



The potential of quantum annealing for rapid solution structure identification

Yuchen Pang¹ · Carleton Coffrin² · Andrey Y. Likhov² · Marc Vuffray²

Accepted: 15 October 2020 / Published online: 18 November 2020
© The Author(s) 2021, corrected publication 2021

Abstract

The recent emergence of novel computational devices, such as quantum computers, coherent Ising machines, and digital annealers presents new opportunities for hardware-accelerated hybrid optimization algorithms. Unfortunately, demonstrations of unquestionable performance gains leveraging novel hardware platforms have faced significant obstacles. One key challenge is understanding the algorithmic properties that distinguish such devices from established optimization approaches. Through the careful design of contrived optimization tasks, this work provides new insights into the computation properties of quantum annealing and suggests that this model has the potential to quickly identify the structure of high-quality solutions. A meticulous comparison to a variety of algorithms spanning both complete and local search suggests that quantum annealing's performance on the proposed optimization tasks is distinct. This result provides new insights into the time scales and types of optimization problems where quantum annealing has the potential to provide notable performance gains over established optimization algorithms and suggests the development of hybrid algorithms that combine the best features of quantum annealing and state-of-the-art classical approaches.

Keywords Discrete optimization · Ising model · Quadratic unconstrained binary optimization · Local search · Quantum annealing · Large neighborhood search · Integer programming · Belief propagation

1 Introduction

As the challenge of scaling traditional transistor-based Central Processing Unit (CPU) technology continues to increase, experimental physicists and high-tech companies have begun

This article belongs to the Topical Collection: *Special Issue on Constraint Programming, Artificial Intelligence, and Operations Research*
Guest Editors: Emmanuel Hebrard and Nysret Musliu

✉ Carleton Coffrin
cjc@lanl.gov

Extended author information available on the last page of the article.

to explore radically different computational technologies, such as quantum computers [14, 41, 62], quantum annealers [43, 45] and coherent Ising machines [40, 47, 59]. The goal of all of these technologies is to leverage the dynamical evolution of a physical system to perform a computation that is challenging to emulate using traditional CPU technology, the most notable example being the simulation of quantum physics [29]. Despite their entirely disparate physical implementations, optimization of quadratic functions over binary variables (e.g., the Quadratic Unconstrained Binary Optimization (QUBO) and Ising models [13]) has emerged as a challenging computational task that a wide variety of novel hardware platforms can address. As these technologies mature, it may be possible for this specialized hardware to rapidly solve challenging combinatorial problems, such as Max-Cut [38] or Max-Clique [53], and preliminary studies have suggested that some classes of Constraint Satisfaction Problems can be effectively encoded in such devices because of their combinatorial structure [8, 9, 67, 72].

At this time, understanding the computational advantage that these hardware platforms may bring to established optimization algorithms remains an open question. For example, it is unclear if the primary benefit will be dramatically reduced runtimes due to highly specialized hardware implementations [31, 76, 77] or if the behavior of the underlying analog computational model will bring intrinsic algorithmic advantages [3, 26]. A compelling example is gate-based quantum computation (QC), where a significant body of theoretical work has found key computational advantages that exploit quantum properties [18, 34, 71]. Indeed, such advantages have recently been demonstrated on quantum computing hardware for the first time [5]. Highlighting similar advantages on other computational platforms, both in theory and in practice, remains a central challenge for novel physics-inspired computing models [36, 46, 51].

Focusing on quantum annealing (QA), this work provides new insights on the properties of this computing model and identifies problem structures where it can provide a computational advantage over a broad range of established solution methods. The central contribution of this work is the analysis of *tricky* optimization problems (i.e., Biased Ferromagnets, Frustrated Biased Ferromagnets, and Corrupted Biased Ferromagnets) that are challenging for established optimization approaches but are easy for QA hardware, such as D-Wave's 2000Q platform. This result suggests that there are classes of optimization problems where QA can effectively identify global solution structure while established heuristics struggle to escape local minima. Two auxiliary contributions that resulted from this pursuit are the identification of the Corrupted Biased Ferromagnet problem, which appears to be a useful benchmark problem beyond this particular study, and demonstration of the most significant performance gains of a quantum annealing platform to the established state-of-the-art alternatives, to the best of our knowledge.

This work begins with a brief introduction to both the mathematical foundations of the Ising model, Section 2, and quantum annealing, Section 3. It then reviews a variety of algorithms than can be used to solve such models in Section 4. The primary result of the paper is presented in carefully designed structure detection experiments in Section 5. Open challenges relating to developing hybrid algorithms are discussed in Section 6, and Section 7 concludes the paper.

2 A brief introduction to ising models

This section introduces the notations of the paper and provides a brief introduction to Ising models, a core mathematical abstraction of QA. The Ising model refers to the class

of graphical models where the nodes, $\mathcal{N} = \{1, \dots, N\}$, represent *spin* variables (i.e., $\sigma_i \in \{-1, 1\} \forall i \in \mathcal{N}$), and the edges, $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$, represent pairwise *interactions* of spin variables (i.e., $\sigma_i \sigma_j \forall i, j \in \mathcal{E}$). A local *field* $\mathbf{h}_i \forall i \in \mathcal{N}$ is specified for each node, and an interaction strength $\mathbf{J}_{ij} \forall i, j \in \mathcal{E}$ is specified for each edge. The *energy* of the Ising model is then defined as:

$$E(\sigma) = \sum_{i,j \in \mathcal{E}} \mathbf{J}_{ij} \sigma_i \sigma_j + \sum_{i \in \mathcal{N}} \mathbf{h}_i \sigma_i \tag{1}$$

Originally introduced in statistical physics as a model for describing phase transitions in ferromagnetic materials [32], the Ising model is currently used in numerous and diverse application fields such as neuroscience [39, 68], bio-polymers [63], gene regulatory networks [55], image segmentation [64], statistical learning [52, 74, 75], and sociology [25].

This work focuses on finding the lowest possible energy of the Ising model, known as a *ground state*, that is, finding the globally optimal solution of the following discrete optimization problem:

$$\begin{aligned} \min : & E(\sigma) \\ \text{s.t.:} & \sigma_i \in \{-1, 1\} \forall i \in \mathcal{N} \end{aligned} \tag{2}$$

The coupling parameters of Ising models are categorized into two groups based on their sign: the *ferromagnetic* interactions $\mathbf{J}_{ij} < 0$, which encourage neighboring spins to take the same value, i.e., $\sigma_i \sigma_j = 1$, and *anti-ferromagnetic* interactions $\mathbf{J}_{ij} > 0$, which encourage neighboring spins to take opposite values, i.e., $\sigma_i \sigma_j = -1$.

Frustration The notion of frustration is central to the study of Ising models and refers to any instance of (2) where the optimal solution does not achieve the minimum of all local interactions [19]. Namely, the optimal solution of a frustrated Ising model, σ^* , satisfies the following property:

$$E(\sigma^*) > \sum_{i,j \in \mathcal{E}} -|\mathbf{J}_{ij}| - \sum_{i \in \mathcal{N}} |\mathbf{h}_i| \tag{3}$$

Gauge Transformations A valuable property of the Ising model is the gauge transformation, which characterizes an equivalence class of Ising models. Consider the optimal solution of Ising model S , σ^s . One can construct a new Ising model T where the optimal solution is the target state σ^t by applying the following parameter transformation:

$$\mathbf{J}_{ij}^t = \mathbf{J}_{ij}^s \sigma_i^s \sigma_j^s \sigma_i^t \sigma_j^t \forall i, j \in \mathcal{E} \tag{4a}$$

$$\mathbf{h}_i^t = \mathbf{h}_i^s \sigma_i^s \sigma_i^t \forall i \in \mathcal{N} \tag{4b}$$

This S -to- T manipulation is referred to as a gauge transformation. Using this property, one can consider the class of Ising models where the optimal solution is $\sigma_i = -1 \forall i \in \mathcal{N}$ or any arbitrary vector of $-1, 1$ values without loss of generality.

Classes of Ising Models Ising models are often categorized by the properties of their optimal solutions with two notable categories being Ferromagnets (FM) and Spin glasses. Ferromagnetic Ising models are unfrustrated models possessing one or two optimal solutions. The traditional FM model is obtained by setting $\mathbf{J}_{ij} = -1, \mathbf{h}_i = 0$. The optimal

solutions have a structure with all spins *pointing* in the same direction, i.e., $\sigma_i = 1$ or $\sigma_i = -1$, which mimics the behavior of physical magnets at low temperatures. In contrast to FMs, Spin glasses are highly frustrated systems that exhibit an intricate geometry of optimal solutions that tend to take the form of a hierarchy of isosceles sets [61]. Spin glasses are challenging for greedy and local search algorithms [7] due to the nature of their energy landscape [24, 60]. A typical Spin glass instance can be achieved using random interactions graphs with $P(\mathbf{J}_{ij} = -1) = 0.5$, $P(\mathbf{J}_{ij} = 1) = 0.5$, and $\mathbf{h}_i = 0$.

Bijection of Ising and Boolean Optimization It is valuable to observe that there is a bijection between Ising optimization (i.e., $\sigma \in \{-1, 1\}$) and Boolean optimization (i.e., $x \in \{0, 1\}$). The transformation of σ -to- x is given by:

$$\sigma_i = 2x_i - 1 \quad \forall i \in \mathcal{N} \quad (5a)$$

$$\sigma_i \sigma_j = 4x_i x_j - 2x_i - 2x_j + 1 \quad \forall i, j \in \mathcal{E} \quad (5b)$$

and the inverse x -to- σ is given by:

$$x_i = \frac{\sigma_i + 1}{2} \quad \forall i \in \mathcal{N} \quad (6a)$$

$$x_i x_j = \frac{\sigma_i \sigma_j + \sigma_i + \sigma_j + 1}{4} \quad \forall i, j \in \mathcal{E} \quad (6b)$$

Consequently, any results from solving Ising models are also immediately applicable to the class of optimization problems referred to as Pseudo-Boolean Optimization or Quadratic Unconstrained Binary Optimization (QUBO):

$$\begin{aligned} \min : & \sum_{i,j \in \mathcal{E}} c_{ij} x_i x_j + \sum_{i \in \mathcal{N}} c_i x_i + \mathbf{c} \\ \text{s.t.} : & x_i \in \{0, 1\} \quad \forall i \in \mathcal{N} \end{aligned} \quad (7)$$

In contrast to gate-based QC, which is Turing complete, QA specializes in optimizing Ising models. The next section provides a brief introduction of how quantum mechanics are leveraged by QA to perform Ising model optimization.

3 Foundations of quantum annealing

Quantum annealing is an analog computing technique for minimizing discrete or continuous functions that takes advantage of the *exotic* properties of quantum systems. This technique is particularly well-suited for finding optimal solutions of Ising models and has drawn significant interest due to hardware realizations via controllable quantum dynamical systems [43]. Quantum annealing is composed of two key elements: leveraging quantum state to lift the minimization problem into an exponentially larger space, and slowly interpolating (i.e., annealing) between an initial easy problem and the target problem. The quantum lifting begins by introducing for each spin $\sigma_i \in \{-1, 1\}$ a $2^N \times 2^N$ dimensional matrix $\widehat{\sigma}_i$ expressible as a Kronecker product of N matrices of dimension 2×2 :

$$\widehat{\sigma}_i = \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \cdots \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}_{1 \text{ to } i-1} \otimes \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}}_{i^{\text{th}} \text{ term}} \otimes \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \cdots \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}_{i+1 \text{ to } N} \quad (8)$$

In this lifted representation, the value of a spin σ_i is identified with the two possible eigenvalues 1 and -1 of the matrix $\widehat{\sigma}_i$. The quantum counterpart of the energy function defined in

(1) is the $2^N \times 2^N$ matrix obtained by substituting spins with the $\hat{\sigma}$ matrices in the algebraic expression of the energy:

$$\hat{E} = \sum_{i,j \in \mathcal{E}} J_{ij} \hat{\sigma}_i \hat{\sigma}_j + \sum_{i \in \mathcal{N}} h_i \hat{\sigma}_i \tag{9}$$

Notice that the eigenvalues of the matrix in (9) are the 2^N possible energy values obtained by evaluating the energy $E(\sigma)$ from (1) for all possible configurations of spins. This implies that finding the lowest eigenvalue of \hat{E} is tantamount to solving the minimization problem in (2). This lifting is clearly impractical from the classical computing context as it transforms a minimization problem over 2^N configurations into computing the minimum eigenvalue of a $2^N \times 2^N$ matrix. The key motivation for this approach is that it is possible to construct quantum systems with only N quantum bits that attempt to find the minimum eigenvalue of this matrix.

The annealing process provides a way of steering a quantum system into the a priori unknown eigenvector that minimizes the energy of (9) [28, 45]. The core idea is to initialize the quantum system at the minimal eigenvector of a simple energy matrix \hat{E}_0 , for which an explicit formula is known. After the system is initialized, the energy matrix is interpolated from the easy problem to the target problem slowly over time. Specifically, the energy matrix at a point during the anneal is given by $\hat{E}_a(\Gamma) = (1 - \Gamma)\hat{E}_0 + \Gamma\hat{E}$, with Γ varying from 0 to 1. When the anneal is complete, $\Gamma = 1$ and the interactions in the quantum system are described by the target energy matrix. The annealing time is the physical time taken by the system to evolve from $\Gamma = 0$ to $\Gamma = 1$. For suitable starting energy matrices \hat{E}_0 and a sufficiently slow annealing time, theoretical results have demonstrated that a quantum system continuously remains at the minimal eigenvector of the interpolating matrix $\hat{E}_a(\Gamma)$ [3] and therefore achieves the minimum energy (i.e., a global optima) of the target problem. Realizing this optimality result in practice has proven difficult due to corruption of the quantum system from the external environment. Nevertheless, quantum annealing can serve as a heuristic for finding high-quality solutions to the Ising models, i.e., (2).

3.1 Quantum annealing hardware

Interest in the QA model is due in large part to D-Wave Systems, which has developed the first commercially available QA hardware platform [43]. Given the computational challenges of classically simulating QA, this novel computing device represents the only viable method for studying QA at non-trivial scales, e.g., problems with more than 1000 qubits [11, 22]. At the most basic level, the D-Wave platform allows the user to program an Ising model by providing the parameters \mathbf{J}, \mathbf{h} in (1) and returns a collection of variable assignments from multiple annealing runs, which reflect optimal or near-optimal solutions to the input problem.

This seemingly simple interface is, however, hindered by a variety of constraints imposed by D-Wave’s 2000Q hardware implementation. The most notable hardware restriction is the *Chimera* connectivity graph depicted in Fig. 1, where each edge indicates if the hardware supports a coupling term J_{ij} between a pair of qubits i and j . This sparse graph is a stark contrast to traditional quadratic optimization tools, where it is assumed that every pair of variables can interact.

The second notable hardware restriction is a limited coefficient programming range. On the D-Wave 2000Q platform the parameters are constrained within the continuous parameter ranges of $-1 \leq J_{ij} \leq 1$ and $-2 \leq h_i \leq 2$. At first glance these ranges may

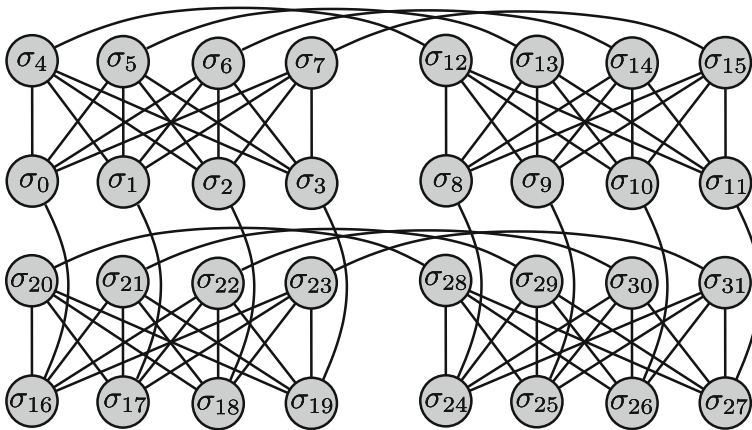


Fig. 1 A 2-by-2 Chimera graph illustrating the variable product limitations of D-Wave's 2000Q processor

not appear to be problematic because the energy function (1) can be rescaled into the hardware's operating range without any loss of generality. However, operational realities of analog computing devices make the parameter values critically important to the overall performance of the hardware. These challenges include: persistent coefficient biases, which are an artifact of hardware slowly drifting out of calibration between re-calibration cycles; programming biases, which introduce some minor errors in the \mathbf{J} , \mathbf{h} values that were requested; and environmental noise, which disrupts the quantum behavior of the hardware and results in a reduction of solution quality. Overall, these hardware constraints have made the identification of QA-based performance gains notoriously challenging [16, 42, 54, 58, 65].

Despite the practical challenges in using D-Wave's hardware platform, extensive experiments have suggested that QA can outperform some established local search methods (e.g., simulated annealing) on carefully designed Ising models [4, 22, 49]. However, demonstrating an unquestionable computational advantage over state-of-the-art methods on contrived and practical problems remains an open challenge.

4 Methods for ising model optimization

The focus of this work is to compare and contrast the behavior of QA to a broad range of established optimization algorithms. To that end, this work considers three core algorithmic categories: (1) *complete search* methods from the mathematical programming community; (2) *local search* methods developed by the statistical physics community; and (3) *quantum annealing* as realized by D-Wave's hardware platform. The comparison includes both state-of-the-art solution methods from the D-Wave benchmarking literature (e.g., Hamze-Freitas-Selby [69], Integer Linear Programming [16]) and simple straw-man approaches (e.g., Greedy, Glauber Dynamics [33], Min-Sum [30, 60]) to highlight the solution quality of minimalist optimization approaches. This section provides high-level descriptions of the algorithms; implementation details are available as open-source software [17, 69].

4.1 Complete search

Unconstrained Boolean optimization, as in (7), has been the subject of mathematical programming research for several decades [10, 12]. This work considers the two most canonical formulations based on Integer Quadratic Programming and Integer Linear Programming.

Integer Quadratic Programming (IQP) This formulation consists of using black-box commercial optimization tools to solve (7) directly. This model was leveraged in some of the first QA benchmarking studies [58] and received some criticism [66]. However, the results presented here suggest that this model has become more competitive due to the steady progress of commercial optimization solvers.

Integer Linear Programming (ILP) This formulation is a slight variation of the IQP model where the variable products $x_i x_j$ are lifted into a new variable x_{ij} and constraints are added to capture the conjunction $x_{ij} = x_i \wedge x_j$ as follows:

$$\min : \sum_{i,j \in \mathcal{E}} c_{ij} x_{ij} + \sum_{i \in \mathcal{N}} c_i x_i + c \tag{10a}$$

s.t.:

$$x_{ij} \geq x_i + x_j - 1, \quad x_{ij} \leq x_i, \quad x_{ij} \leq x_j \quad \forall i, j \in \mathcal{E} \tag{10b}$$

$$x_i \in \{0, 1\} \quad \forall i \in \mathcal{N}, \quad x_{ij} \in \{0, 1\} \quad \forall i, j \in \mathcal{E}$$

This formulation was also leveraged in some of the first QA benchmarking studies [20, 66] and [10], which suggest this is the best formulation for sparse graphs, as is the case with the D-Wave Chimera graph. However, this work indicates that IQP solvers have improved sufficiently and this conclusion should be revisited.

4.2 Local search

Although complete search algorithms are helpful in the validation of QA hardware [6, 16], it is broadly accepted that local search algorithms are the most appropriate point of computational comparison to QA methods [1]. Given that a comprehensive enumeration of local search methods would be a monumental undertaking, this work focuses on representatives from four distinct algorithmic categories including greedy, message passing, Markov Chain Monte Carlo, and large neighborhood search.

Greedy (GRD) The first heuristic algorithm considered by this work is a Steepest Coordinate Decent (SCD) greedy initialization approach. This algorithm assigns the variables one-by-one, always taking the assignment that minimizes the objective value. Specifically, the SCD approach begins with unassigned values, i.e., $\sigma_i = 0 \quad \forall i \in \mathcal{N}$, and then repeatedly applies the following assignment rule until all of the variables have been assigned a value of -1 or 1 :

$$i, v = \underset{i \in \mathcal{N}, v \in \{-1, 1\}}{\operatorname{argmin}} E(\sigma_1, \dots, \sigma_{i-1}, v, \sigma_{i+1}, \dots, \sigma_N) \tag{11a}$$

$$\sigma_i = v \tag{11b}$$

In each application, ties in the argmin are broken at random, giving rise to a potentially stochastic outcome of the heuristic. Once all of the variables have been assigned, the algorithm is repeated until a runtime limit is reached and only the best solution found is returned.

Although this approach is very simple, it can be effective in Ising models with minimal amounts of frustration.

Message Passing (MP) The second algorithm considered by this work is a message-based Min-Sum (MS) algorithm [30, 60], which is an adaptation of the celebrated Belief Propagation algorithm for solving minimization problems on networks. A key property of the MS approach is its ability to identify the global minimum of cost functions with a tree dependency structure between the variables; i.e., if no cycles are formed by the interactions in \mathcal{E} . In the more general case of loopy dependency structures [60], MS provides a heuristic minimization method. It is nevertheless a popular technique favored in communication systems for its low computational cost and notable performance on random tree-like networks [73].

For the optimization model considered here, as in (2), the MS messages, $\epsilon_{i \rightarrow j}$, are computed iteratively along *directed* edges $i \rightarrow j$ and $j \rightarrow i$ for each edge $(i, j) \in \mathcal{E}$, according to the Min-Sum equations:

$$\epsilon_{i \rightarrow j}^{t+1} = \text{SSL}(2\mathbf{J}_{ij}, 2\mathbf{h}_i + \sum_{k \in \mathcal{E}(i) \setminus j} \epsilon_{k \rightarrow j}^t) \quad (12a)$$

$$\text{SSL}(x, y) = \min(x, y) - \min(-x, y) - x \quad (12b)$$

Here, $\mathcal{E}(i) \setminus j$ denotes the neighbors of i without j and SSL denotes the Symmetric Saturated Linear transfer function. Once a fix-point of (12a) is obtained or a prescribed runtime limit is reached, the MS algorithm outputs a configuration based on the following formula:

$$\sigma_i = -\text{sign} \left(2\mathbf{h}_i + \sum_{k \in \mathcal{E}(i)} \epsilon_{k \rightarrow j} \right) \quad (13)$$

By convention, if the argument of the *sign* function is 0, a value of 1 or -1 is assigned randomly with equal probability.

Markov Chain Monte Carlo (MCMC) MCMC algorithms include a wide range of methods to generate samples from complex probability distributions. A natural Markov Chain for the Ising model is given by Glauber dynamics, where the value of each variable is updated according to its conditional probability distribution. Glauber dynamics is often used as a method for producing samples from Ising models at *finite temperature* [33]. This work considers the so-called *Zero Temperature* Glauber Dynamics (GD) algorithm, which is the optimization variant of the Glauber dynamics sampling method, and which is also used in physics as a simple model for describing avalanche phenomena in magnetic materials [23]. From the optimization perspective, this approach is a single-variable greedy local search algorithm.

A step t of the GD algorithm consists in checking each variable $i \in \mathcal{N}$ in a random order and comparing the objective cost of the current configuration σ^t to the configuration with the variable σ_i^t being flipped. If the objective value is lower in the flipped configuration, i.e., $E(\underline{\sigma}^t) > E(\sigma_1^t, \dots, -\sigma_i^t, \dots, \sigma_N^t)$, then the flipped configuration is selected as the new current configuration $\sigma^{t+1} = (\sigma_1^t, \dots, -\sigma_i^t, \dots, \sigma_N^t)$. When the objective difference is 0, the previous or new configuration is selected randomly with equal probability. If after visiting all of the variables, no one single-variable flip can improve the current assignment, then the configuration is identified as a local minimum and the algorithm is restarted with a new randomly generated configuration. This process is repeated until a runtime limit is reached.

Large Neighborhood Search (LNS) The state-of-the-art meta-heuristic for benchmarking D-Wave-based QA algorithms is the Hamze-Freitas-Selby (HFS) algorithm [37, 70]. The core idea of this algorithm is to extract low treewidth subgraphs of the given Ising model and then use dynamic programming to quickly compute the optimal configuration of these subgraphs. This extract and optimize process is repeated until a specified time limit is reached. This approach has demonstrated remarkable results in a variety of benchmarking studies [16, 44, 48, 49, 65]. The notable success of this solver can be attributed to three key factors. First, it is highly specialized to solving Ising models on the Chimera graphs (i.e., Fig. 1), a topological structure that is particularly amenable to low treewidth subgraphs. Second, it leverages integer arithmetic instead of floating point, which provides a significant performance improvement but also leads to notable precision limits. Third, the baseline implementation is a highly optimized C code [69], which runs at near-ideal performance.

4.3 Quantum annealing

Extending the theoretical overview from Section 3, the following implementation details are required to leverage the D-Wave 2000Q platform as a reliable optimization tool. The QA algorithm considered here consists of programming the Ising model of interest and then repeating the annealing process some number of times (i.e., *num_reads*) and then returning the lowest energy solution that was found among all of those replicates. No correction or solution polishing is applied in this solver. By varying the number of reads considered (e.g., from 10 to 10,000), the solution quality and total runtime of the QA algorithm increases. It is important to highlight that the D-Wave platform provides a wide variety of parameters to control the annealing process (e.g., annealing time, qubit offsets, custom annealing schedules, etc.). In the interest of simplicity and reproducibility, this work does not leverage any of those advanced features and it is likely that the results presented here would be further improved by careful utilization of those additional capabilities [2, 50, 56].

Note that all of the problems considered in this work have been generated to meet the implementation requirements discussed in Section 3.1 for a specific D-Wave chip deployed at Los Alamos National Laboratory. Consequently, no problem transformations are required to run the instances on the target hardware platform. Most notably, no embedding or rescaling is required. This approach is standard practice in QA evaluation studies and the arguments for it are discussed at length in [15, 16].

5 Structure detection experiments

This section presents the primary result of this work. Specifically, it analyzes three crafted optimization problems of increasing complexity—the Biased Ferromagnet, Frustrated Biased Ferromagnet, and Corrupted Biased Ferromagnet—all of which highlight the potential for QA to quickly identify the global structural properties of these problems. The algorithm performance analysis focuses on two key metrics, solution quality over time (i.e., performance profile) and the minimum hamming distance to any optimal solution over time. The hamming distance metric is particularly informative in this study as the problems have been designed to have local minima that are very close to the global optimum in terms of objective value, but are very distant in terms of hamming distance. The core finding is that QA produces solutions that are close to global optimality, both in terms of objective value and hamming distance.

Problem generation All problems considered in this work are defined by simple probabilistic graphical models and are generated on a specific D-Wave hardware graph. To avoid bias towards one particular random instance, 100 instances are generated and the mean over this collection of instances is presented. Additionally, a random gauge transformation is applied to every instance to obfuscate the optimal solution and mitigate artifacts from the choice of initial condition in each solution approach.

Computation Environment The CPU-based algorithms are run on HPE ProLiant XL170r servers with dual Intel 2.10GHz CPUs and 128GB memory. Gurobi 9.0 [35] was used for solving the Integer Programming (ILP/IQP) formulations. All of the algorithms were configured to only leverage one thread and the reported runtime reflects the wall clock time of each solver's core routine and does not include pre-processing or post-processing of the problem data.

The QA computation is conducted on a D-Wave 2000Q quantum annealer deployed at Los Alamos National Laboratory. This computer has a 16-by-16 Chimera cell topology with random omissions; in total, it has 2032 spins (i.e., \mathcal{N}) and 5924 couplers (i.e., \mathcal{E}). The hardware is configured to execute 10 to 10,000 annealing runs using a 5-microsecond annealing time per run and a random gauge transformation every 100 runs, to mitigate the various sources of bias in the problem encoding. The reported runtime of the QA hardware reflects the amount of *on-chip time* used; it does not include the overhead of communication or scheduling of the computation, which takes about one to two seconds. Given a sufficient engineering effort to reduce overheads, on-chip time would be the dominating runtime factor.

5.1 The biased ferromagnet

$$\begin{aligned} J_{ij} &= -1.00 \quad \forall i, j \in \mathcal{E}; \\ P(\mathbf{h}_i = 0.00) &= 0.990, P(\mathbf{h}_i = -1.00) = 0.010 \quad \forall i \in \mathcal{N} \end{aligned} \quad (\text{BFM})$$

Inspired by the Ferromagnet model, this study begins with Biased FerroMagnet (BFM) model—a toy problem to build an intuition for a type of structure that QA can exploit. Notice that this model has no frustration and has a few linear terms that bias it to prefer $\sigma_i = 1$ as the global optimal solution. W.h.p. $\sigma_i = 1$ is a unique optimal solution and the assignment of $\sigma_i = -1$ is a local minimum that is sub-optimal by $0.02 \cdot |\mathcal{N}|$ in expectation and has a maximal hamming distance of $|\mathcal{N}|$. The local minimum is an attractive solution because it is nearly optimal; however, it is hard for a local search solver to escape from it due to its hamming distance from the true global minimum. This instance presents two key algorithmic challenges: first, one must effectively detect the global structure (i.e., all the variables should take the same value); second, one must correctly discriminate between the two nearly optimal solutions that are very distant from one another.

Figure 2 presents the results of running all of the algorithms from Section 4 on the BFM model. The key observations are as follows:

- Both the greedy (i.e., SCD) and relaxation-based solvers (i.e., IQP/ILP/MS) correctly identify this problem's structure and quickly converge on the globally optimal solution (Fig. 2, top-right).
- Neighborhood-based local search methods (e.g., GD) tend to get stuck in the local minimum of this problem. Even advanced local search methods (e.g., HFS) may miss the global optimum in rare cases (Fig. 2, top).
- The hamming distance analysis indicates that QA has a high probability (i.e., greater than 0.9) of finding the exact global optimal solution (Fig. 2, bottom-right). This

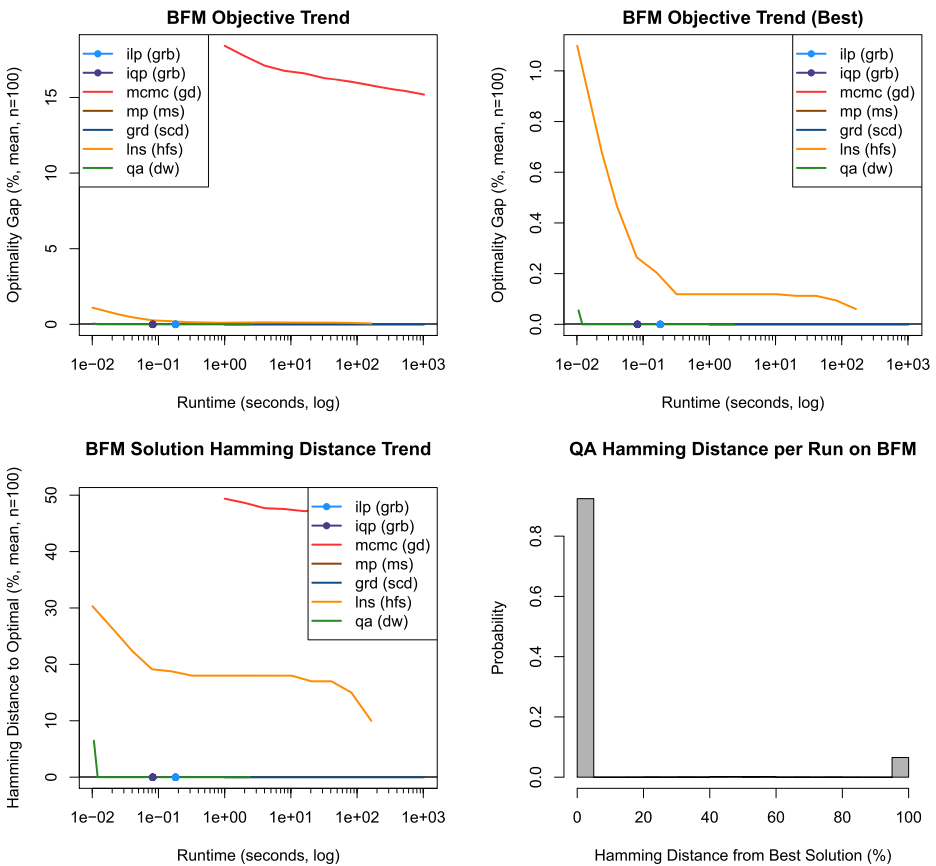


Fig. 2 Performance profile (top) and Hamming Distance (bottom) analysis for the Biased Ferromagnet instance

explains why just 20 runs is sufficient for QA to find the optimal solution w.h.p. (Fig. 2, top-right).

A key observation from this toy problem is that making a continuous relaxation of the problem (e.g., IQP/ILP/MS) can help algorithms detect global structure and avoid local minima that present challenges for neighborhood-based local search methods (e.g., GD/LNS). QA has comparable performance to these relaxation-based methods, both in terms of solution quality and runtime, and does appear to detect the global structure of the BFM problem class.

Howeve encouraging these results are, the BFM problem is a straw-man that is trivial for five of the seven solution methods considered here. The next experiment introduces frustration to the BFM problem to understand how that impacts problem difficulty for the solution methods considered.

5.2 The frustrated biased ferromagnet

$$J_{ij} = -1.00 \forall i, j \in \mathcal{E}$$

$$P(\mathbf{h}_i = 0.00) = 0.970, P(\mathbf{h}_i = -1.00) = 0.020, P(\mathbf{h}_i = 1.00) = 0.010 \forall i \in \mathcal{N} \text{ (FBFM)}$$

The next step considers a slightly more challenging problem called a Frustrated Biased Ferromagnet (FBFM), which is a specific case of the random field Ising model [21] and similar in spirit to the Clause Problems considered in [57]. The FBFM deviates from the BFM by introducing frustration among the linear terms of the problem. Notice that on average 2% of the decision variables locally prefer $\sigma_i = 1$ while 1% prefer $\sigma_i = -1$. Throughout the optimization process these two competing preferences must be resolved, leading to frustration. W.h.p. this model has the same unique global optimal solution as the BFM that occurs when $\sigma_i = 1$. The opposite assignment of $\sigma_i = -1$ remains a local minimum that is sub-optimal by $0.02 \cdot |\mathcal{N}|$ in expectation and has a maximal hamming distance of $|\mathcal{N}|$. By design, the energy difference of these two extreme assignments is consistent with BFM, to keep the two problem classes as similar as possible.

Figure 3 presents the same performance analysis for the FBFM model. The key observations are as follows:

- When compared to BFM, FBFM presents an increased challenge for the simple greedy (i.e., SCD) and local search (i.e., GD/MS) algorithms.

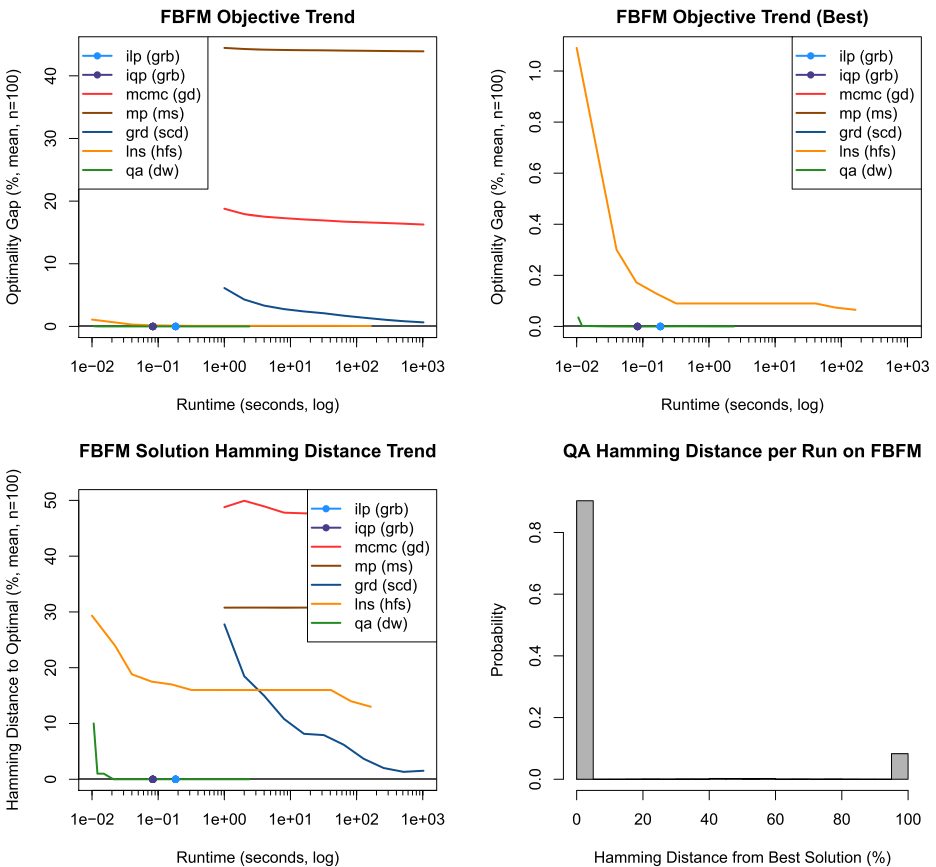


Fig. 3 Performance profile (top) and Hamming Distance (bottom) analysis for the Frustrated Biased Ferromagnet instance

- Although the SCD algorithm is worse than HFS in terms of objective quality, it is comparable or better in terms of hamming distance (Fig. 3, bottom-left). This highlights how these two metrics capture different properties of the underlying algorithms.
- The results of QA and the relaxation-based solvers (i.e., IQP/ILP), are nearly identical to the BFM case, suggesting that this type of frustration does not present a significant challenge for these solution approaches.

These results suggest that frustration in the linear terms alone (i.e., \mathbf{h}) is not sufficient for building optimization tasks that are non-trivial for a wide variety of general purpose solution methods. In the next study, frustration in the quadratic terms (i.e., \mathbf{J}) is incorporated to increase the difficulty for the relaxation-based solution methods.

5.3 The corrupted biased ferromagnet

$$P(\mathbf{J}_{ij} = -1.00) = 0.625, P(\mathbf{J}_{ij} = 0.20) = 0.375 \forall i, j \in \mathcal{E}$$

$$P(\mathbf{h}_i = 0.00) = 0.970, P(\mathbf{h}_i = -1.00) = 0.020, P(\mathbf{h}_i = 1.00) = 0.010 \forall i \in \mathcal{N} \quad (\text{CBFM})$$

The inspiration for this instance is to leverage insights from the theory of Spin glasses to build more computationally challenging problems. The core idea is to carefully corrupt the ferromagnetic problem structure with frustrating anti-ferromagnetic links that obfuscate the ferromagnetic properties without completely destroying them. A parameter sweep of different corruption values yields the Corrupted Biased FerroMagnet (CBFM) model, which retains the global structure that $\sigma_i = 1$ is a near globally optimal solution w.h.p., while obfuscating this property with misleading anti-ferromagnetic links and frustrated local fields.

Figure 4 presents a similar performance analysis for the CBFM model. The key observations are as follows:

- In contrast to the BFM and FBFM cases, solvers that leverage continuous relaxations, such as IQP and ILP, do not immediately identify this problem's structure and can take between 50 to 700 seconds to identify the globally optimal solution (Fig. 4, top-left).
- The advanced local search method (i.e., HFS) consistently converges to a global optimum (Fig. 4, top-right), which does not always occur in the BFM and FBFM cases.
- Although the MS algorithm is notably worse than GD in terms of objective quality, it is notably better in terms of hamming distance. This further indicates how these two metrics capture different properties of the underlying algorithms (Fig. 4, bottom-left).
- Although this instance presents more of a challenge for QA than BFM and FBFM, QA still finds the global minimum with high probability; 500-1000 runs is sufficient to find a near-optimal solution in all cases. This is 10 to 100 times faster than the next-best algorithm, HFS (Fig. 4, top-right).
- The hamming distance analysis suggests that the success of the QA approach is that it has a significant probability (i.e., greater than 0.12) of returning a solution that has a hamming distance of less than 1% from the global optimal solution (Fig. 4, bottom-right).

The overarching trend of this study is that QA is successful in detecting the global structure of the BFM, FBFM, and CBFM instances (i.e., low hamming distance to optimal, w.h.p.). Furthermore, it can do so notably faster than all of the other algorithms considered here. This suggests that, in this class of problems, QA brings a unique value that is not captured by the

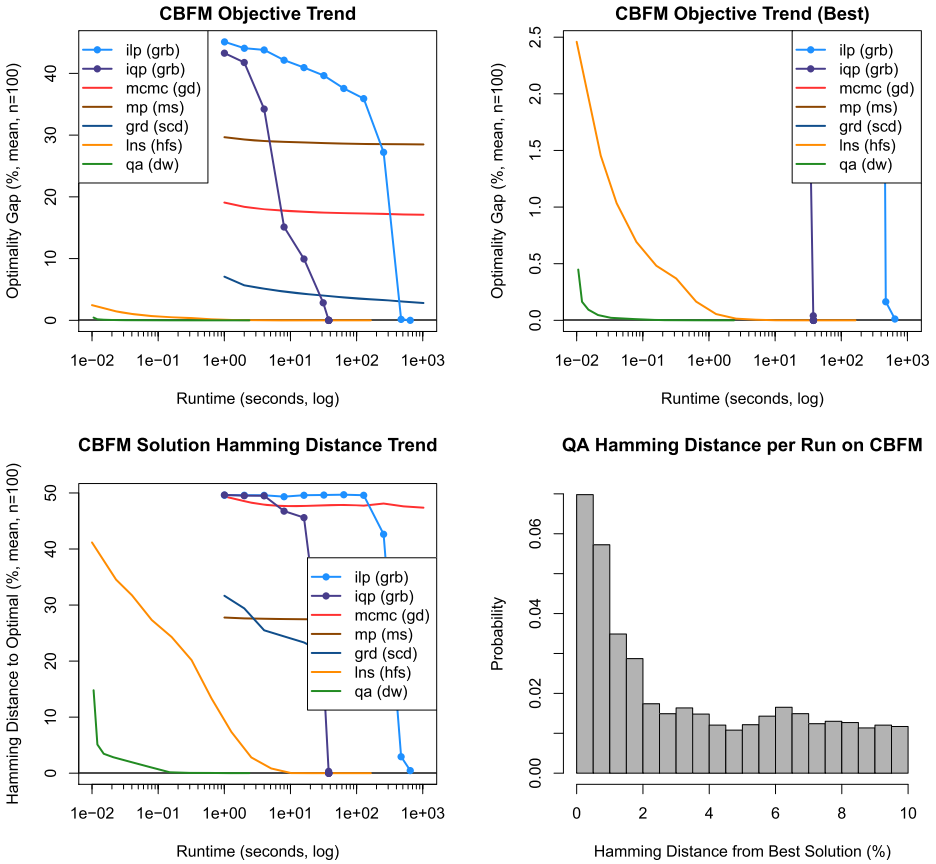


Fig. 4 Performance profile (top) and Hamming Distance (bottom) analysis for the Corrupted Biased Ferromagnet instance

other algorithms considered. Similar to how the relaxation methods succeed at the BFM and FBFM instances, we hypothesize that the success of QA on the CBFM instance is driven by the solution search occurring in a smooth high-dimensional continuous space as discussed in Section 3. In this instance class, QA may also benefit from so-called *finite-range tunnelling* effects, which allows QA to change the state of multiple variables simultaneously (i.e., global moves) [22, 27]. Regardless of the underlying cause, QA’s performance on the CBFM instance is particularly notable and worthy of further investigation.

5.4 Bias structure variants

As part of the design process uniform field variants of the problems proposed herein were also considered. These variants featured weaker and more uniform distributed bias terms. Specifically, the term $P(\mathbf{h}_i = -1.00) = 0.010$ was replaced with $P(\mathbf{h}_i = -0.01) = 1.000$. Upon continued analysis, it was observed that the stronger and less-uniform bias terms resulted in more challenging cases for all of the solution methods considered, and hence, were selected as the preferred design for the problems proposed by this work. In the interest of completeness, Appendix A provides a detailed analysis of the uniform-field variants of

the BFM, FBFM, and CBFM instances to illustrate how this problem variant impacts the performance of the solution methods considered here.

5.5 A comparison to other instance classes

The CBFM problem was designed to have specific structural properties that are beneficial to the QA approach. It is important to note that not all instance classes have such an advantageous structure. This point is highlighted in Fig. 5, which compares three landmark problem classes from the QA benchmarking literature: Weak-Strong Cluster Networks (WSCN) [22], Frustrated Cluster Loops with Gadgets (FCLG) [4], and Random Couplers and Fields (RANF-1) [16, 20]. These results show that D-Wave’s current 2000Q hardware platform can be outperformed by local and complete search methods on some classes of problems. However, it is valuable to observe that these previously proposed instance classes are either relatively easy for local search algorithms (i.e., WSCN and RANF) or relatively easy for complete search algorithms (i.e., WSCN and FCLG), both of which are not ideal properties for conducting benchmarking studies. To the best of our knowledge, the proposed CBFM problem is the first instance class that presents a notable computational challenge for both local search and complete search algorithms.

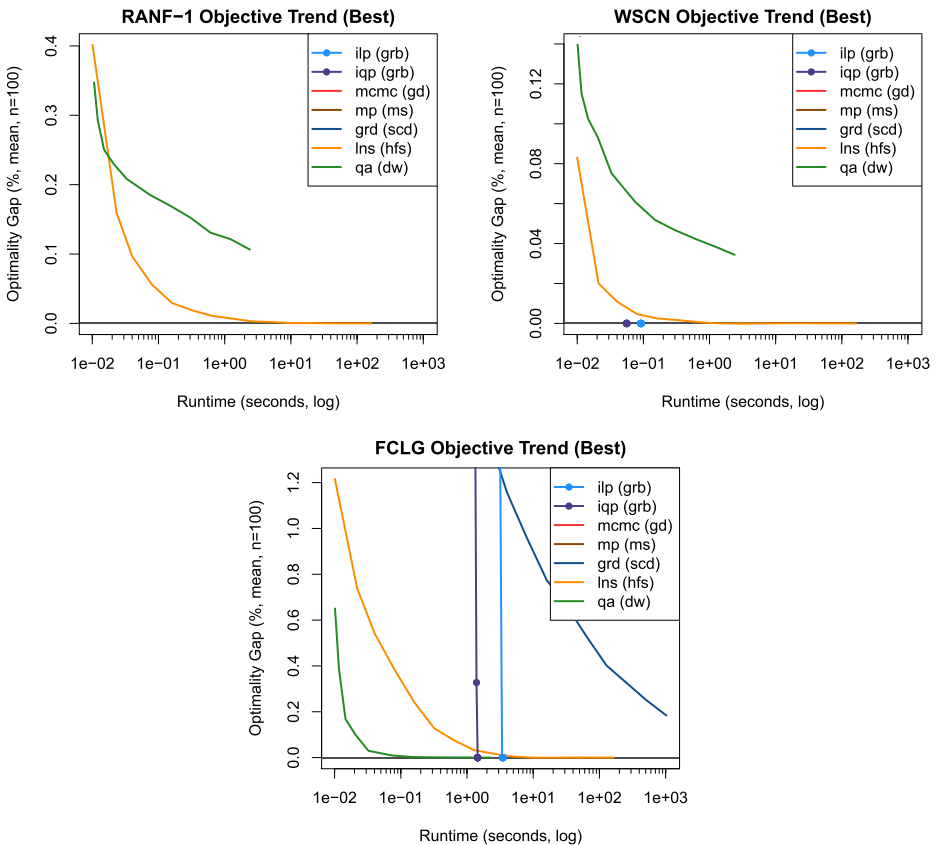


Fig. 5 Performance profiles of other problem classes from the literature

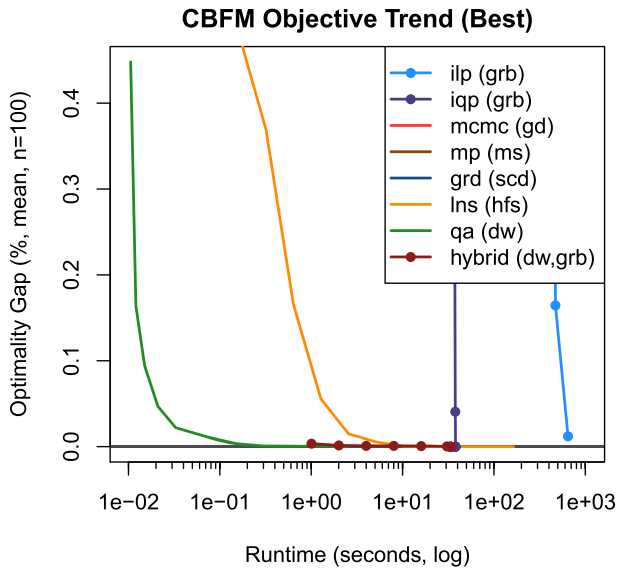


Fig. 6 Performance profile of Warm-Starting IQP with QA solutions

6 Quantum annealing as a primal heuristic

QA's notable ability to find high-quality solutions to the CBFM problem suggests the development of hybrid algorithms, which leverage QA for finding upper bounds within a complete search method that can also provide global optimality proofs. A simple version of such an approach was developed where 1000 runs of QA were used to warm-start the IQP solver with a high-quality initial solution. The results of this hybrid approach are presented in Fig. 6. The IQP solver clearly benefits from the warm-start on short time scales. However, it does not lead to a notable reduction in the time to producing the optimality proof. This suggests that a state-of-the-art hybrid complete search solver needs to combine QA for finding upper bounds with more sophisticated lower-bounding techniques, such as those presented in [6, 44].

7 Conclusion

This work explored how quantum annealing hardware might be able to support heuristic algorithms in finding high-quality solutions to challenging combinatorial optimization problems. A careful analysis of quantum annealing's performance on the Biased Ferromagnet, Frustrated Biased Ferromagnet, and Corrupted Biased Ferromagnet problems with more than 2,000 decision variables suggests that this approach is capable of quickly identifying the structure of the optimal solution to these problems, while a variety of local and complete search algorithms struggle to identify this structure. This result suggests that integrating quantum annealing into meta-heuristic algorithms could yield unique variable assignments and increase the discovery of high-quality solutions.

Although demonstration of a runtime advantage was not the focus of this work, the success of quantum annealing on the Corrupted Biased Ferromagnet problem compared to other

solution methods is a promising outcome for QA and warrants further investigation. An in-depth theoretical study of the Corrupted Biased Ferromagnet case could provide deeper insights into the structural properties that quantum annealing is exploiting in this problem and would provide additional insights into the classes of problems that have the best chance to demonstrate an unquestionable computational advantage for quantum annealing hardware. It is important to highlight that while the research community is currently searching for an unquestionable computational advantage for quantum annealing hardware by any means necessary, significant additional research will be required to bridge the gap between contrived hardware-specific optimization tasks and practical optimization applications.

Appendix A: Uniform fields

This appendix presents the results of the uniform-field variants of the BFM, FBFM, and CBFM instances and illustrates how uniform fields improve the performance of all solution methods considered. Specifically the uniform-field variants replace the bias term, $P(\mathbf{h}_i =$

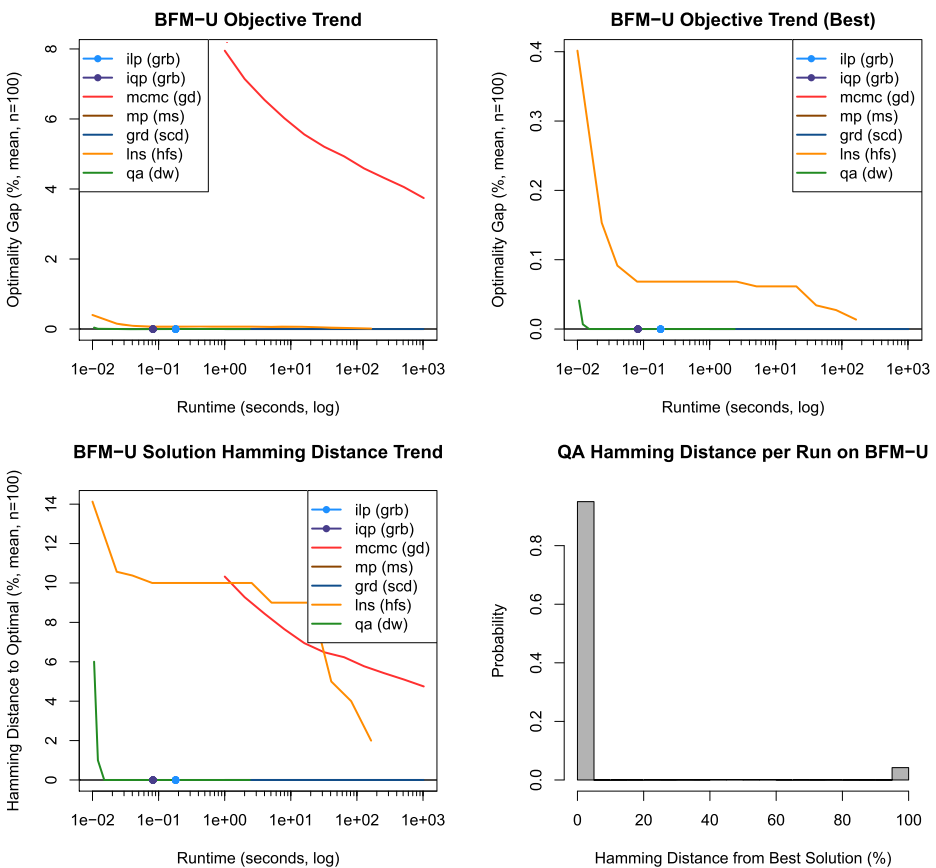


Fig. 7 Performance profile (top) and Hamming Distance (bottom) analysis for the Biased Ferromagnet with Uniform Fields instance

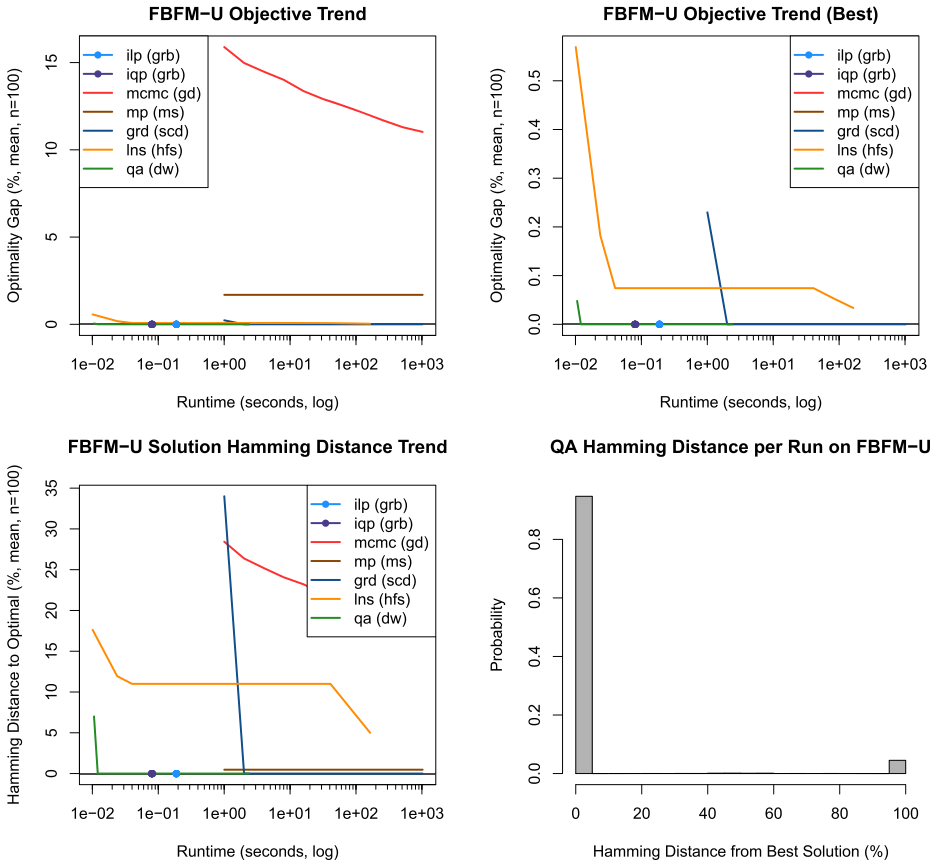


Fig. 8 Performance profile (top) and Hamming Distance (bottom) analysis for the Frustrated Biased Ferromagnet with Uniform Fields instance

$-1.00) = 0.010$, with the uniform variant $P(\mathbf{h}_i = -0.01) = 1.000$. Throughout this study the field’s probability distribution is modified such that there are no zero-value fields (i.e., $P(\mathbf{h}_i = 0.00) = 0.000$) and, for consistency with the BFM, FBFM, and CBFM cases presented in Section 5, the mean of the fields is selected to be -0.01 (i.e., $\mu_{\mathbf{h}} = -0.01$) in all problems considered.

A1: The biased ferromagnet with uniform fields

$$\begin{aligned}
 J_{ij} &= -1.00 \quad \forall i, j \in \mathcal{E}; \\
 \mathbf{h}_i &= -0.01 \quad \forall i \in \mathcal{N}
 \end{aligned}
 \tag{BFM-U}$$

The Biased Ferromagnet with Uniform Fields (BFM-U) is similar to the BFM case, but all of the linear terms are set identically to $\mathbf{h}_i = -0.01$. All of the solution methods considered here perform well on this BFM-U case (see Fig. 7). However, the BFM-U case does appear to reduce both the optimality gap and hamming distance metrics by a factor of two compared to the BFM case. This suggests that BFM-U is easier than BFM based on the metrics considered by this work.

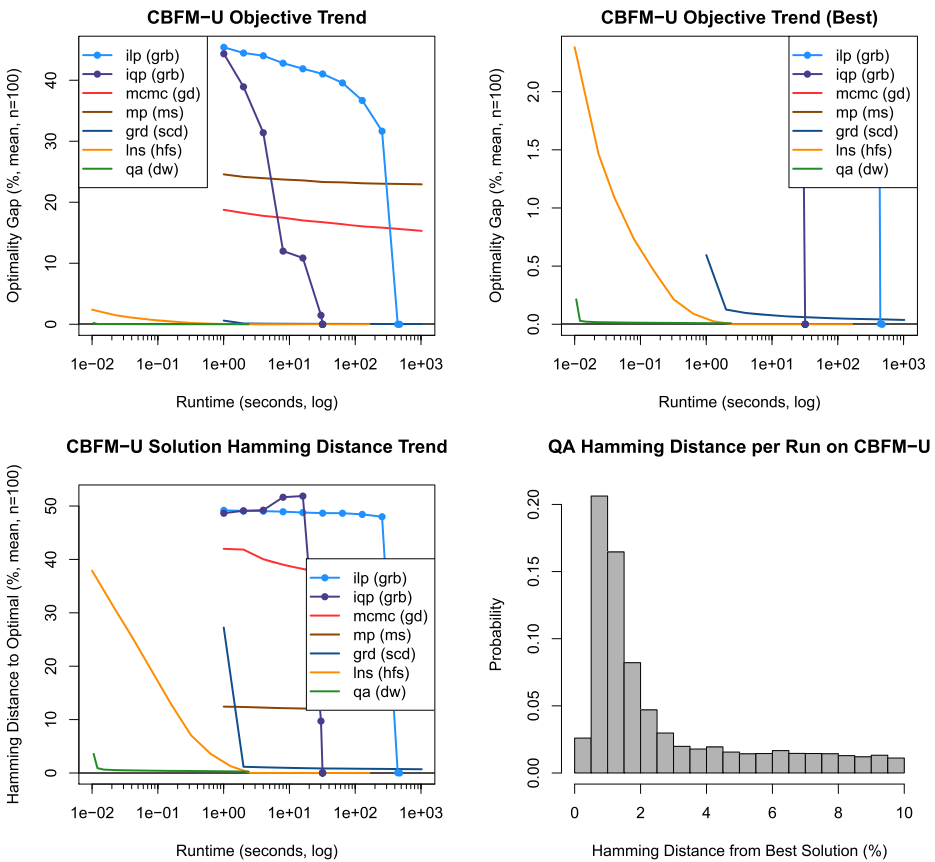


Fig. 9 Performance profile (top) and Hamming Distance (bottom) analysis for the Corrupted Biased Ferromagnet with Uniform Fields instance

A.2: The frustrated biased ferromagnet with uniform fields

$$J_{ij} = -1.00 \forall i, j \in \mathcal{E}$$

$$P(\mathbf{h}_i = -0.03) = 0.666, P(\mathbf{h}_i = 0.03) = 0.334 \forall i \in \mathcal{N} \quad (\text{FBFM-U})$$

The Frustrated Biased Ferromagnet with Uniform Fields (FBFM-U) is similar to the FBFM case, but two-thirds of the linear terms are set to $\mathbf{h}_i = -0.03$ and one-third is set to $\mathbf{h}_i = 0.03$. Although the performance of most of the algorithms on FBFM-U is similar to FBFM (see Fig. 8), there are two notable deviations. The performance of MS and SCD algorithms improves significantly in the FBFM-U case. This also suggests that the FBFM-U is easier than FBFM based on the metrics considered by this work.

A.3: The corrupted biased ferromagnet with uniform fields

$$P(\mathbf{J}_{ij} = -1.00) = 0.625, P(\mathbf{J}_{ij} = 0.20) = 0.375 \forall i, j \in \mathcal{E}$$

$$P(\mathbf{h}_i = -0.03) = 0.666, P(\mathbf{h}_i = 0.03) = 0.334 \forall i \in \mathcal{N} \quad (\text{CBFM-U})$$

The Corrupted Biased Ferromagnet with Uniform Fields (CBFM-U) is similar to the CBFM case, but two-thirds of the linear terms are set to $\mathbf{h}_i = -0.03$ and one-third is set

to $h_i = 0.03$. This case exhibits the most variation from the CBFM alternative (see Fig. 9). The key observations are as follows:

- In CBFM-U, QA has a higher probability of finding a near-optimal solution (i.e., > 0.50) than CBFM (i.e., < 0.20). However, it has a lower probability of finding the true-optimal solution (Fig. 9, bottom-right). Due to this effect, QA finds a near-optimal solution to CBFM-U faster than CBFM but never manages to converge to the optimal solution, as it does in CBFM.
- The performance of the SCD algorithm improves significantly in the CBFM-U case. The SCD algorithm is among the best solutions for CBFM-U ($< 0.5\%$ optimality gap), while it has more than a 2% optimality gap in the CBFM case.

Overall, these results suggest that CBFM-U is easier than CBFM based on the metrics considered by this work. However, the subtle differences in the performance of QA between CBFM and CBFM-U suggest that varying the distribution of the linear terms in the CBFM family of problems could be a useful tool for developing a deeper understanding of how QA responds to different classes of optimization tasks.

Appendix B: Reference implementations

B.1: D-Wave instance generator (DWIG)

The problems considered in this work were generated with the open-source D-Wave Instance Generator tool, which is available at <https://github.com/lanl-ansi/dwig>. DWIG is a command line tool that uses D-Wave’s hardware API to identify the topology of a specific D-Wave device and uses that graph for randomized problem generation. The following list provides the mapping of problems in this paper to the DWIG command line interface:

CBFM:

```
dwig.py cbfm -rgt
```

CBFM-U:

```
dwig.py cbfm -rgt
-j1-val -1.00 -j1-pr 0.625 -j2-val 0.02 -j2-pr 0.375
-h1-val -0.03 -h1-pr 0.666 -h2-val 0.03 -h2-pr 0.334
```

FBFM:

```
dwig.py cbfm -rgt
-j1-val -1.00 -j1-pr 1.000 -j2-val 0.00 -j2-pr 0.000
-h1-val -1.00 -h1-pr 0.020 -h2-val 1.00 -h2-pr 0.010
```

FBFM-U:

```
dwig.py cbfm -rgt
-j1-val -1.00 -j1-pr 1.000 -j2-val 0.00 -j2-pr 0.000
-h1-val -0.03 -h1-pr 0.666 -h2-val 0.03 -h2-pr 0.334
```

BFM:

```
dwig.py cbfm -rgt
```

```
-j1-val -1.00 -j1-pr 1.000 -j2-val 0.00 -j2-pr 0.000
-h1-val -1.00 -h1-pr 0.010 -h2-val 0.00 -h2-pr 0.000
```

BFM-U:

```
dwig.py cbfm -rgt
-j1-val -1.00 -j1-pr 1.000 -j2-val 0.00 -j2-pr 0.000
-h1-val -0.01 -h1-pr 1.000 -h2-val 0.00 -h2-pr 0.000
```

B.2: Ising model optimization methods

The problems considered in this work were solved with the open-source *Ising-Solvers* scripts that are available at <https://github.com/lanl-ansi/ising-solvers>. These scripts include a combination of calls to executables, system libraries, and handmade heuristics. Each script conforms to a standard API for measuring runtime and reporting results. The following commands were used for each of the solution approaches presented in this work:

ILP (GRB):

```
ilp_gurobi.py -ss -rtl <time_limit> -f <case file>
```

IQP (GRB):

```
iqp_gurobi.py -ss -rtl <time_limit> -f <case file>
```

MCMC (GD):

```
mcmc_gd.py -ss -rtl <time_limit> -f <case file>
```

MP (MS):

```
mp_ms.py -ss -rtl <time_limit> -f <case file>
```

GRD (SCD):

```
grd_scd.jl -s -t <time_limit> -f <case file>
```

LNS (HFS):

```
lns_hfs.py -ss -rtl <time_limit> -f <case file>
```

QA (DW):

```
qa_dwave.py -ss -nr <number of reads> -at 5 -srtr 100 -f
<case file>
```

Acknowledgments The research presented in this work was supported by the Laboratory Directed Research and Development program of Los Alamos National Laboratory under project numbers 20180719ER and 20190195ER.

Funding This work was funded by Los Alamos National Laboratory’s Laboratory Directed Research and Development program as part of the projects 20180719ER and 20190195ER.

Availability of data and material The data used to generate the figures in this work is not explicitly archived. It can be recreated using the software that is available as open-source.

Code availability The core software tools that were used in this work are available as open-source under a BSD license. The `dwig` software is available at <https://github.com/lanl-ansi/dwig> and can be used to generate instances of the problem classes considered in this work. The optimization methods considered in this work are archived in the `ising-solvers` repository, which is available at <https://github.com/lanl-ansi/ising-solvers>. One of the solution methods considered in this work requires a commercial software licence.

Declarations

Conflict of interests None.

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

References

1. Aaronson, S. (2017). Insert d-wave post here. Published online at <http://www.scottaaronson.com/blog/?p=3192>. Accessed 28 Apr 2017.
2. Adame, J.I., & McMahon, P.L. (2020). Inhomogeneous driving in quantum annealers can result in orders-of-magnitude improvements in performance. *Quantum Science and Technology*, 5(3), 035011. <https://doi.org/10.1088/2058-9565/ab935a>. <https://iopscience.iop.org/article/10.1088/2058-9565/ab935a>.
3. Albash, T., & Lidar, D.A. (2018). Adiabatic quantum computation. *Reviews of Modern Physics*, 90(1), 015,002.
4. Albash, T., & Lidar, D.A. (2018). Demonstration of a scaling advantage for a quantum annealer over simulated annealing. *Physical Review X*, 8(031), 016. <https://doi.org/10.1103/PhysRevX.8.031016>.
5. Arute, F., Arya, K., Babbush, R., Bacon, D., Bardin, J.C., Barends, R., Biswas, R., Boixo, S., Brandao, F.G.S.L., et al. (2019). Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779), 505–510. <https://doi.org/10.1038/s41586-019-1666-5>.
6. Baccari, F., Gogolin, C., Wittek, P., Acín, A. (2018). Verification of quantum optimizers. arXiv:1808.01275.
7. Barahona, F. (1982). On the computational complexity of ising spin glass models. *Journal of Physics A: Mathematical and General*, 15(10), 3241.
8. Bian, Z., Chudak, F., Israel, R., Lackey, B., Macready, W.G., Roy, A. (2014). Discrete optimization using quantum annealing on sparse ising models. *Frontiers in Physics*, 2, 56. <https://doi.org/10.3389/fphy.2014.00056>.
9. Bian, Z., Chudak, F., Israel, R.B., Lackey, B., Macready, W.G., Roy, A. (2016). Mapping constrained optimization problems to quantum annealing with application to fault diagnosis. *Frontiers in ICT*, 3, 14. <https://doi.org/10.3389/fict.2016.00014>.
10. Billionnet, A., & Elloumi, S. (2007). Using a mixed integer quadratic programming solver for the unconstrained quadratic 0-1 problem. *Mathematical Programming*, 109(1), 55–68. <https://doi.org/10.1007/s10107-005-0637-9>.
11. Boixo, S., Ronnow, T.F., Isakov, S.V., Wang, Z., Wecker, D., Lidar, D.A., Martinis, J.M., Troyer, M. (2014). Evidence for quantum annealing with more than one hundred qubits. *Nature Physics*, 10(3), 218–224. <https://doi.org/10.1038/nphys2900>. Article.
12. Boros, E., & Hammer, P.L. (2002). Pseudo-boolean optimization. *Discrete Applied Mathematics*, 123(1), 155–225. [https://doi.org/10.1016/S0166-218X\(01\)00341-9](https://doi.org/10.1016/S0166-218X(01)00341-9). <http://www.sciencedirect.com/science/article/pii/S0166218X01003419>.
13. Brush, S.G. (1967). History of the lenz-ising model. *Reviews of Modern Physics*, 39, 883–893. <https://doi.org/10.1103/RevModPhys.39.883>.
14. Chmielewski, M., Amini, J., Hudek, K., Kim, J., Mizrahi, J., Monroe, C., Wright, K., Moehring, D. (2018). Cloud-based trapped-ion quantum computing. In *APS Meeting abstracts*.

15. Coffrin, C., Nagarajan, H., Bent, R. (2016). Challenges and successes of solving binary quadratic programming benchmarks on the DW2x QPU. Tech. rep. Los Alamos National Laboratory (LANL).
16. Coffrin, C., Nagarajan, H., Bent, R. (2019). Evaluating ising processing units with integer programming. In Rousseau, L.M., & Stergiou, K. (Eds.) *Integration of constraint programming, artificial intelligence, and operations research* (pp. 163–181). Cham: Springer International Publishing.
17. Coffrin, C., & Pang, Y. (2019). ising-solvers. <https://github.com/lanl-ansi/ising-solvers>.
18. Coles, P.J., Eidenbenz, S., Pakin, S., Adedoyin, A., Ambrosiano, J., Anisimov, P., Casper, W., Chen-nupati, G., Coffrin, C., Djidjev, H., et al. (2018). Quantum algorithm implementations for beginners. arXiv:1804.03719.
19. Cugliandolo, L.F. (2018). Advanced statistical physics: Frustration. <https://www.lpthe.jussieu.fr/leticia/TEACHING/master2018/frustration18.pdf>.
20. Dash, S. (2013). A note on qubo instances defined on chimera graphs. arXiv:1306.1202.
21. d’Auriac, J.A., Preissmann, M., Rammal, R. (1985). The random field ising model: algorithmic complexity and phase transition. *Journal de Physique Lettres*, 46(5), 173–180.
22. Denchev, V.S., Boixo, S., Isakov, S.V., Ding, N., Babbush, R., Smelyanskiy, V., Martinis, J., Neven, H. (2016). What is the computational value of finite-range tunneling?. *Physical Review X*, 6, 031,015. <https://doi.org/10.1103/PhysRevX.6.031015>.
23. Dhar, D., Shukla, P., Sethna, J.P. (1997). Zero-temperature hysteresis in the random-field ising model on a bethe lattice. *Journal of Physics A: Mathematical and General*, 30(15), 5259.
24. Ding, J., Sly, A., Sun, N. (2015). Proof of the satisfiability conjecture for large k . In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pp. 59–68. ACM.
25. Eagle, N., Pentland, A.S., Lazer, D. (2009). Inferring friendship network structure by using mobile phone data. *Proceedings of the national academy of sciences*, 106(36), 15,274–15,278.
26. Fabio L., & Traversa, M.D.V. (2018). Memcomputing integer linear programming. arXiv:1808.09999.
27. Farhi, E., Goldstone, J., Gutmann, S., Lapan, J., Lundgren, A., Preda, D. (2001). A quantum adiabatic evolution algorithm applied to random instances of an np-complete problem. *Science*, 292(5516), 472–475. <https://doi.org/10.1126/science.1057726>. <http://science.sciencemag.org/content/292/5516/472>.
28. Farhi, E., Goldstone, J., Gutmann, S., Sipser, M. (2018). Quantum computation by adiabatic evolution. arXiv:quant-ph/0001106.
29. Feynman, R.P. (1982). Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6), 467–488.
30. Fossorier, M.P., Mihaljevic, M., Imai, H. (1999). Reduced complexity iterative decoding of low-density parity check codes based on belief propagation. *IEEE Transactions on communications*, 47(5), 673–680.
31. Fujitsu (2018). Digital annealer. Published online at <http://www.fujitsu.com/global/digitalannealer/>. Accessed 26 Feb 2019.
32. Gallavotti, G. (2013). *Statistical mechanics: A short treatise*. Berlin: Springer Science & Business Media.
33. Glauber, R.J. (1963). Time-dependent statistics of the ising model. *Journal of mathematical physics*, 4(2), 294–307.
34. Grover, L.K. (1996). A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pp. 212–219. ACM.
35. Gurobi Optimization, Inc. (2014). Gurobi optimizer reference manual Published online at <http://www.gurobi.com>.
36. Hamerly, R., Inagaki, T., McMahon, P.L., Venturelli, D., Marandi, A., Onodera, T., Ng, E., Langrock, C., Inaba, K., Honjo, T., et al. (2019). Experimental investigation of performance differences between coherent ising machines and a quantum annealer. *Science Advances*, 5(5), eaau0823.
37. Hamze, F., & de Freitas, N. (2004). From fields to trees. In *Proceedings of the 20th conference on uncertainty in artificial intelligence, UAI '04*, pp. 243–250. AUAI Press, Arlington, Virginia, United States. <http://dl.acm.org/citation.cfm?id=1036843.1036873>.
38. Haribara, Y., Utsunomiya, S., Yamamoto, Y. (2016). A coherent ising machine for MAX-CUT problems: performance evaluation against semidefinite programming and simulated annealing, pp. 251–262. Springer Japan, Tokyo. https://doi.org/10.1007/978-4-431-55756-2_12.
39. Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8), 2554–2558.
40. Inagaki, T., Haribara, Y., Igarashi, K., Sonobe, T., Tamate, S., Honjo, T., Marandi, A., McMahon, P.L., Umeki, T., Enbutsu, K., Tadanaga, O., Takenouchi, H., Aihara, K., Kawarabayashi, K.I., Inoue, K., Utsunomiya, S., Takesue, H. (2016). A coherent ising machine for 2000-node optimization problems. *Science*, 354(6312), 603–606. <https://doi.org/10.1126/science.aah4243>. <http://science.sciencemag.org/content/354/6312/603>.
41. International Business Machines Corporation (2017). Ibm building first universal quantum computers for business and science. Published online at <https://www-03.ibm.com/press/us/en/pressrelease/51740.wss>. Accessed 28 Apr 2017.

42. Isakov, S., Zintchenko, I., Rønnow, T., Troyer, M. (2015). Optimised simulated annealing for ising spin glasses. *Computer Physics Communications*, 192, 265–271. <https://doi.org/10.1016/j.cpc.2015.02.015>. <http://www.sciencedirect.com/science/article/pii/S0010465515000727>.
43. Johnson, M.W., Amin, M.H., Gildert, S., Lanting, T., Hamze, F., Dickson, N., Harris, R., Berkley, A.J., Johansson, J., Bunyk, P., et al. (2011). Quantum annealing with manufactured spins. *Nature*, 473(7346), 194–198.
44. Jünger, M., Lobe, E., Mutzel, P., Reinelt, G., Rendl, F., Rinaldi, G., Stollenwerk, T. (2019). Performance of a quantum annealer for ising ground state computations on chimera graphs. arXiv:1904.11965.
45. Kadowaki, T., & Nishimori, H. (1998). Quantum annealing in the transverse ising model. *Physical Review E*, 58, 5355–5363. <https://doi.org/10.1103/PhysRevE.58.5355>.
46. Kalinin, K.P., & Berloff, N.G. (2018). Global optimization of spin hamiltonians with gain-dissipative systems. *Scientific Reports*, 8(1), 1–9.
47. Kieplinski, D., Bose, R., Pelc, J., Vaerenbergh, T.V., Mendoza, G., Tezak, N., Beausoleil, R.G. (2016). Information processing with large-scale optical integrated circuits. In *2016 IEEE International conference on rebooting computing (ICRC)*, pp. 1–4. <https://doi.org/10.1109/ICRC.2016.7738704>.
48. King, A.D., Lanting, T., Harris, R. (2015). Performance of a quantum annealer on range-limited constraint satisfaction problems. arXiv:1502.02098.
49. King, J., Yarkoni, S., Raymond, J., Ozfidan, I., King, A.D., Nevisi, M.M., Hilton, J.P., McGeoch, C.C. (2017). Quantum annealing amid local ruggedness and global frustration. arXiv:1701.04579.
50. Lanting, T., King, A.D., Evert, B., Hoskinson, E. (2017). Experimental demonstration of perturbative anticrossing mitigation using nonuniform driver hamiltonians. *Physical Review A*, 96(042), 322. <https://doi.org/10.1103/PhysRevA.96.042322>.
51. Leleu, T., Yamamoto, Y., McMahon, P.L., Aihara, K. (2019). Destabilization of local minima in analog spin systems by correction of amplitude heterogeneity. *Physical Review Letters*, 122(4), 040,607.
52. Likhov, A.Y., Vuffray, M., Misra, S., Chertkov, M. (2018). Optimal structure and parameter learning of ising models. *Science Advances*, 4(3), e1700,.
53. Lucas, A. (2014). Ising formulations of many np problems. *Frontiers in Physics*, 2, 5. <https://doi.org/10.3389/fphy.2014.00005>.
54. Mandrà, S., Zhu, Z., Wang, W., Perdomo-Ortiz, A., 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. *Physical Review A*, 94(022), 337. <https://doi.org/10.1103/PhysRevA.94.022337>.
55. Marbach, D., Costello, J.C., Küffner, R., Vega, N.M., Prill, R.J., Camacho, D.M., Allison, K.R., Aderhold, A., Bonneau, R., Chen, Y., et al. (2012). Wisdom of crowds for robust gene network inference. *Nature Methods*, 9(8), 796.
56. Marshall, J., Venturelli, D., Hen, I., Rieffel, E.G. (2019). Power of pausing: Advancing understanding of thermalization in experimental quantum annealers. *Physical Review Applied*, 11(044), 083. <https://doi.org/10.1103/PhysRevApplied.11.044083>.
57. McGeoch, C.C., King, J., Nevisi, M.M., Yarkoni, S., Hilton, J. (2017). Optimization with clause problems. Published online at https://www.dwavesys.com/sites/default/files/14-1001A_tr.Optimization_with-Clause-Problems.pdf. Accessed 10 Feb 2020.
58. McGeoch, C.C., & Wang, C. (2013). Experimental evaluation of an adiabatic quantum system for combinatorial optimization. In *Proceedings of the ACM international conference on computing frontiers, CF '13*, pp. 23:1–23:11. ACM, New York, NY, USA. <https://doi.org/10.1145/2482767.2482797>.
59. McMahon, P.L., Marandi, A., Haribara, Y., Hamerly, R., Langrock, C., Tamate, S., Inagaki, T., Takesue, H., Utsunomiya, S., Aihara, K., et al. (2016). A fully-programmable 100-spin coherent ising machine with all-to-all connections. *Science*, p aah5178.
60. Mezard, M., Mezard, M., Montanari, A. (2009). Information, physics, and computation. Oxford: Oxford University Press.
61. Mézard, M., & Virasoro, M.A. (1985). The microstructure of ultrametricity. *Journal de Physique*, 46(8), 1293–1307.
62. Mohseni, M., Read, P., Neven, H., Boixo, S., Denchev, V., Babbush, R., Fowler, A., Smelyanskiy, V., Martinis, J. (2017). Commercialize quantum technologies in five years. *Nature*, 543, 171–174. <http://www.nature.com/news/commercialize-quantum-technologies-in-five-years-1.21583>.
63. Morcos, F., Pagnani, A., Lunt, B., Bertolino, A., Marks, D.S., Sander, C., Zecchina, R., Onuchic, J.N., Hwa, T., Weigt, M. (2011). Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proceedings of the National Academy of Sciences*, 108(49), E1293–E1301.
64. Panjwani, D.K., & Healey, G. (1995). Markov random field models for unsupervised segmentation of textured color images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(10), 939–954.

65. Parekh, O., Wendt, J., Shulenburg, L., Landahl, A., Moussa, J., Aidun, J. (2015). Benchmarking adiabatic quantum optimization for complex network analysis. arXiv:1604.00319.
66. Puget, J.F. (2013). D-wave vs cplex comparison. part 2: Qubo. Published online. Accessed 28 Nov 2018.
67. Rieffel, E.G., Venturelli, D., O’Gorman, B., Do, M.B., Prystay, E.M., Smelyanskiy, V.N. (2015). A case study in programming a quantum annealer for hard operational planning problems. *Quantum Information Processing*, 14(1), 1–36. <https://doi.org/10.1007/s11128-014-0892-x>.
68. Schneidman, E., Berry II, M.J., Segev, R., Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087), 1007.
69. Selby, A. (2013). Qubo-chimera. <https://github.com/alex1770/QUBO-chimera>.
70. Selby, A. (2014). Efficient subgraph-based sampling of ising-type models with frustration. <https://arxiv.org/abs/1409.3934>.
71. Shor, P.W. (1994). Algorithms for quantum computation: Discrete logarithms and factoring. In *Proceedings 35th annual symposium on foundations of computer science*, pp. 124–134. *Ieee*.
72. Venturelli, D., Marchand, D.J.J., Rojo, G. (2015). Quantum annealing implementation of job-shop scheduling. arXiv:1506.08479.
73. Vuffray, M. (2014). The cavity method in coding theory. Tech. rep. EPFL.
74. Vuffray, M., Misra, S., Lokhov, A., Chertkov, M. (2016). Interaction screening: Efficient and sample-optimal learning of ising models. In Lee, D.D., Sugiyama, M., Luxburg, U.V., Guyon, I., Garnett, R. (Eds.) *Advances in neural information processing systems 29*. pp 2595–2603. *Curran Associates, Inc.*
75. Vuffray, M., Misra, S., Lokhov, A.Y. (2019). Efficient learning of discrete graphical models. arXiv:1902.00600.
76. Yamaoka, M., Yoshimura, C., Hayashi, M., Okuyama, T., Aoki, H. (2015). Mizuno, h.: 24.3 20k-spin ising chip for combinational optimization problem with cmos annealing. In *2015 IEEE International solid-state circuits conference - (ISSCC) digest of technical papers*, pp. 1–3. <https://doi.org/10.1109/ISSCC.2015.7063111>.
77. Yoshimura, C., Yamaoka, M., Aoki, H., Mizuno, H. (2013). Spatial computing architecture using randomness of memory cell stability under voltage control. In *2013 European conference on circuit theory and design (ECCTD)*, pp. 1–4. <https://doi.org/10.1109/ECCTD.2013.6662276>.

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

Affiliations

Yuchen Pang¹  · Carleton Coffrin²  · Andrey Y. Lokhov²  · Marc Vuffray² 

Yuchen Pang
yuchenp2@illinois.edu

Andrey Y. Lokhov
lokhov@lanl.gov

Marc Vuffray
vuffray@lanl.gov

¹ Department of Computer Science, University of Illinois at Urbana-Champaign, Champaign, IL, 61801, USA

² Advanced Network Science Initiative, Los Alamos National Laboratory, Los Alamos, NM, 87545, USA