Inter-generational comparison of quantum annealers in solving hard scheduling problems

Bibek Pokharel1 · Zoe Gonzalez Izquierdo2,3 · P. Aaron Lott2,3 · Elena Strbac4 · Krzysztof Osiewalski4 · Emmanuel Papathanasiou4 · Alexei Kondratyev5,6 · Davide Venturelli2,3 · Eleanor Rieffel2

Received: 19 June 2022 / Accepted: 11 July 2023 / Published online: 4 October 2023
© The Author(s) 2023

Abstract
We compare the performance of four quantum annealers, the D-Wave Two, 2X, 2000Q, and Advantage in solving an identical ensemble of a parametrized family of scheduling problems. These problems are NP-complete and, in fact, equivalent to vertex coloring problems. They are also practically motivated and closely connected to planning problems from artificial intelligence. We examine factors contributing to the performance differences while separating the contributions from hardware upgrades, support for shorter anneal times, and possible optimization of ferromagnetic couplings. While shorter anneal times can improve the time to solution (TTS) at any given problem size, the scaling of TTS with respect to the problem size worsens for shorter anneal times. In contrast, optimizing the ferromagnetic coupling improves both the absolute TTS and the scaling. There is a statistically significant improvement in performance between D-Wave Two and 2X and from all older generation annealers to Advantage, even when operated under identical anneal time and ferromagnetic couplings. However, the performance improvement from 2X to 2000Q requires the anneal time and ferromagnetic couplings to be optimized. Overall, owing to these inter-generational hardware improvements and optimizations, the scaling exponent reduces from 1.01 ± 0.01 on Two to 0.259 ± 0.008 on Advantage.
Keywords Quantum annealing · Scheduling · Graph coloring

1 Introduction

Theoretically, quantum computers are superior to classical computers when quantum mechanical laws are carefully exploited in deterministic algorithms [1, 2]. However, in real-world deployment, empirical performance can take precedence over theoretical superiority. In fact, for most classical algorithms that perform well empirically on practical tasks, no proof exists for estimates or bounds on their performance. For this reason, empirical experimentation of quantum heuristics is critical to understanding the breadth of quantum computing’s impact.

Quantum annealing is one of the most prominent quantum metaheuristics for optimization [3–9]. Quantum annealing has become accessible to experimentation in recent years, and the company D-Wave has produced and commercialized multiple devices in the last decade. Starting with the 128-qubit D-wave One, each generation of annealer has aimed to improve performance on optimization problems. The D-Wave Two, 2X, 2000Q, and Advantage systems (the four we use in this study) have 512, 1152, 2048, and 5640 qubits, respectively. The D-Wave 2000Q introduced several features like a shorter minimum anneal time, increased control of the annealing schedule, and a wider range of ferromagnetic coupling strengths. Advantage, the newest device, uses a denser native graph structure with 15 instead of 6 degrees of coupling between qubits. In addition to these increased number of qubits and features, these newer generations of annealers also have reduced noise and better calibration and reduction in analog control errors [10, 11].

This report evaluates how the various updates made to these machines over the last few years impact their effectiveness in solving hard problems of practical interest. While the comparative performance of Advantage and 2000Q was discussed in [12] for exact cover problems, to our knowledge, this is the first investigation of its kind that compares the performance of four different generations of quantum annealers with respect to an identical set of applied problems. Our results can be summarized as follows:

- Under the default operating conditions and fixed anneal times, 2X outperformed Two, and Advantage outperformed all its predecessors. While 2000Q was able to solve larger problems than 2X, there was no statistical difference in the average TTS between the two machines. Improvement from Two to 2X is likely due to reduced specification errors and other hardware level improvements. For Advantage, a large part of the performance increase can be attributed to increased connectivity.
- The shortest possible anneal time gives the shortest TTS for most problem sizes, with the effect being most pronounced for small-sized problems. However, a longer anneal time is necessary to obtain solutions for the largest problem sizes that are addressable only via the Advantage platform. Crucially, the scaling of the TTS with respect to size increases as we decrease the anneal time. Overall, in future architectures with lower anneal times, problems of different sizes might benefit from size-appropriate anneal times.
• In general, optimizing the ferromagnetic coupling beyond the default settings lowers the TTS at each problem size. However, the magnitude of this improvement is device-dependent. On 2X and 2000Q, the change in TTS due to such optimization is nearly negligible. However, on Advantage, there is a substantial reduction in the scaling exponent after ferromagnetic coupling optimization. While our brute-force optimization is informative, we suspect that as the devices and the problems increase in size, a more sophisticated method of optimizing the coupling will have to be employed.

The structure of the paper is as follows. Section 2 briefly reviews quantum annealing. Section 3 describes the scheduling problem and the generation of problem instances. Section 4 discusses the methodological details of our experiments. In Sect. 5 we discuss our results, and summarize their implications in Sect. 6.

2 Review of quantum annealing

In quantum annealing, the device starts by implementing an initial Hamiltonian $H_I$, and the overall Hamiltonian $H$ is parametrically changed until it becomes the final Hamiltonian $H_F$. This process corresponds to the following Hamiltonian evolution

$$H(s) = A(s)H_I + B(s)H_F, \ s \in [0, 1],$$

where $A(s)$ is a monotonically decreasing function and $B(s)$ is a monotonically increasing function such that $A(1) = B(0) = 0$. $H_F$ encodes the cost-function landscape of a combinatorial optimization problem in its eigenvalues, and its ground state maps to the optimal solution to the problem. Quantum annealing utilizes both thermal and quantum fluctuations when exploring the cost-function landscape. So, in principle, it is more expressive and possibly more effective than simulated annealing or parallel tempering methods, where quantum effects like tunneling are not possible [13]. However, due to practical aspects such as finite anneal time, temperature, noise, and sub-optimal parameter setting, the final output of an annealing run might not necessarily be a low-energy state of the problem Hamiltonian. Consequently, quantum annealers are usually treated as black-box optimizers that need to be iteratively run in conjunction with a parameter setting strategy. In benchmarking studies, once a problem instance is programmed into the annealer and a set of parameter values chosen, a run consisting of a large number of anneals is typically performed to obtain statistics (see Sect. 4 and in particular Table 1 for details). Two complementary metrics are commonly used to report results; the first one is the probability of the ground state solution

$$P_{gs} = \frac{\text{Number of ground state solutions}}{\text{Total anneals}},$$

where “Total anneals” refers to the number of anneals in one run as described above. Its counterpart, the time-to-solution (TTS) is defined as the expected time to obtain the ground state solution with 0.99 success probability for a specific anneal time, and
Table 1 Hardware details for all four generations of the D-Wave annealers, the first three hosted at NASA Ames Research Center and the last one accessed through the D-Wave Leap cloud platform, are shown.

| D-Wave        | Two       | 2X        | 2000Q     | Advantage |
|---------------|-----------|-----------|-----------|-----------|
| QPU Image     | ![Image](image1.png) | ![Image](image2.png) | ![Image](image3.png) | ![Image](image4.png) |
| Total qubits  | 512       | 1152      | 2048      | 5640      |
| Working qubits| 509       | 1097      | 2031      | 5436      |
| Architecture  | Chimera   | Chimera   | Chimera   | Pegasus   |
| Min anneal time| 20 μs    | 5 μs      | 1 μs      | 1 μs      |
| Temperature   | 13.2 mK   | 12 mK     | 12.1 mK   | 15.8 ± 0.5 mK |
| Anneals per gauge| 45,000   | 10,000    | 10,000    | 500       |
| Problem size  | 8–16      | 8–20      | 8–24 not optimized | 8–40 |
| Gauges        | 10        | 10        | 100 n=20,22,24 | 20–400 depending on J_F, n and t |

The images of the chips were provided by D-Wave Systems. Note that the number of working qubits for a specific device is typically slightly lower than the number of total qubits since fabrication defects leave a small portion of components unusable. This table also lists the default problem sizes, gauges, number of anneals, and anneal time used for each D-Wave annealer.
is computed as

$$\text{TTS} = \frac{\ln(1 - 0.99)}{\ln(1 - P_{gs})} t,$$

where $t$ is the anneal time for a single annealing repetition. The results presented here focus on TTS since it provides a more direct representation of the resources (i.e., time) required to solve optimization problems, and this data is valuable to researchers. Given that comparisons are presented across devices and performed with the same annealing times when possible (limited by the annealing time range of older devices), the conclusions translate to a $P_{gs}$ comparison. In practice, the total time needed to solve a problem will also be affected by processes outside of the actual annealing, such as programming time (which takes the longest but is only needed once before each run), readout time, etc. (see [14] for QPU timing details). These are not included in the total time used for calculating TTS, in accordance with standard practice in benchmarking literature. It is also worth noting that these additional time requirements are common to all the devices considered for this study, making their relevance much more minor than if the comparison were between quantum and classical devices.

The D-Wave quantum annealers have 2-local architectures, meaning that there are pairwise couplings between the qubits, i.e., the Hamiltonian $H_F$ is an Ising Hamiltonian. Consequently, the optimization problem is formulated as a quadratic unconstrained binary optimization (QUBO) problem [15–17], where the cost function is a quadratic function of binary variables. We will refer to obtaining the QUBO for a problem as mapping the problem to its QUBO. As each qubit in the quantum annealer is connected to a small subset of the other qubits, representing a single binary variable usually requires multiple physical qubits to reproduce the QUBO’s connectivity faithfully. This set of physical qubits representing a single logical variable is called a vertex model, and the process of finding vertex models for each logical qubit in the problem is called embedding of the QUBO in the hardware chip. The physical qubits within a vertex model are subjected to ferromagnetic coupling $J_F$ in the final Hamiltonian. Finding the value $J_F$ that maximizes the TTS is called the parameter-setting problem [18]. For a more detailed review of quantum annealing on D-Wave machines, the reader is invited to read Ref. [19–21].

### 3 Parameterized families of scheduling problems

Scheduling problems deal with the allocation of time and resources under certain constraints. While these NP-complete problems are important on their own, they also have many applications to planning problems from artificial intelligence [22]. Here, we consider the task of assigning $k$ time-slots to $n$ tasks while avoiding any possible time conflicts (double-scheduling). These problems are equivalent to vertex-coloring problems, which are also a class of NP-complete problems. In particular, when we map the time-slots to colors and tasks to vertices, the edges between the vertices correspond to constraints between the tasks (see Fig. 1). So the colorability of the
Each vertex must only be colored once:

$$\sum_{i=1}^{n} \left( 1 - \sum_{c=1}^{k} x_{ic} \right)^2$$

Fig. 1 Example of a schedule for $n = 8$ is shown with a valid 3-coloring. Each scheduling problem can be mapped to a graph-coloring problem by identifying tasks with nodes and task conflicts as edges.

The coloring task can be represented using doubly-indexed binary variables $x_{ic}$, where $x_{ic} = 1$ (0) means that the $i$th vertex is colored (not colored) with color $c$. Each vertex must be colored only once, and adjacent vertices must have different colors. Using these two conditions, we get the QUBO:

$$H(x) = \sum_{i=1}^{n} \left( 1 - \sum_{c=1}^{k} x_{ic} \right)^2 + \sum_{(i,j) \in E} \sum_{c=1}^{3} x_{ic} x_{jc}$$

(4)

For a valid coloring, this objective function will give $H(x) = 0$.

We consider an ensemble of problems that are not easily colorable or trivially uncolorable. To be more specific, we consider the 3-colorability of a specific set of Erdos-Renyi graphs $G_{n,p}$. $G_{n,p}$ are graphs with $n$ vertices such that the probability of having an edge between each pair of vertices is $p$. A combinatorial phase transition for $k$-colorability problems is known in terms of the parameter $d = \frac{e}{n}$ where $e$ is the number of edges [23]. This phase transition occurs when the difficulty of finding the coloring goes from easy to hard to easy. This easy-hard-easy transition of colorability is well studied, and there are upper and lower bounds on the transition parameter $d$. However, finding the exact location of the phase transition is still an open problem.
We use $d = 4.5$ for the generation of our graphs, following the study in [25]. We generate these graphs using a C++ program, which is an extension of the graph-generation method used by Culberson et al. [26]. The graphs used in this paper are identical to the set used by Rieffel et al. [27] in their case study about the performance of quantum annealers in solving planning-type problems.

4 Methods

Quantum annealing runs were performed on the three generations of D-Wave quantum annealers (Two, 2X, and 2000Q) housed at NASA Ames Research Center and the latest generation (Advantage), accessed through the D-Wave Leap cloud platform. The number of qubits, minimum anneal times, and base operating temperature for all four annealers are listed in Table 1. We used 100 problem instances at each problem size while restricting the instances to ones with at least one valid schedule (i.e., coloring). We found the embeddings for the QUBO instances using D-Wave’s native heuristic find_embedding [28].

To mitigate biases toward certain solutions formed by asymmetries in the processors, we used random gauges - bits 0,1 are randomly mapped to either -1,1 or 1,-1. Each gauge corresponds to a specific choice of this bit-to-spin mapping. Due to spin-reversal symmetry, the problem Hamiltonian is unaffected by the gauge choice. Table 1 outlines the number of gauges used for different problem sizes. We check for the number of valid schedules (i.e., colorings) obtained using Eq. (4) and report the corresponding median TTS. We bootstrap the median TTS from 100 problem instances to get 5000 samples. The reported TTS is the mean of this ensemble, and the error bars correspond to 95% confidence intervals. Scheduling problems are NP-complete problems, so we expect our TTS to scale exponentially with size in the asymptotic limit. This exponential scaling is expected both for quantum and classical algorithms, with the scaling varying between different algorithms [25]. So we fit $TTS = T_0 e^{n \alpha}$ [29], where $n$ is the size of the problem, and report the scaling exponent $\alpha$. We plot the median, 35th, and 65th percentiles of the TTS across instances; the error bars for the bootstrapped results are not plotted but used when computing the fits.

Each generation of D-Wave annealers has a larger chip layout structure than their predecessor, increasing the size of problems that can be embedded in each device. Moreover, the Advantage platform has transitioned the topology from a Chimera graph to a Pegasus graph [10, 30] (see Fig. 2), which has a higher connectivity. We were able to embed problems with up to 40 tasks on D-Wave Advantage and D-Wave 2000Q, and up to 32 tasks on D-Wave 2X. However, problems of sizes larger than 16, 20, and 24 on Two, 2X and 2000Q respectively had $P_{gs} = 0$. For each machine, the last plotted dot in the curve for Figs. 3, 4, 5, 6 indicates the largest problem size for which the statistics analysis returned a non-infinite median. Working with larger problem sizes would require millions of runs per instance to obtain sufficient statistics.

In line with the typical practice for applied problems [7, 18], we optimized the ferromagnetic couplings. For this optimization, we set the anneal time to 5, 1, and 20 $\mu$s for 2X, 2000Q, and Advantage, respectively. In particular, for each machine, we chose the anneal time that gave the shortest TTS for the largest problem size we
considered before \( J_F \) optimization. When optimizing the ferromagnetic coupling \( J_F \), we chose \( J_F \in [-0.5, -2.0] \) for 2X and \((-0.625, -1.375)\) for 2000Q, with steps of 0.125. For both machines, relatively few problem sizes had their optimal coupling at \( |J_F| > 1.0 \). \( J_F = -2.0, -1.375 \) were not the optimal coupling for any problem sizes on 2X and 2000Q, respectively. In our plots, ‘i.opt’ refers to \( J_F \) optimized for each problem instance separately. A similar optimization of \( J_F \) was done for D-Wave Two by Rieffel et al. in [27], so we do not repeat those experiments here. For 2000Q data, we restrict \( J_F \) optimization to \( n < 20 \) due to limited computational resources. Moreover, for 2X and 2000Q, we performed coupling optimization at various anneal times and found that the optimal coupling did not depend on the anneal time. Optimization of \( J_F \) on Advantage has been recently studied in detail by some of the same author [31]. Here we explored the range \( J_F \in [-0.4, -1] \) for \( n = 24 \) and \( t = 20\mu s \) and found \( J_F = -0.5 \) to be optimal (labeled ‘opt’). Unless specified otherwise, the default problem sizes, gauges, number of anneals, and anneal times used are detailed in Table 1. We discuss the results in the next section.

5 Results

D-Wave’s quantum annealers allow the user to vary the anneal time while keeping the weight functions in Eq. (1) fixed. Using the shortest possible anneal time increases the number of solutions obtained in a fixed period, but it also affects the quality of the solutions [27]. We evaluate this tradeoff by considering TTS as a function of the anneal time and problem size. Figure 3 shows TTS as a function of problem size at various anneal times for D-Wave 2X, 2000Q, and Advantage. Generally, when varying \( t \) while fixing the problem size \( n \), the lowest possible anneal time gives the lowest TTS for most graph sizes. However, the improvement is less significant for larger problem sizes, and the error bars on the TTS also increase with \( n \). In other words, the improvement
Fig. 3  TTS as a function of problem-size across generations at various anneal times. Ferromagnetic coupling is set to the default value. The scaling exponent is labeled as $\alpha_{\text{machine, anneal time, } J_F}$ optimization offered due to shorter anneal time becomes negligible for larger problem sizes. The scaling exponent $\alpha$ increases slightly for shorter anneal times. This effect is most pronounced in Advantage where the exponent goes from $\alpha_{A,20 \mu s, \text{def}} = 0.351 \pm 0.014$ at 20 $\mu$s to $\alpha_{A,1 \mu s, \text{def}} = 0.508 \pm 0.022$ at 1 $\mu$s. However, note that if we limit ourselves to $n \leq 22$ (like for 2000Q), Advantage’s optimal annealing time would also be 1 $\mu$s, and it would outperform 2000Q for the smaller sizes. Overall, our results suggest that lowering anneal time further in future architectures could benefit small-
Fig. 4  TTS as a function of problem-size across generations. $J_F$ is optimized at the best performing anneal time. Optimizing $J_F$ both reduces the TTS at each size and the exponent $\alpha$. The scaling exponent is labeled as $\alpha_{\text{machine, anneal time, } J_F \text{ optimization}}$.

Sized scheduling problems. Experiments using future annealers featuring lower anneal times and allowing to solve larger problem sizes might benefit from assigning different anneal times to problems of different sizes.

Generally, $J_F$, the ferromagnetic coupling applied to each vertex model, is set to be larger than the coupling between the logical qubits $J_{ij}$ and the local field $h_i$. However, if $J_F$ is arbitrarily large, the coupled qubits may have the same spin, but the solution
might not correspond to the ground state of the QUBO \cite{32, 33}. By default, on D-Wave annealers, $J_F = -1.0$ (where $J_{ij}, h_i \in [-1.0, +1.0]$, with this range expanded for the latest generations). We considered optimizing $J_F$ across problem sizes and problem instances (except for Advantage, where we only optimized for $n = 24$).

In general, we found that setting $J_F = -1.0$, as D-Wave’s native algorithm does, is not the optimal setting. Rieffel et al. \cite{27} had argued for non-optimality of this default setting and further observed that for D-Wave Two, the ratio of the optimal ferromagnetic coupling $|J_F|$ to the maximum internal coupling $|J_{ij}|$ decreases with problem sizes for scheduling-type problems. We do not observe this trend on the newer machines. For 2X and 2000Q, $|J_F| = 0.875$ and $0.75$ generally gives the shortest median TTS (see Appendix B for further details). For Advantage, we found $-0.5$ to be the optimal value after exploring the range $J_F \in [-0.4, -1]$ for $n = 24$ and $t = 20 \mu s$.

As shown in Fig. 4, the optimization of $J_F$ leads to a slight, statistically insignificant reduction in TTS for 2X and 2000Q. Optimizing $J_F$ further across each vertex-model using standard optimization methods, like Nelder-Mead and gradient descent, was unwieldy. Due to the noisiness in the solution probability landscape, we observed that these search methods could not improve upon instance-level optimization. We suspect that as quantum annealers become less noisy, similar or more sophisticated optimization methods might succeed.

The effect of $J_F$ optimization is most pronounced for Advantage. Here for all sizes and annealing times, $J_F = -0.5$ performs better than the default setting. For example, the median TTS for the $n = 40$ instances is infinity for all annealing times (1,5, and 20 $\mu s$) with $|J_F| = 1$, while it is finite for $|J_F| = 0.5$, and as low as $2.8 \times 10^4 \mu s$ when the annealing time is 20 $\mu s$. Moreover, we also see the scaling exponent decrease by a factor of two before and after $J_F$ optimization. We suspect that this change has to do with the improved connectivity of the Pegasus architecture. As the vertex models are much smaller in Advantage compared to the older annealers, a strong $|J_F|$ is no longer necessary to keep these vertex models from breaking. Overall, our observations make a strong case for the non-optimality of the default $J_F$ setting.

![Fig. 5](image-url)
Lastly, we compare the performance of each subsequent generation of annealers by solving identical problems at default and optimized settings (see Fig. 5). In order to avoid artificial flatness in the scaling due to inefficiency at smaller problem sizes, when we consider the optimized setting, we report the lowest TTS across anneal times and JF optimization. The results in both configurations confirm the intuition that as the machine got upgraded, the performance improved. Note that the effect of hardware improvements between 2X and 2000Q was minimal. While optimizing the anneal time and JF was shown to reduce the TTS slightly, these minor improvements add up. Using the optimized JF at the lowest possible anneal time of 1 µs, we do observe shorter TTS in 2000Q than 2X. Besides the obvious opportunity to test larger problem sizes, we observe a change in slope for the median TTS, which is very appreciable between generations and especially striking on Advantage. This performance improvement is primarily, but not entirely due to the smaller vertex model sizes required for embedding a densely connected Ising in the Advantage chip, owing to the increased connectivity of the Pegasus graph (see Appendix A for more details).

6 Conclusion

We analyzed and quantified the comparative performance of four generations of quantum annealers, D-Wave Two, 2X, 2000Q, and Advantage in solving a parameterized family of hard scheduling problems. By solving an identical ensemble of problems on each machine, we highlight how the hardware updates, lower anneal times, and the size and the connectivity of the architectures affect the time-to-solution for these problems. Under the default settings, we found a noticeable improvement in performance between 2X and Two, but not when updating from 2X to 2000Q. Advantage outperforms all of its predecessors. Using the shortest possible anneal time gave the lowest TTS at most problem sizes, but this improvement got less pronounced as the problem size increased. For the largest problem sizes, using a longer anneal time yielded the best results. Optimizing the ferromagnetic coupling across each problem instance also reduced TTS, but the improvement was not pronounced for 2X and 2000Q. On Advantage, we found a substantial reduction in TTS when using an optimized ferromagnetic coupling, both in absolute terms and the scaling exponent. Overall, the exponent improved by nearly a factor of six, from $1.006 \pm 0.01$ on Two to $0.259 \pm 0.008$ on Advantage.

Hardware upgrades and optimization of operational parameters like anneal times and ferromagnetic couplings are crucial for improving quantum annealing performance. Using sophisticated annealing schedules, as demonstrated in [6, 7], are other ways to improve hardware performance that we did not consider in this report. Quantifying the performance of scheduling problems using these advanced annealing schedules would be a natural extension of this work. As the problems used here can be generated systematically for larger problem sizes [25], these problems serve as a valuable and fair method of benchmarking future quantum annealers, both against other annealers and state-of-the-art classical heuristics.
Acknowledgements  We want to thank JF, MD, and NASA QuAIL team members for their valuable input throughout this investigation. We are grateful to Daniel Lidar and Tameem Albash for useful comments on the manuscript. BP did a significant portion of this work as a graduate student at the University of New Mexico (UNM) and as an intern at NASA and Stinger Ghaffarian Technologies (SGT), so he wants to thank UNM Physics and Astronomy Department and SGT for their support. Recent work of ZGI, DV, and PAL is supported by USRA NASA Academic Mission Service (NNA16BD14C). The NASA QuAIL team also appreciates the support from the AFRL Information Directorate under grant F4HBKC4162G001, DARPA under IAA 8839 annex 125, and the Office of the Director of National Intelligence (ODNI) and the Intelligence Advanced Research Projects Activity (IARPA), via IAA 145483. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of ODNI, IARPA, AFRL, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes, notwithstanding any copyright annotation thereon. The computer time on D-Wave Leap for the Advantage system was provided thanks to the support of Standard Chartered Bank.

Funding  Open access funding provided by SCELC, Statewide California Electronic Library Consortium

Data availability  All relevant data are available upon request.

Declarations

Conflict of interest  The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Open Access  This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Appendix A: Effect of embedding size

One of the main differences between D-Wave Advantage and its predecessors is the Pegasus architecture. The older architecture, Chimera, is a subgraph of Pegasus and has substantially lower graph connectivity. This results in a substantial reduction in the chain length for embedded problems. In Fig. 6, we show the average embedding size as a function of problem size. We first compute the average vertex model size for each problem instance and then report the median of this average size across instances. The error bars correspond to the 35th and 65th percentile of this distribution. The absolute and the relative scaling difference in the embedding size is substantial.

In the same figure, we also plot TTS as a function of the number of physical qubits instead of problem size (as in Fig. 5). These results focus on cases where we have already optimized the anneal time and ferromagnetic coupling on both machines. By comparing TTS as a function of problem size, we noted that the relative improvement in the scaling was $\alpha_Q/\alpha_A = 2.37$. The comparison in Fig. 6 shows that a substantial portion of this improvement is due to a smaller embedding size. In particular,
\[ \frac{\alpha_Q^{\text{physical}}}{\alpha_A^{\text{physical}}} = 1.67. \] Crucially, even when accounting for embedding sizes, for larger problem sizes, Advantage still outperforms 2000Q.

**Appendix B: Optimization of J_F over problem sizes**

To optimize the ferromagnetic coupling, we ran each instance at various values of \( J_F \). These couplings can be optimized for each vertex model, problem instance, or problem size. Optimizing at the vertex model level was ineffective. We tried to use various optimization algorithms for this task unsuccessfully, owing to a large number of parameters and the noisiness of the landscape. The results reported in this work correspond to \( J_F \) optimization at the instance level (‘i.opt’) unless otherwise specified. Figure 7 shows a heat map of the TTS for each problem size (‘u.opt’) such that all vertex models of all problem instances have this uniform \( J_F \). The heat map is calibrated such that the darkest color is the best value for that size. The reported TTS is averaged over 100 instances. On D-Wave Two, the optimal \( J_F \) at each problem size seemed to increase with problem size [27]. This trend does not appear on the 2X and 2000Q. In fact, for these two annealers, the setting \( J_F = 0.875 \) and 0.75 seems to work well for most problem sizes. Fig. 8 shows how TTS scales under ‘i.opt’ and ‘u.opt’. On both 2X and 2000Q, we notice a slight reduction in the TTS scaling under ‘i.opt’. We do not report \( J_F \) optimization for Advantage, which is studied in detail in [31]. Here we did optimize for \( J_F \in [-0.4, -1] \) for \( n = 24 \) and \( t = 20\mu s \) and found \( J_F = -0.5 \) to be optimal. For the rest of the problem sizes, we only considered \( J_F = -1.0 \) and \(-0.5\); the latter value is close to Advantage’s ‘u.opt’ value.

**References**

1. Rieffel, E.G., Polak, W.: A Gentle Introduction to Quantum Computing. MIT Press, Cambridge (2011)
2. Nielsen, M., Chuang, I.L.: Quantum Computing and Quantum Information. Cambridge University Press, Cambridge (2001)
Fig. 7: Optimal ferromagnetic coupling for different problem sizes. TTS is reported at different coupling parameter values and various graph sizes. The color scaling is calibrated for each problem size and goes from yellow (large TTS) to orange (medium TTS) to red (little TTS) (Color figure online)

3. Rieffel, E.G., Venturelli, D., O’Gorman, B., Do, M.B., Prystay, E.M., Smelyanskiy, V.N.: A case study in programming a quantum annealer for hard operational planning problems. Quantum Inf. Process. 14(1), 1 (2015)
4. Venturelli, D., Mandrà, S., Knysh, S., O’Gorman, B., Biswas, R., Smelyanskiy, V.: Quantum optimization of fully-connected spin glasses. Phys. Rev. X 5(3), 031040 (2015)
5. Kim, M., Venturelli, D., Jamieson, K.: Leveraging quantum annealing for large MIMO processing in centralized radio access networks, in Proceedings of the ACM special interest group on data communication (Association for Computing Machinery, New York, NY, USA, 2019), SIGCOMM ’19, pp. 241-255
6. Marshall, J., Venturelli, D., Hen, I., Rieffel, E.G.: Power of pausing: advancing understanding of thermalization in experimental quantum annealers. Phys. Rev. Appl. 11, 044083 (2019)
7. Gonzalez Izquierdo, Z., Grabbe, S., Hadfield, S., Marshall, J., Wang, Z., Rieffel, E.: Ferromagnetically shifting the power of pausing. Phys. Rev. Appl. 15, 044013 (2021)
8. Farhi, E., Goldstone, J., Gutmann, S., Sipser, M.: Quantum computation by adiabatic evolution. arXiv:quant-ph/0001106 (2000)
Fig. 8 TTS as a function of problem-size for two different $J_F$ optimization methods. ‘u.opt’ and ‘i.opt’ refer to whether $J_F$ was optimized uniformly across all the problems of the same size or individually for each problem instance, respectively.

9. Smelyanskiy, V.N., Rieffel, E.G., Knysh, S.I., Williams, C.P., Johnson, M.W., Thom, M.C., Macready, W.G., Pudenz, K.L.: A near-term quantum computing approach for hard computational problems in space exploration. arXiv:1204.2821 (2012)
10. McGeoch, C., Farré, P.: The D-Wave Advantage system: an overview. D-Wave Tech. report 14-1049A-A, D-Wave Sys (2020)
11. Error sources for problem representation. https://docs.dwavesys.com/docs/latest/c_qpu_ice.html
12. Willsch, D., Willsch, M., Calaza, C.D.G., Jin, F., Raedt, H.D., Svensson, M., Michielsen, K.: Benchmarking advantage and D-Wave 2000Q quantum annealers with exact cover problems, arXiv:2105.02208 (2021)
13. Finnila, A., Gomez, M., Sebenik, C., Stenson, C., Doll, J.: Quantum annealing: a new method for minimizing multidimensional functions. Chem. Phys. Lett. 219(5–6), 343 (1994)
14. The QPU timing. https://docs.dwavesys.com/docs/latest/c_qpu_timing.html
15. Choi, V.: Minor-embedding in adiabatic quantum computation: i the parameter setting problem. Quantum Inf. Process. 7(5), 193 (2008)
16. Lucas, A.: Ising formulations of many NP problems. arXiv:1302.5843 (2013)
17. Shin, S.W., Smith, G., Smolin, J.A., Vazirani, U.: Comment on “Distinguishing classical and quantum models for the D-Wave device”. arXiv:1404.6499 (2014)
18. Pudenz, K.L.: Parameter setting for quantum annealers, arXiv e-prints arXiv:1611.07552 (2016)
19. Johnson, M., Amin, M., Gildert, S., Lanting, T., Hamze, F., Dickson, N., Harris, R., Berkley, A., Johansson, J., Bunyk, P., et al.: Quantum annealing with manufactured spins. Nature 473(7346), 194 (2011)
20. Job, J., Lidar, D.: Test-driving 1000 qubits, quantum. Sci. Technol. 3(3), 030501 (2018)
21. Boothby, K., Enderud, C., Lanting, T., Molavi, R., Tsai, N., Volkmann, M.H., Altmare, F., Amin, M.H., Babcock, M., Berkley, A.J., Baron Aznar, C., Boschnak, M., Christiani, H., Ejtemaee, S., Evert, B., Gullen, M., Hager, M., Harris, R., Hoskinson, E., Hilton, J.P., Jooya, K., Huang, A., Johnson, M.W., King, A.D., Ladizinsky, E., Li, R., MacDonald, A., Medina Fernandez, T., Neufeld, R., Norouzpour, M., Oh, T., Ozfidan, I., Padden, P., Perminov, I., Poulin-Lamarre, G., Prescott, T., Raymond, J., Reis, M., Rich, C., Roy, A., Sadeghi Esfahani, H., Sato, Y., Sheldan, B., Smirnov, A., Svensson, L.J., Whittaker, J., Yao, J., Yarovoy, A., Bunyk, P.I.: Architectural considerations in the design of a third-generation superconducting quantum annealing processor, arXiv e-prints arXiv:2108.02322 (2021)
22. Chien, S., Rabideau, G., Knight, R., Sherwood, R., Engelhardt, B., Mutz, D., Estlin, T., Smith, B., Fisher, F., Barrett, T., Stebbins, G., Tran, D.: in Proceedings of SpaceOps (2000)
23. Achlioptas, D., Friedgut, E.: A sharp threshold for k-colorability. Random Struct. Algorithms 14(1), 63 (1999)
24. Coja-Oghlan, A.: Upper-bounding the k-colorability threshold by counting covers, upper-bounding the k-colorability threshold by counting covers. arXiv:1305.0177 (2013)
25. Rieffel, E.G., Venturelli, D., Hen, I., Do, M., Frank, J.: Parametrized families of hard planning problems from phase transitions, in Proceedings of the twenty-eighth aaai conference on artificial intelligence (AAAI-14), pp. 2337 – 2343 (2014)
26. Culberson, J., Beacham, A., Papp, D.: in Proceedings of the CP95 workshop on studying and solving really hard problems, pp. 31–42 (1995)
27. Rieffel, E.G., Venturelli, D., O’Gorman, B., Do, M.B., Prystay, E.M., Smelyanskiy, V.N.: A case study in programming a quantum annealer for hard operational planning problems. Quantum Inf. Process. 14(1), 1 (2015)
28. Cai, J., Macready, W.G., Roy, A.: A practical heuristic for finding graph minors, arXiv preprint arXiv:1406.2741 (2014)
29. Kowalsky, M., Albash, T., Hen, I., Lidar, D.A.: 3-regular three-XORSAT planted solutions benchmark of classical and quantum heuristic optimizers. Quantum Sci. Technol. 7(2), 025008 (2022)
30. Boothby, K., Bunyk, P., Raymond, J., Roy, A.: Next-generation topology of d-wave quantum processors. Technical report 14-1026A-C, D-Wave Sys (2019)
31. Izquierdo, Z.G., Grabbe, S., Idris, H., Wang, Z., Marshall, J., Rieffel, E.: The advantage of pausing: parameter setting for quantum annealers, arXiv preprint arXiv:2205.12936 (2022)
32. Könz, M.S., Lechner, W., Katzgraber, H.G., Troyer, M.: Embedding overhead scaling of optimization problems in quantum annealing. PRX Quantum 2(4), 040322 (2021)
33. Fang, Y.L., Warburton, P.: Minimizing minor embedding energy: an application in quantum annealing. Quantum Inf. Process. 19(7), 191 (2020)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.