with results published on the same class of problems with the earlier D-Wave Two quantum annealer, and it is believed to be connected to the spin-glass nature of the SK Ising problem (42). Using the optimal value of J_c , Fig. 2C shows that the performance on the D-Wave depends strongly on the single-run annealing time, with the values $T_{\text{ann}} = (1, 10, 100, 1000) \mu\text{s}$ plotted here. The D-Wave annealing time is restricted to the range $[1, 2000] \mu\text{s}$. We observe that longer annealing times give higher success probabilities, in accordance with the expectations from the adiabatic quantum optimization approach that inspired the design of the D-Wave machine. The data fit well to a square exponential, $P = \exp(-(N/N_0^{\text{DW}})^2)$, where the parameter N_0^{DW} increases slowly, roughly logarithmically, with T_{ann} . For problem sizes $N < 30$, the results in Fig. 2C agree with an extrapolation of the benchmark data for $T_{\text{ann}} = 20 \mu\text{s}$ reported in (42), which used an earlier processor [the 512-qubit D-Wave despite the engineering improvements that have been made in the last two generation chips (2X and 2000Q)].

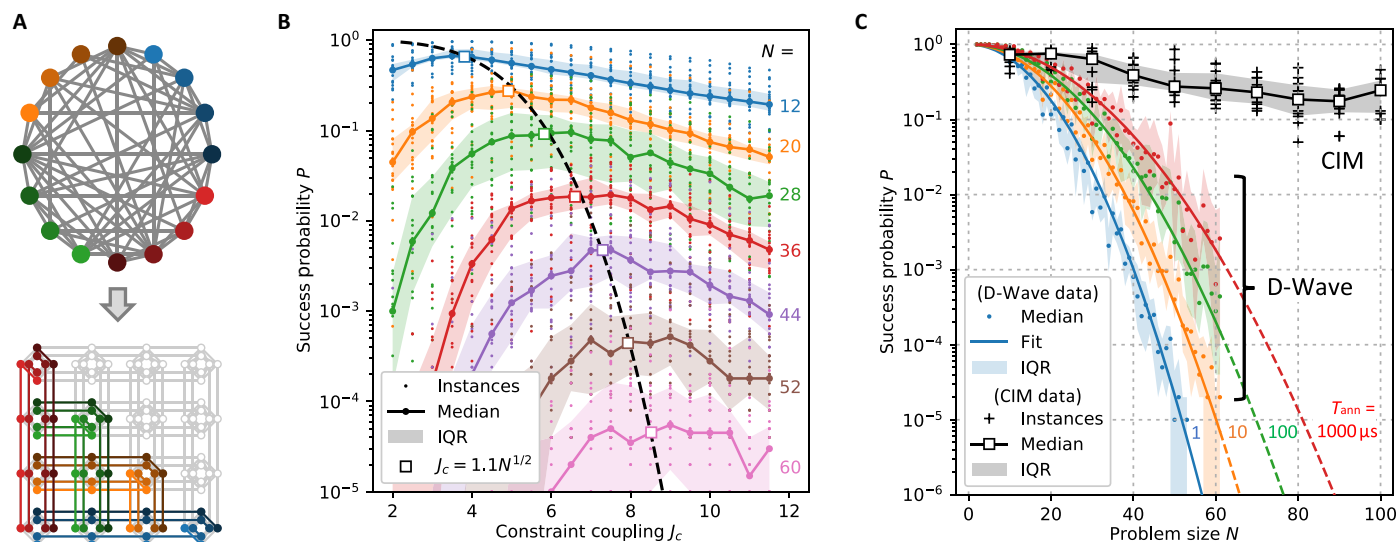


Fig. 2. CIM and D-Wave performance on SK problems. (A) Illustration of clique embedding: An arbitrary $N = 16$ graph is embedded into the D-Wave chimera, each spin mapped to a ferromagnetically coupled line of physical qubits (each color is a logical qubit). (B) D-Wave ground-state probability for SK model as a function of problem size N and embedding parameter J_c (annealing time $T_{\text{ann}} = 20 \mu\text{s}$). Shading indicates interquartile range (IQR; 25th, 75th percentile range of instances). (C) Scaling of ground-state probability for DW2Q (with optimal J_c) and Stanford CIM. D-Wave and CIM ran 20 and 10 instances per problem size, respectively.

The same SK instances for $N = 10, 20, \dots, 60$ were solved on the CIMs hosted at Stanford University in Stanford, CA and Nippon Telegraph and Telephone Corporation (NTT) Basic Research Laboratories in Atsugi, Japan (24, 25). Additional problems with $N \geq 60$ were also solved on the CIM but were too large to be programmed on the DW2Q. The two Ising machines have similar performance (see section S2 for more details). Figure 2C shows a plot of the success probability as a function of problem size: The exponential scaling for the CIM is shallower than the one given by the DW2Q performance. We note that the success probability P for the CIM scales approximately as $\exp(-N/N_0^{\text{CIM}})$, where N_0^{CIM} is a constant. The fact that, for the DW2Q, success probability P scales with an N^2 dependence in the exponential rather than N (as is the case for the CIM) leads to a marked difference in success probability between the quantum annealer and the CIM for problem sizes $N \geq 60$.

MAX-CUT

We next study the DW2Q performance on MAX-CUT for both dense and sparse unweighted graphs. Unweighted MAX-CUT is the problem of finding a partition (called a cut) of the vertices V of a graph $G = (V, E)$, where the partition is defined by two disjoint sets V_1 and V_2 with $V_1 \cup V_2 = V$, and for which the number of edges between the two sets $|\{(v_1 \in V_1, v_2 \in V_2) \in E\}|$ is maximized. Unweighted MAX-CUT is NP-hard for general graphs (2) and can be expressed as an Ising problem by setting the antiferromagnetic couplings $J_{ij} = +1$ along graph edges: $H = \sum_{(ij) \in E} \sigma_i \sigma_j$. Thus, the problem in Fig. 1C is the same as MAX-CUT on the Möbius ladder graph. Previous CIM studies have solved MAX-CUT on problems up to size $N = 2000$ in experiment (24, 25, 30, 37) and $N = 20,000$ in simulation (29, 44).

Random unweighted MAX-CUT graphs at the phase transition (45), with an edge density of 0.5 [i.e., Erdős-Rényi graphs $G(N, \frac{1}{2})$] were tested on DW2Q for problems up to $N = 61$ and on the CIM for $N \leq 150$. For these graphs, clique embeddings were used, but in practice, the performance did not differ from the embedding heuristic provided by the D-Wave API (21). In Fig. 3A, we show that the optimal value of the embedding coupling parameter J_c appears to be correlated with the appearance of defects in the perfect polarization state expected in logical qubits at the end of the anneal. With J_c optimized, the success probability follows the same square exponential ($e^{-O(N^2)}$) trend with N as in the SK model, but the drop-off is even steeper. The CIM success probabilities are also lower than for the SK model but are now orders of magnitude higher than the DW2Q for $N \geq 40$.

To test the effect of sparseness, we plot in Fig. 3C the performance on unweighted regular graphs of degree $d = 3, 4, 5, 7, 9$, where the degree of a graph is the number of edges per vertex. Despite their sparseness, MAX-CUT on these restricted graph classes is also NP-hard (46). The CIM shows no performance difference between $d = 3$ (cubic) and dense graphs. For DW2Q, the sparse graphs are embedded using the graph minor heuristic, which allows problems of up to size $N = 200$ to be embedded in the DW2Q (21). In addition, the found embeddings require significantly fewer qubits (for the sparse graphs) than the clique embeddings (compare Figs. 3D and 2A and see also fig. S3B). For cubic graphs, the DW2Q achieves slightly better performance than the CIM, while the CIM's advantage is noticeable for $d \geq 5$.

The CIM achieves similar success probabilities for cubic and dense graphs, suggesting that dense problems are not intrinsically

harder than sparse ones for this class of annealer. D-Wave's strong dependence on edge density is most likely a consequence of embedding compactness: It is known that more compact embeddings (fewer physical qubits per chain) tend to give better annealing performance after all optimization and parameter settings are considered (21). Since qubits on the D-Wave chimera graph have at most six connections, the minimum chain length is $\ell = \lceil (d-2)/4 \rceil$, so embeddings grow less compact with increasing graph degree (see section S3). Since degree-1 and degree-2 vertices can be pruned from a graph in polynomial time [a variant of cut-set conditioning (47)], $d = 3$ is the minimum degree required for NP hardness. Of NP-hard MAX-CUT instances, Fig. 3C suggests that there is only a very narrow region ($d = 3, 4$) where D-Wave matches or outperforms the CIM in success probability; for the remainder of the graphs, the CIM dominates.

Time to solution and optimal annealing time

Although success probability is a helpful metric to understand scaling for fixed T_{ann} , the key computational figure of merit is the time to solution $T_{\text{soln}} = T_{\text{ann}} \lceil \log(0.01) / \log(1-P) \rceil$. This figure multiplies the expected number of independent runs to solve a problem with 99% probability with the time of a single run, T_{ann} . When evaluating the time to solution of a physical annealer, it is important to optimize, as much as possible, the run parameters, in particular considering the machine's T_{soln} at the optimal annealing time. This avoids a common pitfall of fixed-anneal time analysis, where if the chosen anneal time is too large, near-flat T_{soln} scaling for small problem sizes gives the illusion of speedup (18, 48).

In the CIM, the anneal time is set by the pump turn-on schedule and is an integer number of round trips. The experiments in this paper were conducted with $T_{\text{ann}} = 1000$ round trips, but shorter or longer times are also possible. To assess the effect of the anneal time on CIM performance, we simulate the CIM with c -number stochastic differential equations (c -SDEs) using the truncated Wigner representation (36). The algorithm, which is based on (38), is described in section S3. Figure 4A compares the performance of the experimental CIMs to the c -SDE model for dense MAX-CUT instances, indicating that the model reasonably reproduces the behavior of the experimental CIMs (similar agreement is found for SK and cubic MAX-CUT instances; see fig. S9).

Figure 4B plots the (c -SDE) CIM success probability and time to solution (in round trips) as a function of anneal time. Consistent with the experimental results, we see an exponential scaling with N in the large- N limit [the curves are fit to $P(N) = (a + (1-a)e^{bN})^{-1}$, which becomes exponential for large N]. The time-to-solution plot is a series of (nearly) linear intersecting curves, where curves with shorter anneal time have a lower intercept but a larger slope. Optimizing T_{soln} reveals a tradeoff between the annealing time of a single run and success probability: Short anneals are preferred for small problems where the success probability is always close to unity and insensitive to the annealing time, and long anneals are preferred for large problems where the success probability dominates. Thus, the optimal anneal time depends on problem size and increases with N . Figure 4C shows the analogous D-Wave data for the same problem class (dense MAX-CUT). Here, the fixed- T_{ann} curves scale quadratically with N rather than linearly. The same scaling is observed in SK problems (see figs. S10 and S11).

It has been observed empirically on D-Wave quantum annealers that, for Chimera graph spin glasses, optimal time to solution scales as $T_{\text{soln}} \propto \exp(O(N^{1/2}))$ (18, 19, 49), while T_{soln} at fixed anneal times increases more steeply (18). The lower envelope of the curves in Fig. 4B

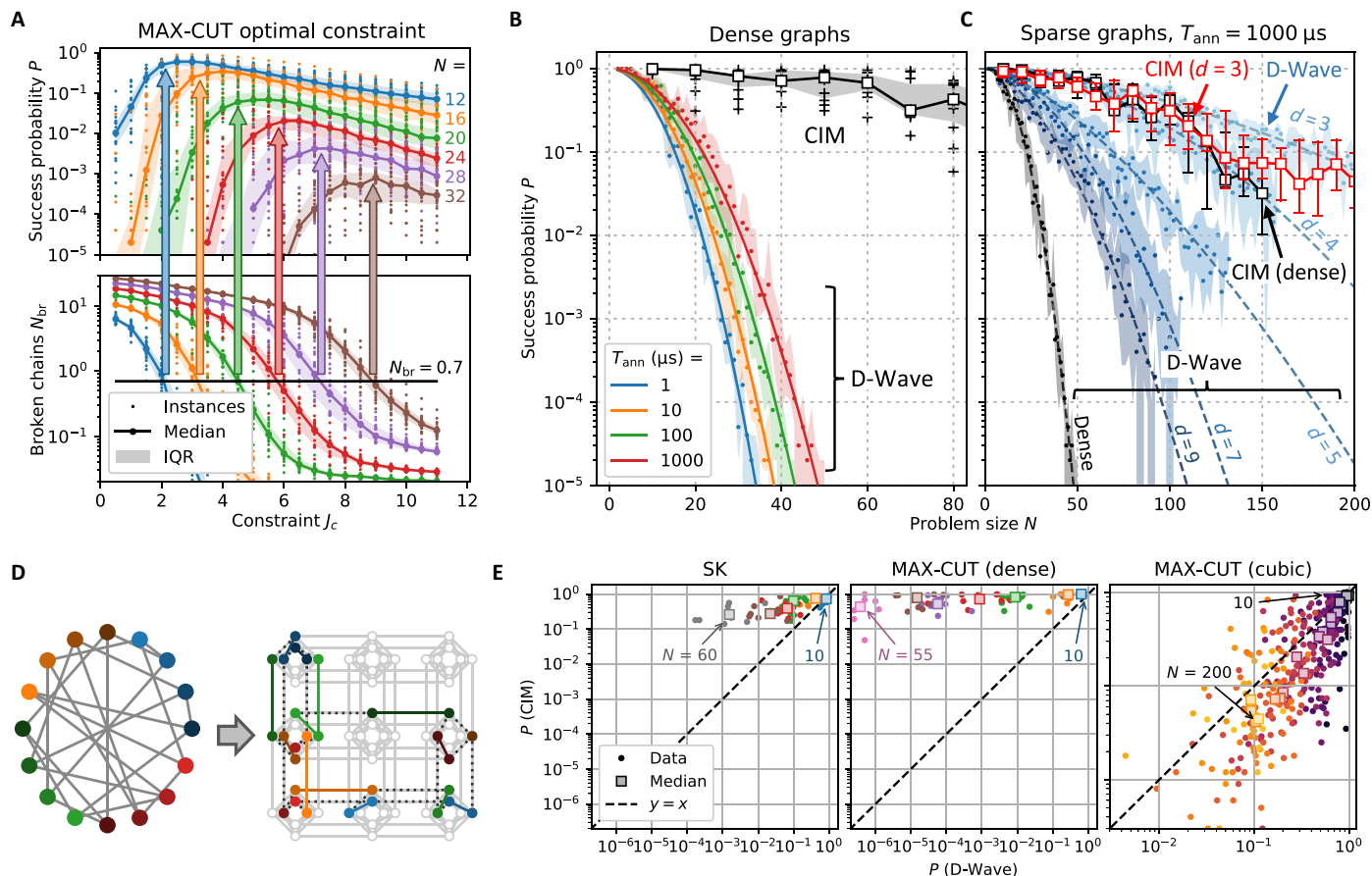


Fig. 3. Success probability for MAX-CUT on dense and sparse graphs. (A) D-Wave performance on dense MAX-CUT problems with (edge density of 0.5), showing that optimal performance occurs when the J_c coupling is strong enough to make it unlikely that logical qubits (chains) become “broken” (see also figs. S1 and S2). (B) D-Wave and NTT CIM success probability for dense MAX-CUT as a function of problem size (for T_{soln} see fig. S6). (C) D-Wave (annealing time $T_{\text{ann}} = 1000 \mu\text{s}$) and NTT CIM success probability for sparse graphs of degree $d = 3, 4, 5, 7, 9$ and dense graphs. (D) Example of a cubic graph embedding found with the heuristic. (E) Success probability scatterplots comparing D-Wave ($T_{\text{ann}} = 1000 \mu\text{s}$) and CIM.

can be reasonably fit to this form, suggesting a similar scaling for the CIM, although the CIM is based on an entirely different computational principle. Moreover, when solving embedded problems in quantum annealers, the optimal embedding parameters are believed to be associated with the emergence of a spin-glass phase of the embedded problem (42). Since for dense graphs the embedded problem has $N_{\text{ph}} \propto N^2$ qubits, this would suggest a scaling $T_{\text{soln}} \propto \exp(O(N^2)^{1/2}) = \exp(O(N))$, as shown in Fig. 4C. Both the CIM scaling $T_{\text{soln}} \propto \exp(O(N^{1/2}))$ (dashed curves in Fig. 4, B and C) and the DW2Q scaling $T_{\text{soln}} \propto \exp(O(N))$ (solid curve in Fig. 4C) are consistent with the hypothesis that a physical annealer’s time to solution (at optimal annealing time) should scale exponentially with $N_{\text{ph}}^{1/2}$, where N_{ph} is the number of physical qubits (or bits) required to encode the Ising problem.

We note that our claims are only suggestive, but not conclusive, of $\exp(O(N_{\text{ph}}^{1/2}))$ scaling at the optimal annealing time. Only for a limited range of problem sizes ($25 \leq N \leq 50$) is the optimal annealing time accessible with the DW2Q, and the data are noisy enough that other curves would also fit the lower envelope. Thus, we caution against naively extrapolating these curves to large problem sizes. However, a scaling advantage for the CIM does exist at measured problem sizes, a conclusion also observed (figs. S10 and S11) for SK and cubic (i.e., $d = 3$) MAX-CUT instances (although the DW2Q nonetheless outper-

forms the CIM in absolute terms for all measured cubic MAX-CUT problems).

While the optimal-annealing time analysis is important theoretically, in realistic machines, T_{ann} is limited by parameter misspecification, finite temperature, and various noise sources and cannot be increased arbitrarily. Therefore, it is relevant to consider the achievable performance for practically realizable choices of T_{ann} . Table 1 shows the experimental times to solution for the NTT CIM (fixed $T_{\text{ann}} = 1000 \mu\text{s}$) and the DW2Q (for range, $T_{\text{ann}} \in [1, 1000] \mu\text{s}$). This allows a comparison of the optimal time to solution in experimentally accessible parameters, which may be more a useful benchmark for near-term annealing machines. The DW2Q outperforms the CIM by a factor of 10 to 100× at cubic MAX-CUT problems, although this factor shrinks with increasing problem size. For the SK and dense MAX-CUT problems, on the other hand, the CIM outperforms D-Wave by several orders of magnitude when $N \geq 40$. For MAX-CUT, by $N = 55$, the factor is 10^7 , and when extrapolated to $N = 100$, it exceeds 10^{20} .

Graph density and performance

Earlier in Fig. 3, we compared MAX-CUT on dense graphs and sparse regular graphs to show that the DW2Q’s performance depends strongly on graph sparseness. Fixing the problem size and varying the edge

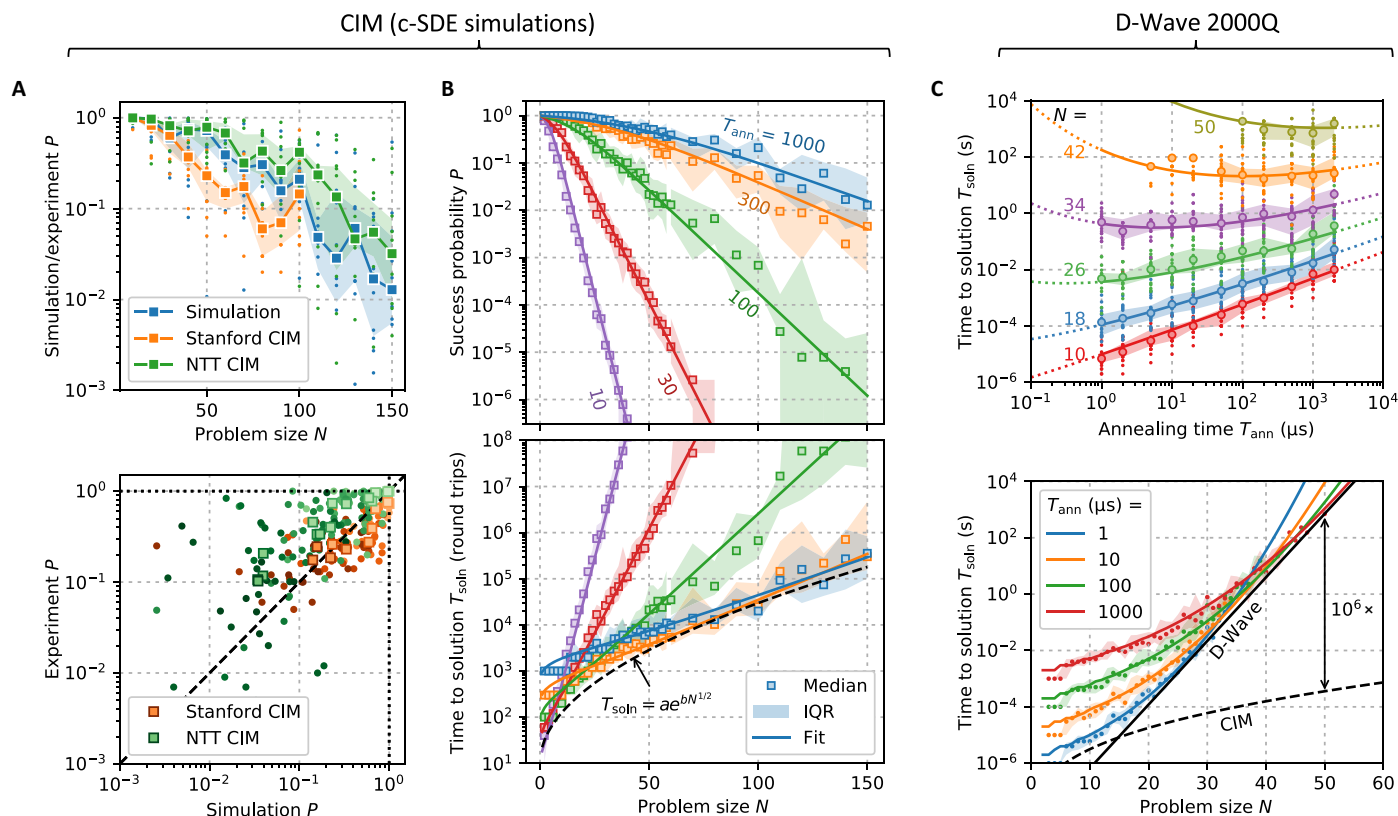


Fig. 4. Time to solution for D-Wave and CIM at optimal annealing time dense MAX-CUT instances. (A) CIM experimental performance versus c-SDE simulations. (B) CIM success probability and time to solution (given in terms of the number of round trips) as a function of problem size N . The effective round-trip time for the NTT CIM [(2.5N) ns; see section S2] is used to convert this figure to seconds. (C) D-Wave time to solution as a function of annealing time T_{ann} and problem size N . Dashed curve shows optimal CIM T_{soln} from (B) for comparison.

Table 1. Time to solution T_{soln} for SK, dense MAX-CUT, and $d = 3$ MAX-CUT problems on D-Wave and NTT CIM (see section S2). The annealing time for D-Wave runs was chosen (in the range [1, 1000] μ s) to optimize T_{soln} (see section S4). All CIM data are for fixed anneal times (1000 round trips). "Factor" refers to the ratio of solution times $T_{soln}^{(DW)} / T_{soln}^{(CIM)}$.

SK				MAX-CUT (dense)				MAX-CUT ($d = 3$)			
N	DW2Q	CIM	Factor	N	DW2Q	CIM	Factor	N	DW2Q	CIM	Factor
10	6.0 μ s	25 μ s	0.2	10	6.0 μ s	25 μ s	0.2	10	1.0 μ s	50 μ s	0.02
20	35 μ s	100 μ s	0.3	20	0.4 ms	100 μ s	4	20	3.0 μ s	100 μ s	0.03
40	6.1 ms	0.4 ms	15	40	6.1 s	0.4 ms	10^4	50	12 μ s	0.4 ms	0.03
60	1.4 s	0.6 ms	2000	55	10^4 s	1.2 ms	10^7	100	100 μ s	3.3 ms	0.03
80*	(400 s)	1.8 ms	(10^5)	80*	$(10^{11}$ s)	1.8 ms	(10^{13})	150	2.8 ms	22 ms	0.1
100*	$(10^5$ s)	3.0 ms	(10^7)	100*	$(10^{19}$ s)	2.3 ms	(10^{21})	200	11 ms	51 ms	0.2

*D-Wave solution times extrapolated using $P = e^{(W/N_0)^2}$ fits in Figs. 2C and 3B. Note that dense problems with $N > 61$ are not embeddable in the DW2Q.

density, we see the same effect and can fill in the gap between sparse graphs and dense graphs. We constructed random unweighted graphs of degree $d = 1, 2, \dots, (N - 2)$ for each graph size $N = 20, 30, 40, 50, 60$. The success probabilities for DW2Q and the CIM are shown in Fig. 5A (for clarity only, $N = 40$ CIM data are shown). In this case, we used clique embeddings for all problems, so for a given N , all the embeddings are the same. Even with the embeddings fixed, the DW2Q finds sparse problems

easier to solve than dense ones. The reason is that, consistent with (42), the optimal constraint coupling is weaker for sparse problems than for dense problems (Fig. 5B). In general, we find that $J_c \propto d$ for fixed N . Having a large constraint coupling could be problematic because the physical quantum annealer scales the largest coupling coefficient to the maximum coupling strength on the chip; the constraints max out this coupling and cause the logical couplings to be downscaled proportionally

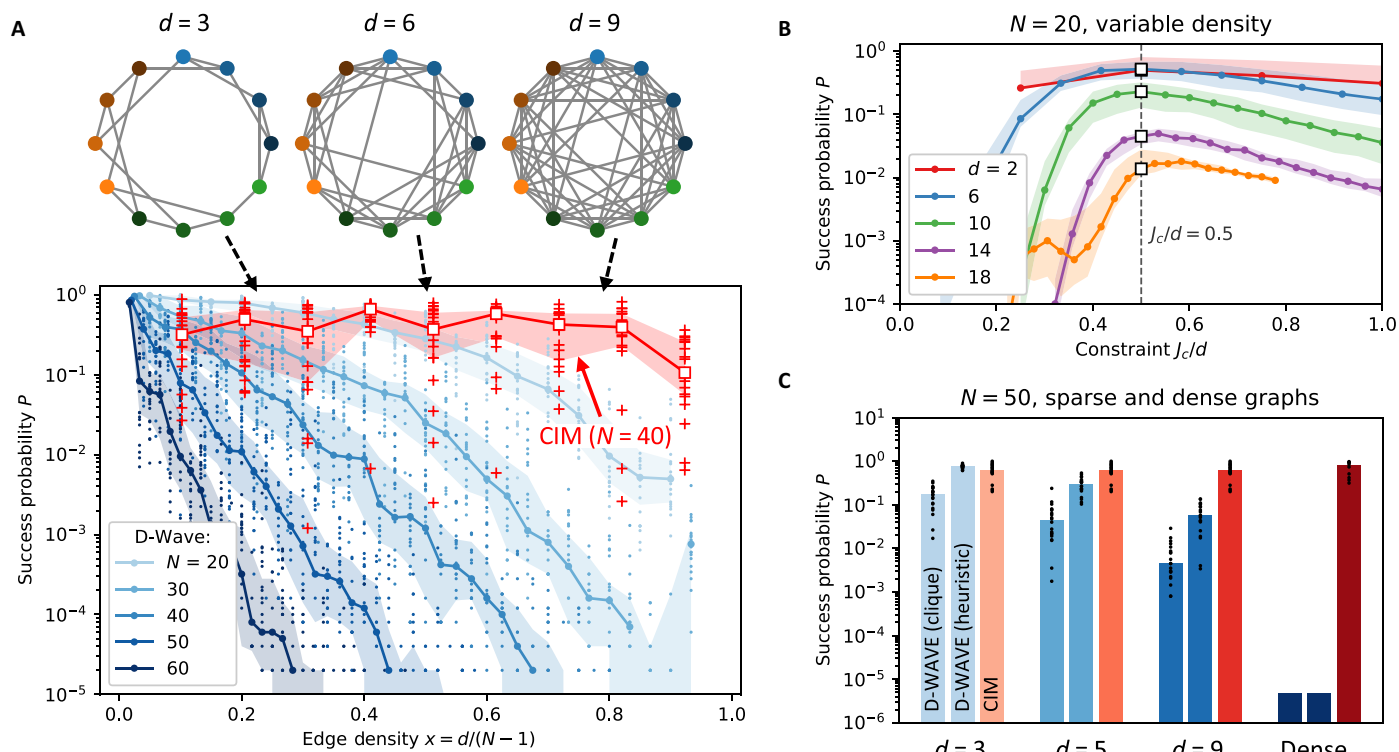


Fig. 5. Relation between edge density and annealer performance. (A) Success probability as a function of edge density. Native clique embeddings used for D-Wave. Optimal embedding parameter [see subgraph (B)] is used, with $T_{\text{ann}} = 1000 \mu\text{s}$. (B) D-Wave success probability as a function of graph degree, showing that the optimal J_c scales as $J_c \propto d$ for fixed N . For fixed edge density, the N dependence was determined previously to be $J_c \propto N^{3/2}$; see Fig. 3A. (C) Comparison of D-Wave and NTT CIM success probabilities for $N = 50$ using both clique embeddings and heuristically determined embeddings ($T_{\text{ann}} = 1000 \mu\text{s}$; dense D-Wave bars are extrapolation from $e^{-(N/N_0)^2}$ fit in Fig. 3B).

as J_c^{-1} . Thus, dense graphs have weaker logical couplings in the embedded problem, hindering the annealer's ability to find the ground state because of parameter misspecification or "intrinsic control errors" (42, 50).

The CIM has only weak dependence on the edge density $x = d/(N-1)$. Earlier work on $N = 100$ graphs (24), as well as the CIM data plotted in Fig. 3C, is consistent with this result. This suggests that the CIM has promise as a general-purpose Ising solver, achieving good performance on a large class of problems, irrespective of connectivity.

Comparing Figs. 5A and 3C, we can glean some insight regarding the effect of embedding overhead on the D-Wave quantum annealer's performance. The heuristic embeddings in Fig. 3C are designed to minimize the overhead factor (ratio of physical qubits to logical qubits). This ratio is much larger for the native-clique embeddings, growing linearly, i.e., as $O(N)$ (see section S1). Figure 5C compares these two D-Wave settings against the CIM at $N = 50$; while the CIM outperforms on all graphs with $d \geq 5$, the difference between the success probabilities using clique and heuristic embeddings suggests that performance is heavily dependent on embedding overhead and the difference grows with edge density (and graph size). This illustrates an additional tradeoff in quantum annealing: poor-performing but easy-to-find embeddings versus well-performing embeddings that require substantial precomputation. This tradeoff is expected to favor the well-performing embeddings when the number of qubits (or connections) becomes large.

DISCUSSION

In conclusion, we have benchmarked the DW2Q system hosted at NASA Ames and measurement-feedback CIMs hosted at Stanford

University and NTT Basic Research Laboratories, focusing on MAX-CUT problems on random graphs and SK spin-glass models, and found that the merits of each machine are highly problem dependent. Connectivity appears to be a key factor in performance differences between these machines. Problems with sparse connectivity, such as one-dimensional chains [compare (51) and (52)] and MAX-CUT on cubic graphs (Fig. 3), can be embedded into the DW2Q with little or no overhead, resulting in similar performance between the quantum annealer and the CIMs. However, the embedding overhead for dense problems such as SK is very steep, requiring $O(N^2)$ physical qubits to represent a size- N graph and resulting in large embedded problems that decrease the performance of the quantum annealer. The ability to avoid an embedding overhead likely contributes to the CIM's performance advantage on SK models that grows exponentially with the square of the problem size. For problems of intermediate sparseness, such as MAX-CUT on regular graphs of small degree $d \geq 5$, the CIM is still faster by a large factor.

Ultimately, it is overall quantities such as wall clock time or energy usage that are of practical interest. Read-in and read-out times, classical pre- and postprocessing, and energy usage must be included in a comprehensive evaluation. Both CIMs and superconducting qubit quantum annealers are in early stages of development, with these quantities currently in flux. Moreover, to beat state-of-the-art classical techniques (section S5) on the problems studied in this paper, advances will be required. D-Wave has recently implemented features allowing unconventional control of the annealing process that can significantly improve results, and as mentioned above, efforts to improve the connectivity are under way. A key question will be the extent to which

these technologies can harness quantum effects for computational purposes. Signatures of entanglement have been seen in D-Wave quantum annealers, although it remains open the extent to which the computation makes use of entanglement-related effects. CIMs, already interesting as semiclassical computational devices, can, in principle, also have entanglement (53) by either building scalable all-optical couplings (30, 54) (albeit with low losses being required) or by creating entanglement in the measurement-feedback architecture, e.g., by performing entanglement swapping.

While the path forward for designing improved CIMs and quantum annealers involves many different aspects, this paper has primarily observed results that can be interpreted as being related to connectivity differences between the machines that were benchmarked. It has been conjectured often that increased internal connectivity in quantum annealers can improve performance (55, 56), and there are large projects under way to realize higher-connectivity quantum annealers [including efforts by D-Wave and the IARPA Quantum Enhanced Optimization (QEO) program (57) and Google (58)]. Our results provide strong experimental justification for this line of development.

MATERIALS AND METHODS

Sample problems

For fully connected SK and MAX-CUT on dense graphs, 20 random instances were created of each size $N = 2, 3, \dots, 61$ for the D-Wave. Of these, the $N = 2, 10, 20, \dots, 60$ instances were also used for the CIM. An additional set of random instances were created for $N = 70, 80, \dots, 150$ for the CIM using the same algorithm.

For the sparse-graph analysis, we computed regular graphs of size $N = 2, 4, \dots, 300$ and degree $d = 3, 4, \dots, 20$, with 20 instances for each pair (N, d) . The algorithm randomly assigns edges to eligible vertices until all reach the required degree (and backtracks if it gets stuck). The same algorithm was also used for the variable-density graphs: $d = 1, 2, \dots, (N - 2)$ for $N = 20, 30, 40, 50, 60$, creating 20 instances per pair (N, d) .

Exact SK ground states were found with the Spin Glass Server (59), which uses BiqMac (60), an exact branch-and-bound algorithm. For SK instances of size $N \leq 100$, the algorithm obtained proven ground states. For $N > 100$, the solver timed out before exhausting all branches (runtime $T = 3000$ s), so the result is not a guaranteed ground state; however, we believe that it reaches the ground state with high probability for $N \leq 150$ because multiple runs of the algorithm give the same state energy and none of the CIM runs found an Ising energy lower than the Spin Glass Server result. MAX-CUT ground states for $N \leq 30$ were found by brute-force search on a personal computer; for $20 \leq N \leq 150$, a BLS algorithm was used (5). Although BLS is a heuristic solver, for $N \leq 150$, it finds the ground state with nearly 100% probability, giving us high confidence that the BLS solutions are ground states. While the brute-force solver, D-Wave, and the CIM found states of equal energy to the BLS solution (if run long enough), they never found states of lower energy.

D-wave annealers

Initial D-Wave experiments were performed on the D-Wave 2X at NASA Ames Research Center and the D-Wave 2X online system at D-Wave Systems Inc. Later runs were made on the DW2Q at NASA Ames, once that machine came online. The 2X and 2000Q systems use a C12 (12 cells \times 12 cells \times 4 qubits) and C16 (16 \times 16 \times 4) Chimera, respectively. For all-to-all graphs, D-Wave 2X supports $N \leq 48$ and 2000Q supports $N \leq 64$ (the number is slightly smaller because of broken qubits). All $N \leq 48$ runs were consistent across the three ma-

chines and with extrapolation of data in (42) from runs performed on a different set of instances on the earlier-generation machine D-Wave Two. All data reported in this paper came from the DW2Q.

Embeddings were precomputed for all problems (heuristic embeddings for sparse MAX-CUT and native clique embeddings for SK, dense MAX-CUT, and variable-density MAX-CUT) so that runs in different conditions (e.g., annealing times and constraint couplings) would use the same embeddings. For each problem type, the optimal annealing parameter J_c is found as a function of problem size N by sweeping J_c (section S1). The optimal J_c was found to be independent of the annealing time. The standard annealing schedule was used in all experiments, but the annealing time was tuned. Most instances were run 10^4 to 10^5 times in total depending on the observed success rate (the especially hard $N \geq 50$ MAX-CUT instances were run up to 4×10^6 times). Five to 10 different embeddings were used per instance, and the success probability was averaged. Spin-reversal transformations were used to avoid spurious effects. After an anneal, each logical qubit value was determined by taking the majority vote of all qubits in the chain.

In all figures, the shaded regions give the [25th, 75th] percentile range [interquartile range (IQR)] for the data. Figures 2B, 3A, and 5A show individual instances as dots and the solid line gives the median. Figures 2C, 3 (B and C), and 5B are too crowded to show D-Wave instances; the dots give medians and the smooth lines give analytic fits. For CIM data, medians and IQR are shown in Figs. 2C and 3B, while Fig. 3C only shows medians and IQR due to crowding.

Coherent Ising machine

CIM experiments were performed on the 100 OPO CIM at Ginzton Laboratory of Stanford University and the 2048 OPO CIM at NTT Basic Research Laboratories. The Stanford and NTT devices are described in (24, 25), respectively. Computation time of the Stanford CIM is 1.6 ms, which is the time for 1000 round trips of the 320-m fiber ring cavity. Since the NTT CIM processes a 2000-node problem in 5.0 ms, which is the time for 1000 round trips of the 1-km fiber ring cavity, we can solve up to $\lfloor 2000/N \rfloor$ problems in parallel per computation time.

The CIM's reliable operation depends on relative phases between the OPO pulses, injection pulses, and measurement local oscillator pulses being kept stable and well calibrated. Such phase stabilization is imperfect in the experimental setups used in this study, and consequently, post-selection procedures have been applied to both the Stanford and NTT CIM experimental data. This is described in detail in section S2. Computation times have been reported in terms of annealing times; as with the DW2Q, these times exclude the time required to transfer data to and from the CIM.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/5/5/eaau0823/DC1>

Section S1. D-Wave embeddings and J_c optimization

Section S2. CIM data and post-selection

Section S3. C-SDE simulations of CIM

Section S4. Optimal-annealing time analysis

Section S5. Performance of parallel tempering

Fig. S1. D-Wave success probability for SK problems and MAX-CUT problems of edge density 0.5 as a function of problem size N and embedding parameter J_c .

Fig. S2. MAX-CUT on graphs with an edge density of 0.5.

Fig. S3. Properties of heuristic embeddings for fixed-degree graphs.

Fig. S4. Choice of optimal coupling for sparse graphs using the heuristic embedding.

Fig. S5. Data filtering and post-selection in NTT CIM.

Fig. S6. Comparison of Stanford and NTT CIM performance for SK and dense MAX-CUT problems.

Fig. S7. Abstract schematic of measurement-feedback CIM.

Fig. S8. Simulated CIM success probability as a function of F_{\max} for the SK, dense MAXCUT, and cubic MAX-CUT problems in this paper.

Fig. S9. Comparison of c-SDE simulations with experimental CIM data.

Fig. S10. Simulated CIM success probability and time to solution (in round trips) for SK and MAX-CUT problems.

Fig. S11. Time-to-solution analysis for D-Wave at optimal annealing time.

Fig. S12. CIM time to solution compared against the parallel tempering algorithm implemented in the Unified Framework for Optimization.

Table S1. Seven steps in a single round trip for the measurement-feedback CIM and the appropriate truncated Wigner description.

Table S2. Problem-dependent constants α and β used in the relation $N_0 = \alpha + \beta \log_{10}(T/\mu\text{s})$ for the success probability exponential $P = e^{-(N/N_0)^2}$.

References (61–64)

REFERENCES AND NOTES

- S. Rudich, A. Wigderson, *Computational Complexity Theory* (American Mathematical Society, 2004), vol. 10.
- R. M. Karp, *Complexity of Computer Computations* (Springer, 1972), pp. 85–103.
- M. X. Goemans, D. P. Williamson, Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *JACM* **42**, 1115–1145 (1995).
- S. Kirkpatrick, C. D. Gelatt Jr., M. P. Vecchi, Optimization by simulated annealing. *Science* **220**, 671–680 (1983).
- U. Benlic, J.-K. Hao, Breakout local search for the Max-Cut problem. *Eng. Appl. Artif. Intell.* **26**, 1162–1173 (2013).
- C. Yoshimura, M. Yamaoka, M. Hayashi, T. Okuyama, H. Aoki, K.-i. Kawarabayashi, H. Mizuno, Uncertain behaviours of integrated circuits improve computational performance. *Sci. Rep.* **5**, 16213 (2015).
- S. Tsukamoto, M. Takatsu, S. Matsubara, H. Tamura, An accelerator architecture for combinatorial optimization problems. *FUJITSU Sci. Tech. J.* **53**, 8–13 (2017).
- B. Sutton, K. Y. Camsari, B. Behin-Aein, S. Datta, Intrinsic optimization using stochastic nanomagnets. *Sci. Rep.* **7**, 44370 (2017).
- A. Parihar, N. Shukla, M. Jerry, S. Datta, A. Raychowdhury, Vertex coloring of graphs via phase dynamics of coupled oscillatory networks. *Sci. Rep.* **7**, 911 (2017).
- X. Yin, B. Sedighi, M. Varga, M. Ercsey-Ravasz, Z. Toroczkai, X. S. Hu, Efficient analog circuits for boolean satisfiability. *IEEE Trans. Very Large Scale Integr. VLSI Syst.* **26**, 155–167 (2018).
- A. N. Tait, T. F. de Lima, E. Zhou, A. X. Wu, M. A. Nahmias, B. J. Shastri, P. R. Prucnal, Neuromorphic photonic networks using silicon photonic weight banks. *Sci. Rep.* **7**, 7430 (2017).
- E. Farhi, J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren, D. Preda, A quantum adiabatic evolution algorithm applied to random instances of an NP-complete problem. *Science* **292**, 472–475 (2001).
- T. Kadowaki, H. Nishimori, Quantum annealing in the transverse Ising model. *Phys. Rev. E* **58**, 5355–5363 (1998).
- G. E. Santoro, E. Tosatti, Optimization using quantum mechanics: quantum annealing through adiabatic evolution. *J. Phys. A Math. Gen.* **39**, R393–R431 (2006).
- S. Boixo, T. F. Rønnow, S. V. Isakov, Z. Wang, D. Wecker, D. A. Lidar, J. M. Martinis, M. Troyer, Evidence for quantum annealing with more than one hundred qubits. *Nat. Phys.* **10**, 218–224 (2014).
- G. E. Santoro, R. Martoňák, E. Tosatti, R. Car, Theory of quantum annealing of an Ising spin glass. *Science* **295**, 2427–2430 (2002).
- R. D. Somma, D. Nagaj, M. Kieferová, Quantum speedup by quantum annealing. *Phys. Rev. Lett.* **109**, 050501 (2012).
- T. F. Rønnow, Z. Wang, J. Job, S. Boixo, S. V. Isakov, D. Wecker, J. M. Martinis, D. A. Lidar, M. Troyer, Defining and detecting quantum speedup. *Science* **345**, 420–424 (2014).
- T. Albash, D. A. Lidar, Demonstration of a scaling advantage for a quantum annealer over simulated annealing. *Phys. Rev. X* **8**, 031016 (2018).
- V. Choi, Minor-embedding in adiabatic quantum computation: I. The parameter setting problem. *Quantum Inf. Process* **7**, 193–209 (2008).
- J. Cai, W. G. Macready, A. Roy, A practical heuristic for finding graph minors. arXiv:1406.2741 [quant-ph] (10 June 2014).
- V. Choi, Minor-embedding in adiabatic quantum computation: II. minor-universal graph design. *Quantum Inf. Process* **10**, 343–353 (2011).
- C. Klymko, B. D. Sullivan, T. S. Humble, Adiabatic quantum programming: Minor embedding with hard faults. *Quantum Inf. Process* **13**, 709–729 (2014).
- P. L. McMahon, A. Marandi, Y. Haribara, R. Hamerly, C. Langrock, S. Tamate, T. Inagaki, H. Takesue, S. Utsunomiya, K. Aihara, R. L. Byer, M. M. Fejer, H. Mabuchi, Y. Yamamoto, A fully programmable 100-spin coherent Ising machine with all-to-all connections. *Science* **354**, 614–617 (2016).
- T. Inagaki, Y. Haribara, K. Igarashi, T. Sonobe, S. Tamate, T. Honjo, A. Marandi, P. L. McMahon, T. Umeki, K. Enbutsu, O. Tadanaga, H. Takenouchi, K. Aihara, K.-i. Kawarabayashi, K. Inoue, S. Utsunomiya, H. Takesue, A coherent Ising machine for 2000-node optimization problems. *Science* **354**, 603–606 (2016).
- F. Barahona, On the computational complexity of Ising spin glass models. *J. Phys. A Math. Gen.* **15**, 3241–3253 (1982).
- S. Kirkpatrick, D. Sherrington, Solvable model of a spin-glass. *Phys. Rev. Lett.* **35**, 1792–1796 (1975).
- R. W. Boyd, *Nonlinear Optics* (Academic Press, 2003).
- Z. Wang, A. Marandi, K. Wen, R. L. Byer, Y. Yamamoto, Coherent Ising machine based on degenerate optical parametric oscillators. *Phys. Rev. A* **88**, 063853 (2013).
- A. Marandi, Z. Wang, K. Takata, R. L. Byer, Y. Yamamoto, Network of time-multiplexed optical parametric oscillators as a coherent Ising machine. *Nat. Photonics* **8**, 937–942 (2014).
- H. Takesue, T. Inagaki, 10 GHz clock time-multiplexed degenerate optical parametric oscillators for a photonic Ising spin network. *Opt. Lett.* **41**, 4273–4276 (2016).
- D. F. Walls, G. J. Milburn, *Quantum Optics* (Springer Science & Business Media, 2007).
- G. Milburn, D. Walls, Production of squeezed states in a degenerate parametric amplifier. *Opt. Commun.* **39**, 401–404 (1981).
- P. Drummond, K. McNeil, D. Walls, Non-equilibrium transitions in sub/second harmonic generation. *J. Mod. Opt.* **28**, 211–225 (1981).
- L.-A. Wu, H. J. Kimble, J. L. Hall, H. Wu, Generation of squeezed states by parametric down conversion. *Phys. Rev. Lett.* **57**, 2520–2523 (1986).
- P. Kinsler, P. D. Drummond, Quantum dynamics of the parametric oscillator. *Phys. Rev. A* **43**, 6194–6208 (1991).
- K. Takata, A. Marandi, R. Hamerly, Y. Haribara, D. Maruo, S. Tamate, H. Sakaguchi, S. Utsunomiya, Y. Yamamoto, A 16-bit coherent Ising machine for one-dimensional ring and cubic graph problems. *Sci. Rep.* **6**, 34089 (2016).
- R. Hamerly, K. Inaba, T. Inagaki, H. Takesue, Y. Yamamoto, H. Mabuchi, Topological defect formation in 1D and 2D spin chains realized by network of optical parametric oscillators. *Int. J. Mod. Phys. B* **30**, 1630014 (2016).
- W. R. Clements, J. J. Renema, Y. H. Wen, H. M. Chrzanowski, W. S. Kolthammer, I. A. Walmsley, Gaussian optical Ising machines. *Phys. Rev. A* **96**, 043850 (2017).
- Y. Yamamoto, K. Aihara, T. Leleu, K.-i. Kawarabayashi, S. Kako, M. Fejer, K. Inoue, H. Takesue, Coherent Ising machines—Optical neural networks operating at the quantum limit. *npj Quantum Inf.* **3**, 49 (2017).
- T. Boothby, A. D. King, A. Roy, Fast clique minor generation in chimera qubit connectivity graphs. *Quantum Inf. Process* **15**, 495–508 (2016).
- D. Venturelli, S. Mandrà, S. Knysch, B. O’Gorman, R. Biswas, V. Smelyanskiy, Quantum optimization of fully connected spin glasses. *Phys. Rev. X* **5**, 031040 (2015).
- Y. Fu, P. W. Anderson, Application of statistical mechanics to NP-complete problems in combinatorial optimisation. *J. Phys. A Math. Gen.* **19**, 1605–1620 (1986).
- Y. Haribara, S. Utsunomiya, Y. Yamamoto, Computational principle and performance evaluation of coherent Ising machine based on degenerate optical parametric oscillator network. *Entropy* **18**, 151 (2016).
- D. Coppersmith, D. Gamarnik, M. Hajiaghayi, G. B. Sorkin, Random max sat, random max cut, and their phase transitions. *Random Struct. Algorithms* **24**, 502–545 (2004).
- P. Alimonti, V. Kann, Hardness of approximating problems on cubic graphs, in *Proceedings of the Third Italian Conference on Algorithms and Complexity (CIAC)*, Berlin, Heidelberg, 12 to 14 March 1997, pp. 288–298.
- R. Dechter, J. Pearl, *The Cycle-Cutset Method for Improving Search Performance in AI Applications* (University of California, Computer Science Department, 1986).
- C. C. McGeoch, W. Bernoudy, J. King, Comment on “Scaling advantages of all-to-all connectivity in physical annealers: The Coherent Ising Machine vs D-Wave 2000Q”. arXiv:1807.00089 [quant-ph] (29 June 2018).
- S. Mandrà, H. G. Katzgraber, A deceptive step towards quantum speedup detection. *Quantum Sci. Technol.* **3**, 04LT01 (2018).
- D-Wave Systems Inc., Technical description of the D-Wave quantum processing unit, Part of D-Wave API documentation (2016).
- B. Gardas, J. Dziarmaga, W. H. Zurek, M. Zwolak, Defects in quantum computers. *Sci. Rep.* **8**, 4539 (2018).
- T. Inagaki, K. Inaba, R. Hamerly, K. Inoue, Y. Yamamoto, H. Takesue, Large-scale Ising spin network based on degenerate optical parametric oscillators. *Nat. Photonics* **10**, 415–419 (2016).
- D. Maruo, S. Utsunomiya, Y. Yamamoto, Truncated Wigner theory of coherent Ising machines based on degenerate optical parametric oscillator network. *Phys. Scr.* **91**, 083010 (2016).
- Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, M. Soljačić, Deep learning with coherent nanophotonic circuits. *Nat. Photonics* **11**, 441–446 (2017).
- H. G. Katzgraber, F. Hamze, R. S. Andrist, Glassy chimeras could be blind to quantum speedup: Designing better benchmarks for quantum annealing machines. *Phys. Rev. X* **4**, 021008 (2014).

56. E. G. Rieffel, D. Venturelli, B. O’Gorman, M. B. Do, E. M. Prystay, V. N. Smelyanskiy, A case study in programming a quantum annealer for hard operational planning problems. *Quantum Inf. Process* **14**, 1–36 (2015).
57. S. Weber, G. Samach, D. Rosenberg, J. Yoder, D. Kim, A. Kerman, W. Oliver, Hardware considerations for high-connectivity quantum annealers. *Bull. Am. Phys. Soc.* (2018).
58. Y. Chen, C. Quintana, D. Kafri, A. Shabani, B. Chiaro, B. Foxen, Z. Chen, A. Dunsworth, C. Neill, J. Wenner, H. Neven, J. Martinis; Google Quantum Hardware Team, Progress towards a small-scale quantum annealer I: Architecture. *Bull. Am. Phys. Soc.* (2017).
59. F. Liers, M. Jünger, Spin glass server; <http://informatik.uni-koeln.de/spinglass/>.
60. F. Rendl, G. Rinaldi, A. Wiegele, Solving Max-Cut to optimality by intersecting semidefinite and polyhedral relaxations. *Math. Program.* **121**, 307–335 (2010).
61. D-Wave Systems Inc., Chimera embedding (2016); <https://github.com/dwavesystems/chimera-embedding>.
62. A. D. King, W. Bernoudy, J. King, A. J. Berkley, T. Lanting, Emulating the coherent Ising machine with a mean-field algorithm. arXiv:1806.08422 [quant-ph] (21 Jun 2018).
63. S. Mandrà, H. G. Katzgraber, C. Thomas, The pitfalls of planar spin-glass benchmarks: raising the bar for quantum annealers (again). *Quantum Sci. Technol.* **2**, 038501 (2017).
64. S. Mandrà, Z. Zhu, W. Wang, A. Perdomo-Ortiz, H. G. Katzgraber, Strengths and weaknesses of weak-strong cluster problems: A detailed overview of state-of-the-art classical heuristics versus quantum approaches. *Phys. Rev. A* **94**, 022337 (2016).

Acknowledgments: We acknowledge S. Mandrà for useful discussions and parallel-tempering simulation results and D. Lidar, A. King, and C. McGeoch for helpful correspondence.

Funding: This research was funded by the Impulsing Paradigm Change through Disruptive Technologies (ImPACT) Program of the Council of Science, Technology and Innovation (Cabinet Office, Government of Japan). R.H. is supported by an IC Postdoctoral Research Fellowship at MIT, administered by ORISE through U.S. DOE and ODNI. P.L.M. was partially supported by a Stanford Nano- and Quantum Science and Engineering Postdoctoral Fellowship. D.V. acknowledges funding from NASA Academic Mission Services contract no. NNA16BD14C. H.M., E.N., and T.O. acknowledge funding from NSF award no. PHY-1648807. D.E. acknowledges support from the U.S. ARO through the ISN at MIT (no. W911NF-18-2-0048) and

the SRC-NSF E2CDA program. **Author contributions:** Y.Y., P.L.M., and E.R. proposed the project. R.H. performed D-Wave experiments and data analysis and prepared the figures. R.H., P.L.M., D.V., and T.I. wrote the manuscript. T.I. and P.L.M. performed NTT and Stanford CIM experiments, respectively. D.V. helped with D-Wave experiments and data analysis. A.M., C.L., R.L.B., M.M.F., and H.M. contributed to building the Stanford CIM. K.I., T.H., K.E., T.U., R.K., and H.T. contributed to building the NTT CIM. R.H. performed simulations of the CIM, adapting code from P.L.M., E.N., and T.O. E.R., Y.Y., and A.M. assisted with preparation of the manuscript. S.U., S.K., K.-i.K., and D.E. assisted with interpretation of the results. **Competing interests:** T.I., H.T., T.H., S.U., Y.Y. are inventors on patent JP6429346 awarded in November 2018 to National Institute of Informatics (NII), NTT, and Osaka University that covers an OPO pulse sequence for calculation and stabilization of CIM. A.M., Y.Y., R.L.B., and S.U. are inventors on patent US9830555 awarded in November 2017 to Stanford University that covers a CIM based on a network of OPOs. S.U., Y.Y., and H.T. are inventors on patent US10140580 awarded in November 2018 to NII and NTT that covers a CIM using measurement feedback. T.I., K.I., H.T., and T.H. are inventors on patent application PCT/JP2018/038994 submitted by NTT that covers a phase checking scheme for the CIM. T.U. and K.E. are inventors on patent JP5856083 awarded in February 2016 to NTT that covers phase-sensitive amplifiers based on periodically poled lithium niobate waveguides. P.L.M. is an advisor to QC Ware Corp. The authors declare no other competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 4 May 2018

Accepted 17 April 2019

Published 24 May 2019

10.1126/sciadv.aau0823

Citation: R. Hamerly, T. Inagaki, P. L. McMahon, D. Venturelli, A. Marandi, T. Onodera, E. Ng, C. Langrock, K. Inaba, T. Honjo, K. Enbutsu, T. Umeki, R. Kasahara, S. Utsunomiya, S. Kako, K.-i. Kawarabayashi, R. L. Byer, M. M. Fejer, H. Mabuchi, D. Englund, E. Rieffel, H. Takesue, Y. Yamamoto, Experimental investigation of performance differences between coherent Ising machines and a quantum annealer. *Sci. Adv.* **5**, eaau0823 (2019).