

Effective optimization using sample persistence: A case study on quantum annealers and various Monte Carlo optimization methods

Hamed Karimi,^{1,2} Gili Rosenberg,¹ and Helmut G. Katzgraber^{1,3,4}

¹*IQB Information Technologies (IQBit), 458-550 Burrard Street, Vancouver, British Columbia, Canada V6C 2B5*

²*Department of Computer Science, University of British Columbia, 2366 Main Mall, Vancouver, British Columbia, Canada V6T 1Z4*

³*Department of Physics and Astronomy, Texas A&M University, College Station, Texas 77843-4242, USA*

⁴*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA*

(Received 13 July 2017; revised manuscript received 18 September 2017; published 31 October 2017)

We present and apply a general-purpose, multistart algorithm for improving the performance of low-energy samplers used for solving optimization problems. The algorithm iteratively fixes the value of a large portion of the variables to values that have a high probability of being optimal. The resulting problems are smaller and less connected, and samplers tend to give better low-energy samples for these problems. The algorithm is trivially parallelizable since each start in the multistart algorithm is independent, and could be applied to any heuristic solver that can be run multiple times to give a sample. We present results for several classes of hard problems solved using simulated annealing, path-integral quantum Monte Carlo, parallel tempering with isoenergetic cluster moves, and a quantum annealer, and show that the success metrics and the scaling are improved substantially. When combined with this algorithm, the quantum annealer's scaling was substantially improved for native Chimera graph problems. In addition, with this algorithm the scaling of the time to solution of the quantum annealer is comparable to the Hamze-de Freitas-Selby algorithm on the weak-strong cluster problems introduced by Boixo *et al.* Parallel tempering with isoenergetic cluster moves was able to consistently solve three-dimensional spin glass problems with 8000 variables when combined with our method, whereas without our method it could not solve any.

DOI: [10.1103/PhysRevE.96.043312](https://doi.org/10.1103/PhysRevE.96.043312)

I. INTRODUCTION

When solving optimization problems, it is a common strategy to run a heuristic solver multiple times and keep the best solution found. If all of the found solutions are aggregated into a sample, one could ask if there is any additional information in this sample, aside from the solution with the best value. In this paper, we present results that suggest that it is indeed possible to use the sample more efficiently. In particular, the idea is that if we observe which variables have the same value in all solutions, and fix those variables to those values, we have fixed them to their values in at least one optimum with a high degree of confidence. Once this has been done, the remaining problem tends to be much smaller and simpler to solve.

Significant research has been done on solving Ising problems, and the equivalent quadratic unconstrained binary optimization (QUBO) problems,

$$\begin{aligned} \operatorname{argmin}_s [s^T J s + h^T s] \\ \text{s.t. } s \in \{-1, 1\}^N \end{aligned} \quad \text{or} \quad \begin{aligned} \operatorname{argmin}_x [x^T Q x] \\ \text{s.t. } x \in \{0, 1\}^N, \end{aligned}$$

using different solvers. Numerous well known NP-hard problems can be formulated in this way, such as the traveling salesman problem, the quadratic assignment problem, the maximum cut problem, the maximum clique problem, the set packing problem, and the graph coloring problem (for a selection of formulations, see [1]).

An idea that is prevalent in genetic algorithms is that of iteratively looking at a pair of solutions to find the common part, which is then fixed, and searching the remaining solution space. Although it is more common to breed two solutions, some research has suggested that it can be beneficial to breed a larger pool of solutions [2–5]. In addition, some algorithms

maintain a reference set of elite solutions (typically obtained by performing a local search), and rely on finding the variables that are often set to the same value in the elite solutions, the idea being that they are likely to be set to the same value in the optimum [6–8]. Another related concept is that of short- and long-term memory, in which local search algorithms such as tabu 1-opt guide their future exploration based on information gained from past exploration [9,10]. The closest existing work to the method described in this paper is Chardaire *et al.* [11], which fixes variables whose values remain constant as the temperature is decreased during simulated annealing.

Quantum annealers have recently become commercially available [12,13]. Manufactured by D-Wave Systems, Inc., they are designed to heuristically find low-energy states of an Ising problem. It has been suggested that quantum annealers have an advantage over classical optimizers due to quantum tunneling, which allows an optimizer to search the solution space of an optimization problem by passing through energy barriers instead of traversing them. For certain problem classes, this might provide a quantum speedup [14–20]. For this reason, there has been much recent interest in benchmarking quantum annealers against classical solvers, although conclusive evidence of quantum speedup has not been shown to date [21–32]. The classical solvers that are most usually benchmarked against quantum annealers are simulated annealing and simulated quantum annealing, both of which were included in our benchmarks.

Our research is based on an idea first proposed specifically for use with quantum annealers [33], and studied only on a narrow problem set. In this work, we utilize a modified and improved version of the algorithm outlined in [33], and show that it is effective for a range of solvers and hard problem sets. The original algorithm has several limitations that are

mitigated in the method presented in this work (see Sec. II B). The main focus of our study is not the benchmarking of the solvers against one another, but to gain an understanding for which problems and samplers our method is effective.

II. METHOD

In this paper, we study a multistart version of the sample persistence variable reduction algorithm (SPVAR) recently proposed by Karimi and Rosenberg [33]. For completeness, we briefly describe that algorithm in Sec. II A, after which, in Sec. II B, we describe in more detail the multistart variant that we used in this work. In Sec. II C, we discuss special considerations related to our method.

A. SPVAR

The SPVAR algorithm is based on the idea that, if a variable has the same value in all of the states obtained from a low-energy sampler, then it is more likely that the variable has that value in (at least) one optimum. This algorithm has two parameters: *fixing_threshold*, which controls what percentage of the solutions should share a value for a variable to be fixed; and *elite_threshold*, which controls the fraction of the low-energy part of the sample on which the algorithm should be applied. The idea is illustrated in Fig. 1. For further details, see Algorithm 1 and the original paper [33].

It is worth mentioning that finding and fixing these variables in a given sample is a very cheap computational operation, but as our results will show (Sec. III), it significantly improves the performance of the underlying sampler. Using the SPVAR algorithm in a single run has some drawbacks, the most important being that, as this method is a heuristic method, there is a probability that some of the variables will be fixed to a value that does not occur in any optimal solution. This was the motivation for using this method in a multistart fashion, as we describe in more detail in the following section.

Algorithm 1 SPVAR

Require: Ising problem (J, h) , *sampler*, *fixing_sample_size*, *fixing_threshold*, *elite_threshold*

Obtain sample of *fixing_sample_size* from *sampler*
 Record energies from sample
 Narrow down solutions to *elite_threshold* percentile
 Find mean value of each variable in all solutions
 Fix variables for which mean absolute value is larger than *fixing_threshold*
 Update J and h
return J, h , recorded energies, and a mapping from fixed variables to values to which they were fixed

B. Multistart SPVAR

The multistart SPVAR algorithm consists of *num_starts* independent starts of the SPVAR algorithm, using a sample size of *fixing_sample_size*. Each of these starts is followed by calling the sampler on the modified problem, with a sample size of *solving_sample_size*. The output of the algorithm is a collection of all of the energies encountered during this

process. The steps of the multistart SPVAR algorithm are summarized in Algorithm 2. For considerations related to parameter choice, see [33].

Multistart SPVAR has several advantages over single-start SPVAR. First, running SPVAR has a finite but unknown probability of fixing variables incorrectly. This is mitigated by restarting the algorithm multiple times. Second, this version allows one to choose the fraction of the sample size that will be used for fixing variables (via SPVAR), out of the total sample size. This could be done with a single start as well, but doing so often results in a wastefully large sample being used for fixing variables. In addition, this algorithm is trivially parallelizable, allowing for a speedup if multiple samples can be collected in parallel. Finally, for problems with a degenerate ground state, SPVAR might fix variables to their values in an optimum, but contrary to their values in other optima. This would make it impossible to observe these other optima in the modified problem. Multiple restarts make the resulting sample less biased.

We remark that ISPVAR, the iterative version of SPVAR, could also be used as the building block for this algorithm. The iterative algorithm can be useful if the objective is to fix more variables, and simplify the problem more drastically. However, every additional step of SPVAR that is applied incurs additional risk of fixing variables incorrectly. If a variable is fixed incorrectly at an early step, it is not unfixed at a later step. We suggest, instead, to set the thresholds more aggressively, which also results in the fixing of more variables. The price again is an increased risk of fixing variables incorrectly, but this risk can be mitigated by increasing *num_starts*.

Algorithm 2 Multistart SPVAR

Require: Ising problem (J, h) , *sampler*, *fixing_sample_size*, *solving_sample_size*, *fixing_threshold*, *elite_threshold*, *num_starts*

[Optional] Apply preprocessing to find modified J, h
 foreach start of *num_starts* do
 Apply SPVAR with sample of *fixing_sample_size* to find modified J, h
 Record energies from sample
 [Optional] Apply preprocessing to find modified J, h
 [Optional] Fix variables via correlations to find modified J, h
 Obtain sample of size *solving_sample_size* for modified J, h
 Record energies from sample
end For
return Recorded energies, and a mapping from fixed variables to values to which they were fixed

C. Special considerations

For Ising problems with zero bias (such as those in Secs. III C and III D), there is a twofold degeneracy of all states, which leads to many states appearing in a given sample with their reversed state. For this reason, if the method is applied exactly as described in Sec. II B, it will generally fix no variables and fail to result in an improvement. As suggested in [33], a possible solution is to break the degeneracy by arbitrarily fixing a single variable. However, as was found in that work, it is possible to use correlations in the sample

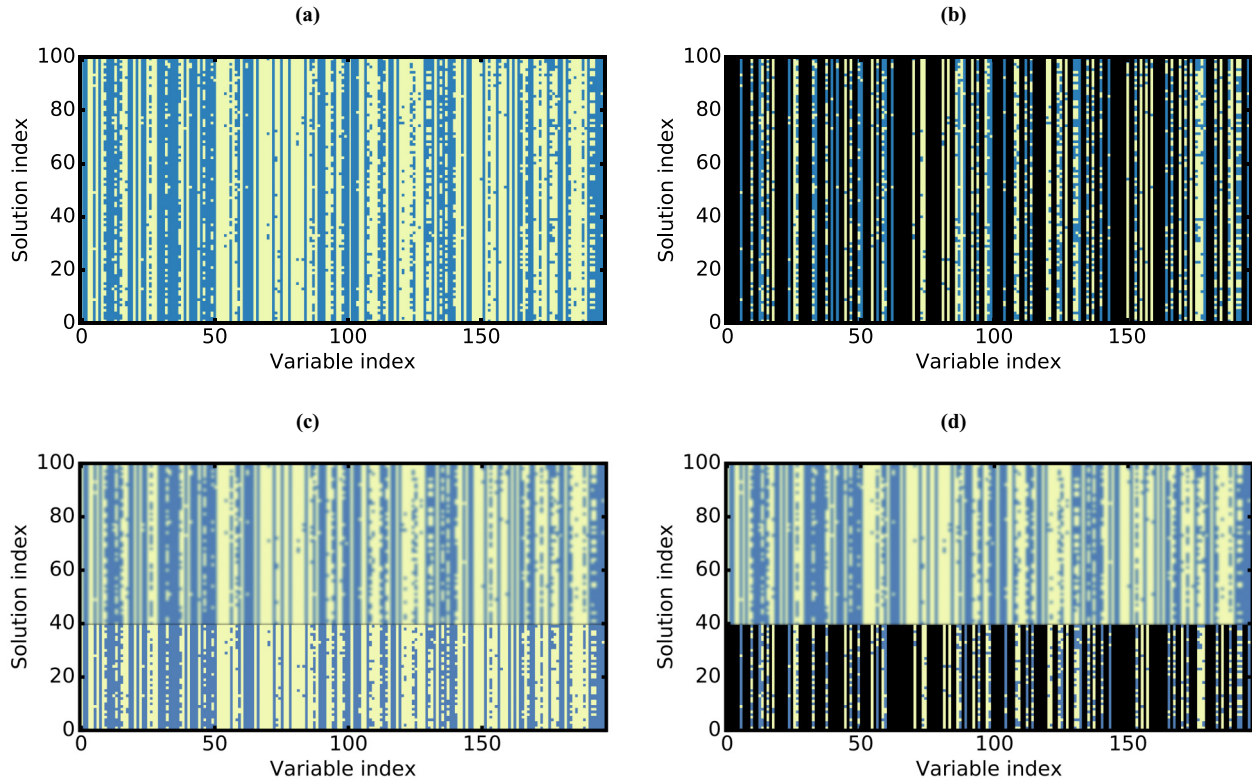


FIG. 1. A graphical illustration of the SPVAR algorithm. (a) The full sample, with 100 solutions (the rows) and 200 variables (the columns). In each row, a dark dot (blue) indicates that that variable was $+1$ in that solution, and a light dot (beige) indicates that that variable was -1 . (b) The full sample, but with variables that had the same value in all solutions marked in black. (c) The elite sample, formed by keeping only the 20% lowest-energy solutions. (d) The elite sample, but with variables that had the same value in all solutions in the elite sample marked in black. In all figures, the solution index is sorted with respect to the energy, such that the solutions with the lowest energies are at the bottom.

to fix a cluster of correlated variables instead, which is the method we employed in this study.

In cases in which the modified problem (after applying SPVAR) was disconnected, we took advantage of this fact to boost our results. In such cases, it is wasteful to assess the sampler’s performance based on the energies for the whole problem, for the reason that even if the sampler did not manage to solve all of the disconnected problems at once, it may have solved all of them at least once. For this reason, we separated the solutions in a given sample into a partial sample for each connected problem. We then evaluated the partial energies for each connected component, for each solution; sorted the partial sample based on the partial energies; merged all of the partial (but sorted) solutions; and, finally, summed the partial energies to find the new sample. In this way, if the sampler solved each connected component in at least one solution, the best solution in the new sample would be a ground state of the whole problem.

The optional preprocessing step can encompass many different methods for fixing variables. In this case, we used the *fix_variables* function in D-Wave’s SAPI 2. We then added an additional method which deals efficiently with trees. For nodes (variables) with degree one, if the absolute value of the coupling is smaller than the absolute value of the bias for that variable, the value of the variable is determined by the sign of the bias. On the other hand, if the absolute value of the coupling is larger than the absolute value of the bias, the value

of the variable is determined by the sign of the product of the coupling and the value of the neighboring variable. This method can be applied recursively to fix, in the former case, or infer, in the latter case, the values of variables in trees, leading to a reduction in the number of variables equal to the number of variables in the tree.

III. RESULTS AND DISCUSSION

This section is organized as follows. In Sec. III A, we describe the benchmarking procedure we used in order to collect the results that are presented in this section. In the sections that follow, we present results for different problem sets and solvers, and discuss them. In Sec. III B, we present results for weak-strong cluster problems. In Sec. III C, we present results for reduced-degeneracy Chimera graph problems with zero and nonzero bias; in Sec. III D, we present results for three-dimensional (3D) spin glass problems; and in Sec. III E, we present results for fault diagnosis problems. In Sec. III F, we present results for Max-2-SAT and Max-3-SAT problems.

A. Benchmarking procedure

For each problem set and each parameter value (when applicable), we generated or procured 50–100 instances, which were solved using a variety of samplers, with and

without multistart SPVAR. We used the following heuristic solvers as samplers in this study: “SA,” an implementation of simulated annealing [34]; “SQA,” an implementation of discrete-time path-integral quantum Monte Carlo as a simulation of the quantum annealing process, which we refer to as SQA in this study [16,19,35]; “DW,” D-Wave’s quantum annealer (we had access to the DW2X_SYS4 and the DW_2000Q); and “PTICM,” parallel tempering Monte Carlo [36–38] with isoenergetic cluster moves, also known as *borealis* [39,40]. In each case, the version that used multistart SPVAR is indicated by the suffix “SPVAR,” for example, “SA_SPVAR.”

In addition, we solved, or procured best known solutions for, all problem instances with an additional solver which either guarantees finding the optimum, or finds it with a high probability. Specifically, the reduced-degeneracy Chimera graph problems were solved with the Hamze–de Freitas–Selby (HFS) algorithm [41], and the Max- k -SAT problems were solved with the CCLS [42] and *ahmaxsat* [43] algorithms. For the 3D spin glass problems and the fault diagnosis problems, we procured best known solutions using PTICM. For the 3D spin glass problems with planted solutions, the ground state was known [44].

Using the energies returned by multistart SPVAR, we evaluated several success metrics: “Fraction of problems solved”; “Gap,” the difference in value between the best energy found and the best known solution’s energy; “Residual,” the relative difference (in percent) in value between the best energy value found and the best known solution’s energy; and “R99,” the size of sample required to observe the ground state with 99% confidence. When calculating R99 values for results obtained with SPVAR, care is needed to get the correct value. We first calculated the mean success rate for each start and the number of starts required to observe the ground state with 99% confidence. We then multiplied this value by the total size of the sample used in each start, to finally obtain the R99 value.

B. Weak-strong cluster problems

The weak-strong cluster problems were introduced by Boixo *et al.* [45] as a toy model for studying the role of multiqubit tunneling in a quantum annealer. They are Chimera graph problems in which the variables in each unit cell are ferromagnetically coupled to one another, and all have an equal bias which is +1 for the “strong” clusters and $h_w < 0$ for the “weak” clusters. For $h_w = -0.5$, the ground state is doubly degenerate, corresponding to a state in which the weak cluster is either aligned with, or opposite to, the strong cluster. The case studied is $-0.5 < h_w < 0$, which exhibits a false minimum in which the weak cluster is the opposite of the strong cluster. It has been shown, both theoretically and empirically, that SA tends to end up in the false minimum, unless cluster moves are allowed, whereas SQA and QA are able to tunnel out of the false minimum.

The weak-strong problems were subsequently studied by Denchev *et al.* [22]. It was shown that, on these problems, QA outperforms SA both in time and in scaling, and outperforms SQA in time (but not in scaling). Mandrà *et al.* [27] subsequently showed that these problems can, in fact, be solved more

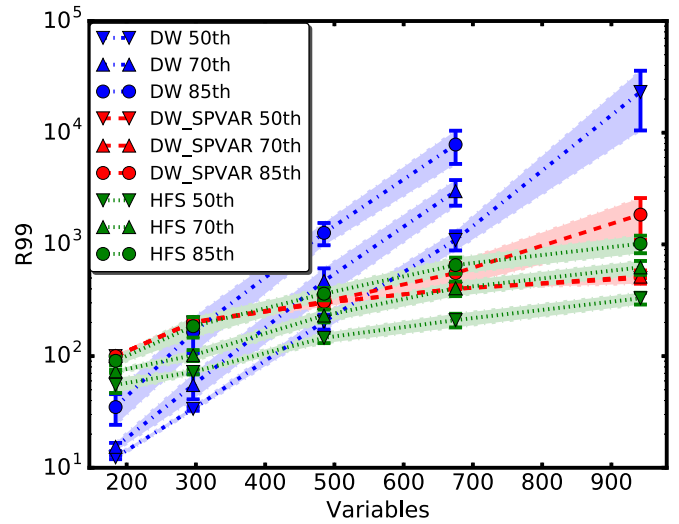


FIG. 2. Results for weak-strong cluster problems. We present the R99 values for different samplers, as a function of the problem size (number of variables), for different percentiles. The error bars were calculated using bootstrapping.

efficiently both in time and scaling by algorithms such as HFS and others. Later, Mandrà *et al.* [32] also showed that these problems can be solved exactly in polynomial time, due to the planar structure of the logical problem. Our motivation for studying these problems was to check whether the application of our method would close the performance gap between the quantum annealer and HFS.

We obtained the original problem instances from the above studies [22,27]. Since those instances were constructed for a different D-Wave chip, and each chip has a small number of arbitrarily located inactive qubits, we modified the instances to account for this. In addition, due to our having a limited amount of quantum annealer time, we had to change the weak bias h_w from -0.44 (as in the original instances) to -0.42 , which results in problems that are slightly easier for the quantum annealer. We expect that our results would hold qualitatively also for the original instances, run on the same chip they were benchmarked on in the past studies.

We present the R99 values for HFS and for the quantum annealer used with and without our method in Fig. 2. In this case, the R99 value reported is the time required to find the ground state with 99% confidence, in units of $20 \mu\text{s}$, which was the annealing time we used for the quantum annealer. We were unable to solve these problems with SA or SQA due to our having insufficient resources. The quantum annealer we used in this section was the DW2X_SYS4 [46].

Our results show a clear improvement in the R99 values when running the quantum annealer with our method, compared to running it without our method. Notably, the improvement is of two orders of magnitude for the largest problems. Furthermore, the scaling of the R99 value with the number of variables is clearly significantly improved, making it qualitatively comparable to HFS’s scaling, and the R99 values themselves are also comparable with those of HFS. We expect that qualitatively similar results would be drawn for the problems benchmarked in [22,27].

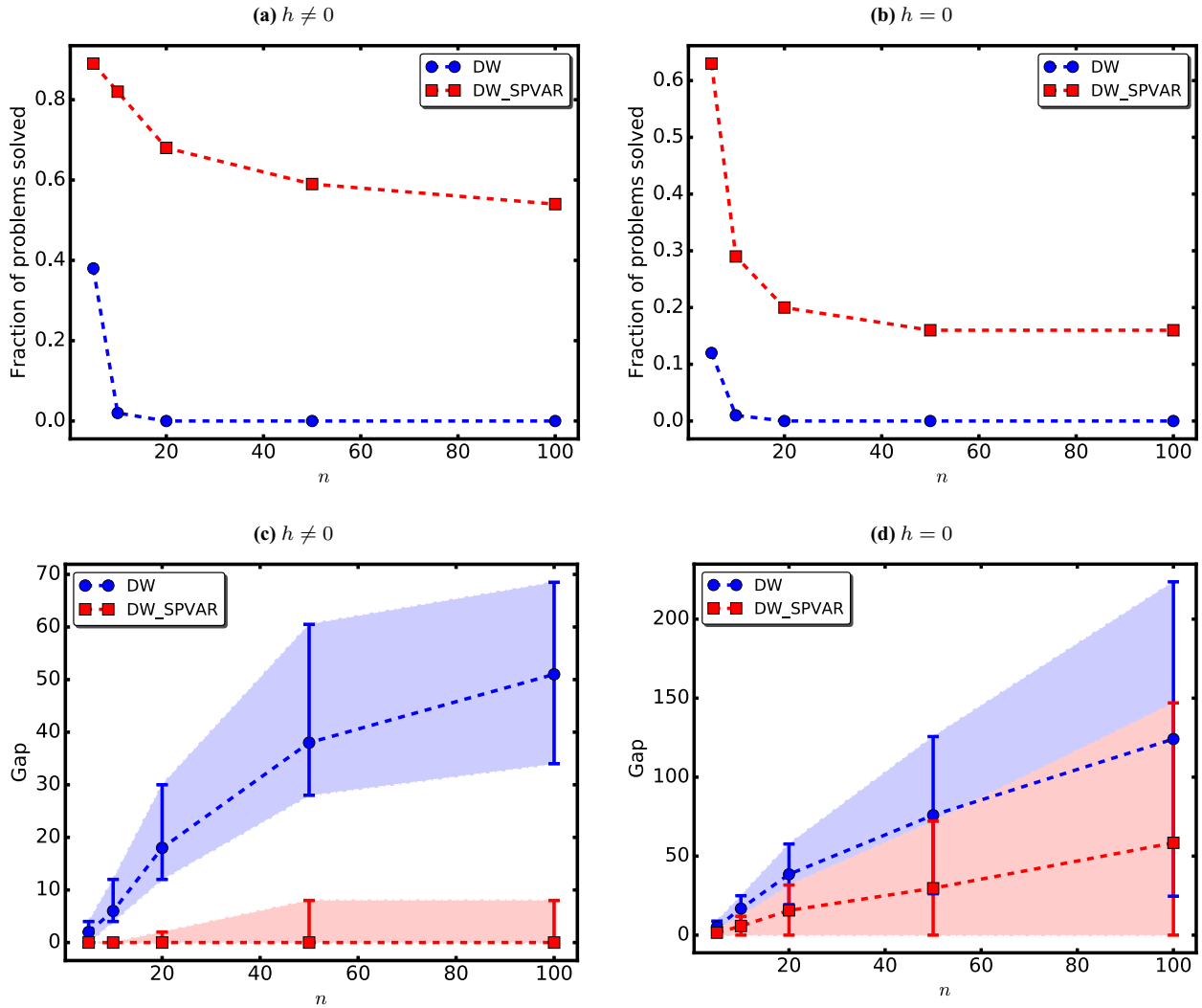


FIG. 3. Success metrics for $U_{n,n-1,n-2}$ Chimera graph problems with reduced degeneracy. (a), (b) The fraction of problems solved, as a function of n , for nonzero-bias problems and zero-bias problems, respectively. (c), (d) The median gap, which is the energy difference between the best solution found and the best known solution, as a function of n , for nonzero-bias problems and zero-bias problems, respectively. In both figures, each point represents data from 100 random instances with the respective n .

C. Chimera graph problems

1. Reduced-degeneracy Chimera graph problems

Random Chimera graph problems have been benchmarked thoroughly in the past [23,24,26,28,30], after which it was shown that they are not expected to show a quantum speedup [24]. However, they remain a well studied test bed, and are easy to construct. It was suggested that harder instances could be constructed by choosing the couplers and biases from a set with a large range compared to the spacing. Our instances were constructed by choosing the couplers and biases from a uniform probability distribution over the set $\{-n, -(n-1), -(n-2), n-2, n-1, n\}$, which we denote by $U_{n,n-1,n-2}$. Such instances challenge a noisy quantum annealer, due to the effects of intrinsic control error (ICE). In addition, they challenge our method since the mean change in energy due to an incorrectly flipped variable is much larger than in an instance in which the couplers and biases are chosen from the complete range $-n$ to n .

In addition, in the interest of creating hard instances, it is helpful to reduce the ground state degeneracy [30]. To this end, we eliminated local degeneracies; doing so does not guarantee that the ground state will not be degenerate, but reduces the probability of this occurring. An instance was created via the following process. First, an initial set of couplers and biases was selected. Then, for each variable, we checked if there is a possible configuration of the neighboring variables that would result in the effective field on the central variable being zero. If such a configuration was found, one of the couplers was changed to a different and randomly chosen value. This process was repeated until no more local degeneracies remained. We remark also that for small n , the probability of local degeneracies occurring is significant, but it falls rapidly as n is increased.

We present results for the success metrics of $U_{n,n-1,n-2}$ Chimera graph problems with reduced degeneracy with nonzero and zero bias in Fig. 3 for the quantum annealer, with and without applying multistart SPVAR. We present

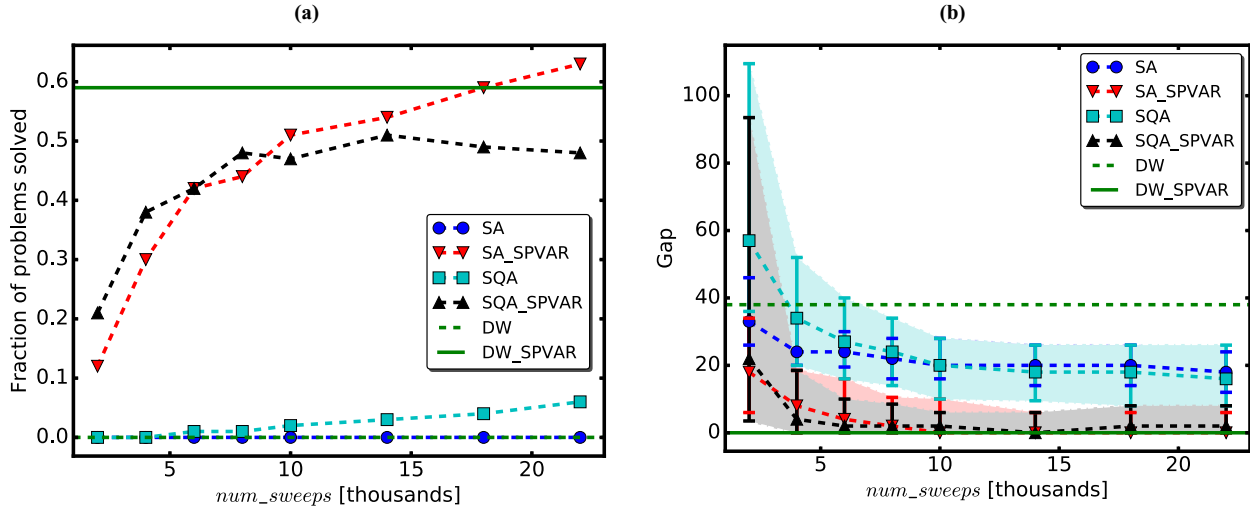


FIG. 4. Success metrics for $U_{50,49,48}$ Chimera graph problems with nonzero bias and reduced degeneracy, versus the number of sweeps. (a) The fraction of problems solved. (b) The median gap, which is the energy difference between the best solution found and the best known solution. The figures show median values for 100 random instances, solved with differing numbers of sweeps. Horizontal lines indicate the median values for D-Wave’s quantum annealer, without and with multistart SPVAR (DW and DW_SPVAR, respectively).

results comparing success metrics for SA, SQA, and the quantum annealer (the DW2X_SYS4) for $U_{50,49,48}$ problems with nonzero bias and reduced degeneracy in Fig. 4. For the zero-bias problems, we set the *elite_threshold* adaptively as suggested in [33].

The results presented in Fig. 3 show that these problems are very hard for the quantum annealer, with a negligible success rate for $n > 5$ for both the nonzero-bias and zero-bias problems. We attribute this difficulty to ICE, given that the accuracy required to distinguish different couplers scales with $1/n$ for these problems. In addition, the reduction of the ground state degeneracy by eliminating local degeneracies is likely a contributing factor. Zero-bias problems are known to be harder, in general, and indeed for $n = 5$, the results obtained from the quantum annealer are better for the nonzero-bias problems.

Despite the very low success rate of the quantum annealer used alone, when coupled with our method, the quantum annealer solved over 60% of each of the problem sets for the nonzero-bias problems, and close to 20% for the zero-bias problems. As noted in [33], our intuition is that the quantum annealer is better at finding low-energy states than it is at finding optima. We hypothesize that when considering only the state with the lowest energy value in the sample, additional information about the low-energy landscape contained within the rest of the sample is discarded. Utilizing this information allows our method to substantially improve the success metrics.

Figure 4 shows that the success metrics of SA and SQA improve as the number of sweeps is increased, albeit with diminishing returns. The results also show that both SA and SQA are able to match the quantum annealer’s performance using a relatively modest 2000 sweeps. However, if our method is applied to all samplers, SA and SQA require a significantly larger number of sweeps to match the performance of the quantum annealer when using our method. Therefore, in this case, the quantum annealer seems to benefit more from the application of our method. It is worth mentioning that,

combined with SPVAR, QA was able to solve 60% of the problems roughly three orders of magnitude faster than either of SA or SQA combined with SPVAR. The median R99 for SA was 43,709, and for QA it was 66,745, and the time required to obtain a single solution was 20 ms for SA (for this problem size, 1100 variables) and 20 μ s for QA (the annealing time).

2. Scaling analysis

In this section, we investigate the scaling of success metrics when using SPVAR. We generated four native Chimera graph problem sets from $U_{10} \in \{-10, \dots, 10\}$ (0 was excluded for the couplers but included for the biases), each consisting of 100 problems with increasing sizes, from a Chimera graph of size 8 to 16 (the latter referring to a 16×16 grid of blocks). The quantum annealer we used for this study was the D-Wave 2000Q, which has 2048 qubits, with a qubit yield of almost 99% [47]. In Fig. 5, we show the dependence of success metrics on the problem size.

Figure 5(a) shows that the fraction of problems solved decreases exponentially for the quantum annealer used alone, but scales better when used in conjunction with SPVAR. Note that for size 16, none of the problems were solved using the quantum annealer alone, but 46% of problems were solved when using the quantum annealer in conjunction with SPVAR.

We also observed significant improvement in the R99, which is a measure of the time to solution [48], shown in Fig. 5(b). The reported R99 value is the mean of the median R99 in 1000 bootstrapped samples (using only the R99s that could be measured). In cases in which less than 50% of the problems were solved, this value is a lower bound of the actual median R99. In particular, for size 12, the quantum annealer solved only 6 problems, so the R99 value is most likely a very loose lower bound. The scaling when using SPVAR is clearly improved.

Figure 5(c) shows a clear advantage in scaling when our method is utilized. In Fig. 5(d), the number of variables fixed

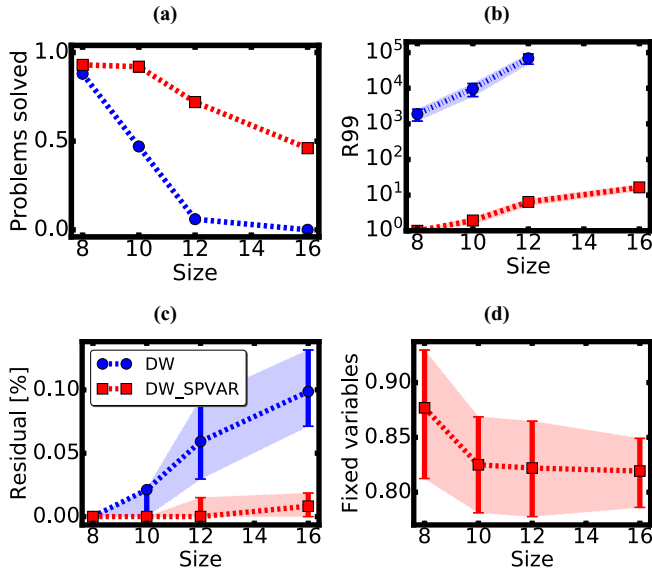


FIG. 5. Success metrics for multistart SPVAR on Chimera graph-structured problems in U_{10} vs size. (a) The fraction of problems solved at each size. (b) The median R99 values, calculated using bootstrapping, with and without using SPVAR. (c) The median energy residual. (d) The median fraction of fixed variables at each size.

appears to plateau, such that even at large sizes the method fixes more than 80% of the variables.

D. 3D spin glass problems

Ising problem instances generated on a 3D cubic lattice have been shown to be relatively hard instances, for example, by benchmarking with parallel tempering and population annealing [49,50]. We obtained and benchmarked a selection of instances with $L = 4, 6, 8, 10$ from the authors of those studies, where the number of variables is L^3 . For each L , we chose the 20 hardest instances from a larger instance pool and

selected an additional 80 instances uniformly randomly from that same pool. We present results for the success metrics as a function of L for different samplers in Fig. 6, and results for the dependence of the success metrics on num_starts in Fig. 7. As these problems could not be embedded on the quantum annealer chip to which we had access, we present the results of only SA and SQA for these problems.

In Fig. 6, we see that the application of our method provides a substantial improvement in the success metrics, especially for the larger and harder problem instances. For example, for $L = 10$, no problems are solved by SA and SQA, but with the addition of our method, they are both able to solve almost 40% of the problems, with a greatly reduced residual. Increasing num_starts , as in Fig. 7, improves the results of all methods, albeit with diminishing returns. For the largest problems ($L = 10$), increasing num_starts barely improves the results for the samplers when our method is not applied, but gives a substantial improvement when our method is applied. The performance is comparable with population annealing and parallel tempering [49].

1. Parallel tempering with isoenergetic cluster moves

For all solvers, except for the parallel tempering with isoenergetic cluster moves algorithm (PTICM), the sample was formed by multistarting the given algorithm and aggregating the results. This is a natural choice for algorithms that return a single state (or a finite number of states) per start. PTICM returns a single state per replica. However, forming the sample in this way suffers from the disadvantage that each replica might have been in many lower-energy states during the optimization process, which would be ignored. In addition, since the temperatures of the replicas typically span a large range, many of the replicas return relatively high-energy states, such that the sample formed would not be a good low-energy sample. For this reason, we chose instead to keep track of the 10 lowest-energy states found for each of the replicas in the lower half of the temperature range.

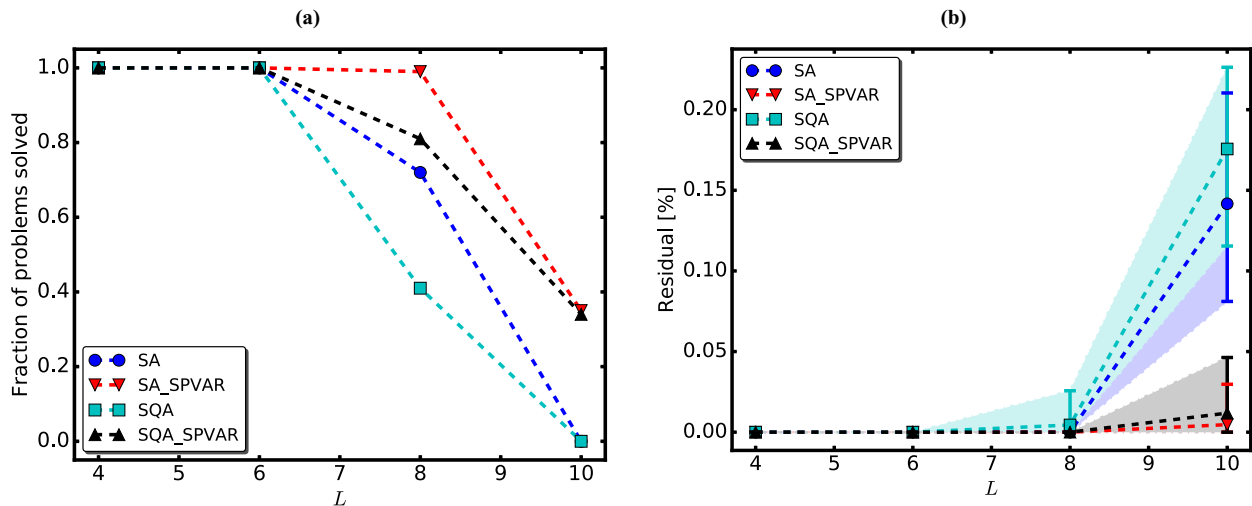


FIG. 6. Success metrics for 3D spin glass problems vs size. (a) The fraction of problems solved, as a function of L , where the size is L^3 . (b) The median residual, which is the relative energy difference (in percent) between the best solution found and the best known solution, as a function of L . In both figures, each point represents data from 100 random instances with the respective L .

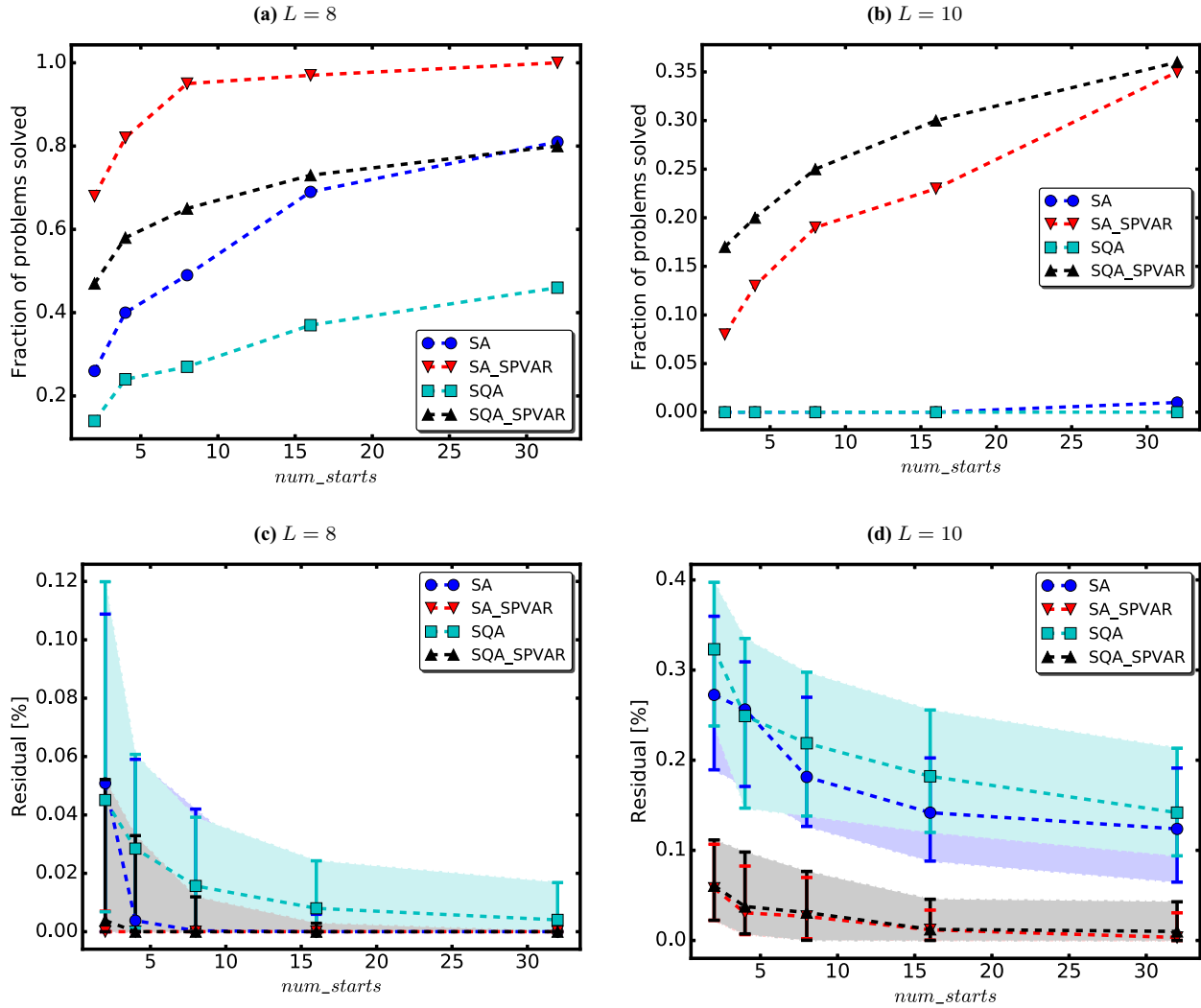


FIG. 7. Success metrics for 3D spin glass problems vs num_starts . (a), (b) The fraction of problems solved, for $L = 8$ and 10 (respectively). (c), (d) The median residual, which is the relative energy difference (in percent) between the best solution found and the best known solution, for $L = 8$ and 10 (respectively). In all figures, each point represents data from 100 random instances.

The results for PTICM are presented separately from the other solvers. One reason for this is that the sample was formed in an inherently different way; for the other solvers, the sample consisted of independently sampled states. In addition, there is no obvious way to equate the computational effort used by PTICM to that of SA or SQA since a single sweep is a combination of a Metropolis update, a parallel tempering move, and an isoenergetic cluster move. Finally, the performance of PTICM was superior on the instances benchmarked with the other solvers, leaving little room for improvement by SPVAR. For example, the 3D spin glass problems of sizes $L \leq 8$ were all solved by PTICM without SPVAR, and even for $L = 10$, most were solved.

The results for the same pool of $L = 10$ problems from Sec. III D are presented in Fig. 8. For $L \leq 8$, all instances were solved by PTICM as well as by PTICM with SPVAR, so we do not present results for those problems. In order to probe the performance of PTICM on larger problems and to study the scaling, we procured 3D spin glass instances with $L = 10, 12, 16, 20$ with planted solutions [44]. This was

necessary, as for the larger problems, a huge computational effort is required to ensure that the ground state is found with high confidence using heuristics. We present results for the success metrics as a function of num_sweeps in Fig. 9(a), and results for the dependence of the success metrics on L in Fig. 9(b).

In Fig. 8, we see that our method provides a modest improvement for the non-planted 3D spin glass problems of size $L = 10$. In Fig. 9(a), we see that for $L \geq 12$, for almost all values of num_sweeps , the fraction of problems solved is substantially larger. In Fig. 9(b), for the planted 3D spin glass problems, we see that our method provides a substantial improvement in both the success metrics and the scaling. For example, for $L = 16$ and 20, no problems are solved by PTICM, but with the addition of our method, it is able to solve almost all of the problems with $num_sweeps = 10^4$. Interestingly, the $L = 16$ and 20 planted problems are harder for PTICM than the $L = 10$ nonplanted problems, but for PTICM with SPVAR this is reversed: it solved all of the $L = 16$ planted problems, and almost all of the $L = 20$ planted

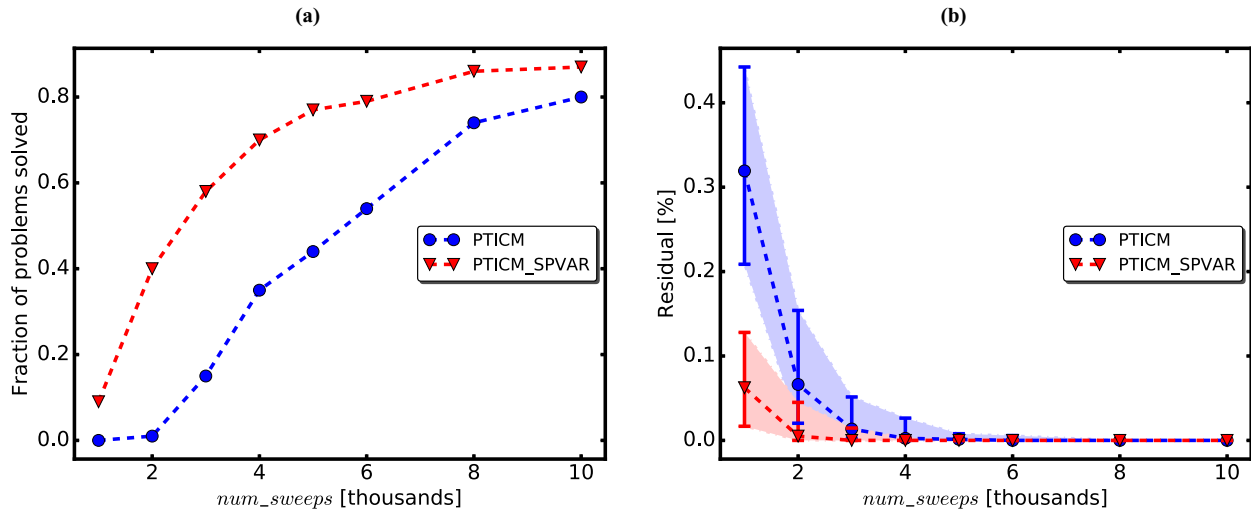


FIG. 8. Success metrics for PTICM for 3D spin glass problems with $L = 10$ vs num_sweeps . (a) The fraction of problems solved. (b) The median residual, which is the relative energy difference (in percent) between the best solution found and the best known solution. In both figures, each point represents data from 100 random instances.

problems. We suspect that SPVAR is able to exploit the structure of the planted-solution problems.

E. Fault diagnosis problems

Fault diagnosis problems are one of the leading candidates for benchmarking quantum annealers [51–53]. We obtained a single problem set of 100 instances from 4-bit \times 4-bit multiplier circuits as generated in Ref. [53], where it was shown that these instances are at least one order of magnitude harder than conventionally used random spin glass instances. Aside from their intrinsic hardness in terms of time to solution, these fault diagnosis instances are also harder from

an asymptotic scaling perspective, making them an interesting test bed on which to apply SPVAR.

In order to solve these problems with D-Wave’s quantum annealer, special treatment was required since their adjacency matrix differs from the hardware graph of the quantum annealer (a Chimera graph). A common method is to find an *embedding*, which is a mapping from each logical variable to one or more qubits, referred to as a *chain*. The identification of multiple qubits with a single logical variable results in a higher effective connectivity, but requires additional couplers in order to try to force all of the qubits in a chain to take the same state. This is commonly done by connecting the qubits with a strong ferromagnetic coupling. However, in a sample

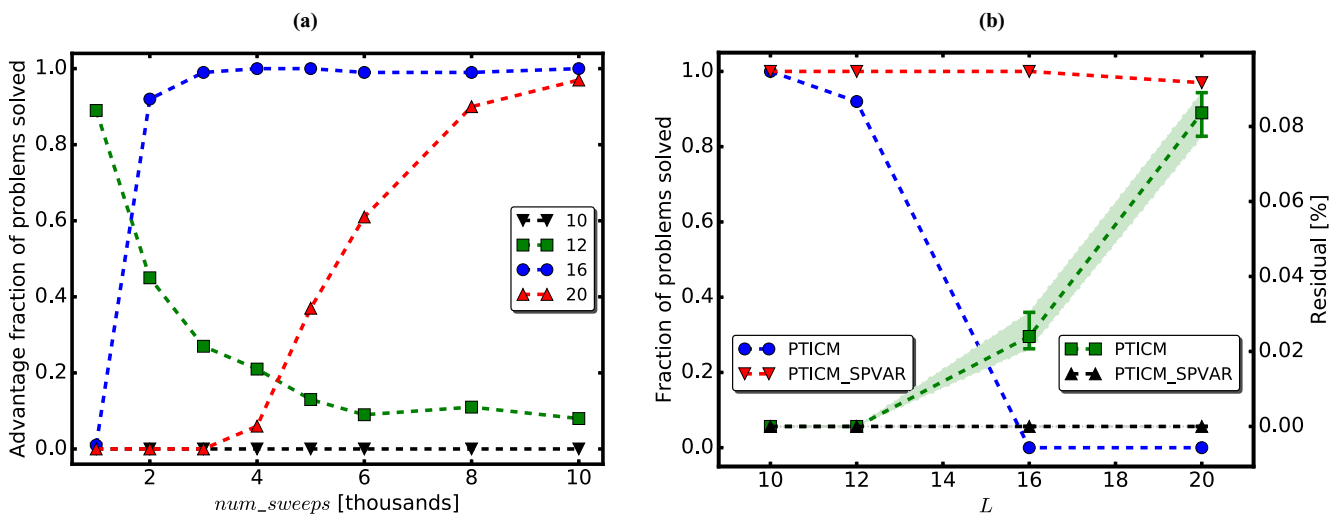


FIG. 9. Fraction of problems solved for 3D spin glass problems with planted solutions vs num_sweeps and L . (a) Shows the difference in the fraction of problems solved for $L = 10$ – 20 between PT+ICM with SPVAR and PT+ICM without SPVAR. Note that for $L = 10$, PT+ICM solved all problems both with and without SPVAR. (b) Shows the fraction of problems solved, and the median residual, which is the relative energy difference (in percent) between the best solution found and the best known solution, as a function of L . In both panels, each point represents data from 100 random instances with the respective L , solved with $num_sweeps = 10^4$.

TABLE I. Success metrics for the (4,4) fault diagnosis problems for three samplers: simulated annealing (SA), simulated quantum annealing (SQA), and D-Wave’s quantum annealer (DW). For each sampler, we present the success metrics for solving using the sampler alone compared with solving using SPVAR, with the same computational effort. For DW, we also show the results with post-processing (PP). The success metrics are as follows: “Solved,” the percentage of problems solved; and “Residual,” the mean difference (in percent) in energy between the sample’s best result and the best known solution.

Sampler	Without Solved	SPVAR Residual	With Solved	SPVAR Residual
SA (2000 sweeps)	21	0.167	80	0.032
SA (20000 sweeps)	67	0.056	96	0.005
SQA (2000 sweeps)	0	2.055	4	0.447
SQA (20000 sweeps)	2	0.348	45	0.115
DW	0	4.660	0	1.278
DW + PP	19	0.189	88	0.017

obtained from the quantum annealer, it is still possible that not all qubits in a chain would have the same state. In that case, a “decoding” technique must be utilized to decide the value of the logical variable corresponding to those qubits. There are various decoding techniques, such as majority vote, and local and global energy minimization [54]. In this study, we employed local energy minimization for decoding. More specifically, we assigned an effective field to each broken chain, selected the chain with the strongest effective field, set its state to be opposite to that of the direction of the field (to minimize the energy), updated the effective fields, and repeated the process until no broken chains remained. This is a quick method of local error correction. We remark that finding an embedding for a graph is formally known as graph-minor embedding, an NP-hard problem [55,56] commonly solved by heuristic methods, although for some classes of problems deterministic embeddings can be found [57].

We present the success metrics for the fault diagnosis problems in Table I, and the dependence of the R99 on the percentile for SA with and without our method in Fig. 10.

For the fault diagnosis problems, the application of our method results in a large improvement to the success metrics for SA and SQA, as can be seen in Table I. The application of the method significantly improves the quality of solutions found by the quantum annealer, as seen in the residual, but it was still unable to solve any of those problems. We attribute this to the low quality of the samples obtained from the quantum annealer, a result of these problems being extremely hard for it to solve. With the addition of post-processing, the quantum annealer achieves results that are in line with the raw results obtained by SA and SQA, and shows a large improvement when our method is applied. We were not able to apply the same post-processing to SA and SQA since the post-processing is internal to the quantum annealer’s SAPI.

F. Max- k -SAT problems

It is known that the size of the backbone, defined as the number of variables that are true in all of the optima, is an order

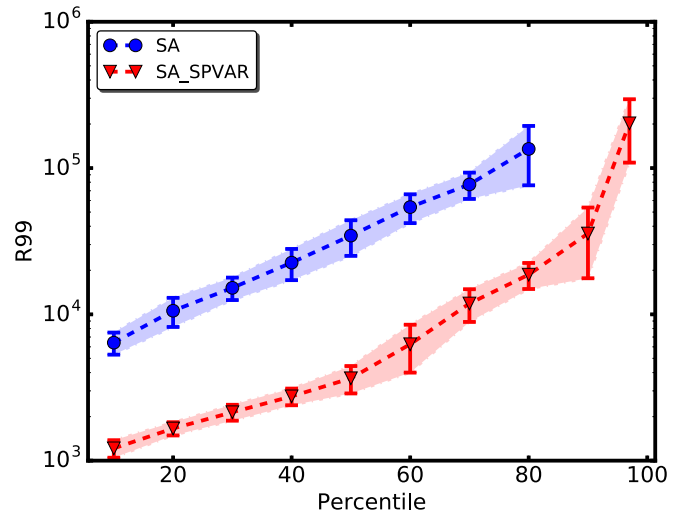


FIG. 10. R99 as a function of percentile for the (4,4) fault diagnosis problems, for simulated annealing, with and without SPVAR. The error bars were calculated using bootstrapping.

parameter for Max- k -SAT problems [58,59]. k -SAT and Max- k -SAT problems undergo a phase transition (second order), with the order parameter defined as the size of the backbone, at some critical point. At this critical point, Max- k -SAT problems undergo an “easy-hard” phase transition. It is natural to ask how our method would perform near this phase transition for Max- k -SAT problems. The critical point has been proven to be at $\phi_c = 1$ for $k = 2$ [60–62], and has been shown to be at $\phi_c \simeq 4.267$ for $k = 3$, where ϕ is the ratio of the clauses to literals [63–65].

We generated random sets of Max- k -SAT problems with a given number of literals and a varying number of variables. For the Max-3-SAT instances, the resulting objective function is of order three, necessitating the addition of auxiliary variables and corresponding terms in order to reduce it to a quadratic objective function since the samplers we used were all implemented to solve quadratic unconstrained problems. A comparison of the number of fixed variables for different methods is presented in Fig. 11, for Max-2-SAT and Max-3-SAT problems. In addition, for the Max-3-SAT problems, the fraction of problems solved and the gap are presented in Fig. 12.

Based on our observations, our method performs well near the phase transition, and continues to perform well as the number of clauses over literals is increased, which corresponds to an increase in hardness. This can be seen both in the number of fixed variables remaining high to the right of the phase transition in Fig. 11, as well as in the fact that the method continues to provide a boost to success metrics, as seen in Fig. 12.

To the left of the phase transition, it is easy to satisfy all of the clauses, and the problems are expected to have a large number of ground states, making them easy to solve for local search methods, such as tabu 1-opt search. In this regime, for $k = 2$, the method without the optional calls to `fix_variables` fixes very few variables. This could be explained by the intuition that, due to the large number of ground states, low-energy samplers tend to give a large number of

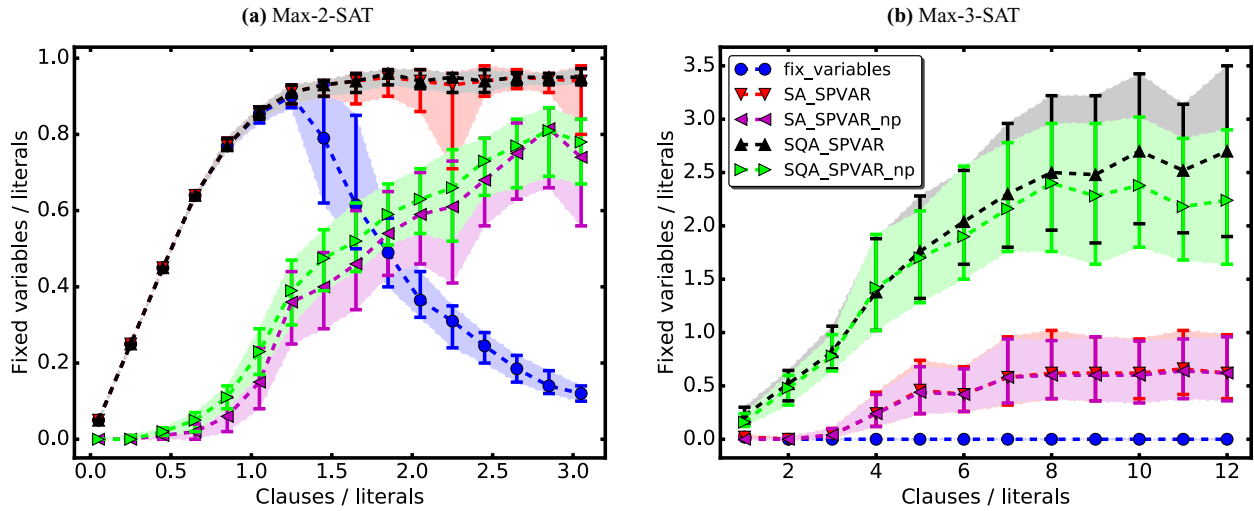


FIG. 11. Number of fixed variables for Max- k -SAT problems. (a) The median number of fixed variables for Max-2-SAT problems with 100 literals as a function of the number of clauses. SA_SPVAR refers to simulated annealing run in conjunction with SPVAR, SQA_SPVAR refers to simulated quantum annealing run in conjunction with SPVAR, and *fix_variables* refers to persistence-based fixing of variables, via the *fix_variables* function in D-Wave’s SAPI 2. The suffix “np” refers to switching off the optional calls to *fix_variables* in multistart SPVAR. (b) The median number of fixed variables for Max-3-SAT problems with 50 literals as a function of the number of clauses. In both figures, the means were taken over 50 random instances with the respective number of clauses.

distinct states with low overlap between them. However, this is the regime in which *fix_variables* performs well. To the right of the phase transition, *fix_variables* rapidly breaks down, whereas our method continues to fix a large number of variables. It appears that using our method in conjunction with *fix_variables* allows one to benefit from the strengths of each.

The number of variables that are fixed by our algorithm is related to the size of the backbone, which we refer to as an approximate backbone. For this reason, it is not surprising that we observed a distinct change in behavior of the number of fixed variables near the phase transition.

For $k = 3$, *fix_variables* is not able to fix any variables. We hypothesize that the reduction of the third-order polynomial to a quadratic one results in a structure that is not well suited to the type of persistence fixing that *fix_variables* employs. However, it is still able to provide a small increase in the number of variables fixed when used in conjunction with our method. For $k = 2$, SA and SQA perform similarly, but for $k = 3$, SQA fixes substantially more variables than SA. There are at least two factors that might contribute to this. First, on the left of the phase transition, the curve for SQA is much steeper for $k = 3$ than for $k = 2$. This might seem

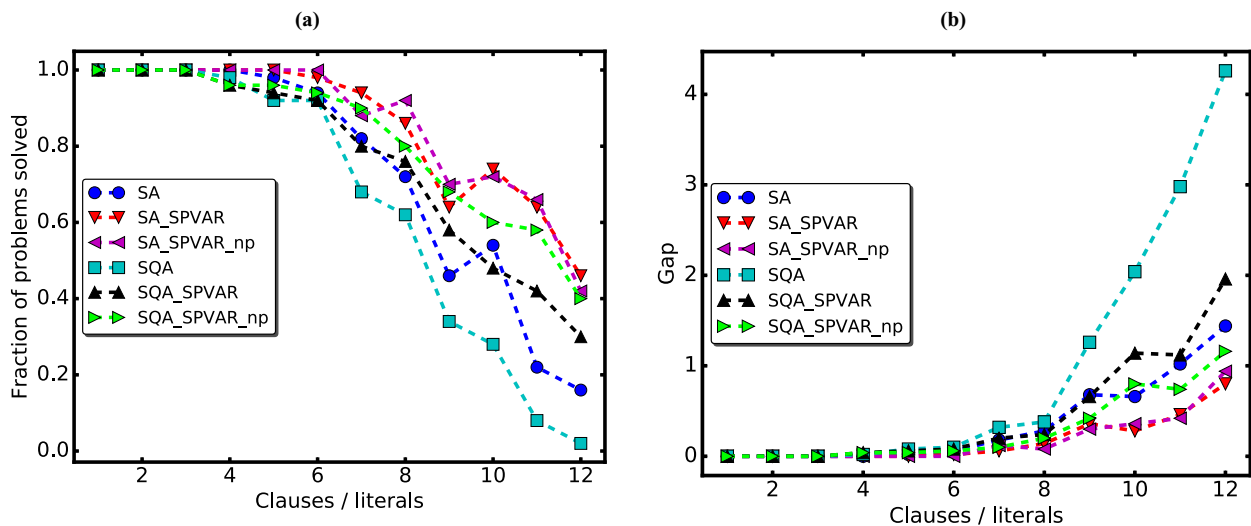


FIG. 12. Success metrics for Max-3-SAT problems with 50 literals. (a) The fraction of problems solved, as a function of the number of clauses. (b) The median gap, which is the energy difference between the best solution found and the best known solution, as a function of the number of clauses. For clarity, error bars are not shown. In both figures, each point represents data from 50 random instances with the respective number of clauses.

unintuitive, due to the observation that in this regime samples are expected to contain states with low overlap. However, recent work on SQA [66] and the experimental results on the D-Wave quantum annealer [67] suggest that their samples are biased towards few of the ground states, which might help to explain this. Second, it has been argued that higher-order polynomials provide a rougher landscape, which could lead to an advantage for SQA over SA, at least if the barriers are thin [22].

IV. CONCLUSIONS AND FUTURE WORK

We set out two objectives for this work: to study the performance of SPVAR on hard problems with different samplers, and to study what classes of problems or samplers might not benefit from the application of our method. The results show that for almost all problems studied, even extremely hard problems, the application of our method results in a significant improvement in the success metrics, for all samplers studied.

In general, SPVAR can result in a reduction in the number of variables. In some cases, the reduction of variables might also result in a computationally simpler problem (e.g., a simplified topology). Therefore, when comparing the numerical effort as a function of the size of the input of a particular heuristic to the same heuristic with SPVAR, an advantage in the scaling can occur via an improved exponential (due to the simplification of a problem) or a reduced prefactor (due to the reduction of the number of variables). Figure 5 depicts a case where the scaling of the algorithm is improved via the addition of SPVAR. A characterization of the conditions and mechanism under which SPVAR can result in an improvement in scaling would be an interesting future study, however, larger problems would be needed to carefully probe the asymptotic regime.

We have also identified several types of problems that appear to challenge SPVAR. First, zero-bias problems tend to have large correlation lengths, such that clusters of spins cannot be fixed locally. The reduced-degeneracy Chimera graph problems with zero bias fall under this category. As shown in Sec. III C, setting the *elite_threshold* adaptively can help with this. It is also possible to iteratively apply SPVAR sequentially, which leads to the problems gradually accumulating increasing numbers of variables with nonzero bias, which makes it easier to fix variables, as shown in [33].

Second, problems in which the “cost” in value due to an incorrectly fixed variable is large compared with the range of couplers. Increasing *num_starts* can help with this, by reducing the risk of fixing variables incorrectly. Finally, if a sampler

gives a low-quality result, it is possible that the application of our method will not improve the results, although it is not expected to make them worse. By setting *elite_threshold* adaptively, it appears to be possible to improve the results, although the risk of fixing variables incorrectly in this case could be high.

In the future, it might be worthwhile to study the performance of a multistart iterative application of SPVAR, as well as the possible application of a similar algorithm to other discrete optimization problems, such as mixed-integer programming.

ACKNOWLEDGMENTS

The authors would like to thank Z. Zhu for the use of his implementation of the *borealis* algorithm (PTICM) [40], Matthias Troyer’s group for the use of their implementation of the simulated annealing (SA) algorithm [34], and A. Selby for the use of his implementation of the Hamze–de Freitas–Selby (HFS) algorithm, available for public use on GitHub [41]. We would also like to thank A. Abramé and D. Habet for the use of their Max-*k*-SAT solver *ahmaxsat* [43] and C. Luo for the use of the Max-*k*-SAT solver CCLS [42]. In addition, we would like to thank W. Wang for supplying us with the 3D spin glass instances, S. Boixo for supplying us with the weak-strong cluster problem instances, and A. Perdomo-Ortiz for providing us with the fault diagnosis problem instances. This work was supported by IQBit and Mitacs. The research of H.G.K. was supported by the National Science Foundation (Grant No. DMR-1151387) and is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via MIT Lincoln Laboratory Air Force Contract No. FA8721-05-C-0002. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purpose notwithstanding any copyright annotation thereon.

APPENDIX: PARAMETERS USED

In Table II, we list the parameter values used for the benchmarking, for each problem set. In addition, *fixing_threshold* was always set to 1.0, as was *correlation_threshold*, and the chain repairing method used for quantum annealer runs was

TABLE II. Parameter values of multistart SPVAR and the various problem sets used in the benchmarking.

Problems	<i>elite_thresholds</i>	<i>num_starts</i>	<i>total_sample_size</i>	<i>num_sweeps</i>
Weak-strong	0.2, 0.2	20	100–500	20×10^3
Red. deg. Chimera	0.2, 0.2	20	1000	20×10^3
3D spin glass	0.2, 0.3	20	500	20×10^3
3D spin glass (PTICM)	0.2, —	10	500	10×10^3
Fault diagnosis	0.1, 0.4	40	500	$2, 20 \times 10^3$
Max- <i>k</i> -SAT	0.2, 0.2	40	500	20×10^3

local energy minimization. The thresholds column shows first the *elite_threshold* and then the *correlation_ elite_threshold*.

For the nonzero-bias problems, we used half of the *total_sample_size* for the SPVAR fixing step, and half of the

sample size to solve the problems. For the zero-bias problems, we used 40% of the total sample size for the correlation-based prefixing step, and divided the rest equally between the fixing step and solving step.

-
- [1] A. Lucas, Ising formulations of many NP problems, *Front. Phys.* **2**, 5 (2014).
- [2] A. E. Eiben, P.-E. Raue, and Z. Ruttkay, in *International Conference on Parallel Problem Solving from Nature* (Springer, Berlin, 1994), pp. 78–87.
- [3] A. E. Eiben, C. H. van Kemenade, and J. N. Kok, in *European Conference on Artificial Life* (Springer, Berlin, 1995), pp. 934–945.
- [4] J. Lis and A. E. Eiben, in *IEEE International Conference on Evolutionary Computation, 1997* (IEEE, Piscataway, NJ, 1997), pp. 59–64.
- [5] C.-K. Ting, in *European Conference on Artificial Life* (Springer, Berlin, 2005), pp. 403–412.
- [6] Y. Wang, Z. Lü, F. Glover, and J.-K. Hao, Effective variable fixing and scoring strategies for binary quadratic programming, in *Evolutionary Computation in Combinatorial Optimization* (Springer, Cham, 2011), pp. 72–83.
- [7] Y. Wang, Z. Lü, F. Glover, and J.-K. Hao, Backbone guided tabu search for solving the UBQP problem, *J. Heuristics* **19**, 679 (2013).
- [8] W. Zhang, Configuration landscape analysis and backbone guided local search.: Part I: Satisfiability and maximum satisfiability, *Artif. Intell.* **158**, 1 (2004).
- [9] F. Glover, Tabu search—Part I, *ORSA J. Comput.* **1**, 190 (1989).
- [10] F. Glover, Tabu search—Part II, *ORSA J. Comput.* **2**, 4 (1990).
- [11] P. Chardaire, J. L. Lutton, and A. Sutter, Thermostatistical persistence: A powerful improving concept for simulated annealing algorithms, *Eur. J. Operat. Res.* **86**, 565 (1995).
- [12] M. Johnson, M. Amin, S. Gildert, T. Lanting, F. Hamze, N. Dickson, R. Harris, A. Berkley, J. Johansson, P. Bunyk *et al.*, Quantum annealing with manufactured spins, *Nature (London)* **473**, 194 (2011).
- [13] P. I. Bunyk, E. M. Hoskinson, M. W. Johnson, E. Tolkacheva, F. Altomare, A. J. Berkley, R. Harris, J. P. Hilton, T. Lanting, A. J. Przybysz *et al.*, Architectural considerations in the design of a superconducting quantum annealing processor, *IEEE Trans. Appl. Supercond.* **24**, 1 (2014).
- [14] P. Ray, B. K. Chakrabarti, and A. Chakrabarti, Sherrington-Kirkpatrick model in a transverse field: Absence of replica symmetry breaking due to quantum fluctuations, *Phys. Rev. B* **39**, 11828 (1989).
- [15] A. Finnila, M. Gomez, C. Sebenik, C. Stenson, and J. Doll, Quantum annealing: A new method for minimizing multidimensional functions, *Chem. Phys. Lett.* **219**, 343 (1994).
- [16] T. Kadowaki and H. Nishimori, Quantum annealing in the transverse Ising model, *Phys. Rev. E* **58**, 5355 (1998).
- [17] G. E. Santoro, R. Martoňák, E. Tosatti, and R. Car, Theory of quantum annealing of an Ising spin glass, *Science* **295**, 2427 (2002).
- [18] D. A. Battaglia, G. E. Santoro, and E. Tosatti, Optimization by quantum annealing: Lessons from hard satisfiability problems, *Phys. Rev. E* **71**, 066707 (2005).
- [19] B. Heim, T. F. Rønnow, S. V. Isakov, and M. Troyer, Quantum versus classical annealing of Ising spin glasses, *Science* **348**, 215 (2015).
- [20] S. Muthukrishnan, T. Albash, and D. A. Lidar, Tunneling and Speedup in Quantum Optimization for Permutation-Symmetric Problems, *Phys. Rev. X* **6**, 031010 (2016).
- [21] S. Boixo, T. F. Rønnow, S. V. Isakov, Z. Wang, D. Wecker, D. A. Lidar, J. M. Martinis, and M. Troyer, Evidence for quantum annealing with more than one hundred qubits, *Nat. Phys.* **10**, 218 (2014).
- [22] V. S. Denchev, S. Boixo, S. V. Isakov, N. Ding, R. Babbush, V. Smelyanskiy, J. Martinis, and H. Neven, What is the Computational Value of Finite-Range Tunneling? *Phys. Rev. X* **6**, 031015 (2016).
- [23] C. C. McGeoch and C. Wang, in *Proceedings of the ACM International Conference on Computing Frontiers* (ACM, New York, 2013), p. 23.
- [24] H. G. Katzgraber, F. Hamze, and R. S. Andrist, Glassy Chimeras Could be Blind to Quantum Speedup: Designing Better Benchmarks for Quantum Annealing Machines, *Phys. Rev. X* **4**, 021008 (2014).
- [25] A. D. King, T. Lanting, and R. Harris, Performance of a quantum annealer on range-limited constraint satisfaction problems, [arXiv:1502.02098](https://arxiv.org/abs/1502.02098).
- [26] J. King, S. Yarkoni, M. M. Nevisi, J. P. Hilton, and C. C. McGeoch, Benchmarking a quantum annealing processor with the time-to-target metric, [arXiv:1508.05087](https://arxiv.org/abs/1508.05087).
- [27] S. Mandrà, Z. Zhu, W. Wang, A. Perdomo-Ortiz, and H. G. Katzgraber, Strengths and weaknesses of weak-strong cluster problems: A detailed overview of state-of-the-art classical heuristics versus quantum approaches, *Phys. Rev. A* **94**, 022337 (2016).
- [28] T. F. Rønnow, Z. Wang, J. Job, S. Boixo, S. V. Isakov, D. Wecker, J. M. Martinis, D. A. Lidar, and M. Troyer, Defining and detecting quantum speedup, *Science* **345**, 420 (2014).
- [29] I. Hen, J. Job, T. Albash, T. F. Rønnow, M. Troyer, and D. A. Lidar, Probing for quantum speedup in spin-glass problems with planted solutions, *Phys. Rev. A* **92**, 042325 (2015).
- [30] H. G. Katzgraber, F. Hamze, Z. Zhu, A. J. Ochoa, and H. Muñoz-Bauza, Seeking Quantum Speedup Through Spin Glasses: The Good, the Bad, and the Ugly, *Phys. Rev. X* **5**, 031026 (2015).
- [31] J. King, S. Yarkoni, J. Raymond, I. Ozfidan, A. D. King, M. M. Nevisi, J. P. Hilton, and C. C. McGeoch, Quantum annealing amid local ruggedness and global frustration, [arXiv:1701.04579](https://arxiv.org/abs/1701.04579).
- [32] S. Mandrà, H. G. Katzgraber, and C. Thomas, The pitfalls of planar spin-glass benchmarks: Raising the bar for quantum annealers (again), *Quantum Sci. Technol.* **2**, 038501 (2017).
- [33] H. Karimi and G. Rosenberg, Boosting quantum annealer performance via sample persistence, *Quantum Inf. Process.* **16**, 166 (2017).

- [34] S. V. Isakov, I. N. Zintchenko, T. F. Rønnow, and M. Troyer, Optimised simulated annealing for Ising spin glasses, *Comput. Phys. Commun.* **192**, 265 (2015).
- [35] R. Martoňák, G. E. Santoro, and E. Tosatti, Quantum annealing by the path-integral Monte Carlo method: The two-dimensional random Ising model, *Phys. Rev. B* **66**, 094203 (2002).
- [36] H. G. Katzgraber, Introduction to Monte Carlo methods, [arXiv:0905.1629](https://arxiv.org/abs/0905.1629).
- [37] C. J. Geyer and E. A. Thompson, Constrained Monte Carlo maximum likelihood for dependent data, *J. R. Stat. Soc. Ser. B* **54**, 657 (1992).
- [38] K. Hukushima and K. Nemoto, Exchange Monte Carlo method and application to spin glass simulations, *J. Phys. Soc. Jpn.* **65**, 1604 (1996).
- [39] Z. Zhu, A. J. Ochoa, and H. G. Katzgraber, Efficient Cluster Algorithm for Spin Glasses in Any Space Dimension, *Phys. Rev. Lett.* **115**, 077201 (2015).
- [40] Z. Zhu, C. Fang, and H. G. Katzgraber, borealis—A generalized global update algorithm for Boolean optimization problems, [arXiv:1605.09399](https://arxiv.org/abs/1605.09399).
- [41] A. Selby, QUBO-Chimera, <https://github.com/alex1770/QUBO-Chimera>
- [42] C. Luo, S. Cai, W. Wu, Z. Jie, and K. Su, CCLS: An efficient local search algorithm for weighted maximum satisfiability, *IEEE Trans. Comput.* **64**, 1830 (2015).
- [43] A. Abramé and D. Habet, AHMAXSAT: Description and evaluation of a branch and bound Max-SAT solver, *J. Satisfiab., Boolean Model. Computat.* **9**, 89 (2015).
- [44] W. Wang, S. Mandrà, and H. G. Katzgraber, Patch-planting spin-glass solution for benchmarking, *Phys. Rev. E* **96**, 023312 (2017).
- [45] S. Boixo, V. N. Smelyanskiy, A. Shabani, S. V. Isakov, M. Dykman, V. S. Denchev, M. H. Amin, A. Y. Smirnov, M. Mohseni, and H. Neven, Computational multiqubit tunneling in programmable quantum annealers, *Nat. Commun.* **7**, 10327 (2016).
- [46] The chip at our disposal has 1100 active qubits, a working temperature of 26 ± 5 mK, and a minimum annealing time of 20 μ s.
- [47] The chip at our disposal has 2023 active qubits, a working temperature of 14 ± 1 mK, and a minimum annealing time of 5 μ s.
- [48] The R99 is the number of annealing cycles required to find the best known energy with a confidence of 99%.
- [49] W. Wang, J. Machta, and H. G. Katzgraber, Comparing Monte Carlo methods for finding ground states of Ising spin glasses: Population annealing, simulated annealing, and parallel tempering, *Phys. Rev. E* **92**, 013303 (2015).
- [50] W. Wang, J. Machta, and H. G. Katzgraber, Population annealing: Theory and application in spin glasses, *Phys. Rev. E* **92**, 063307 (2015).
- [51] A. Perdomo-Ortiz, J. Fluegemann, S. Narasimhan, R. Biswas, and V. N. Smelyanskiy, A quantum annealing approach for fault detection and diagnosis of graph-based systems, *Eur. Phys. J. Special Topics* **224**, 131 (2015).
- [52] Z. Bian, F. Chudak, R. Israel, B. Lackey, W. G. Macready, and A. Roy, Mapping constrained optimization problems to quantum annealing with application to fault diagnosis, [arXiv:1603.03111](https://arxiv.org/abs/1603.03111).
- [53] A. Perdomo-Ortiz, A. Feldman, A. Ozaeta, S. V. Isakov, Z. Zhu, B. O’Gorman, H. G. Katzgraber, A. Diedrich, H. Neven, J. de Kleer *et al.*, On the readiness of quantum optimization machines for industrial applications, [arXiv:1708.09780](https://arxiv.org/abs/1708.09780).
- [54] W. Vinci, T. Albash, G. Paz-Silva, I. Hen, and D. A. Lidar, Quantum annealing correction with minor embedding, *Phys. Rev. A* **92**, 042310 (2015).
- [55] V. Choi, Minor-embedding in adiabatic quantum computation: I. The parameter setting problem, *Quantum Inf. Process.* **7**, 193 (2008).
- [56] V. Choi, Minor-embedding in adiabatic quantum computation: II. Minor-universal graph design, *Quantum Inf. Process.* **10**, 343 (2011).
- [57] A. Zaribafiyani, D. J. J. Marchand, and S. S. C. Rezaei, Systematic and deterministic graph minor embedding for Cartesian products of graphs, *Quantum Inf. Process.* **16**, 136 (2017).
- [58] R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman, and L. Troyansky, Determining computational complexity from characteristic ‘phase transitions’, *Nature (London)* **400**, 133 (1999).
- [59] D. Achlioptas, A. Chtcherba, G. Istrate, and C. Moore, The phase transition in 1-in-k SAT and NAE 3-SAT, in *Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms* (ACM Press, New York, 2001), Vol. 1, pp. 721–722.
- [60] V. Chvátal and B. Reed, in *Proceedings, 33rd Annual Symposium on Foundations of Computer Science, 1992* (IEEE, Piscataway, NJ, 1992), pp. 620–627.
- [61] W. F. de La Vega, On random 2-SAT (unpublished).
- [62] A. Goerdt, A threshold for unsatisfiability, *J. Comput. Syst. Sci.* **53**, 469 (1996).
- [63] M. Mézard and R. Zecchina, The random K -satisfiability problem: From an analytic solution to an efficient algorithm, *Phys. Rev. E* **66**, 056126 (2002).
- [64] M. Mézard, G. Parisi, and R. Zecchina, Analytic and algorithmic solution of random satisfiability problems, *Science* **297**, 812 (2002).
- [65] S. Mertens, M. Mézard, and R. Zecchina, Threshold values of random K -SAT from the cavity method, *Random Struct. Alg.* **28**, 340 (2006).
- [66] Y. Matsuda, H. Nishimori, and H. G. Katzgraber, Ground-state statistics from annealing algorithms: Quantum versus classical approaches, *New J. Phys.* **11**, 073021 (2009).
- [67] S. Mandrà, Z. Zhu, and H. G. Katzgraber, Exponentially Biased Ground-State Sampling of Quantum Annealing Machines with Transverse-Field Driving Hamiltonians, *Phys. Rev. Lett.* **118**, 070502 (2017).