

Quantum Science and Technology



PAPER

OPEN ACCESS

RECEIVED
3 December 2022

REVISED
3 April 2023

ACCEPTED FOR PUBLICATION
28 April 2023

PUBLISHED
10 May 2023

Logical qubit implementation for quantum annealing: augmented Lagrangian approach

Hristo N Djidjev

Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Sofia, Bulgaria
Los Alamos National Laboratory, CCS-3, Los Alamos, NM 87545, United States of America

E-mail: djidjev@parallel.bas.bg

Keywords: quantum annealing, augmented Lagrangian method, quadratic penalty method, D-Wave, Ising problem, logical qubits, broken chains

Original Content from
this work may be used
under the terms of the
Creative Commons
Attribution 4.0 licence.

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.



Abstract

Solving optimization problems on quantum annealers (QA) usually requires each variable of the problem to be represented by a connected set of qubits called a logical qubit or a chain. Chain weights, in the form of ferromagnetic coupling between the chain qubits, are applied so that the physical qubits in a chain favor taking the same value in low energy samples. Assigning a good chain-strength value is crucial for the ability of QA to solve hard problems, but there are no general methods for computing such a value and, even if an optimal value is found, it may still not be suitable by being too large for accurate annealing results. In this paper, we propose an optimization-based approach for producing suitable logical qubits representations that results in smaller chain weights and show that the resulting optimization problem can be successfully solved using the augmented Lagrangian method. Experiments on the D-Wave Advantage system and the maximum clique problem on random graphs show that our approach outperforms both the default D-Wave method for chain-strength assignment as well as the quadratic penalty method.

1. Introduction

Quantum annealing (QA) is a method for solving optimization problems on special quantum devices that is targeting NP-hard problems that are difficult to solve on conventional computers [10]. The largest commercially available QA, which are designed by the company D-Wave Systems, have thousands of qubits, and can propose high-quality solutions of quadratic optimization problems of the type

$$\text{minimize } Is(\mathbf{s}) = \sum_{i < j} J_{ij} s_i s_j + \sum_i h_i s_i, \quad (1)$$

called Ising problem, where $\mathbf{s} = (s_1, \dots, s_N)$ and $s_i \in \{-1, 1\}$ are the variables. When variables of problem (1) are restricted to $s_i \in \{0, 1\}$, the corresponding formulation is called a quadratic unconstrained binary optimization (QUBO) problem. The two formulations can easily be converted into each other by using a linear transformation of the variables, but D-Wave hardware natively implements the Ising version.

Problem (1) is NP-hard to solve and many other NP-hard problems can be easily reformulated as Ising and QUBO problems [14].

In order to solve an Ising problem (1) on a D-Wave QA, the coefficients of the problem are sent to the device, which solves it multiple times using QA and return the obtained solutions called *samples*, which are analyzed in postprocessing, and the sample resulting in a lowest value of the Ising function is chosen as a solution. This is justified by the fact that the QA itself takes a small portion of the total solution time, which also includes the time to map the problem on the hardware and to program it.

For mapping the Ising problem on the quantum processing unit (QPU), the coefficients h_i , called linear biases, have to be mapped to distinct qubits of the QPU, while each coefficient J_{ij} , called a quadratic bias, has to be mapped to a coupler linking the qubits corresponding to i and j . Such a coupler may not exist in the

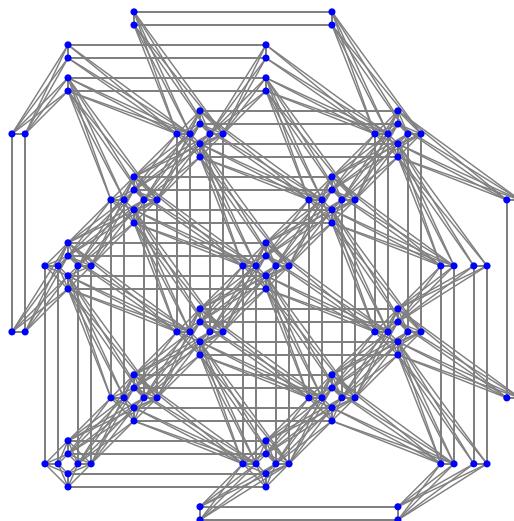


Figure 1. The connection pattern of a Pegasus graph, $P(3)$, of order 3. Blue dots are qubits and connecting edges are couplers. Advantage_system4.1 computer, used for experiments in this paper, is based on the $P(16)$ Pegasus graph and has 5627 working qubits and 40 279 working couplers (some qubits and couplers are disabled due to defects).

QPU, however, as the qubits in current D-Wave devices are not fully connected. Figure 1 shows the connection pattern of the Pegasus graph of the current largest QA, the Advantage™ system, for which the average number of couplers for a qubit is about 15. In contrast, for many optimization problems, the graph corresponding to the quadratic biases is very dense, meaning that a nonzero coupler J_{ij} exists for most of the possible pairs (i, j) . To address this issue, the Ising problem is minor-embedded into the QPU. Specifically, consider the graph with vertices $\{1, \dots, N\}$ and undirected edges $\{(i, j) \mid J_{ij} \neq 0\}$, which we refer to as the Ising graph. Define as a hardware graph the graph with vertices the qubits of the QPU and an edge between each pair of vertices whose corresponding pair of qubits are connected by a coupler. Then, a minor-embedding of the Ising graph into the hardware graph is a mapping M of the vertices of the Ising graph into connected subsets of vertices of the hardware graphs that are mutually disjoint and, for each edge (i, j) of the Ising graph, there is an edge (i', j') of the hardware graph such as $i' \in M(i)$ and $j' \in M(j)$. The sets of qubits corresponding to the sets vertices $M(i)$ are called logical qubits or chains for short. Finding such an embedding is NP-hard in general, but good heuristic algorithms are available [4, 5].

In order for the set of qubits corresponding to a logical qubit to faithfully represent a single qubit, the qubits should be required to take the same value at the end of the QA process. For that end, each coupler (i, j) from a chain is assigned a negative bias, say $-\mu$, favoring physical qubits i and j to take the same value in a minimization problem. The magnitude μ of that bias, controlled by D-Wave's parameter `chain_strength`, is very important. If μ is too low, then it will not be able to enforce that the physical qubits in a chain take the same value. If μ is too large, then the accuracy of the representation of the Ising problem coefficients in the hardware will be reduced since, after normalization to the ranges accepted by the hardware, some problem coefficients may become too small for the limited resolution of the QPU. Picking a suitable value for the chain strength is usually a trial-and-error process and, although the D-Wave software offers a heuristic based method, called `uniform_torque_compensation`, for setting it, it does not always work well, resulting in so-called broken chains, or logical qubits whose physical qubits have obtained both values $+1$ and -1 . Such broken chains are resolved in postprocessing by a heuristic method such as the D-Wave's default method `majority_vote`, which we use in our experimental analysis. For more advanced methods for resolving broken chains, see [15].

Getting broken chains is a major obstacle to improving the accuracy and scalability of the QA [6]. In this paper, we propose an optimization-based approach for finding good representations of the chains that uses not only quadratic biases on the couplers, as the previous methods, but also linear ones on the qubits. For this end, we represent the problem of minimizing an Ising function (1) under the condition that all qubits in a chain should take the same value as a constrained optimization problem. However, the Ising model (1) does not allow constraints, so the constraints have to be added to the objective function. The typical method for solving constrained problems on D-Wave is the penalty method (PM), which adds a constraint $\mathbf{a}\mathbf{s} = \mathbf{b}$ as a quadratic penalty $\mu/2(\mathbf{a}\mathbf{s} - \mathbf{b})^2$. In the case of logical qubit representation, this results in using negative bias $-\mu$ on the chain's couplers, and hence the PM produces a similar representation of the logical qubits as the one discussed above and suffers similar drawbacks. We show in this paper that better results can be achieved

if the augmented Lagrangian method (ALM) is used instead of the penalty one. The ALM, which we discuss in more detail in a next section, adds a linear Lagrangian term $\lambda^T(\mathbf{a}\mathbf{s} - \mathbf{b})$ to the quadratic penalty, and uses an iterative procedure to find optimal or near-optimal values for λ and μ .

The main contributions of this paper are the following.

- (i) We show that a good representation of the logical qubits can be constructed by defining both linear and quadratic biases and using the ALM to find the values of such biases.
- (ii) We test our implementation on the maximum clique (MC) problem (MCP), an important NP-hard problem, and on graphs of up to 100 vertices.
- (iii) We show that our implementation outperforms both the default D-Wave representation of logical qubits as well as the quadratic PM.
- (iv) We show that one can optimize and store values of the augmented Lagrangian coefficients for a whole class of graphs and then use them to solve problems on any graph of the class.

The rest of the paper is organized as follows. In section 2, we overview previous results related to this work. In section 3, we describe our methods and present the pseudocodes of their implementation. In section 4, we analyze the proposed implementation of logical qubits and compare it with other algorithms. Finally, we conclude with a summary and a list of open problems.

2. Previous work

Several studies have focused on empirically determining good chain strength values for certain problems. King and McGeoch [12] consider the algorithm-engineering aspects of solving optimization problems on QA including searching for a best value for the chain-strength parameter, which they empirically analyze on a subclass of 3-SAT instances. Pudenz [17] uses empirical analysis to compare different strategies for setting parameters of logical qubits such as the distribution of the Ising problem biases on logical qubits and couplers and setting chain weights. Grant and Humble [9] benchmark on a set of instances of portfolio optimization problems the probability of chain-breaks and how they are impacted by the specific embedding. Lee [13] use simulated annealing and exact optimization methods to estimate optimal chain strength for hypothetical 2D $L \times L$ architectures for $L = 4, 6$, and 8 , where a third dimension is added to the 2D architecture to allow chains between the main qubits. Venturelli *et al* [20] study optimal parameter settings on fully-connected Ising models corresponding to the Sherrington–Kirkpatrick spin-glass model. They establish that, for their problem, the optimal chain strength scales as \sqrt{N} , where N is the number of the logical variables. For general Ising problems, the same method implies that the optimal chain strength scales proportionally to $\tau\sqrt{N}$, where τ is the root mean square (RMS) of the quadratic couplers, which formula is used in the D-Wave Ocean software. Such a setting for the chain strength, along with the anneal duration and the minor embedding parameters, were analyzed in [18] for spin glasses and channel communication problems, and interdependence and strategies for tuning these three parameters were proposed. Andriyash *et al* [1] address the disparate tunneling dynamics of chains with different lengths in a problem by utilizing a technique called annealing offsets, where the anneal of qubits belonging to longer chains is delayed, thereby increasing the relative tunneling energy of such chains. For simulated-annealing-based Ising machines, Shirai *et al* [19] propose a formula based on statistical mechanics for the coupling strength of each chain and experimentally show that their proposed methods outperforms two existing methods on a set of benchmarking problems.

3. Methods

We start this section by formulating the logical qubits representation problem as an optimization problem and describe the generic quadratic penalty and ALMs.

3.1. Representing logical qubits as constraints

An alternative way of thinking about enforcing the condition that all physical qubits representing a given logical qubit Q take the same value when solving an optimization problem P is to consider it as an (additional) constraint added to P . In the case P is a constrained problem, one can first convert it into an unconstrained one using the method described in the previous section. Hence, we can assume without loss of generality that P is itself an Ising problem of the type (1), i.e. unconstrained. Assume a minor embedding for P maps each variable s_i onto a set $C_i = \{s_i^1, \dots, s_i^{k_i}\}$, where s_i^k are the physical qubits, and let the set of active couplers joining pairs of vertices from C_i , which we call chain couplers, be E_i . We assume that the chains, i.e. each graph (C_i, E_i) , is connected. Define $\mathcal{E} = \bigcup_i E_i$ denote by C_{ij} the set

$\{(k, l) \in \mathcal{E} \mid s_i^k \in C_i, s_j^l \in C_j\}$, i.e. the set of couplers between the chains C_i and C_j , which we call a logical coupler. Replacing each s_i by C_i , $i = 1, \dots, N$, in $Is(\mathbf{s})$ from (1), we get the Ising function

$$Is'(\mathbf{s}') = \sum_{i < j} \sum_{(k, l) \in C_{ij}} J_{ij}^{kl} s_i^k s_j^l + \sum_i \sum_{s_i^k \in C_i} h_i^k s_i^k, \quad (2)$$

where J_{ij}^{kl} and h_i^k are appropriately chosen coefficients and \mathbf{s}' is the vector of all variables s_i^k , $k \in C_i$, $1 \leq i \leq N$. The *uniform spreading approach* [18], used by D-Wave's Ising embedding tool, which we also use in the paper, assigns equal weights for each logical qubit and coupler. Hence, we define

$$J_{ij}^{kl} = J_{ij}/|C_{ij}|, \quad h_i^k = h_i/|C_i|. \quad (3)$$

For enforcing that all variables s_i^k in a logical qubit take the same value in an optimal solution, we can add the constraints

$$s_x = s_y \text{ for each } (s_x, s_y) \in \bigcup_i E_i, \quad (4)$$

i.e. that the qubits at the endpoints of each chain coupler should take the same value. Since the graph (C_i, E_i) is connected, (4) ensures that all variables corresponding to any given chain will take the same value.

However, the resulting problem ‘minimize $Is'(\mathbf{s}')$ subject to (4)’ cannot be solved on D-Wave, as it is, since it is not unconstrained. To convert it into an unconstrained one, (4) is usually added to the objective function as a penalty term, resulting in a problem

$$\text{minimize } Is'(\mathbf{s}') + \Pi(\mathbf{s}'), \quad \mathbf{s}' \in \{-1, 1\}^n, \quad (5)$$

where $\Pi(\mathbf{s}')$ is a penalty function that takes a value zero, if all constraints are satisfied, and a positive value large enough to prevent assignments that violate some of the constraints to be an optimal solution, otherwise. In the next subsection, we will discuss two methods for defining a suitable function Π .

3.2. Penalty and ALMs for constrained optimization

The most common constraints for binary optimization problems are equalities and inequalities between variables. We consider here domain $\mathcal{D} = \{-1, 1\}^N$ and equality-constrained problems since this is the type of our problem (5), but the described methods were originally designed for continuous domains and it can be generalized for inequality-constrained or mixed problems. We consider in this subsection the constrained problem

$$\text{minimize } f(\mathbf{s}) \text{ subject to } c(\mathbf{s}) = 0, \quad \mathbf{s} \in \mathcal{D}. \quad (6)$$

3.2.1. Penalty method

The quadratic PM uses a penalty function $\Pi(\mathbf{s}) = \mu/2 \|c(\mathbf{s})\|^2$, transforming (6) into the unconstrained problem

$$\text{minimize } f(\mathbf{s}) + \frac{\mu}{2} \|c(\mathbf{s})\|^2, \quad \mathbf{s} \in \mathcal{D}. \quad (7)$$

Choosing the penalty factor $\mu > 0$ large enough will enforce $c(\mathbf{s}) = 0$ in each optimal solution of (7) (assuming (6) does have a feasible solution) thereby producing an optimal solution of (6). But if the penalty factor is chosen too large, that will result in some very small in absolute value coefficients of the Ising problem submitted to the QA, since all coefficients are normalized in order to satisfy hardware-imposed limits. Combined with the analog nature of the QA and the finite digital-to-analogue converter quantization step size, such large values of μ would negatively affect the quality of the solution of problem (7). Even in the continuous case, large values of μ result in ill-conditioning. Therefore, finding a suitable penalty factor is important for the success of the method.

We will use in our implementation of the PM an iterative procedure that computes a suitable penalty factor, which starts with a small value of μ and increases that value in each iteration until all constraints are satisfied. A pseudocode for the method is given in algorithm 1.

Algorithm 1. Quadratic PM.

Input: initial $\mu > 0$, increase factor $\alpha > 1$
Output: final μ for penalty function

```

1: repeat
2:    $s \leftarrow \arg \min_{s \in \mathcal{D}} \left( f(s) + \frac{\mu}{2} \|c(s)\|^2 \right)$             $\triangleright$  Solve on QA
3:   if  $c(s) > 0$ 
4:      $\mu \leftarrow \alpha \mu$ 
5:   end if
6: until  $c(s) = 0$  or iterations limit reached            $\triangleright$  Other stopping criteria possible

```

3.2.2. Augmented Lagrangian method

The ALM combines the idea of the PM with the method of the Lagrangian multipliers. The Lagrangian multipliers method tries to solve problem (6) by constructing the function $\mathcal{L}(s, \lambda) = f(s) + \lambda^T c(s)$, where λ is a vector of coefficients called Lagrangian multipliers, and looks at its stationary points, i.e. points where all partial derivatives are zero, including the one with respect to the Lagrangian multipliers λ . These points contain all optimal solutions of the original problem (6). But the Lagrangian multipliers method cannot be directly used with D-Wave, since $\mathcal{L}(s, \lambda)$ is not an Ising problem as λ consists of real numbers, which have to be computed by an iterative method. However, a simple procedure like that of Algorithm 1, which increases the value of the coefficient at each iteration, will not work since (i) the exact value of λ has to be used, unlike PM where the value of μ just needs to be large enough; (ii) unlike μ , which is a single number, λ is a vector, which makes its exact estimation more difficult. For that reason, the ALM, which augments $\mathcal{L}(s, \lambda)$ with a quadratic penalty term to produce the unconstrained problem

$$\text{minimize } f(s) + \lambda^T c(s) + \frac{\mu}{2} \|c(s)\|^2, \quad s \in \mathcal{D}, \quad (8)$$

where λ is a vector of dimension the number of logical qubits, is a better alternative. It also results in lower values of μ , compared with PM. A pseudocode of a standard implementation of ALM [2] is given in algorithm 2.

Algorithm 2. ALM.

Input: initial $\mu > 0$, initial λ , increase factor $\alpha > 1$
Output: final μ, λ for the augmented Lagrangian function

```

1: repeat
2:    $s \leftarrow \arg \min_{s \in \mathbb{B}^N} \left( f(s) + \lambda^T c(s) + \frac{\mu}{2} \|c(s)\|^2 \right)$             $\triangleright$  Solve on QA
3:   If  $c(s) \neq 0$  then
4:      $\lambda \leftarrow \lambda + \mu c(s)$ 
5:      $\mu \leftarrow \alpha \mu$ 
6:   end if
7: until  $c(s) = 0$  or iteration limit reached            $\triangleright$  Other stopping criteria possible

```

In the next subsection, we describe the application of the methods to the problem of implementing logical qubits.

3.3. Applying the methods to QA and logical qubits implementation

Assume that we are given to solve, on a QA, the constrained problem determined by equations (2)–(4), i.e. the problem

$$\text{minimize } Is'(\mathbf{s}') = \sum_{i < j} \frac{J_{ij}}{|C_{ij}|} \sum_{(k,l) \in C_{ij}} s_i^k s_j^l + \sum_i \frac{h_i}{|C_i|} \sum_{s_i^k \in C_i} s_i^k \quad (9)$$

$$\text{subject to } s_x = s_y \text{ for each } (s_x, s_y) \in \bigcup_i E_i. \quad (10)$$

Converting (9)–(10) into an unconstrained problem of type (8) results into the problem

$$\text{minimize } Is^*(\mathbf{s}') = Is'(\mathbf{s}') + \sum_{(s_x, s_y) \in \mathcal{E}} \lambda_{xy} (s_x - s_y) + \frac{\mu}{2} \sum_{(s_x, s_y) \in \mathcal{E}} (s_x - s_y)^2, \quad (11)$$

where $\mathcal{E} = \bigcup_i E_i$. Applying the square formula and using the fact that $s^2 = 1$ for $s \in \{-1, 1\}$, we simplify $Is^*(s')$ as

$$Is^*(s') = Is'(s') + \sum_{(s_x, s_y) \in \mathcal{E}} (\lambda_{xy} s_x - \lambda_{xy} s_y - \mu s_x s_y) + \mu |E|. \quad (12)$$

Note $\mu |E|$ is a constant that can be ignored when minimizing with respect to s' .

Formula (12) leads to the following implementation of algorithm 2 for the case of logical qubits given on algorithm 3. The specific stopping criterion we use is that there are no broken chains or the number of iteration exceeds 20.

Algorithm 3. ALM.

Input: Ising model $Is(s)$, initial $\mu_0 > 0$, initial λ_0 , increase factor $\alpha > 1$

Output: Final μ, λ for the augmented Lagrangian function

1. Find an embedding of $Is(s)$ and construct the transformed Ising $Is'(s')$ using the D-Wave's Ocean software function `embed_ising` but setting the parameter `chain_strength` to 0 (so that we can define chain strength by ALM). Let $J'(s_x, s_y)$ and $h'(s_x)$ denote the coefficients in front of $s_x s_y$ and s_x in $Is'(s')$, respectively, and let $J^*(s_x, s_y)$ and $h^*(s_x)$ denote the corresponding coefficients of $Is^*(s')$.
2. Set suitable initial values to the parameters of Algorithm 2. In our implementation, we set $\mu \leftarrow \mu_0 = 1.5$, $\lambda \leftarrow \lambda_0 = 0$, $\alpha \leftarrow 1.1$.
3. Construct the set \mathcal{E} of pairs of same-chain qubits that are connected by a coupler.
4. Initialize $Is^*(s') \leftarrow Is'(s')$ and repeat steps (i)–(v) until a stopping criterion has been met.
 - (i) For each $(s_x, s_y) \in \mathcal{E}$ make the updates

$$h^*(s_x) \leftarrow h'(s_x) + \lambda_{xy}$$

$$h^*(s_y) \leftarrow h'(s_y) - \lambda_{xy}$$

$$J^*(s_x, s_y) = \mu$$
 - (ii) Submit $Is^*(s')$ to the QA and get a list of samples (proposed solutions).
 - (iii) Find the sample σ resulting in a best solution.
 - (iv) Update λ as follows


```
for (s_x, s_y) in E do
  (sigma_x, sigma_y) <- values of (s_x, s_y) from sigma
  lambda_xy <- lambda_xy + mu * (sigma_x - sigma_y)
end for
```
 - (v) $\mu \leftarrow \alpha \mu$

ALM is considered a very good choice in practice for a large set of applications [3], but its application to logical qubits has distinctive aspects that may affect its effectiveness. Such features include the following.

1. The original method has been developed for continuous variables. In that case and under mild conditions, ALM converges to a global optimum regardless of the initial values of the Lagrange multipliers [8]. However, in the case with discrete variables like ours, there are no theoretical results available that guarantee optimality.
2. The number of constraints in our case can be quite large (it is equal to the number of variables). ALM has been shown to work well for number of constraints of order hundreds [7] in the *continuous* case, which gives a hope that it may work in our case as well. But it needs to be checked.
3. QA are stochastic samplers, which means that the outcome of measuring the quantum state at the end of the annealing is random. As a consequence, the result of each iteration depends not only on the parameters λ and μ , but also on chance. This is further exacerbated by the fact that the annealer is susceptible to noises from the environment, hardware leaks, and other interactions that may substantially affect the outcome of the QA process [16].

For that reason, it is important to experimentally analyze the practical effectiveness of our algorithm, which we address in the section 4.

3.4. Using ALM on a class of problems

So far, we have applied ALM on a single problem to find a representation of the logical qubits for that particular problem. Here we propose a modification that allows ALM to construct implementations of logical qubits that can then be used to solve a set of problems from a given class. In order to allow that, the problems in the class should have the same number of variables (logical qubits). The algorithm is similar to algorithm 3, but we solve each problem in the set separately and, when we update the Lagrangian coefficients λ , we take into account the violated constraints (broken chains) in *all* of the solutions. The details are given in algorithm 4.

Algorithm 4. ALM for a set of problem of the same size.**Input:** A set \mathcal{S} of Ising models $Is_1(s), \dots, Is_k(s)$, initial $\mu_0 > 0$ and λ_0 , increase factor $\alpha > 1$ **Output:** Final μ, λ for the set of Ising models1. Find an embedding of a complete graph of $\dim(s)$ vertices and use it to embed each graph of \mathcal{S} , and construct the set \mathcal{S}' of transformed Ising models $Is'_1(s), \dots, Is'_k(s)$ using Ocean's `embed_ising` function while setting the parameter `chain_strength` to 0. Let J', h', J^* , and h^* be defined as in Algorithm 3.2–3. Set μ, λ, α , and \mathcal{E} as in Algorithm 3.4. Initialize $Is_i^*(s') \leftarrow Is'_i(s')$ for $1 \leq i \leq k$ and denote $\mathcal{S}^* = \{Is_1^*(s), \dots, Is_k^*(s)\}$.

Repeat steps (i)–(v) until a stopping criterion has been met.

(i) Update h^* and J^* as in Algorithm 3.(ii) Submit each problem from \mathcal{S}^* to the QA and get a list of samples for each of them.(iii) Find the sample σ_i resulting in the best solution of Is_i^* for $1 \leq i \leq k$.(iv) Update λ as follows **for** $(s_x, s_y) \in \mathcal{E}$ **do** $(\sigma_x^i, \sigma_y^i) \leftarrow$ values of (s_x, s_y) from σ_i for $1 \leq i \leq k$ $\lambda_{xy} \leftarrow \lambda_{xy} + \mu \sum_{i=1}^k (\sigma_x^i - \sigma_y^i) / k$ **end for**(v) $\mu \leftarrow \alpha \mu$

4. Results

4.1. Experimental set-up

We test the proposed method on instances of the MCP. A clique is a graph that is fully connected, i.e. such that there is an edge between each pair of vertices. A MC of a graph G is a subgraph of G that is a clique of maximum size. The MCP asks to find a MC in G and it is an *NP*-hard problem [11].

For constructing input graphs for MCP, we generate Erdős–Rényi random graphs using probability parameter $p = 0.5$ and number of vertices $n \in \{50, 75, 100\}$. For each combination (n, p) of values, we generate five random $G(n, p)$ graphs.

In order to solve an MCP on a D-Wave annealer, we have to encode it as a QUBO or Ising problem. For this end, we define a binary variable $x_i \in \{0, 1\}$ for each vertex i of G that indicates whether i is included in the proposed clique or not. Then, the MCP can be formulated as

$$\text{minimize } Q = -A \sum_{i \in V} x_i + B \sum_{(i,j) \in \bar{E}} x_i x_j, \quad (13)$$

where \bar{E} is the set of edges of the complement of G and the positive constants A and B should satisfy $A < B$ [14]. In (13), the first term encodes the objective function, which is the clique size (weighted by $-A$), and the second one encodes the constraint function as a penalty, which takes value at least $B = B \cdot 1$ if the set of vertices i with $x_i = 1$ does not form a clique. In our implementation, we choose $A = 1$ and $B = 2$. Since problem (13) is a QUBO while algorithm 3 requires an Ising representation, we transform (13) into an Ising problem using D-Wave's Ocean function `qubo_to_ising`.

For our experiments, we use `Advantage_system4.1` QA available through the D-Wave's Leap quantum cloud service. The annealing settings we use are 1000 for the number of samples and 100 μ s for the annealing time. The chain strength for the ALM and PM implementations are determined as a result of running algorithms 3 and 1, respectively, and for the default logical qubits implementation we use as a chain strength the value computed by D-Wave Ocean's recommended `uniform_torque_compensation` function, which determines a chain strength that usually results in a small number of broken chains, with a prefactor set to the default value of 1.414. For handling broken chains, we use the default `majority_vote` method, which assigns a value (+1 or -1) to a logical qubit equal to the the most common value of its physical qubits.

Next, we compare the performance of ALM against PM and the default D-Wave method with uniform torque compensation.

4.2. Number of chain breaks per iteration

As the goal of the algorithms we described is to deal with chain breaks (CB) in logical qubits, our first goal is the examine how the number of CB changes during the iterations and how ALM compares with the PM implementation. Figure 2 shows the average number of CB per iteration on the left, and that number for PM on the right. Red 'X' indicates the optimal iteration in terms of the average value of the MC number. We observe that the number of CB is quite large in the beginning, and that number goes down to zero in all cases, more steeply in the beginning and flattening out for the ALM method, while the reduction is more gradual and with roughly constant slope for PM. Also, for both methods and for all graph sizes, the best value for the

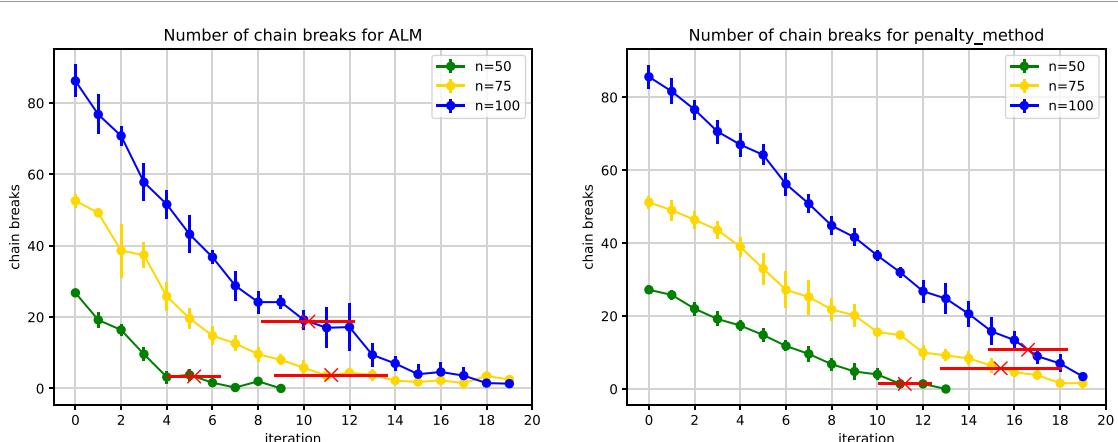


Figure 2. Number of chain breaks per iteration for ALM (left) and for PM (right) for random graphs of various sizes averaged over five input problems. Dots correspond to the average values and the bars indicate the standard deviation of the values, while 'x' indicates the average iteration number at which the highest value for the clique number is obtained.

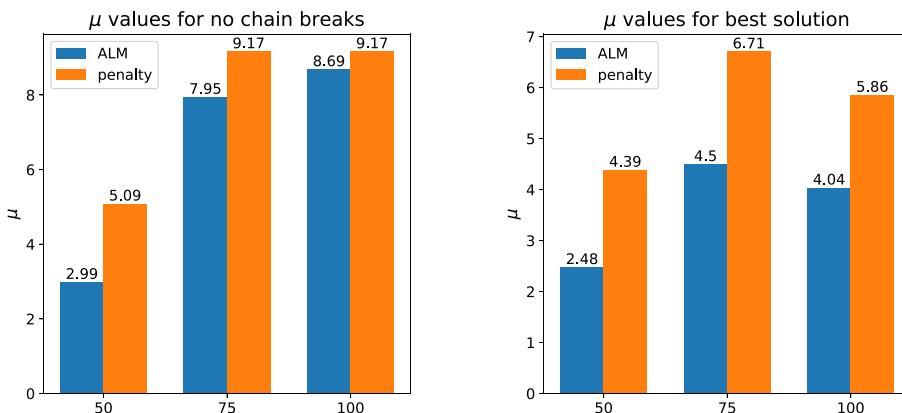


Figure 3. Average values of the penalty factor, μ , at the completion of ALM and PM, i.e. when the number of CB becomes zero (left), and the corresponding values at the iterations where a best value for the MC number is found (right).

MC is found a few iterations before the number of CB gets down to zero, rather than at the last iteration. That can be explained with the trade-off between reducing the number of CB and increasing the value of the penalty parameter μ . Specifically, on one side, the elimination of the broken chains makes the Ising model a better match to the original problem (1). On the other side, each iteration increases the value of μ by a factor of 1.1, and such an increase negatively impacts the accuracy of the Ising model (12) solution. Hence, the best solution may be found at an iteration at which the number of CB is not zero, but μ is only moderately large.

Finally, comparing the number of iterations between ALM and PM to get to a best solution, we see ALM uses fewer iterations than PM in all three cases (5.2 vs 11.2 for $n = 50$, 11.2 vs 15.4 for $n = 75$, and 10.2 vs 12.8 for $n = 100$). Note that 'best' here means the best solution over all iterations of a particular run. While, for $n = 100$ case, 10.2 vs 12.8 looks like ALM has only a modest advantage over PM in terms of number of iterations, the difference in the quality of solution is much more substantial, as we will see in section 4.4.

4.3. Values of the penalty factor

Next we compare the values of the penalty factor, μ , that ALM and PM obtain at the completion of the algorithms, as well as those values at the iteration where the best value of the MC number is achieved, averaged over five input problems. The results are shown in figure 3. We can see that ALM gets consistently better (lower) values, with reductions over PM 41%, 13%, 5%, for $n = 50$, 75, and 100 and the last-iteration case, and reductions of 44%, 33%, 31%, for the best-MC iteration.

4.4. Values of the largest clique numbers found

Probably the most important comparison criterion is whether and how much the new algorithm can help to get better solutions than the previous ones. In this subsection, we show comparison between ALM, PM, and

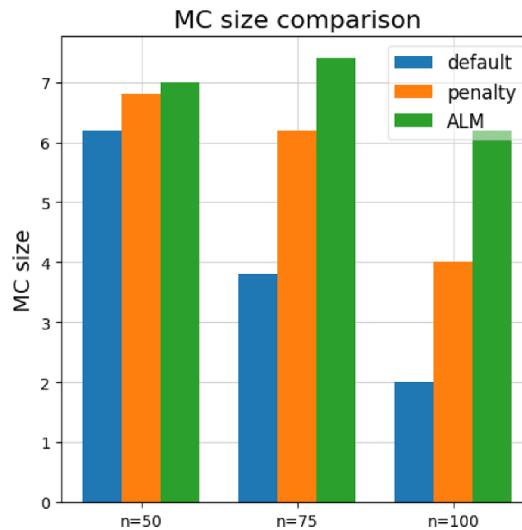


Figure 4. Comparison between default, penalty, and ALM methods on random test graphs of various sizes.

the standard default method (SM) offered by the D-Wave software. The default `uniform_torque_compensation` (UTC) method by D-Wave's Ocean software, used by SM, is a heuristic procedure that assigns a chain strength (penalty factor) as $c\tau\sqrt{N}$, where τ is the RMS of the quadratic couplers, N is the number of logical qubits, and c is a user-defined constant prefactor. We use the default value of $c = 1.414$.

Figure 4 shows the results of comparing SM, PM, and ALM on random test graphs of various sizes. Each bar shows the average MC size found by the corresponding algorithm and graph size. For instances where no valid clique is returned, we assign MC size of zero. We can see that ALM consistently gets better values for the MC number than the other methods. The biggest difference is for the largest size ($n = 100$), where the average clique size found by ALM is twice larger than the one found by PM and more than three times larger than the one found by SM. The large difference in performance between ALM and PM is due to the fact that ALM arrives at smaller values of μ , which helps in the accuracy of the QA, and, additionally, it uses Lagrangian coefficients λ to help enforce the chain constraints, while PM relies only on μ .

Comparing SM and PM we observe that the default values for μ used by SM are relatively good, but only for the smallest and easiest case $n = 50$. The performance of SM relative to PM deteriorates as the problem gets more difficult with the size increase. That means that SM and the UTC heuristic do not use the full potential of PM and improvement may be possible with a more sophisticated and better heuristic or by just using an iterative algorithm of the type of algorithm 1 for penalty factor estimation.

4.5. ALM on a class of problems

In order to use algorithm 4 on a set of problems, all of the problems in the set need to have the same number of variables, which in the case of the MC problem means that the input graphs have to be of the same size. In our experiments, we choose our set \mathcal{S}_n to be the set of Erdős–Rényi random graphs $R(n, p)$ with probability parameter $p = 0.5$ and a fixed n . As in the other experiments in this section, n takes values in $\{50, 75, 100\}$, and we generate five graphs for each n . To compute values for μ and λ , for each $n \in \{50, 75, 100\}$, we apply algorithm 4 on a set of ten randomly chosen graphs from \mathcal{S}_n and store the values for later use.

To give an idea about a typical range of values for the Lagrangian coefficients λ , we round their values to the closest integer and put them in bins, for each of the three graph sizes. The resulting histograms are shown in figure 5. We can see that zero is the most common value and that most of the values are between $[-4, 4]$.

Furthermore, to illustrate the distribution of λ values on the edges of a typical chain, we picked one chain from the embedding for $n = 100$ and plotted on each edge of that chain the corresponding λ value, as shown on figure 6. While there are both positive and negative values on the edges, the signs are irrelevant and only the magnitudes matter, because the signs depend on the directions of the corresponding edges, which are arbitrarily oriented. These λ values, along with the μ value, which in this particular case is -3.9 , are used to compute the additional linear and quadratic biases on qubits and couplers corresponding to the chain constraints, using equation (12). The resulting additional biases are illustrated on figure 7. As one might expect, there are portions of the chain with mostly positive biases and portions where the negative biases dominate. Note that these linear biases are not necessarily the ones in the final Ising problem sent to the QA as in addition to the chain biases there are also linear biases coming from the problem Ising (9), and the two

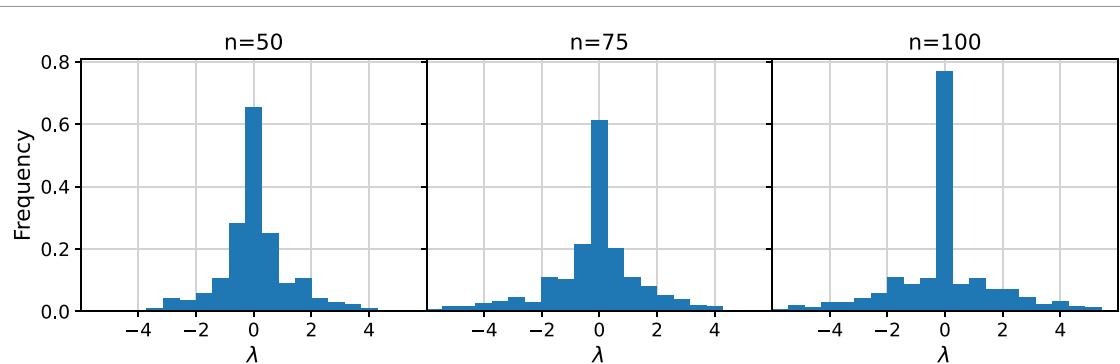


Figure 5. Distribution of the λ values for the ALM_set algorithm. The density parameter of the `matplotlib.pyplot.hist` method was used to normalize bin heights so that the integral of the histogram equals one.

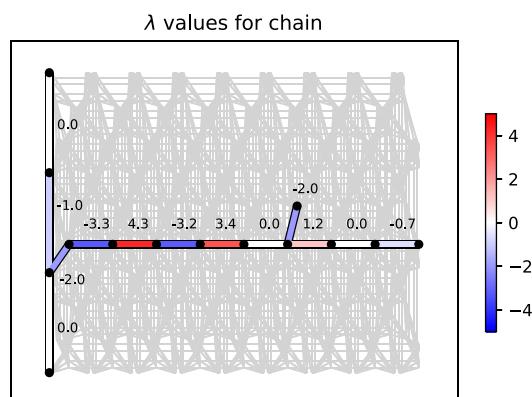


Figure 6. Values of the λ parameter computed for a specific randomly picked chain from the embedding for $n = 100$ rounded to one decimal place.

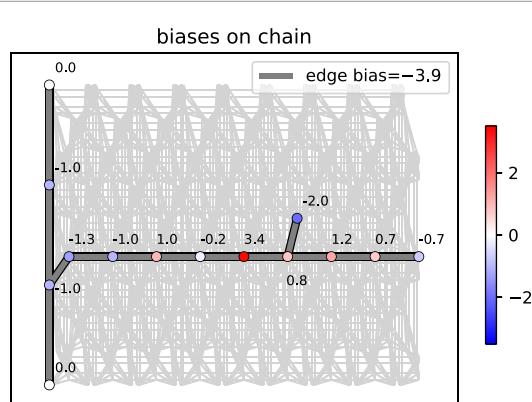
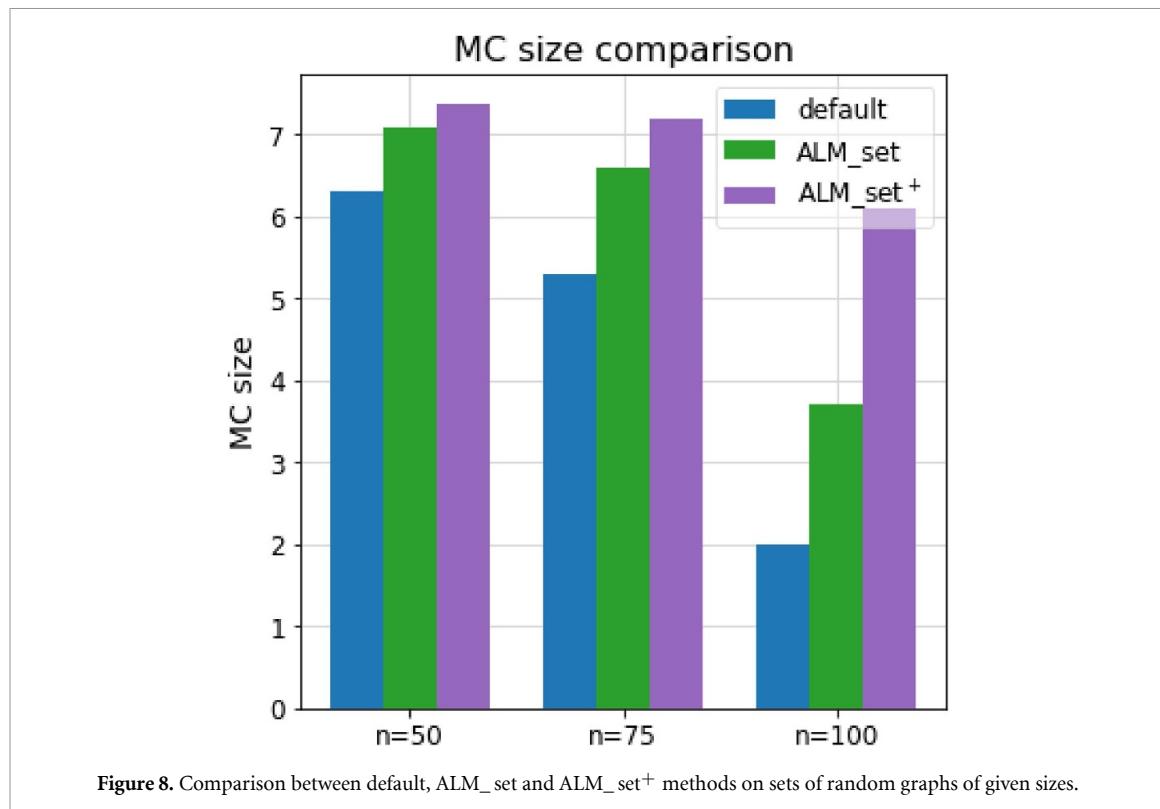


Figure 7. Linear and quadratic biases for the same chain as in figure 6. Biases do not sum up to zero in the figure due to the rounding errors.

biases are added together. On the other hand, the quadratic chain biases, set to -3.9 for that particular chain, are the final ones.

In order to compare the accuracy of the algorithms, we run them on different sets S_n of random graphs and compare the sizes of the MCs found. The first of these algorithms is the default D-Wave algorithm, SM, considered in the previous subsection. The second algorithm, ALM_set, uses the representations of the logical qubits defined by the stored values of μ and λ . The difference between ALM from the previous subsection and ALM_set is that the former uses a number of iteration to arrive at suitable values for μ and λ for the particular graph, while the latter takes just a single annealing call. The third algorithm we compare, ALM_set⁺, adds a single augmented Lagrangian iteration after running ALM_set. The rationale for this is that the stored values of μ and λ are optimized for the class S_n as whole, but they may not be well-suited for



any particular graph in that class. We conjecture that, by using μ and λ as initial values, a single iteration may help to adjust the logical qubit representation so that they better match the particular features of the specific graph, and result in better results than ALM_set.

Figure 8 shows the results of this experiment. Since ALM_set⁺ takes the better of two MC values, one from the ALM_set run and one from the single augmented Lagrangian iteration, in order to make the comparison fair, we also run SM and ALM_set twice and take the better of the two MC values. We observe that storing μ and λ values (ALM_set) does lead to significant improvements over SM. However, despite the improvement, ALM_set results for $n = 100$ are significantly worse than ALM results, figure 4. Adding the extra iteration (ALM_set⁺), however, leads to a significant improvement and MC values that roughly match the ones from ALM.

5. Conclusion

In this paper, we showed that the ALM can be used to determine a representation of logical qubits for a given Ising problem in terms of quadratic and linear biases on the chain couplers and qubits by an iterative optimization procedure. We experimentally analyzed the method on the MCP on random graphs of various sizes and observed that our proposed method outperforms quadratic-penalty based methods. With increasing the graph size and the difficulty, the performance gap becomes more significant. We also propose a modification that allows values for the chains' biases for entire classes of graphs to be generated and saved. While this method outperforms the method that uses default chain weights, the version of the method that is adding a single augmented Lagrangian iteration matches the performance of the full augmented Lagrangian version (with 5–11 iterations).

The paper leaves several open problems for future research. Our formulation assumed that, for optimization problem that originally have had constraints, they have already been included in the objective function as penalty terms. But since we are using ALM to find representations of the logical qubits, we can use the same iterations to find better representations of the constraints as well. The problem we used in our experiments, the MCP, has a relatively low penalty factor of two for its constraint, but many other optimization problems need much larger penalty factors and using the ALM method can make a difference. Another possible improvement is use an individual penalty factor μ_i for the i th logical qubit, than a single one as in our implementation.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

This work was supported by grant number KP-06-DB/1 of the Bulgarian National Science Fund and by Grant Number BG05M2OP001-1.001-0003, financed by the Science and Education for Smart Growth Operational Program (2014–2020) and co-financed by the European Union through the European Structural and Investment Funds. The work was also partially supported by the U.S. Department of Energy through the Los Alamos National Laboratory. Los Alamos National Laboratory is operated by Triad National Security, LLC, for the National Nuclear Security Administration of U.S. Department of Energy (Contract No. 89233218CNA000001).

ORCID iD

Hristo N Djidjev  <https://orcid.org/0000-0001-9286-8824>

References

- [1] Andriyash E, Bian Z, Chudak F, Drew-Brook M, King A D, Macready W G and Roy A 2016 Boosting integer factoring performance via quantum annealing offsets *D-Wave Technical Report Series* vol 14
- [2] Bertsekas D P 2014 *Constrained Optimization and Lagrange Multiplier Methods* (New York: Academic)
- [3] Birgin E G and Mario Martínez J 2014 *Practical Augmented Lagrangian Methods for Constrained Optimization* (Philadelphia, PA: SIAM)
- [4] Choi V 2008 Minor-embedding in adiabatic quantum computation: I. The parameter setting problem *Quantum Inf. Process.* **7** 193–209
- [5] Choi V 2011 Minor-embedding in adiabatic quantum computation: II. Minor-universal graph design *Quantum Inf. Process.* **10** 343–53
- [6] Coffrin C J 2019 Challenges with chains: testing the limits of a D-Wave quantum annealer for discrete optimization *Technical Report* Los Alamos National Lab (LANL), Los Alamos, NM
- [7] Evtushenko Y G, Golikov A I and Mollaverdy N 2005 Augmented Lagrangian method for large-scale linear programming problems *Optim. Methods Softw.* **20** 515–24
- [8] Fernández D and Solodov M V 2012 Local convergence of exact and inexact augmented Lagrangian methods under the second-order sufficient optimality condition *SIAM J. Optim.* **22** 384–407
- [9] Grant E and Humble T S 2022 Benchmarking embedded chain breaking in quantum annealing *Quantum Sci. Technol.* **7** 025029
- [10] Kadokawa T and Nishimori H 1998 Quantum annealing in the transverse ising model *Phys. Rev. E* **58** 5355
- [11] Karp R M 1972 Reducibility among combinatorial problems *Complexity of Computer Computations (The IBM Research Symposia Series)*, eds E M Raymond and W T James (New York: Plenum Press) pp 85–103
- [12] King A D and McGeoch C C 2014 Algorithm engineering for a quantum annealing platform (arXiv:1410.2628)
- [13] Lee H 2022 Determination of chain strength induced by embedding in D-Wave quantum annealer (arXiv:2209.12166)
- [14] Lucas A 2014 Ising formulations of many NP problems *Front. Phys.* **2** 1–15
- [15] Pelofske E, Hahn G and Djidjev H 2020 Advanced unembedding techniques for quantum annealers 2020 *Int. Conf. on Rebooting Computing (ICRC)* (IEEE) pp 34–41
- [16] Pelofske E, Hahn G, and Djidjev H N 2022 Noise dynamics of quantum annealers: estimating the effective noise using idle qubits (arXiv:2209.05648)
- [17] Pudenz K L 2016 Parameter setting for quantum annealers 2016 *IEEE High Performance Extreme Computing Conf. (HPEC)* pp 1–6
- [18] Raymond J, Ndiaye N, Rayaprolu G and King A D 2020 Improving performance of logical qubits by parameter tuning and topology compensation 2020 *IEEE Int. Conf. on Quantum Computing and Engineering (QCE)* (IEEE Computer Society, Los Alamitos, CA, USA) pp 295–305
- [19] Shirai T, Tanaka S and Togawa N 2020 Guiding principle for minor-embedding in simulated-annealing-based ising machines *IEEE Access* **8** 210490–502
- [20] Venturelli D, Mandrà S, Knysh S, O’Gorman B, Biswas R and Smelyanskiy V 2015 Quantum optimization of fully connected spin glasses *Phys. Rev. X* **5** 031040