Inhomogeneous driving in quantum annealers can result in orders-of-magnitude improvements in performance

Juan I Adame\textsuperscript{1,3} and Peter L McMahon\textsuperscript{1,2}

\textsuperscript{1} QC Ware Corp., 125 University Ave., Suite 260, Palo Alto, CA 94301, United States of America, \textsuperscript{2} Ginzton Laboratory, Stanford University, Stanford, CA 94305, United States of America

3 Author to whom any correspondence should be addressed.
E-mail: juan.adame@qcware.com and pmcmahon@stanford.edu

Keywords: quantum computing, adiabatic quantum computing, quantum annealing, inhomogeneous quantum annealing

Abstract
Quantum annealers are special-purpose quantum computers that primarily target solving Ising optimization problems. Theoretical work has predicted that the probability of a quantum annealer ending in a ground state can be dramatically improved if the spin driving terms, which play a crucial role in the functioning of a quantum annealer, have different strengths for different spins; that is, they are inhomogeneous. In this paper we describe a time-shift-based protocol for inhomogeneous driving and demonstrate, using an experimental quantum annealer, the performance of our protocol on a range of hard Ising problem instances that have been well-studied in the literature. Compared to the homogeneous-driving case, we find that we are able to improve the probability of finding a ground state by up to $10^7 \times$ for Weak–Strong–Cluster problem instances, and by up to $10^3 \times$ for more general spin-glass problem instances. In addition to being of practical interest as a heuristic speedup method, inhomogeneous driving may also serve as a useful tool for investigations into the physics of experimental quantum annealers.

1. Introduction
Quantum annealing is a form of quantum computation that is primarily targeted at solving Ising combinatorial optimization problems [1–4]. In recent years, there has been great interest in finding whether or not an experimental quantum annealer (QA) can deliver a speedup over the best classical heuristic optimization methods [5–11]. Considerable effort has been put into understanding both what classes of problems might be most amenable to speedup on current experimental systems, as well as the design of modifications to current quantum annealing systems and protocols that may result in improved performance [12–17]. The prospect of experimental quantum annealers delivering a speedup has resulted in a large volume of work exploring potential applications for future quantum annealers, ranging from particle physics [18], to statistics [19], to bioinformatics [20].

A quantum annealer operates [1] by starting with strong quantum-fluctuation terms, called driving terms, that are slowly brought to zero by the end of the computation. Simultaneously, spin–spin couplings and external-field terms, which encode the problem to be solved, are increased from zero between the beginning and the end of the computation. Ideally the QA will end in a ground state of the encoded optimization problem. In practice, the probability of a QA finding a ground state at the end of any particular annealing run is far less than 100\% [5]—probabilities of $10^{-7}$, or even smaller, are routinely observed for many problems on current experimental systems. A longstanding goal of the quantum-annealing community is to discover principles and methods that result in the probability of finding a ground state being maximized. Theoretical work has predicted that dramatic improvements in the success probability can be achieved if the driving terms are applied with different strengths to different spins; that is, they are inhomogeneous [13, 17, 21–25]. Recently it has become possible to experimentally test a restricted form of inhomogeneous driving, in which one does not have arbitrary control over the
are an exponential (in number of qubits, $N$) number of possible choices for advancing or delaying the qubits’ driving schedules, which provides scope for the investigation of a wide variety of strategies for using so-called anneal offsets (AO) to improve the performance of quantum annealers. In this paper we describe one particular strategy, which is distinct from the strategies reported in references [26, 27], and we present results from experiments performed on the D-Wave 2000Q (DW2000Q) QA [28] hosted at NASA Ames in which we show how this AO strategy improves the performance of the annealer across a range of benchmark problems that have previously been studied in the quantum-annealing literature [5, 7, 9, 29, 30].

We now introduce more formally the concept of anneal offsets, and the strategy that we investigate in this paper.

The canonical form of quantum annealing, which involves homogeneous driving terms, is described by the following time-dependent Hamiltonian for $N$ qubits:

$$H(s) = A(s)H_D + B(s)H_P, \quad H_D = \sum_{i=1}^{N} \sigma_x^{(i)}, \quad H_P = \sum_{i=1}^{N} h_i \sigma_z^{(i)} + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} I_{ij} \sigma_x^{(i)} \otimes \sigma_x^{(j)}, \quad (1)$$

where $s \in [0, 1]$ is the normalized time parameter ($s = 0$ is the start of the computation, and $s = 1$ is the end), and superscripts $(i)$ and $(j)$ are qubit indices. The Hamiltonian $H_D$ is the so-called problem Hamiltonian, and its ground states encode the solutions to the classical energy-minimization problem of the Ising model, $\arg\min_{x} \sum_{i=1}^{N} h_i x_i + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} J_{ij} x_i x_j$, where $x := (x_1, \ldots, x_N)$. The time-dependent terms in equation (1) are the schedules: $A(s)$ controls how the driving terms, given in $H_D$, are turned off between time $s = 0$ and $s = 1$, and $B(s)$ controls how the problem-Hamiltonian terms are turned on between time $s = 0$ and time $s = 1$. For this form of quantum annealing, the driving is homogeneous: the driving term for every qubit is turned off using the same schedule $(A(s))$.

The D-Wave 2000Q quantum annealer, when operating with homogeneous driving, has both schedules $A(s)$ and $B(s)$ controlled by a single time-dependent signal $c(s)$. Inhomogeneous driving is made possible by allowing this signal $c(s)$ to be qubit-dependent in a specific way [26, 27]: for each qubit $i$, the signal defining its annealing schedules can be perturbed by an offset $\delta_i$, which results in its driving-term schedule and problem-Hamiltonian-terms schedule being modified. Formally there are now $N$ signals $c_i(s)$, which are set to $c_i(s) := c(s) + \delta_i$, and these result in qubit-independent schedules $A_i(s) := A(s, \delta_i)$ and $B_i(s) := B(s, \delta_i)$. Illustrations showing the relationships between the normalized annealing time $s$, the signals $c_i(s)$, and the schedules $A_i(s, \delta_i)$ and $B_i(s, \delta_i)$, are given in figure A6. We can define a vector specifying the anneal offsets for each qubit, $\delta = (\delta_1, \ldots, \delta_N)$, and the homogeneous-driving quantum annealing Hamiltonian in equation (1) is modified to become the following inhomogeneous-driving Hamiltonian:

$$H_\delta(s) = \sum_{i=1}^{N} A_i(s) \sigma_x^{(i)} + \sum_{i=1}^{N} B_i(s) h_i \sigma_z^{(i)} + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \sqrt{B_i(s)B_j(s)} I_{ij} \sigma_x^{(i)} \otimes \sigma_x^{(j)}. \quad (2)$$

The driving terms $\sigma_x^{(i)}$ now have per-qubit schedules $A_i(s)$. Figure 1 shows the annealing schedules for a problem with two clusters of qubits, where in one case homogeneous driving is used (figure 1(A)), and in the other, inhomogeneous driving—where the schedules for the qubits in the second cluster are delayed—is used (figure 1(B)).

How should the anneal offsets $\delta$ be chosen to increase (ideally maximally) the probability of successfully finding a ground state? In appendix C we show that the problem of finding optimal offsets can be viewed as the problem of optimizing a non-linear, high-dimensional functional that itself depends on the solution of the encoded minimization problem. Thus a formal approach seems impractical. Instead we may resort to heuristic approaches. Two prior experimental studies have outlined two different approaches for different classes of problem instances. Andriyash et al. [26] report a strategy for problems that are embedded in a physical QA graph such that each logical variable in the original problem is represented by a chain of physical qubits, which occurs on the D-Wave QA when the problem to be solved is not a subgraph of the QA’s Chimera physical hardware graph. Their strategy is to apply to each qubit a delay that increases monotonically with the length of the chain that the qubit forms part of. They tested their strategy on integer-factoring problems embedded in the Chimera graph, and reported improvements of up to $\sim 10^3 \times$. Lanting et al. [27] present an iterative approach to choosing the anneal offsets based on discovering the flappiness of each qubit (which they relate to the classical notion of a floppy spin—a spin that does not change the system energy if it is flipped), and setting the offsets based on flappiness. They tested their strategy on a class of crafted instances of size up to $N = 24$ qubits; larger problem instances were not explored.
One prominent hypothesis for why quantum annealers may fail to find a ground state is the phenomenon of freeze-out, in which certain qubits are thought to become frozen long before the end of the computation [29, 31, 32]. A hypothesized intuition for how anneal offsets can improve success probabilities is by delaying the reduction of the driving terms for qubits that are prone to early freeze-out. The aforementioned reference [26] applied this intuition in the context of problems with qubit chains, where qubits in long chains are claimed to freeze earlier than those in short chains. In this paper we experimentally explore an AO strategy that relies on three hypotheses. These hypotheses are: 1.) early freeze-out of some qubits is a cause of QA failure; 2.) early freeze-out of a qubit can be mitigated by delaying annealing schedules for that qubit, and 3.) qubits that are more strongly coupled to the rest of the system freeze out earlier than those that are only weakly coupled [33]. We have designed an AO protocol that is designed to delay the annealing schedule for each qubit in proportion to how strongly coupled each qubit is to the rest of the system. This protocol is motivated by the fact that if the three hypotheses are all correct, then our AO strategy should increase the QA’s success probability...

We now give a precise definition of the strategy that we experimentally tested. One key component of the strategy is that we quantify how strongly coupled a spin is to the rest of the system via an effective field. Let $j_1, \ldots, j_N$ denote an enumeration of the spins that are coupled to spin $i$, and let $s_{j_1}, \ldots, s_{j_N} \in \{+1, -1\}$ be some configuration of these spins. We denote the effective field on spin $i$, as a...
function of the value of the spins neighboring spin $i$, by $\mathcal{F}_i(s_{j_1}, \ldots, s_{j_{N_i}})$, where this is defined as

$$\mathcal{F}_i(s_{j_1}, \ldots, s_{j_{N_i}}) := h_i + \sum_{j = j_1, \ldots, j_{N_i}} J_{ij} s_j.$$  

(3)

We will primarily be interested specifically in the absolute value of this quantity, $|\mathcal{F}(s_{j_1}, \ldots, s_{j_{N_i}})|$, and in particular, the average of this over all possible neighboring spin values:

$$\overline{|\mathcal{F}_i|} := \frac{1}{2^{N_i}} \sum_{s_{j_1} \in \{+1, -1\}} \cdots \sum_{s_{j_{N_i}} \in \{+1, -1\}} |\mathcal{F}_i(s_{j_1}, \ldots, s_{j_{N_i}})|.$$  

(4)

Next, we normalize these averages to obtain values in the interval $[0, 1]$:

$$r_i := \frac{\overline{|\mathcal{F}_i|}}{\max_{k \in \{1, \ldots, N\}} |\mathcal{F}_k|}.$$  

(5)

A function that delays qubits based on the magnitude of their average effective fields is

$$\delta_i(r_i) := |\delta|_{\text{max}} (1 - 2r_i),$$

where $|\delta|_{\text{max}} > 0$ is the maximum magnitude of offset that is applied on any qubit according to this method, and is a parameter that we are free to choose. One detail we need to consider is that different offset ranges are available for different qubits, and it can be the case that this function assigns an offset to a qubit that the hardware cannot physically realize. If $|\delta|_{\text{max}}$ is the maximum offset value that can be applied on qubit $i$ allowed by the hardware, and $|\delta|_{\text{min}}$ is the minimum such offset value, we can ensure that $\delta_i(r_i) \in [|\delta|_{\text{min}}, |\delta|_{\text{max}}]$ by using the following prescription, which is the formal definition of the strategy that we explore in this paper:

$$\delta_i(r_i) := \begin{cases} 
\max\{|\delta|_{\text{max}} (1 - 2r_i), |\delta|_{\text{min}}\} & \text{if } r_i \geq \frac{1}{2}, \\
\min\{|\delta|_{\text{max}} (1 - 2r_i), |\delta|_{\text{max}}\} & \text{if } r_i < \frac{1}{2}.
\end{cases}$$  

(6)

We note that the averages in equation (4) can be computed without difficulty for any graph that has low maximum degree, such as that of the DW2000Q (which has a maximum degree of 6, so only at most $2^6$ spin configurations have to be enumerated per qubit to calculate $|\mathcal{F}_i|$). The computation defined in equation (4) is intractable when the underlying QA hardware graph has vertices with high degree. However, in appendix A, we present a modified version of the heuristic that is tractable even for fully connected graphs. Since in this paper we only work with the DW2000Q, we can directly perform the computations specified by equation (4).

We experimentally tested the strategy in equation (6) on four different classes of problems by comparing the performance of DW2000Q using baseline settings (no anneal offsets) against the performance of the DW2000Q using anneal offsets with various choices of $|\delta|_{\text{max}}$. Success probabilities were estimated by taking the number of observed successes, and dividing by the number of annealing runs performed to observe this number of successes. From these observed success probabilities, $p$, the corresponding time-to-solution (TTS) with desired probability $p_{\text{d}}$ (which we chose to be 99%), was computed using the formula

$$\text{TTS} := t_{\text{ann}} \log\left(\frac{1 - p_{\text{d}}}{p}\right) / \log(1 - p),$$

where $t_{\text{ann}}$ is the annealing time used in each run. We used the default value $t_{\text{ann}} = 20\mu s$ for all problems classes except for the alternating-sectors-chain problems, for which we used $t_{\text{ann}} = 5\mu s$ to allow for direct comparison with results from a recent baseline experimental study [30]. Furthermore, all problem classes were defined on the Chimera graph (native to the DW2000Q), except for the alternating sectors chain, which is a 1D chain that embeds directly into the native Chimera graph.

2. Results

2.1. Uniform-range-$k$-disorder (URkD) problems

The uniform-range-$k$-disorder (URkD) class of problems is defined (figure 2(A)) on the Chimera graph as those for which $h_i = 0$ for all $i \in \{1, \ldots, N\}$, and each $J_{ij}$ is chosen at random, with uniform probability, from the $2k$ discrete values in the set $U_k := \{-k, -k+1, \ldots, -1, 1, \ldots, k-1, k\}$. These problems have previously been studied in the context of quantum annealing in references [5, 6, 34]. The generic nature of this problem class makes it a good candidate for getting a sense for how useful the heuristic strategy in equation (6) might be on optimization problems that arise in a variety of application areas. We generated random instances of problems in the URkD class, for various graph sizes $N$ and coupling ranges $k$, and measured the success probability of finding a ground state for each instance, both with baseline DW2000Q
Figure 2. Results for the uniform-range-$k$ disorder (URD) problem class. (A) A zoom-in on one of the Chimera cells in a URD problem instance. The coupler values $J_{ij}$ are chosen at random, with uniform probability, from the set $U_k = \{-k, -k+1, \ldots, 1, \ldots, k-1, k\}$. Note that $0 \notin U_k$. The local fields $h_v$ are all set to 0 (i.e., $h_v = 0$ for all $v \in V$). (B) Instance-by-instance comparison of the observed success probability when using anneal offsets, $p_{AO}$, versus the observed success probability with the baseline schedule, $p_{BL}$. (C) The corresponding times-to-solution, calculated directly from the success probabilities in (B); TTS$^{AO}$ denotes the time-to-solution when using anneal offsets, and TTS$^{BL}$ denotes the time-to-solution with the baseline schedule. The instance for which the maximum improvement was observed is emphasized by a gray, dashed circle. Instances in the white zone correspond to those for which the schedule with anneal offsets resulted in better performance, whereas those in the gray zone correspond to those for which the baseline schedule resulted in better performance. The color indicates the relative difficulty of the instance as measured by the performance with the baseline schedule. (D) Percentage of instances for which an improvement in the success probability when using anneal offsets was observed, versus the maximum magnitude of the applied offsets, $|\delta|_{max}$. (E) Median speed-up observed when using anneal offsets versus $|\delta|_{max}$. (F) Percentage of instances for which an improvement in the success probability when using anneal offsets was observed, versus problem size $N$. (G) Median TTS speed-up observed when using anneal offsets versus $N$. (H) The maximum TTS speed-up observed when using anneal offsets versus $N$, for all the different values of $|\delta|_{max}$ used. (I) Median TTS speed-up observed when using anneal offsets versus the spin-spin coupling range, $k$. Note that $k$ is held fixed (at $k = 8$) in the rest of the figure.

operation, and when using anneal offsets as prescribed by the heuristic. Our experimental results, summarized in figure 2, show that anneal offsets chosen using the heuristic typically improve performance as measured by several different metrics.

Figure 2(B) shows a scatter plot of the observed success probability when using anneal offsets, $p_{AO}$, versus the observed success probability with the baseline schedule, $p_{BL}$, for 200 randomly generated instances of problem size $N = 400$ and maximum magnitude of spin–spin coupling $k = 8$, using a maximum magnitude of applied offset $|\delta|_{max} = 0.03$. Figure 2(C) shows the corresponding times-to-solution, calculated directly from the success probabilities in figure 2(B); TTS$^{AO}$ denotes the time-to-solution when using anneal offsets, and TTS$^{BL}$ denotes the time-to-solution with the baseline schedule. In figures 2(B) and (C), we have colored the instances by the difficulty with which the DW2000Q solves them using baseline settings. In particular, we have divided the instances into four difficulty groups, which we formally defined using ranges of the percentile-rank calculated for each instance based on its baseline success probability $p_{BL}$. The grouping of instances by difficulty occurs in subsequent panels too, and in subsequent figures, and has been performed the same way throughout. We note that the more difficult instances are both more likely to benefit from the use of anneal offsets, as well as more likely to benefit to a larger degree, relative to the easier instances in this problem class. One can see this more clearly in figures 2(D) and (E), which show the percentage of instances for which using anneal offsets resulted in...
improved performance compared to baseline, and the median time-to-solution ratio (i.e., the median of $TT^{BL}/TT^{AO}$), respectively, both as functions of $|\delta|_{\text{max}}$.

It is natural to ask what value of $|\delta|_{\text{max}}$ results in the ‘best’ performance. An answer to this question is not straightforward, and figures 2(D) and (E) give some insight into the trade-offs encountered through various choices of $|\delta|_{\text{max}}$. To start, determining the best choice of $|\delta|_{\text{max}}$ depends on the performance metric used. Another fact that further complicates this question is that different instances will be affected differently for the same value of $|\delta|_{\text{max}}$. For this problem class, the primary trade-offs to be balanced are that smaller values of $|\delta|_{\text{max}}$ will generally result in increased performance over a larger percentage of the instances, but smaller median time-to-solution speed-ups for the more difficult instances, relative to larger values of $|\delta|_{\text{max}}$. For a clearer picture of these trade-offs, see the scatter plots in (figure A1).

Another question that arises naturally is how the performance of the heuristic depends on problem size. To that end, figure 2(F) shows the percentage of instances for which the use of anneal offsets resulted in improved performance for different problem sizes, $N$, with $|\delta|_{\text{max}}$ fixed at a value chosen based on figure 2(D) such that performance averaged over all instances is improved ($|\delta|_{\text{max}} = 0.03$). While the change in this metric as a function of $N$ for each difficulty group is different, the overall impression is that as $N$ increases, the benefit obtained from using anneal offsets over baseline is either constant or increasing, depending on the difficulty group. Furthermore, across all problem sizes there is a tendency for performance to be improved on a larger percentage of the more difficult instances, relative to the easier instances. Figure 2(G) shows the median time-to-solution ratio for various problem sizes, with $|\delta|_{\text{max}}$ fixed at a value chosen based on figure 2(E) such that performance averaged over all instances is improved ($|\delta|_{\text{max}} = 0.05$). While the behavior is again dependent on which difficulty group is being considered, in general there appears to be a tendency for the median time-to-solution ratio to increase slightly with problem size for the hardest 50% of instances, whereas for the easiest 50% it appears to be constant. Figure 2(H) shows the maximum observed speed-up (i.e., the maximum of $TT^{BL}/TT^{AO}$) for various problem sizes, for all values of $|\delta|_{\text{max}}$ tested. In general, there appears to be a tendency for the maximum speed-up to increase with increasing $N$, up to $N = 400$, at which point the maximum speed-up plateaus. The largest speed-up observed across all instances was just over $2000\times (N = 300, |\delta|_{\text{max}} = 0.1)$. We note that higher values of $|\delta|_{\text{max}}$ tend to be more likely to benefit in the maximum speed-up, compared to the smaller values of $|\delta|_{\text{max}}$.

The results discussed thus far have been for instances with $k = 8$, where $k$ is the maximum magnitude of the spin–spin coupling values. We also performed experiments for a range of different $k$ values, and found the results to be broadly similar (figure 2(I)), suggesting that the use of anneal offsets provides benefit generally for UR$k$D instances, at least when the $U_k$ values are well within the precision limits of the QA ($2 \leq k \leq 14$ tested here).

2.2. 4-Modal-range-$k$-disorder (4MR$k$D) problems

The 4-modal-range-$k$-disorder (4MR$k$D) class of problems is defined (figure 3(A)) on the Chimera graph as those for which $h_i = 0$ for all $i \in \{1, \ldots, N\}$, and $J_{ij}$ is chosen at random, with uniform probability, from the four discrete values in the set $M_k := \{-k, -1, 1, k\}$. As with the UR$k$D class, we generated random instances of the 4MR$k$D class for various problem sizes $N$ and coupling ranges $k$, and measured the success probability of finding a ground state for each instance, with and without the use of anneal offsets. Our experimental results, summarized in figure 3, show that, again, anneal offsets chosen using the heuristic in equation (6) typically result in improved performance.

Figure 3(B) shows a scatter plot of the observed success probability when using anneal offsets, $p_{\text{AO}}$, versus the observed success probability with the baseline schedule, $p_{\text{BL}}$, for 200 randomly generated instances of problem size $N = 392$ and maximum magnitude of spin–spin coupling $k = 8$, using a maximum magnitude of applied offset $|\delta|_{\text{max}} = 0.05$. Figure 3(C) shows the corresponding times-to-solution, calculated directly from the success probabilities in figure 3(B); $TT^{AO}$ denotes the time-to-solution when using anneal offsets, and $TT^{BL}$ denotes the time-to-solution with the baseline schedule. As with the UR$k$D problem class, we note that it is again the case that the more difficult instances (for the baseline solver, as defined by $p_{\text{BL}}$) are both more likely to benefit from the use of anneal offsets, as well as more likely to benefit to a larger degree (relative to the easier instances in this problem class). One can see this more clearly in figures 3(D) and (E), which show the percentage of instances for which using anneal offsets resulted in improved performance compared to baseline and the median time-to-solution ratio (i.e., the median of $TT^{BL}/TT^{AO}$), respectively, both as a function of $|\delta|_{\text{max}}$.

It is again difficult to pick a single value of $|\delta|_{\text{max}}$ that results in the ‘best’ performance for the 4MR$k$D problem class, since determining the best choice of $|\delta|_{\text{max}}$ depends on the performance metric used to measure its optimality, and because different instances will be affected differently for the same value of $|\delta|_{\text{max}}$ (figures 3(D) and (E)). The primary trade-offs to be balanced are again that smaller values of $|\delta|_{\text{max}}$
Figure 3. Results for the 4-modal range-$k$ disorder (4MR$k$D) problem class. (A) A zoom-in on one of the Chimera cells in a 4MR$k$D disorder problem instance. The coupler values $J_{ij}$ are chosen at random, with uniform probability, from the set $M_k := \{-k, -1, 1, k\}$. Note that $0 \not\in M_k$. The fields $h_v$ are all set to 0 (i.e., $h_v = 0$ for all $v \in V$). (B) Instance-by-instance comparison of the observed success probability when using anneal offsets, $p_{AO}$, versus the observed success probability with the baseline schedule, $p_{BL}$. (C) The corresponding times-to-solution, calculated directly from the success probabilities in (B); $TTSAO$ denotes the time-to-solution when using anneal offsets, and $TTSBL$ denotes the time-to-solution with the baseline schedule. The instance for which the maximum improvement was observed is emphasized by a gray, dashed circle. Instances in the white zone correspond to those for which the schedule with anneal offsets resulted in better performance, whereas those in the gray zone correspond to those for which the baseline schedule resulted in better performance. The color indicates the relative difficulty of the instance as measured by the performance with the baseline schedule. (D) Percentage of instances for which an improvement in the success probability when using anneal offsets was observed, versus the maximum magnitude of the applied offsets, $|δ_{\text{max}}|$. (E) Median TTS speed-up observed when using anneal offsets versus $|δ_{\text{max}}|$. (F) Percentage of instances for which an improvement in the success probability when using anneal offsets was observed, versus $N$. (G) Median TTS speed-up observed when using anneal offsets versus $N$. (H) Median TTS speed-up observed when using anneal offsets for different probability distributions with which the $J_{ij}$ are drawn from the set $M_k$. (I) The maximum TTS speed-up observed when using anneal offsets versus $N$, for all the different values of $|δ_{\text{max}}|$ used. (J) Median TTS speed-up observed when using anneal offsets versus the spin–spin coupling range, $k$. Note that $k$ is held fixed (at $k = 8$) in the rest of the figure.

will generally result in increased performance over a larger percentage of the instances compared to larger values of $|δ_{\text{max}}|$; however, smaller values of $|δ_{\text{max}}|$ will generally result in smaller median time-to-solution speed-ups compared to larger values of $|δ_{\text{max}}|$ (figure A2).

We now discuss how the performance of the heuristic depends on problem size. Figure 3(F) shows the percentage of instances for which the heuristic improved performance for various problem sizes, $N$, with $|δ_{\text{max}}|$ fixed at a value chosen based on figure 3(D) such that performance averaged over all instances is improved ($|δ_{\text{max}}| = 0.03$). While the metric as a function of $N$ for each difficulty group is different, the overall impression is that as $N$ increases, the benefit obtained from using anneal offsets over baseline is
either constant or increasing. Furthermore, across all problem sizes there is a tendency for performance to be improved on a larger percentage of the more difficult instances, relative to the easier instances. Figure 3(G) shows the median time-to-solution ratio for various problem sizes, with $|\delta|_{\text{max}}$ fixed at a value chosen based on figure 3(E) such that performance averaged over all instances is improved ($|\delta|_{\text{max}} = 0.05$). While the behavior is again dependent on the difficulty group, in general there appears to be a tendency for the median time-to-solution to increase slightly with problem size, with the trend slightly more pronounced than for the URkD problem class. Figure 3(H) shows the maximum observed speed-up (i.e., the maximum of TTSBLO/TTTSBLO) for various problem sizes, for all values of $|\delta|_{\text{max}}$ tested. In general, the maximum speed-up appears to be roughly constant with problem size. The largest speed-up observed across all instances was just over $2000 \times (N = 200, |\delta|_{\text{max}} = 0.1)$.

Furthermore, we note that there is a tendency for the higher values of $|\delta|_{\text{max}}$ to be more likely to result in the maximum speed-up, compared to the smaller values of $|\delta|_{\text{max}}$.

The results discussed thus far have concerned the case when $k = 8$, where $k$ is the maximum magnitude of the spin–spin coupling values; additionally, we have focused on the case where the edge weights $J_{0j}$ are drawn with uniform probability from the set $M_6$. We now discuss the performance of the heuristic when these features of the problem class are modified. Still drawing the $J_{0j}$ uniformly from $M_6$, the results appear to be generally consistent across a broad range of different $k$ values (figure 3(I)). Figure 3(J) shows how the median time-to-solution ratio changes when we change the probability distribution with which the instances was just over $2000 \times (N = 200, |\delta|_{\text{max}} = 0.1)$.

2.3. Alternating-sectors-chain (ASC) problems

The alternating sectors chain (ASC) class of problems has been studied in the context of quantum annealing and adiabatic quantum computation in references [30, 35]. An alternating sectors chain is a 1D chain of $N$ spins divided into equally sized sectors of length $n$ sectors alternate between having so-called ‘heavy’ ferromagnetic spin–spin couplings, $W_1$, and ‘light’ ferromagnetic spin–spin couplings, $W_2$, where $|W_1| > |W_2| > 0$. Formally, if the spins are indexed by $\{1, \ldots, N\}$, then ASC problems are given by $h_i = 0$ for all $i \in \{1, \ldots, N\}$, and for all $i \in \{1, \ldots, N-1\} J_{i,i+1} = W_1$ if $i/n$ is odd, and $J_{i,i+1} = W_2$ otherwise. Furthermore, there is a technical restriction that there be $b + 1$ heavy sectors and $b$ light sectors, which results in a limitation on the possible combinations of $N$ and $n$. Figure 4(A) shows an example of an ASC instance with $N = 10, n = 3$. Because all the couplings are ferromagnetic, the problem is trivial to solve: the two degenerate ground states are the fully aligned states, with all spins pointing either up or down.

Nevertheless, it is known that this problem exhibits an exponentially small gap in the sector size $n$ [35], which implies an exponential computation time in the AQC framework. Recently, however, reference [30] showed experimentally that the performance of a quantum annealer solving ASC problems differed substantially from what one would expect based purely on the scaling of the minimum gap.

We briefly paraphrase the intuitive argument of reference [30] for why quantum annealing fails to efficiently solve this problem. For $N \gg 1$ and $n \gg 1$, any given sector approximates a 1D ferromagnetic chain with all-equivalent couplings. Such a chain encounters a quantum phase transition separating the ordered phase from the disordered phase when $A(s) = B(s)J_{01}$ [36]. Therefore, the heavy sectors order independently before the light sectors during the anneal. Since the transverse field generates only local spin flips, quantum annealing is likely to get stuck in a local minimum (with domain walls at the boundaries between heavy and light sectors) unless the annealing time is scaled exponentially with $n$.

The intuitive argument above suggests the following remedy via the use of anneal offsets. Let $s_1$ be defined as the value such that $A(s_1) = B(s_2)W_1$, and let $s_2$ be the value such that $A(s_2) = B(s_2)W_2$. In other words, $s_1$ is the normalized time when the heavy sectors order, and $s_2$ is the normalized time when the light sectors order, under homogeneous driving. Let $\delta_1$ be an offset such that $A(s_2, \delta_1) = B(s_2, \delta_1)W_1$. Applying this offset to the heavy sectors, one can make it so that both the light and heavy sectors order at the same time, $s_2$. If it is the case that the sectors independently ordering at different times makes the problem more difficult to solve via quantum annealing, then we should see an increase in the success probability by applying this offset. It turns out that anneal offsets calculated according to the prescription above are very similar to the anneal offsets that the heuristic in equation (6). To further demonstrate the generality of the heuristic in equation (6), in this section we present results from the application of it to ASC problems, instead of the above problem-specific prescription. Our experimental results, summarized in figures 4(B)–(H), show that the application of anneal offsets chosen using the heuristic in equation (6) improves performance on ASC instances for nearly all of the problem class parameter space that was tested.

Figure 4(B) shows the success probability $p_{\text{success}}$ as a function of $|\delta|_{\text{max}}$, for various problem sizes $N$ ($W_1 = -1.0, W_2 = -0.5, n = 4$). A performance improvement is observed for a broad range of $|\delta|_{\text{max}}$. 

8
values, with a peak at $|\delta|_{\text{max}} \approx 0.056$, and the success probabilities dropping roughly symmetrically for smaller and greater values of $|\delta|_{\text{max}}$. Note that the value of $|\delta|_{\text{max}}$ that maximizes the success probability does not change appreciably with $N$. This is because the times at which the light and heavy sectors order depend primarily on the parameters $W_1, W_2, n$, and approximately the same value of $|\delta|_{\text{max}}$ should synchronize the dynamics of the two kinds of sectors if these three parameters are kept fixed. In figure 4(C), we can see how the time-to-solution speed-up, $TTS_{\text{BL}}/TTS_{\text{AO}}$, scales with problem size both for a fixed value of $|\delta|_{\text{max}}$ chosen based on figure 4(B) such that performance is improved for every $N$ ($|\delta|_{\text{max}} = 0.056$), and how it scales when at each $N$ we use the value of $|\delta|_{\text{max}}$ that maximizes the success probability for that $N$. In both cases, the speed-up appears to scale exponentially with $N$. We note that $|\delta|_{\text{max}} = 0.056$ either maximizes or very nearly maximizes the success probability for every $N$ tested.

Figure 4(D) shows $p_{\text{success}}$ as a function of $|\delta|_{\text{max}}$ for various values of the light coupling, $W_2$ ($W_1 = -1.0, N = 150, n = 4$). A performance improvement is observed for every value of $W_2$ for some $|\delta|_{\text{max}}$. While there is no clear relationship between the optimal choice of $|\delta|_{\text{max}}$ and $W_2$, in figure 4(E) we can see that one can find a fixed value of $|\delta|_{\text{max}}$ such that performance is either maximized or nearly maximized ($|\delta|_{\text{max}} = 0.056$) for all values of $W_2$. Naively, one might expect $p_{\text{BL}}$ to increase monotonically with $W_2/W_1$, for $0 < W_2/W_1 \leq 1$. Intuitively, the smaller $W_2/W_1$, the more inhomogeneous the dynamics of the light and heavy sectors, suggesting the problem might be more difficult to solve under homogeneous driving. Indeed, this is what we see in figure 4(D). Similarly, one might intuitively expect the maximum speed-up obtained with the anneal offsets heuristic to decrease monotonically with $W_2/W_1$, for $0 < W_2/W_1 \leq 1$: the smaller $W_2/W_1$, the more inhomogeneous the dynamics of the light and heavy sectors, and therefore, potentially, the more room for there is for the use of anneal offsets to provide an improvement. While such monotonic behavior is indeed observed for $2/8 \leq W_2/W_1 \leq 7/8$, there is a single, stark exception to this intuition when $W_2/W_1 = 1/8$. It is unclear what accounts for this exception.

Figure 4(F) shows the success probability as a function of $|\delta|_{\text{max}}$ for various values of the sector size, $n$ ($W_1 = -1.0, W_2 = -0.5, N = 200$) (for clarity of presentation, figure 4(F) shows the data for only 4 of the
18 different sector sizes tested). A performance improvement is seen for some values of \( n \), with the degree of improvement being very strongly correlated with the baseline success probability for the instance (figure 4(G) and (H)). As was previously reported in reference [30], instead of the \( p_{\text{BL}} \) dropping monotonically with \( n \), instead we see in figure 4(G) that \( p_{\text{BL}} \) achieves a minimum for some intermediate \( n^* \) \( (n^* = 5) \), and then rises again for sector sizes larger than \( n^* \). In figure 4(H) we can see that there is a trend for the more difficult instances to benefit to a larger degree from the use of anneal offsets.

### 2.4. Weak-strong-cluster (WSC) problems

The weak-strong cluster (WSC) class of problems was studied in the context of quantum annealing in references [7, 9, 29]. The weak-strong cluster (WSC) problem class was designed so that multi-qubit tunneling strongly impacts the success probability. The building block of this problem is a pair of strongly connected spins, also referred to as a pair of clusters. One cluster is referred to as a strong cluster, and the other as a weak cluster; each cluster corresponds to a cell in the Chimera graph. Within a weak-strong cluster pair, all the couplings are set ferromagnetically \( (J_{hi} = -1) \); all the local fields \( h_i \) in the strong cluster are set to \( h_{\text{strong}} = -1 \), and all the local fields in the weak cluster are set to \( h_{\text{weak}} = -\lambda h_{\text{strong}} \), for some \( 0 < \lambda < 1 \) (depicted graphically in figure 5(A)).

For \( \lambda < 0.5 \), the global minimum of a weak-strong cluster pair corresponds to the configuration in which all spins point in the direction of the strong local field. As explained in reference [7], early in the anneal, however, the local-field terms dominate, so each spin orients itself along the direction of its own local field. Later in the anneal, the coupling terms dominate, and the spins in the weak cluster must tunnel through an energy barrier to escape the local minimum into which they are lead during this initial phase of the anneal. This problem class is interesting because it was designed to benefit from a computational strength of quantum annealing (multi-spin cotunneling [7, 29]), while simultaneously being difficult for classical simulated annealing to solve. This has made it a well-studied problem class for which a quantum speed-up might be obtained, although a speed-up against the best classical methods has not yet been achieved [9]. Our experimental results, summarized in figure 5, show that anneal offsets chosen using the heuristic typically improve performance of the DW2000Q on the WSC class as measured by several different metrics.

Figure 5(B) shows a scatter plot of the observed success probability when using anneal offsets, \( p_{\text{AO}} \), versus the observed success probability with the baseline schedule, \( p_{\text{BL}} \), for 80 randomly generated instances with problem size \( N = 966 \) spins and weak local field \( h_{\text{weak}} = -0.44 \) \( (i.e., \lambda = 0.44) \); this value of \( \lambda \) is chosen because references [7, 9, 29] focused on instances with this choice of \( \lambda \), using \( |\delta|_{\text{max}} = 0.08 \). Figure 5(C) shows the corresponding times-to-solution, calculated directly from the success probabilities in figure 5(B). Before we continue to discuss the results in more detail, we note that 10/80 instances were not solved with the baseline schedule after \( 10^7 \) runs. Furthermore, not all of these instances were solved with the anneal offsets heuristic; depending on the value of \( |\delta|_{\text{max}} \) used, different fractions of them were solved. For all values of \( |\delta|_{\text{max}} \), the instances not solved using the heuristic were a strict subset of the instances not solved with the baseline schedule. Because we were unable to obtain concrete estimates of some of the success probabilities (both \( p_{\text{AO}} \) and \( p_{\text{BL}} \)), we are forced to present a somewhat tailored interpretation of the results for the corresponding instances.

With this in mind, we now continue on to discuss the results. A performance improvement is observed for nearly all instances in figure 5(B) and (C) \( (\text{with performance worsening on only 5/80 instances, all of which lie in the easiest 50% of instances}) \). Figure 5(D) shows the percentage of instances for which using anneal offsets resulted in improved performance compared to baseline, versus \( |\delta|_{\text{max}} \). We can see that, when \( |\delta|_{\text{max}} \) is chosen appropriately \( (e.g., |\delta|_{\text{max}} = 0.08) \), one can improve performance on a larger percentage of the more difficult instances, relative to the easier instances. Similarly, we can see in figure 5(E), which shows the median times-to-solution speed-ups \( (i.e., \text{the median of } TTS_{\text{BL}}/TTS_{\text{AO}} \text{ versus } |\delta|_{\text{max}} \text{, that one can improve performance on the more difficult instances to a greater degree, relative to the easier instances, with appropriate choice of } |\delta|_{\text{max}} \text{. Note that the values reported for the hardest 10/80 instances represent lower bounds for the median time-to-solution speed-up for the subset of the instances solved with the heuristic; the subset is in general different for different values of } |\delta|_{\text{max}} \text{.}

We now discuss how the use of the anneal offsets heuristic affects performance as a function of problem size, \( N \). Figure 5(F) shows the percentage of instances for which the use of the heuristic improved performance for various problem sizes, with \( |\delta|_{\text{max}} \) fixed at a value chosen based on figure 5(D) such that performance averaged over all instances is improved \( (|\delta|_{\text{max}} = 0.08) \). While the behavior is different for each difficulty group, over all instances the overall impression is that the benefit obtained from using anneal offsets increases up to some intermediate size \( (N = 507) \), beyond which the results remain roughly constant \( (\text{with performance increasing on } \approx 90\% \text{ of instances}) \). Figure 5(G) shows the median time-to-solution ratio for various problem sizes, with \( |\delta|_{\text{max}} \) fixed at a value chosen based on figure 5(E) such that performance
Figure 5. Results for the weak-strong cluster (WSC) problem class. (A) (Left) weak-strong cluster pair. (Right) WSC problem example instance. Gray blocks represent weak clusters, and black blocks represent strong clusters. Blue lines represent ferromagnetic couplings, and red lines represent antiferromagnetic couplings. (B) Instance-by-instance comparison of the observed success probability when using anneal offsets versus the observed success probability with the baseline schedule. (C) The corresponding times-to-solution, calculated directly from the success probabilities in (B). The instance for which the maximum improvement was observed is emphasized by a gray, dashed circle. Instances in the white zone correspond to those for which the schedule with anneal offsets resulted in better performance, whereas those in the light gray zone correspond to those for which the baseline schedule resulted in better performance. The color indicates the relative difficulty of the instance as measured by the performance with the baseline schedule. Note that the instances in the darker shade of red were not solved after 10^7 runs neither using anneal offsets nor the baseline schedule. Instances in the lighter shade of red were solved using the anneal offsets heuristic, but not with the baseline schedule. (D) Percentage of instances for which an improvement in the success probability when using anneal offsets was observed, versus the maximum magnitude of the applied offsets, |δ|_{max}. (E) Percentile of instances for which an improvement in the success probability when using anneal offsets was observed, versus the number of problems N. (F) Maximum time-to-solution speed-up observed when using anneal offsets versus N. (H) Maximum time-to-solution speed-up observed when using anneal offsets versus for all the different values of |δ|_{max} used. (I) The median time-to-solution speed-up observed when using anneal offsets versus the weak local fields, h_{weak}. Note that h_{weak} is held fixed (at h_{weak} = −0.44) in the rest of the figure.

averaged over all instances is improved (|δ|_{max} = 0.08). While the behavior again depends on the difficulty group being considered, over all instances there is a trend for the benefit obtained from using anneal offsets to increase with problem size. Note that for the largest problem size tested (N = 966), the median speed-up reported for the hardest 12.5% of instances (∼160×) is in fact a lower bound for the actual median time-to-solution speed-up for a subset of the instances, namely, the subset of instances solved with the heuristic. Figure 5(H) shows the maximum observed speed-up (i.e., the maximum of TTS_{AO}/TTS_{BL}) for
quantum annealer (a 1000-qubit QA) on WSC problem instances reported [9] that the QA was very close to achieving a quantum advantage against a battery of state-of-the-art classical solvers across all tested problem sizes. A re-evaluation using the DW2000Q with anneal offsets is an interesting prospect for future study, especially since the WSC problems benchmarked in reference [9] were of the same subclass we show results for in figure 5(B)–(H) (namely $h_{\text{weak}} = -0.44$; $h_{\text{strong}} = 1$; $J_l = -1$), and in this work we have found that substantial performance improvements can be achieved on this subclass.

3. Discussion

The heuristic we have tested has a single parameter, $|\delta|_{\text{max}}$, which is freely chosen. We found that different instances benefit differently for the same value of $|\delta|_{\text{max}}$. In general, it appears that the more difficult instances of a problem class are more likely to benefit across a wider range of $|\delta|_{\text{max}}$ values, as well as more likely to benefit to a larger degree, relative to the easier instances in that problem class. Of course, in practice, one does not in general know the difficulty of a particular instance a priori. It would be helpful to develop a method capable of predicting the optimal $|\delta|_{\text{max}}$ value to use for a particular instance based on its $h_i$, $J_l$ values; this is an avenue for future work. For the present, our empirical results provide a guideline for choosing $|\delta|_{\text{max}}$: if one has no other information, begin with $|\delta|_{\text{max}} = 0.05$. This choice provides a good balance between improving the performance for difficult instances and not excessively decreasing the performance for easy instances; this was true across all four broad problem classes we explored. The scatter plots in figures A1, A2, and A4 allow one to build intuition for the trade-offs obtained by various choices of $|\delta|_{\text{max}}$. Relatively, in appendix B, we present a hybrid strategy that mitigates the risk that a chosen $|\delta|_{\text{max}}$ results in detrimental performance for a particular instance, and ensures that the strategy’s TTS is at worst two times longer than the baseline TTS.

A quantitative understanding of how anneal offsets applied using the approach in this paper improves performance is currently lacking. The alternating-sectors-chain problem instances are analytically tractable, which has made it possible to explore how anneal offsets change the minimum gap and the number of thermally accessible excited states, even for large problems sizes. In appendix D3 we show that neither change explains the improvement in performance that we experimentally observed for this problem class. A more detailed analysis of the dynamics of the computation process might ultimately be necessary to develop a predictive model for the impact of anneal offsets; such a model could aid efforts to design optimal strategies for applying anneal offsets.

In conclusion, in this paper we demonstrated a heuristic strategy for tuning the anneal offsets on a DW2000Q quantum annealer that results in improved time-to-solution over baseline DW2000Q performance for a broad range of problems. For the most generic problem class we investigated, the uniform-range-$k$-disorder problems, one can improve the performance for up to 74% of instances overall, and for 85% of the hardest 10% of instances, with speed-ups of up to $10^7 \times$ observed. For more structured problem classes, like weak-strong-cluster problems, one can improve the performance for up to 94% of instances overall, and 100% of the hardest 10% of instances, with speed-ups of up to $10^7 \times$ observed. Furthermore, for certain parameter regimes of WSC and ASC problems we found that the speed-ups...
achieved by using anneal offsets increased exponentially as a function of problem size \( N \), suggesting that speed-ups orders of magnitude larger than even \( 10^7 \times \) may be achievable. We anticipate that the strategy we described, and derivatives thereof, will form a useful part of the toolbox in experimental quantum annealing. The speedups we obtain for broad classes of problems naturally suggest this is a technique of potential practical relevance when attempting to optimize the performance of a quantum annealer. Furthermore, its dramatic impact on the success probabilities of alternating-sectors-chain problem instances, for which we have detailed analytical results, may make this heuristic strategy a useful control knob for studies attempting to elucidate the working mechanisms of experimental quantum annealers.

Note added: A key component of the anneal-offsets heuristic presented in this paper is that a qubit’s average effective field is obtained by summing over all possible neighboring spin configurations, each configuration having equal weight. This procedure can naturally be generalized to a weighted average for the effective field. During preparation of this manuscript we became aware of related unpublished work along this line by D-Wave Systems Inc, wherein it is proposed that these weights be chosen according to the frequency with which they appear in the solutions returned by running the problem instance with the baseline schedule [37].

4. Materials and methods

4.1. D-wave run protocol

Each call to the DW200Q is limited to \( 10^4 \) annealing runs. The success probabilities were estimated by executing up to \( 10^4 \) calls, i.e., \( 10^7 \) annealing runs, and stopping the calls after the instance was solved a minimum of \( n_{\text{success}} = 5 \) times. For some of the more difficult problem instances, \( 10^7 \) runs were not enough to solve the problem 5 times, but in nearly all cases it was enough to solve the problem at least once; the instances that were unsolved after \( 10^7 \) runs are explicitly denoted in our results. A different gauge was used every \( 10^3 \) runs.

4.2. Error analysis

4.2.1. URkD

Error bars in figures 2(D) and (F) correspond to the standard deviation when dividing the data into 4 bins. Error bars in figures 2(E), (G), and (I) indicate the 35th and 65th percentiles. Finally, error bars in figure 2(H) correspond the standard deviation across the 4 largest speed-ups observed.

4.2.2. 4MRkD

Error bars in figures 3(D) and (F) correspond to the standard deviation when dividing the data into 4 bins. Error bars in figures 3(E), (G), (I), and (J) indicate the 35th and 65th percentiles. Finally, error bars in figure 2(H) correspond the standard deviation across the 4 largest speed-ups observed.

4.2.3. ASC

Error bars in figures 4(B), (D), (F), and (G) correspond to 95\% confidence intervals. For figures 4(C), (E) and (H), each data point was collected 4 times. The values presented correspond to the mean, and the error bars correspond to the standard deviation.

4.2.4. WSC

Error bars in figures 5(D) and (F) correspond to the standard deviation when dividing the data into 4 bins. Error bars in figures 5(E), (G), (I), and (J) indicate the 35th and 65th percentiles. Finally, error bars in figure 5(H) correspond the standard deviation across the 4 largest speed-ups observed.

Acknowledgments

We would like to thank M Amin, E Hoskinson, T Lanting, J Dunn, and A Mishra for useful discussions. We would also like to thank B Fuller and J Brahm for a thorough reading of a draft of this paper. Finally, we gratefully acknowledge USRA and NASA for providing us with D-Wave 2000Q machine time on the system installed at NASA Ames, and in particular Davide Venturelli.

Appendix A. Generalizations of the heuristic

In the heuristic used in the paper, the key quantity in determining the anneal offsets to be applied on a particular qubit is its effective field averaged uniformly over all possible neighboring spin values. However, this can be generalized to arbitrary weighted averages in a straightforward way by introducing nonnegative
weights, which sum to 1, for each qubit and each configuration of its neighboring spins. In other words, for each \( i \in \{1, \ldots, N\} \) and each \( s_{i_1}, \ldots, s_{i_N} \in \{-1, +1\} \), we introduce a \( p'_{i_1 \ldots i_N} \) such that
\[
\sum_{s_{i_1} \ldots s_{i_N} \in \{-1, +1\}} p'_{i_1 \ldots i_N} = 1, \tag{A1}
\]
and look the corresponding weighted average
\[
\mathcal{F}_i^{(p')} := \sum_{s_{i_1} \ldots s_{i_N} \in \{-1, +1\}} p'_{i_1 \ldots i_N} \mathcal{F}_i(s_{i_1}, \ldots, s_{i_N}). \tag{A2}
\]

In [37], it is proposed that these weights be chosen according to the frequency with which the corresponding configuration appears in the solutions returned by running the problem instance with the baseline schedule several times. They test their method on one problem instance of size \( N = 72 \), and show that the strategy indeed improves performance. We leave it as an interesting possible route of research for the future to test this strategy on a broader class of problems, as well as larger problem sizes.

We now discuss how the heuristic tested in this paper can be modified so that it is practically implementable even for fully-connected graphs. First, note that the computation defined in equation (A2) can always be made tractable by setting a sufficient number of the weights to zero. Alternatively, note that the computation defined in equation (A2) is tractable whenever \( \max_{i \in \{1, \ldots, N\}} N_i \) is sufficiently small (i.e., the underlying hardware graph is of sufficiently low maximal degree). This suggests the following possible modification. Instead of using each qubit’s effective field, \( \mathcal{F}_i(s_{i_1}, \ldots, s_{i_N}) \), we use an approximation of its effective field in which only its interactions with a subset of its neighbors are taken into account. For example, let \( s_{i_1}, \ldots, s_{i_M} \in \{-1, +1\} \) denote the configuration of a subset of the spins coupled to spin \( i \) for some \( M < N \).

Appendix B. Mitigating the risk of using anneal offsets

In this section, we present a simple hybrid strategy that mitigates the risk that a chosen \( |\delta|_{\text{max}} \) results in detrimental performance for a particular instance, and ensures that the strategy’s TTS is at worst two times longer than the baseline TTS. The strategy consists of alternating between calls to the DW2000Q with baseline settings, and calls using anneal offsets.

Suppose for some problem instance the success probability with the baseline settings is \( p_{BL} \), and the success probability for a given anneal offsets strategy is \( p_{AO} \). Using the alternating strategy described above, the cumulative time-to-solution, \( \text{TTS}^{\text{Hybrid}} \), is given by
\[
\text{TTS}^{\text{Hybrid}} = 2t_{\text{ann}} \frac{\log(1 - p_{AO})}{\log[(1 - p_{BL})(1 - p_{AO})]]. \tag{B1}
\]
Because \( p_{AO} \geq 0 \), this implies the following upper bound:
\[
\text{TTS}^{\text{Hybrid}} \leq 2t_{\text{ann}} \frac{\log(1 - p_{AO})}{\log(1 - p_{BL})} = 2\text{TTS}^{\text{BL}}. \tag{B2}
\]

Appendix C. The challenge of finding optimal anneal offsets

The task of finding optimal offsets is in general difficult. The following intuitive argument partially explains why this is the case.

For simplicity, let us consider this problem in the framework of closed-system adiabatic quantum computing [38]. Let \( |\psi(s)\rangle \) denote the quantum state of the \( N \)-qubit processor at normalized time \( s \in [0, 1] \). Let \( U_S(s) \) be the unitary evolution operator induced by the Hamiltonian \( H_S \) of equation (2) (i.e., \( U_S(s)|\psi(0)\rangle = |\psi(s)\rangle \)). Let \( X = \left[0, \delta_1^{\text{min}}, \delta_1^{\text{max}}\right] \times \cdots \times \left[0, \delta_N^{\text{min}}, \delta_N^{\text{max}}\right] \subset \mathbb{R}^N \) denote the set of offset values the
hardware can physically implement. In principle, the optimal choice of offsets, which we denote $\delta^*$, is then given by

$$\delta^* = \arg\max_{\delta} p_0(\delta) = \arg\max_{\delta} \sum_{d=1}^{d_0} | \langle \psi^0_d | U_\delta(1) | \psi(0) \rangle |^2,$$

where $|\psi^0_d\rangle, \ldots, |\psi^0_{d_0}\rangle$ are the $d_0$-degenerate ground states of the problem Hamiltonian, and $p_0(\delta)$ denotes the probability of success. This is a high-dimensional, non-linear, constrained optimization problem. Furthermore, equation (C1) suggests the success probability depends on the ground states $|\psi^0_d\rangle, \ldots, |\psi^0_{d_0}\rangle$ of the problem. In other words, in order to determine the choice of offsets that will maximize the probability of solving the problem, one has to know the answers in advance. This defeats, at least in practice, the purpose of finding optimal offsets in the first place. Indeed, equation (C1) is not practically solvable in full generality by any single, currently known computational method. Therefore, if one is interested in treating...
the problem analytically, one likely has to assume enough knowledge of the problem so that equation (C1) can be simplified enough that an analytic treatment becomes possible. On the other hand, assuming such knowledge a priori reduces the scope of applicability of the treatment in practice.

Appendix D. Additional Results

In this section we present additional experimental results for each of the problem classes we tested. While the summary statistics presented in the main text have the advantage of being concise, we find that there is nevertheless a lot of additional, more subtle information that they do not quite capture.

D1. Uniform-range-k-disorder (URkD) problems

Figure A1 shows an instance-by-instance comparison of the observed success probability when using anneal offsets, $p_{AO}$, versus the observed success probability with the baseline schedule, $p_{BL}$, for the U8RD problem class of problem size $N = 400$, for every value of $|\delta|_{max}$ tested. We can see that there is a considerable amount of variance both in the baseline success probabilities, as well as the success probabilities when using anneal offsets. Furthermore, there is a significant amount of variance in the degree to which anneal offsets either improves or worsens the success probability, with the variance becoming increasingly pronounced with larger $|\delta|_{max}$. This high-variance nature of the data makes it difficult to summarize in just a few summary statistics.

D2. 4-Modal-range-k-disorder (4MRkD) problems

Figure A2 shows an instance-by-instance comparison of the observed success probability when using anneal offsets, $p_{AO}$, versus the observed success probability with the baseline schedule, $p_{BL}$, for the 4MR8D problem class of problem size $N = 392$, for every value of $|\delta|_{max}$ tested. We can see a very similar situation here as with the results in figure A1, discussed in the previous paragraph: the high-variance nature of the data makes it difficult to succinctly summarize.

D3. Alternating-sectors-chain (ASC) problems

Figure A3 shows additional results for the alternating sectors chain problem class of problem size $N \approx 175$. In figure A3(A) we can see that the minimum gap at the critical point when using the anneal offsets heuristic, which we will denote $\Delta_{AO}^*$, is larger than the baseline minimum gap at the critical point, which we will denote $\Delta_{BL}^*$, for sector sizes $n \geq 5$ (we use $h = 1$ units throughout). While both $\Delta_{AO}^*$ and $\Delta_{BL}^*$ decrease monotonically with $n$, $\Delta_{AO}^*$ appears to decreases more slowly compared to $\Delta_{BL}^*$. One might naively expect
Figure A4. Instance-by-instance comparison of the observed success probability when using anneal offsets, \( p_{AO} \), versus the observed success probability with the baseline schedule, \( p_{BL} \), for the WSC problem class for every value of \( |\delta|_{\text{max}} \) tested; problem size \( N = 966 \). In general, the variance in the ratio \( p_{BL} / p_{AO} \) increases with \( |\delta|_{\text{max}} \). Interestingly, in contrast to the UR8D (figure A1) and 4MR8D (figure A2) problem classes, the increased variance is primarily a consequence of instances benefiting to a larger degree, as opposed to a mix of some instances improving to a larger degree, and others being negatively impacted to a larger degree. Data points with numbers inscribed indicate overlapping data points.

Figure A5. Optimal \( |\delta|_{\text{max}} \) for each of the WSC problem instances of problem size \( N = 72 \) used in this study, and the corresponding success probabilities and TTS speed-ups. Instances for which only an upper bound on \( p_{BL} \) is reported indicate instances that were not solved with at least \( 10^7 \) runs with the baseline schedule. For these instances, we can only report a lower bound on the time-to-solution speed-up.

from figure A3(A) that the largest speed-up observed when using anneal offsets would occur for \( n = 20 \), where \( \Delta_{AO}^*/\Delta_{BL}^* \) is maximized, but we can see in figure A3(B) that this is not the case. Instead, the speed-up increases with sector size for \( 1 \leq n \leq 4 \), and then generally decreases for \( n > 4 \) (with very minor exceptions at \( n = 16, 19 \)). Indeed, there does not appear to exist any clear correlation between the ratio \( \Delta_{AO}^*/\Delta_{BL}^* \) and the observed speed-ups. Note that the energy scale set by the operating temperature is much larger than both \( \Delta_{AO}^* \) and \( \Delta_{BL}^* \) for every sector size.
In figure A3(C) we can see how the minimum gap when using anneal offsets, which we will denote $\Delta_{AO}$, differs from the baseline minimum gap, which we will denote $\Delta_{BL}$, as a function of normalized annealing time $s \in [0, 1]$, when the sector size $n = 4$, which is the value of $n$ at which the greatest time-to-solution speed-up is observed. Given the substantial difference in the times-to-solution, it is somewhat surprising that the only notable difference between $\Delta_{AO}$ and $\Delta_{BL}$ is the shift in the critical points. A similar shift is observed for all sector sizes (figure A3(D)), independent of the time-to-solution ratio observed for that value of the sector size.

In reference [30], it is argued that a key quantity in predicting the success probability of the quantum annealer on this problem class is the number of single-fermion states that lie below the energy scale set by the temperature at the critical point, which we will denote $k^*$. In general, in reference [30] it is shown that the success probability decreases when $k^*$ increases. It is therefore interesting to check what effect, if any,
anneal offsets has on this quantity; one could conjecture that the speed-ups observed when using anneal offsets are a consequence of anneal offsets decreasing $k'$. In figure A3(E) we indeed see that $k'_{\text{AO}} \leq k'_{\text{BL}}$ for all sector sizes tested; here $k'_{\text{AO}}$ is the number of single-fermion states that lie below the energy scale set by the temperature at the critical point using the anneal offsets heuristic, and $k'_{\text{BL}}$ is the same quantity but using the baseline schedule. Interestingly, however, that there is no clear correlation between observed speed-up and the ratio $k'_{\text{AO}}/k'_{\text{BL}}$. For example, the greatest speed-up is observed for sector size $n = 4$, but $k'_{\text{AO}} = k'_{\text{BL}}$ for this sector size.

It is interesting to note that the values of $k'_{\text{BL}}$ (figure A3(E)) differ considerably from those in [30]. This can be attributed in large part to the fact that the annealing schedules of the quantum annealer used in this study are different from the annealing schedules of the quantum annealer in [30]. In fact, by rescaling the energy (in simulation) of the annealing schedules used in this study to match the energy scales of [30], we get exact agreement (figure A3(F)) for nearly all sector sizes (with only a minimal disagreement for sector sizes $n \in \{2, 3, 4\}$). It is still the case at these different energy scales that there is no clear correlation between the time-to-solution ratio and the ratio $k'_{\text{AO}}/k'_{\text{BL}}$; here $k'_{\text{AO}}$ is the number of single-fermion states that lie below the energy scale set by the temperature at the critical point using the anneal offsets heuristic on a DW2000Q with the energy rescaled to match the energy of the quantum annealer used in [30], and $k'_{\text{BL}}$ denotes the analogous quantity with homogeneous annealing schedules. For example, while the greatest speed-up is observed for sector size $n = 4$, $k'_{\text{AO}} > k'_{\text{BL}}$ for this sector size\(\text{enleadertwodots}\)Similarly, $k'_{\text{AO}}(n) < k'_{\text{BL}}(n)$ for $n > 13$, even though the performance of the quantum annealer with the anneal offsets heuristic is in fact worse than with the baseline schedule for those sector sizes.

### D4. Weak-strong-cluster (WSC) problems

Figure A4 shows an instance-by-instance comparison of the observed success probability when using anneal offsets, $p_{\text{AO}}$, versus the observed success probability with the baseline schedule, $p_{\text{BL}}$, for the WSC problem class of problem size $N = 966$ and weak local fields $h_{\text{weak}} = -0.44$, for every value of $|\delta|_{\text{max}}$ tested. Compared to both the UR8D problem instances and the 4MR8D disorder problem instances of figure A1 and figure A2 respectively, we can see that the WSC problem instances in figure A4 are more robust to anneal offsets: they benefit both to a larger degree and across a wider-range of $|\delta|_{\text{max}}$ values. Note that 10/80 instances were not solved with the baseline schedule after $10^7$ runs. Depending on the value of $|\delta|_{\text{max}}$, a different fraction of these instances were solved using anneal offsets, which allows one to derive at least a lower bound on the improvement when using anneal offsets. For all values of $|\delta|_{\text{max}}$, the instances that were not solved with the anneal offsets heuristic were a strict subset of the instances not solved with the baseline schedule (i.e., we did not observe any instance solved with the baseline schedule that was not also solved using the anneal offsets heuristic, independent of $|\delta|_{\text{max}}$). For these particularly difficult instances, perhaps it would be interesting in future work to perform more runs to obtain more concrete estimates on the success probabilities. It is unclear how many runs this would require, or if it would be feasible.

Figure A5 shows an instance-by-instance comparison of the observed success probability when using anneal offsets, $p_{\text{AO}}$, versus the observed success probability with the baseline schedule, $p_{\text{BL}}$, for all 8 instances of the WSC problem class of problem size $N = 72$, and weak local fields $h_{\text{weak}} = -8/15$. For each instance, $p_{\text{AO}}$ indicates the observed success probability using the anneal offsets heuristic with the value of $|\delta|_{\text{max}}$, which resulted in the highest observed success probability; the optimal value of $|\delta|_{\text{max}}$ for each instance is shown. Despite the relatively small problem size, these instances turn out to be surprisingly difficult when using the baseline schedule, with half of the instances (namely, the instances in figures A5(A), (B), (F), and (G)) unsolved after at least $10^7$ runs with the baseline schedule. In particular, the instance in figure A5(A) was not solved with the baseline schedule after $1.12 \times 10^7$ runs, implying an upper bound on the baseline success probability of $p_{\text{BL}} < 9 \times 10^{-8}$. In contrast, using the anneal offsets heuristic used in this study we observed a success probability of up to $p_{\text{AO}} = 0.902$ (with $|\delta|_{\text{max}} = 0.1$), from which we can deduce the lower bounds $p_{\text{AO}}/p_{\text{BL}} \geq 10^7$, which in turn implies a time-to-solution speed-up of on the order of $10^7 \times$, with $\text{TTS}^{\text{AO}}/\text{TTS}^{\text{BL}} \geq 2 \times 10^7$ (figures A6 and A7).

### ORCID iDs

Juan I Adame https://orcid.org/0000-0001-8751-9873

### References

[1] Kadowaki T and Nishimori H 1998 Quantum annealing in the transverse Ising model Phys. Rev. E 58 5355
[2] Farhi E, Goldstone J, Gutmann S and Sipser M 2000 Quantum computation by adiabatic evolution (arXiv:quant-ph/0001106)
[3] Santoro G E and Tosatti E 2006 Optimization using quantum mechanics: quantum annealing through adiabatic evolution J. Phys. A: Math. Gen. 39 R393

[4] Johnson M W et al 2011 Quantum annealing with manufactured spins Nature 473 194

[5] Rannow T F, Wang Z, Job J, Boixo S, Isakov S V, Wecker D, Martinis J M, Lidar D A and Troyer M 2014 Defining and detecting quantum speedup Science 345 420–4

[6] Boixo S, Rannow T F, Isakov S V, Wang Z, Wecker D, Lidar D A, Martinis J M and Troyer M 2014 Evidence for quantum annealing with more than one hundred qubits Nat. Phys. 10 218

[7] Denchev V S, Boixo S, Isakov S V, Ding N, Babbush R, Smelyanskiy V, Martinis J and Neven H 2016 What is the computational value of finite-range tunneling? Phys. Rev. X 6 031015

[8] Hen I, Job J, Albash T, Rannow T F, Troyer M and Lidar D A 2015 Probing for quantum speedup in spin-glass problems with planted solutions Phys. Rev. A 92 042325

[9] Mandrà S, Zhu Z, Wang W, Perdomo-Ortiz A and Katzgraber H G 2016 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

[10] Albash T and Lidar D A 2017 Evidence for a limited quantum speedup on a quantum annealer (arXiv:1705.07452)

[11] Mandrà S and Katzgraber H G 2017 A deceptive step towards quantum speedup detection (arXiv:1711.01368)

[12] Nishimori H and Takada K 2017 Exponential enhancement of the efficiency of quantum annealing by non-stoquastic Hamiltonians Frontiers in ICT 4 2

[13] Susa Y, Yamashiro Y, Yamamoto M and Nishimori H 2018 Exponential speedup of quantum annealing by inhomogeneous driving of the transverse field J. Phys. Soc. Jpn 87 023002

[14] Susa Y, Jodebeck J F and Nishimori H 2017 Relation between quantum fluctuations and the performance enhancement of quantum annealing in a nonstoquastic Hamiltonian Phys. Rev. A 95 042321

[15] Hormozi L, Brown E W, Carleo G and Troyer M 2017 Nonstoquastic Hamiltonians and quantum annealing of an Ising spin glass Phys. Rev. B 95 184416

[16] Vinci W and Lidar D A 2017 Non-stoquastic Hamiltonians in quantum annealing via geometric phases npj Quantum Information 3 38

[17] Mohseni M, Strumpler J and Rams M M 2018 Engineering non-equilibrium quantum phase transitions via causally gapped hamiltonians New J. Phys. 20 105002

[18] Mott A, Job J, Vlamin J-R, Lidar D and Spiropulu M 2017 Solving a Higgs optimization problem with quantum annealing for machine learning Nature 550 375

[19] Lokhov A Y, Vuffray M, Misra S and Cherktov M 2018 Optimal structure and parameter learning of Ising models Sci. Adv. 4 e1700791

[20] Li R Y, Di Felice R, Ro h R and Lidar D A 2018 Quantum annealing versus classical machine learning applied to a simplified computational biology problem npj Quantum Information 4 14

[21] Dziarmaga J and Rams M M 2010 Dynamics of an inhomogeneous quantum phase transition New J. Phys. 12 055007

[22] Del Campo A, Kibble T and Zurek W 2013 Causality and non-equilibrium second-order phase transitions in inhomogeneous systems J. Phys.: Condens. Matter. 25 404210

[23] CAMPO A D and Zurek W H 2014 Universality of phase transition dynamics: Topological defects from symmetry breaking Symmetry and Fundamental Physics: Tom Kibble at 80 (Singapore: World Scientific) pp 31–87

[24] Rams M M, Mohseni M and del Campo A 2016 Inhomogeneous quasi-adiabatic driving of quantum critical dynamics in weakly disordered spin chains New J. Phys. 18 123034

[25] Gómez-Ruiz F and Del Campo A 2019 Universal dynamics of inhomogeneous quantum phase transitions: suppressing defect formation Phys. Rev. Lett. 122 080604

[26] Andriyash E, Bian Z, Chudak F, Drew-Brook M, King A D, Macready W G and Roy A 2016 Boosting Integer Factoring Performance via Quantum Annealing Offsets (D-Wave Technical Report Series)

[27] Lanting T, King A D, Evert B and Hoskinson E 2017 Experimental demonstration of perturbative anticrossing mitigation using nonuniform driver Hamiltonians Phys. Rev. A 96 042322

[28] D-Wave Systems Inc The D-wave 2000QTM system https://dwavesys.com/d-wave-two-system

[29] Boixo S et al 2016 Computational multiqubit tunnelling in programmable quantum annealers Nat. Commun. 7 10327

[30] Mishra A, Albash T and Lidar D A 2018 When finite temperature quantum annealing defies expectations based on gap scaling (arXiv:1801.05511)

[31] Dickson N G et al 2013 Thermally assisted quantum annealing of a 16-qubit problem Nat. Commun. 4 1903

[32] Mohseni M and Neven H 2018 Constructing and programming quantum hardware for quantum annealing processes US Patent Specification 9934468

[33] Amin M 2017 private communication

[34] Zhu Z, Ochoa A J, Schnabel S, Hamze F and Katzgraber H G 2016 Best-case performance of quantum annealers on native spin-glass benchmarks: how chaos can affect success probabilities Phys. Rev. A 93 012317

[35] Reichardt B W 2004 The quantum adiabatic optimization algorithm and local minima Proc. of the 36th Annual ACM Symp. on Theory of Computing (New York: NY ACM) pp 502–10

[36] Sachdev S 2011 Quantum Phase Transitions 2nd edn (Cambridge: Cambridge University Press)

[37] D-Wave Systems Inc D-wave online learning https://cloud.dwavesys.com/learning/user/[user]/notebooks/2000Q/04-feature-examples/anneal-offset-module3.ipynb

[38] It is by no means a given that the results of a closed-system analysis will carry over into the open-system case, but the goal in this section is to develop some intuition rather than establish any rigorous results.