The Quantum Computing Company™

Fast Anneals and Coherent Quantum Annealing

WHITEPAPER

Summary

As of April 2024, *fast anneals* are available on all D-Wave quantum computers through our standard solver API. This whitepaper describes the fast-anneal protocol and presents some examples of new capabilities made possible by fast anneals.

Introduction. Four recent papers by D-Wave researchers and collaborators describe a series of remarkable results on problems arising in quantum simulation of spin glasses:

[1] A.D. King et al., Coherent quantum annealing in a programmable 2,000-qubit Ising chain, *Nature Physics* (2022).

[2] A.D. King et al., Quantum critical dynamics in a 5,000-qubit programmable spin glass, *Nature* (2023).

[3] M. H. Amin et al., Quantum error mitigation in quantum annealing *(in submission)* (2023).

[4] A.D. King et al., Computational supremacy in quantum simulation *(in submission)* (2024).

These results were made possible by the *fast-anneal* protocol, which had been deployed on internal quantum systems to support this research program.

In April 2024, D-Wave announced general availability of fast anneals on all public quantum annealing (QA) systems, including Advantage™ systems and small prototype versions of the Advantage2™ system.[1] The protocol allows users to specify extremely short annealing times as low as 5 nanoseconds (ns), compared to previous minimum times of 0.5 or 1 microseconds (µs), depending on the solver. This whitepaper presents an overview of fast anneals and how they affect quantum performance.

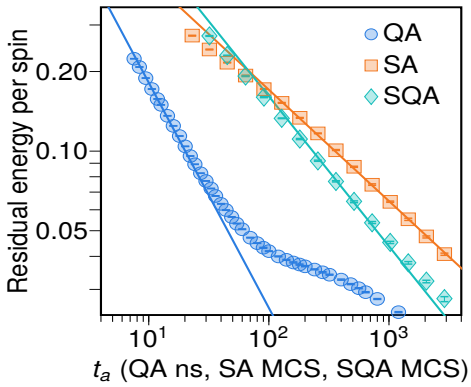

Figure 1: (From Figure 4 of [2].) A plot of QA residual energy versus annealing time. Measured QA data (blue points) matches theoretical predictions (blue line) at faster annealing times, but diverges at slower annealing times. Also shown are plots of solution quality versus computational work, for two classical solvers SA and SQA. (These curves measure work in abstract units and do not reflect true measured runtimes, which are much higher.) The plot demonstrates that QA approaches ground states faster than classical solvers when operating in the coherent quantum annealing regime.

[1] The specific solvers used in experiments described herein were `Advantage_system4.1` and `Advantage2_prototype2.3`.

Scalable coherent quantum annealing. Fast anneals allow the quantum processing unit (QPU) to operate within a computational regime known as *coherent quantum annealing*. In this regime, the effects of thermal noise are negligible, and energy distributions of sampled outputs faithfully match theoretical models of noise-free quantum computations. In layperson's terms, fast anneals can *outrun the noise*, in the sense that the computation is finished before thermal noise has a chance to interact with qubits and degrade solution quality.

Figure 1 presents data from Ref. [2] illustrating this effect. The graph shows measurements of a solution quality metric called residual energy (lower is better), versus annealing time, for the Advantage QPU. At annealing times below 30 ns, the QPU operates in the coherent regime and measurements of residual energy (blue points) closely match theoretical predictions assuming noise-free conditions (blue line). At longer annealing times, the data diverges from theory due to interactions with thermal noise.

Also shown for comparison with their slopes are data curves for two classical methods called simulated annealing (SA) and simulated quantum annealing (SQA). The steeper slope of the QA line indicates that QA algorithm reaches low energies faster than the classical solvers when operating in the coherent regime. This qualifies as a demonstration of *quantum scaling advantage*, for a class of inputs that represent a paradigmatic testbed for combinatorial optimization.

Note that the locations of the SA and SQA curves on the x-axis do not reflect true runtimes, but instead correspond to an abstract measure of computational work called Monte Carlo Sweeps (MCS). Runtime comparisons (in Fig. S11 of Ref. [2]) show that the computation performed by the QPU in 10 ns annealing time requires classical runtimes that are slower by four and nine orders of magnitude, respectively.

Demonstration of quantum supremacy. Ref. [4] describes experimental simulations of nonequilibrium dynamics of a magnetic spin system as it is quenched through a quantum phase transition. In tests on Advantage QPUs and Advantage2 prototype QPUs, using four different spin structures, the authors measure distributional accuracy of certain statistics of spin systems as a function of input size N. Two classical simulation methods, called Matrix Product States (MPS) and Projected Entangled Pair States (PEPS), were run on the Summit and Frontier supercomputers at the Oak Ridge Leadership Computing Facility, and on a cluster computer at the Vector Institute, University of British Columbia. A key feature of the classical simulation methods is that they can trade accuracy for speed via user parameters.

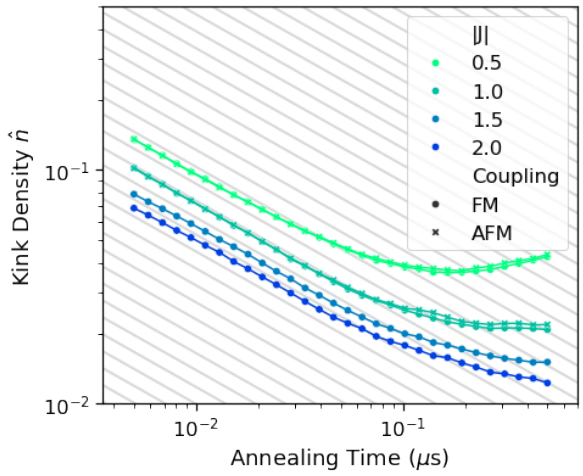

Figure 2: Fast anneals can be used to reproduce results from the published papers. Following Figure 2(a) in Ref. [1], this graph of kink density as a function of annealing time demonstrates that, within the coherent regime, measured data (color-coded points) is consistent with theoretical predictions (slopes of diagonal lines). The colors and markers correspond to coupling weights J, which can be ferromagnetic (FM) or antiferromagnetic (AFM).

The results show that supercomputer-based simulators can generate distributions with equivalent accuracy to those from the QPUs on small inputs, but classical runtimes needed to maintain comparable levels of accuracy grow rapidly with N. The authors estimate that using a supercomputer to produce comparable statistics on the largest problems solved quantumly would require computation times on the scale of millions of years, and power resources that far exceed annual global electricity consumption. This meets Preskill's definition of *quantum supremacy* [5], the demonstration of an efficient quantum computation that could not be reproduced on classical computers within a reasonable time frame.

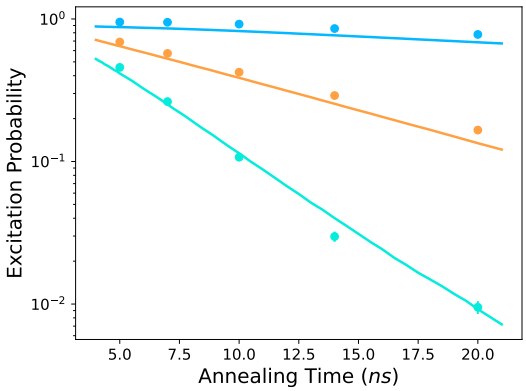

Figure 3: Following Figure 2(c) of Ref. [2], this graph of excitation probabilities versus annealing times shows close agreement between measured data (dots) and predictions of the Schrödinger equation (lines). Color-coding identifies three kinds of random inputs, generated to show a range of possible outcomes.

Using fast anneals to study coherent quantum computation. Availability of the fast-anneal feature on all public QPUs, together with new software tools in the Ocean™ SDK for using this feature, means that researchers who are interested in reproducing or extending the published results, modeling the impact of coherence on quantum performance, developing annealing-based algorithms that leverage coherent-regime quantum dynamics, or otherwise driving quantum simulation research, can begin immediately.

We present a brief overview of the new Ocean tools, with examples showing how they can be used to study properties of coherent quantum computation. Coding details may be found in *D-Wave Solver Docs* [6] and *D-Wave Release Notes* [7]; see also the D-Wave demo *Coherent Annealing: Kibble-Zurek Simulation on a QPU* [8]. Some notes:

- The range of fast-anneal times supported on a given QPU solver can be found by accessing the `fast_anneal_time_range` solver property.
- To invoke the fast-anneal protocol, set the parameter `fast_anneal = True`. The annealing time can be provided using either the `annealing_time` or the `anneal_schedule` parameter.
- Annealing times are specified in units of microseconds, with precision up to six decimal points; see the Solver Docs for details.
- The fast-anneal protocol uses a different anneal schedule than the standard anneal: instead of driving linear growth of the qubits' persistent current, the flux applied to all qubits ($\Phi_{CCJJ}(s)$) is ramped up linearly. This enables the QPU to complete an anneal in nanoseconds and stay within the coherent regime.
- Programming of linear coefficients under the fast-anneal protocol is accomplished by assignment of flux biases rather than by assignment of weights h_i, which are fixed at 0. This changes their dependence on transition time as defined by the normalized anneal fraction; see the Solver Docs for details [9].
- Advancements in the calibration and synchronization of control lines mean that some of the shimming methods required in the papers (see Chern et al. [10]) are no longer necessary for reproducing published results.

Two Examples Inspired by Fig. 2(a) in Ref. [1], Figure 2 shows mean *kink density* as a function of annealing time, for chains of 1024 qubits linked by coupling weights $J \in [-2, -1.5, \pm 1.0, \pm 0.5]$. Negative weights are ferromagnetic (FM) and positive weights are antiferromagnetic (AFM). Kink density correlates with the number of errors in outputs, in which pairs of linked spins do not match the sign of their shared couplings.

The graph demonstrates that measured kink densities (dots) on the Advantage2 prototype QPU match closed-system dynamics predicted by the Kibble-Zurek mechanism (slopes of diagonal sightlines), when operating in the coherent regime. Furthermore, comparison of this graph to that for the `Advantage_system4.1` solver studied in Ref. [1] suggests that the coherent regime on the Advantage2 prototype has been extended by approximately an order of magnitude, due to better noise suppression.

Following Fig. 2(c) in Ref. [2], Figure 3 plots excitation probabilities versus annealing times, measured on the Advantage2 prototype QPU. This experiment uses random inputs of three (color-coded) types, generated to show a range of possible outcomes. The data points show means and error bars (usually smaller than the

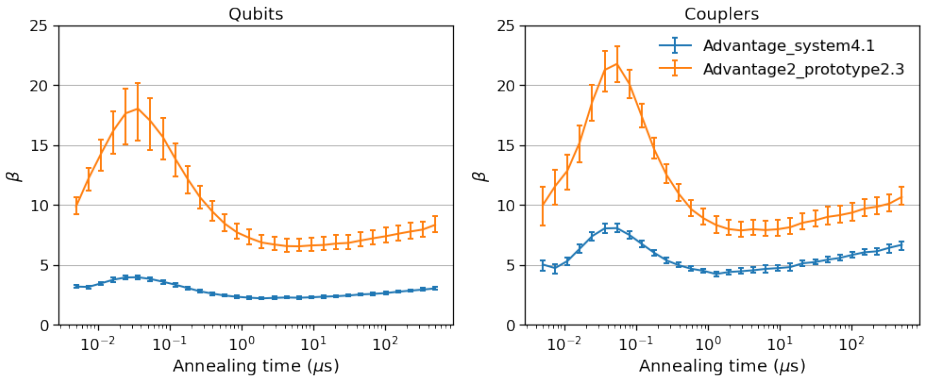

Figure 4: Fast anneals can be used to study effects of noise on solution quality for optimization problems. The panels show how inverse effective temperature β (higher is better), a property of individual qubits and couplers in a specific QPU, depends on annealing time.

markers) for 32 instances and 1000 output samples per input type. The lines show the predicted behavior based on solving the Schrödinger equation.

Comparison of this graph with that for the `Advantage_-system4.1` solver measured in Ref. [2] shows that the lines here are steeper, approaching their asymptotic limits approximately twice as fast, because the energy scale on the Advantage2 prototype QPU is approximately twice as large.

Using fast anneals to study optimization performance. The fast-anneal feature provides a new tool for studying the effects of annealing time and noise on solution quality in quantum optimization. For example, a metric called *inverse effective temperature* (β), which measures excitations in individual qubits and couplers of a given QPU, can be used to characterize baseline effectiveness of the technology at suppressing noise. Higher β corresponds to fewer excitations, which tends to produce lower-energy (better-quality) solutions to optimization problems (see [11] for details).

Figure 4 illustrates the effect of annealing time on β. Data points and error bars show medians and quantile ranges $[10\%, 90\%]$, for individual qubits (left) and couplers (right), measured over the full hardware graphs of an Advantage QPU (blue) and an Advantage2 prototype QPU (orange). Here are some observations:

- The Advantage2 QPU has higher inverse temperature than the Advantage QPU in all tests, due to fabrication improvements that yield better noise reduction.
- Measurements of β on the lower-noise Advantage2 QPU respond strongly to increasing annealing times in the coherent regime, peaking near 30-40 ns. Above that threshold, β decreases due to the growing impact of thermal noise, and then improves slightly because longer anneals partially outweigh the detrimental effects of noise.

It is important to note that fast annealing in the coherent regime does not *necessarily* translate to observable performance improvements on optimization applications. First, because annealing time differences on scales of microseconds to nanoseconds can be negligible when compared to the millisecond-scale access times (for chip programming and readout) experienced in typical QA workflows. Second, because thermally-assisted slow anneals typically yield better solutions than noise-free fast anneals: that is, despite the non-monotonic patterns seen in Figure 4, solution energies for most types of inputs tend to improve monotonically with increasing annealing time.

We have, however, observed some intriguing counter-examples to this general trend. Work is ongoing at D-Wave to leverage improved performance on quantum simulation problems to broader application domains including optimization and machine learning.

We look forward to the new insights, innovations, and performance breakthroughs that are sure to result from wider customer use of this exciting new feature.

Bibliography

[1] A. D. King, S. Suzuki, J. Raymond *et al.*, Coherent quantum annealing in a programmable 2,000 qubit Ising chain, *Nature Physics*, 18, 1324-1328 (2022)

[2] A. D. King, J. Raymond, T. Lanting *et al.*, Quantum critical dynamics in a 5,000-qubit programmable spin glass, *Nature* 617, 61-66 (2023)

[3] A. D. Amin et al., Quantum error mitigation in quantum annealing (manuscript in submission). arXiv:2311.0106 [quant-ph] (2023)

[4] A. D. King, A. Nocera, M.M. Rams *et al.*, Computational supremacy in quantum simulation (manuscript in submission). arXiv:2403.00910 [quant-ph] (2024)

[5] J. Preskill, *Quantum Computing in the NISQ Era and Beyond*. *Quantum* 2, 79 (2018)

[6] D-Wave Solver Docs: Annealing Implementation and Controls: Fast-Anneal Protocol (2024)
docs.dwavesys.com /docs/latest/c_qpu_annealing.html #qpu-annealingprotocol-fast

[7] D-Wave Release Notes: 2024-04-17 Leap Release: Coherent Annealing: Fast-Anneal Protocol Now Generally Available (2024)
docs.dwavesys.com/docs/latest/doc_release_notes.html #coherent-annealing-fast-annealing-protocol-now-generally-available (2024)

[8] D-Wave Demo: Coherent Annealing: Kibble-Zurek Simulation on a QPU (2024)
github.com/dwave-examples/kibble-zurek

[9] D-Wave Solver Docs: Problem-Solving Handbook: QPU Solvers: Configuration (2024)
docs.dwavesys.com/docs/latest/handbook_qpu.html #emulate-linear-biases-with-flux-bias-offsets

[10] K. Chern et al., Tutorial: calibration refinement in quantum annealing, in the Research Topic: Experience with Quantum Annealing Computation, *Frontiers in Computer Science*, Vol. 5 (2023)

[11] J. Raymond et al., Global warming: temperature estimation in annealers, in Sec. Quantum Engineering and Technology, *Frontiers in ICT*, Vol. 3 (2016)