

PAPER • OPEN ACCESS

# Boosting quantum annealing performance through direct polynomial unconstrained binary optimization

To cite this article: Sebastian Nagies *et al* 2025 *Quantum Sci. Technol.* **10** 035008

View the [article online](#) for updates and enhancements.

## You may also like

- [Quantum annealing for industry applications: introduction and review](#)  
Sheir Yarkoni, Elena Raponi, Thomas Bäck *et al.*
- [An integrated programming and development environment for adiabatic quantum optimization](#)  
T S Humble, A J McCaskey, R S Bennink *et al.*
- [An algorithm for solving a 1D seismic data inversion problem on quantum or digital annealers in the presence of a priori information on layer parameters specified by arbitrary functions](#)  
Nickolay V Maletin, Anastasiia M Eremenko and Dmitry V Minaev



 **Delft Circuits**  
Pioneering i/o for advanced technologies

**Connect to  
your ambitions**  
Flexible, high-density, cryogenic i/o

# Quantum Science and Technology



## PAPER

### OPEN ACCESS

#### RECEIVED

31 December 2024

#### REVISED

1 April 2025

#### ACCEPTED FOR PUBLICATION

9 April 2025

#### PUBLISHED

22 April 2025

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.



## Boosting quantum annealing performance through direct polynomial unconstrained binary optimization

Sebastian Nagies<sup>1,2,\*</sup> , Kevin T Geier<sup>1,2,3</sup> , Javed Akram<sup>4</sup> , Dimitrios Bantounas<sup>4</sup>, Michael Johanning<sup>4</sup> and Philipp Hauke<sup>1,2,\*</sup>

<sup>1</sup> Pitaevskii BEC Center and Department of Physics, University of Trento, Via Sommarive 14, 38123 Trento, Italy

<sup>2</sup> INFN-TIFPA, Trento Institute for Fundamental Physics and Applications, Trento, Italy

<sup>3</sup> Quantum Research Center, Technology Innovation Institute, PO Box 9639, Abu Dhabi, United Arab Emirates

<sup>4</sup> eleQtron GmbH, Heeserstr. 5, 57072 Siegen, Germany

\* Authors to whom any correspondence should be addressed.

E-mail: [sebastian.nagies@unitn.it](mailto:sebastian.nagies@unitn.it) and [philipp.hauke@unitn.it](mailto:philipp.hauke@unitn.it)

**Keywords:** PUBO, QUBO, 3-SAT, higher-order interactions, quantum computing, quantum optimization, quantum annealing

### Abstract

Quantum annealing aims at solving optimization problems of practical relevance using quantum-computing hardware. Problems of interest are typically formulated in terms of quadratic unconstrained binary optimization (QUBO) Hamiltonians. However, many optimization problems are much more naturally formulated in terms of polynomial unconstrained binary optimization (PUBO) functions of higher order. As we show with various problem examples, leveraging the PUBO formulation can bring considerable savings in terms of required number of qubits. Moreover, in numerical benchmarks for the paradigmatic 3-SAT problem, we find scenarios where the scaling of the minimum energy gap during the optimization sweep differs significantly, suggesting the possibility of an exponentially faster annealing time when using the PUBO as compared to the QUBO formulation. This advantage persists even when considering the overhead caused by the higher-order interactions necessary for PUBO cost Hamiltonians. As an interesting side effect, the analysis on minimum energy gaps of different 3-SAT instance generators reveals different degrees of hardness, which will be of interest also for classical benchmark calculations. Our findings show a promising path to improving the resource efficiency and sweeping speed of quantum annealing protocols on both analog and digital platforms, which are important prerequisites when aiming at solving larger optimization problems with relevance to industry.

## 1. Introduction

The continuous improvements of quantum computing devices have positioned quantum annealing as a promising candidate for tackling combinatorial optimization problems of practical relevance [1–5]. Experiments on commercially available quantum annealers have recently even claimed quantum advantage [6]. Since most of the existing quantum computers consist of discrete two-level systems coupled by two-qubit gates, the most common approach for solving combinatorial optimization problems on quantum devices is via quadratic unconstrained binary optimization (QUBO), where the cost function is a polynomial of degree two. However, many problems are more naturally expressed as polynomials of higher degrees, i.e. in terms of polynomial unconstrained binary optimization (PUBO), sometimes also referred to as higher-order binary optimization. Application-relevant examples of PUBO problems range from vehicle routing [7] to polymer sampling [8, 9] and lattice protein folding [10, 11]. To implement these on quantum hardware, the PUBO problems are usually reduced to QUBO problems [3, 12], and efforts have been made to render the reduction of PUBO to QUBO instances as efficient as possible [13, 14]. Nevertheless, the QUBO reduction comes at the price of introducing additional ancillary variables, along with constraints that enforce their consistency, both of which is best avoided considering the stringent space and interaction-scale limitations of current NISQ devices. Then again, considering the overhead incurred by the associated

multi-qubit interactions, i.e. slower interaction rates on analog platforms or increased circuit depth on digital architectures, it remains an open question whether a direct PUBO implementation could be advantageous in quantum optimization [15].

Here, we answer this question in the affirmative, by highlighting the potential for a gain in efficiency through a direct implementation of PUBO problems. Paradigmatic problems that are naturally formulated in terms of PUBO range from the Boolean satisfiability problem (SAT), over the hypergraph coloring problem, to the  $p$ -spin model. Computing the number of qubits necessary to implement such problems as QUBO instances, we find a potential resource saving of up to an order of magnitude in the number of qubits for some problems when formulating them in terms of PUBO problems instead. Additionally, we compute the minimum energy gap across a range of problems by employing exact diagonalization on small-scale systems, specifically 3-SAT instances of varying complexity generated using different methods [16]. We find the direct formulations as PUBO instances to have larger gaps than the equivalent QUBO reductions, suggesting an improvement in computing time. While for hard instances the gap closing and thus the time-to-solution nonetheless scale exponentially with the problem size, direct PUBO formulations can exhibit smaller exponents for some classes of 3-SAT problems, which translates to an exponential speedup over the corresponding QUBO formulation. This behavior is agnostic to the underlying quantum computing architecture and holds for analog as well as digital platforms alike.

For gate-based quantum computers, we discuss the possibility to synthesize the necessary 3-body spin interactions with typical gate sets employing single-qubit gates combined with CNOT or  $R_{zz}$  gates. The required overhead of only four two-qubit gates per three-qubit gate is more than compensated by the saving of ancillary qubits as well as the increased minimum energy gap. This shows that it can be of large practical advantage to directly implement PUBO formulations in quantum annealing protocols, thus increasing the scalability of these machines.

On top of that, our studies on the scaling of the minimum energy gap also shed light on the hardness of instances created by various 3-SAT benchmark generators [16]. Among the tested generators, we find completely randomly generated instances to be on average easier to solve (larger minimum energy gap) than the structured instances with fewer logical clauses constructed by one of the benchmark generators. In contrast, for yet another benchmark generator, the formulation as a PUBO shows that it generates only trivially solvable instances, reflected in a large energy gap. Thus, our analysis of solving 3-SAT on quantum computers can also inform benchmark calculations for combinatorial optimization problems on classical computers, where recent studies have demonstrated significant performance gains through quantum-inspired metaheuristic solvers for problems as large as 100 000 variables [17–19].

This paper is organized as follows. Section 2 contains a brief recap on solving combinatorial optimization problems with quantum annealing and its main limitation due to the closing of the minimum energy gap. The behavior of the minimum gap will be our main figure of merit for assessing potential performance gains of PUBO over QUBO. In section 3, we discuss the reduced spatial resource requirements (i.e. number of qubits) of optimization problems that are naturally formulated as PUBO, as compared to equivalent QUBO reductions. Section 4 puts forward our numerical results on the behavior of the minimum energy gap with the example of the paradigmatic 3-SAT problem, illustrating the potential for PUBO speedups due to larger gaps. In section 5, we discuss the synthesis of three-body interactions using single- and two-qubit gates, and compare the resulting overhead with the expected performance gain. Finally, section 6 contains our conclusions on the potential advantages of PUBO over QUBO in terms of spatial and temporal resources.

## 2. Quantum annealing for PUBO

In this section, we introduce the PUBO and QUBO formulations of combinatorial optimization problems, how these can be solved using quantum annealing, and how we assess the potential advantage of PUBO over QUBO in this study.

### 2.1. PUBO

Quantum annealing aims at finding the optimal solution to a combinatorial problem [3–5].

Such optimization problems can be posed as finding the minimum of a cost function  $f: \{0, 1\}^N \rightarrow \mathbb{R}$ , mapping a Boolean vector (or bit string)  $\mathbf{x} = (x_1, \dots, x_N)$  to a real number. The problem falls into the realm of PUBO if the cost function can be written as a (multivariate) polynomial,

$$f_{\text{PUBO}}(\mathbf{x}) = \sum_{i_1 + \dots + i_N \leq P} a_{i_1 \dots i_N} x_1^{i_1} \dots x_N^{i_N}, \quad (1)$$

where  $a_{i_1 \dots i_N}$  are coefficients,  $P$  is the total degree of the polynomial, and the indices vary over non-negative integers. This comprises QUBO as the special case where the cost function is quadratic ( $P = 2$ ),

$$f_{\text{QUBO}}(\mathbf{x}) = \sum_{i,j=1}^N Q_{ij} x_i x_j + \sum_{i=1}^N c_i x_i, \quad (2)$$

where  $Q$  is a symmetric matrix,  $\mathbf{c} = (c_1, \dots, c_N)$  is a vector of local biases, and we generally discard the irrelevant constant term.

Equivalently, the problem can be formulated in terms of classical discrete Ising spin variables  $\mathbf{s} = (s_1, \dots, s_N)$ , taking values in  $\{+1/2, -1/2\}^N$ . Spin and logical variables are related to each other via the one-to-one mapping  $\mathbf{s} = 1/2 - \mathbf{x}$ . Note that this mapping does not change the degree of the optimization problem. Solving the minimization problem then corresponds to finding the ground state configuration of the associated spin system which minimizes the Hamiltonian function  $H_{\text{cost}}(\mathbf{s}) = f(1/2 - \mathbf{s})$ .

## 2.2. The quantum annealing paradigm

In quantum annealing, the classical spins are promoted to quantum spins (or qubits) by identifying them with spin-1/2 operators,  $s_i \rightarrow \hat{s}_i^z$  (note the circumflex denoting operators). The cost function (or Hamiltonian function) now corresponds to an operator, the cost Hamiltonian  $\hat{H}_{\text{cost}} = f(1/2 - \hat{\mathbf{s}})$ , which is diagonal in the computational basis. For the purposes of this work, we will mainly be interested in PUBO with cubic cost functions, where the associated cost Hamiltonians are of the form

$$\hat{H}_{\text{cost}} = - \sum_{ijk} J_{ijk}^{(3)} \hat{s}_i^z \hat{s}_j^z \hat{s}_k^z - \sum_{ij} J_{ij}^{(2)} \hat{s}_i^z \hat{s}_j^z - \sum_i h_i^z \hat{s}_i^z. \quad (3)$$

Here,  $J^{(3)}$  and  $J^{(2)}$  represent three-body and two-body spin interactions, respectively, while the vector  $\mathbf{h}^z = (h_1^z, \dots, h_N^z)$  describes a (typically inhomogeneous) longitudinal field.

The idea of quantum annealing is to evolve the system from the ground state of a suitably chosen driving Hamiltonian  $\hat{H}_{\text{drive}}$  to the desired ground state of the cost Hamiltonian  $\hat{H}_{\text{cost}}$  by adiabatically changing Hamiltonian parameters. In the simplest case of a linear sweep from  $\hat{H}_{\text{drive}}$  to  $\hat{H}_{\text{cost}}$ , the instantaneous Hamiltonian reads

$$\hat{H}(s) = (1 - s) \hat{H}_{\text{drive}} + s \hat{H}_{\text{cost}}, \quad (4)$$

where  $s = t/t_{\text{sweep}} \in [0, 1]$  is the time in units of the sweep duration  $t_{\text{sweep}}$ . The driving Hamiltonian should not commute with the cost Hamiltonian and its ground state should be easy to prepare. In practice, the driving Hamiltonian is therefore often chosen as  $\hat{H}_{\text{drive}} = -h^x \sum_i \hat{s}_i^x$ , where  $\hat{s}_i^x$  is the spin-1/2 operator on the  $i$ th spin in  $x$  direction and  $h^x$  is a uniform transverse field.

According to the adiabatic theorem, the system remains in its instantaneous ground state (and the final state thus corresponds to the desired ground state of  $\hat{H}_{\text{cost}}$ ), provided the Hamiltonian changes sufficiently slowly such that the adiabatic condition holds [20–23]:

$$t_{\text{sweep}} \gg \hbar \max_{s \in [0,1]} \frac{|\langle n(s) | d\hat{H}(s)/ds | 0(s) \rangle|}{\Delta E_{n0}^2(s)}, \quad \forall n \geq 1. \quad (5)$$

Here,  $\Delta E_{n0}(s)$  denotes the energy difference between the  $n$ -th instantaneous excited state  $|n(s)\rangle$  and the instantaneous (non-degenerate) ground state  $|0(s)\rangle$ .

It should be noted that adiabaticity may not strictly be necessary for a successful annealing sweep. In fact, the search for optimized (non-adiabatic) annealing schedules and tighter bounds for the minimal quantum annealing time is an active field of research, see for example [24–26]. For our purposes, we will nonetheless resort to the adiabatic condition in equation (5) to assess the performance of quantum annealing for solving PUBO or QUBO problems.

## 2.3. Estimating PUBO speedup for quantum annealing

An order-of-magnitude estimate of the time required for a successful adiabatic sweep can be obtained from the right-hand side of the adiabatic condition in equation (5). For our purposes, we employ the following simplified sufficient condition for adiabaticity:

$$t_{\text{sweep}} \gg \hbar \frac{V}{\Delta E^2} \equiv T. \quad (6)$$

Here,  $\Delta E = \min_{s \in [0,1]} \Delta E_{10}(s)$  is the minimum energy gap between the instantaneous first excited state and the instantaneous ground state, while the matrix elements in the numerator of equation (5) are incorporated in the quantity  $V = \max_{s \in [0,1]} |\langle 1(s) | d\hat{H}(s)/ds | 0(s) \rangle|$  (if the first excited state is degenerate, we take the maximum over the degenerate state manifold). Throughout this work, we will refer to the right-hand side of equation (6) as the adiabaticity time  $T$ .

In this study, we are primarily interested in hard problems where the minimum gap decreases exponentially with the problem size  $N$ ,

$$\Delta E = \epsilon e^{-\alpha N}, \quad (7)$$

with characteristic energy  $\epsilon$  and exponent  $\alpha > 0$ . Such an exponential closing of the gap is a signature of a first-order quantum phase transition [27]. By contrast, the quantity  $V$  in the numerator of equation (6) scales at most proportional to the norm of the Hamiltonian, i.e. as  $\mathcal{O}(N)$  for extensive Hamiltonians. Consequently, the adiabaticity time is dominated by the behavior of the minimum gap and grows exponentially with the problem size as  $T \propto e^{2\alpha N}$ , up to polynomial corrections. In our analysis, we will therefore use  $\Delta E$  as the main figure of merit for the hardness of solving an optimization problem with quantum annealing.

When comparing the relative annealing performance for solving a problem posed in PUBO form versus its equivalent QUBO reduction, it is important to keep in mind that the energy scale of the Hamiltonian is typically different for the two problem types: the characteristic energy scale of the cost Hamiltonian encoding a PUBO problem is set by the three-body interaction strength  $J^{(3)}$ , while for a QUBO problem it is determined by the two-body interaction strength  $J^{(2)}$ . Typically, three-body interactions are slower than two-body interactions, but the precise overhead depends on the hardware platform and implementation details (see also section 5). One can keep track of this by expressing the quantities in the relations 6 and 7 in terms of these scales, i.e. by defining the dimensionless quantities  $\tilde{V}_P = V_P/J^{(3)}$  and  $\tilde{V}_Q = V_Q/J^{(2)}$ , as well as  $\tilde{\epsilon}_P = \epsilon_P/J^{(3)}$  and  $\tilde{\epsilon}_Q = \epsilon_Q/J^{(2)}$ , where  $P$  and  $Q$  denote PUBO and QUBO encoding, respectively. One can then estimate the potential PUBO speedup with respect to the equivalent QUBO formulation of a problem as the ratio of the corresponding adiabaticity times,

$$\frac{T_Q}{T_P} = \frac{J^{(3)}}{J^{(2)}} \frac{\tilde{V}_Q}{\tilde{V}_P} \left( \frac{\tilde{\epsilon}_P}{\tilde{\epsilon}_Q} \right)^2 e^{2(\alpha_Q - \alpha_P)N}. \quad (8)$$

As equation (8) illustrates, a potential PUBO speedup for solving exponentially hard problems with quantum annealing ( $T_Q/T_P > 1$ ) is primarily determined by the size of the minimum gap and its scaling with the problem size, assuming at most a polynomial scaling for the various prefactors. Specifically, under the above assumption and if  $\alpha_P < \alpha_Q$ , there is always an  $N$  above which PUBO outperforms QUBO. Even if PUBO and QUBO scale similarly with the system size ( $\alpha_P \approx \alpha_Q$ ), assuming a not too small ratio  $\tilde{V}_Q/\tilde{V}_P$ , PUBO may be advantageous if the minimum gap has a larger offset,  $\epsilon_P > \epsilon_Q$ , which in spite of the typically slower three-body interaction rate is often the case due to the higher resource efficiency of the PUBO encoding. In our numerical study of PUBO versus QUBO formulations of 3-SAT problems in section 4, we will encounter examples of both behaviors,  $\alpha_P \approx \alpha_Q$  and  $\alpha_P < \alpha_Q$ , see figures 3(a) and (b), respectively. Together with estimates for the overhead of the three-body relative to the two-body interaction, this allows us to estimate the potential for PUBO speedups for the investigated problem classes, see section 5.

## 2.4. Analog versus digital quantum annealing

Quantum annealing has originally been conceptualized for analog platforms, where the relative strength of the cost and driving Hamiltonian can be tuned continuously. With the increasing availability of digital quantum computers, variants of quantum annealing adapted to gate-based architectures have become popular. Digital annealers are particularly interesting in the light of direct implementations of PUBO as they allow one to synthesize the required higher-order interactions from universal sets of two- and single-qubit gates, whereas engineering native higher-order interactions in analog annealers is more challenging (see discussion in section 5.1). For this reason, we briefly introduce a digital version of quantum annealing in this section. Our analysis of PUBO versus QUBO performance applies to both analog and digital protocols alike as long as they are based on adiabatic evolution (in distinction to certain quantum-annealing-inspired algorithms like the Quantum Approximate Optimization Algorithm [28], where optimal performance requires schedules with large angles corresponding to evolution times outside the adiabatic regime).

Digital quantum annealing (or Trotterized quantum annealing) is accomplished via a Trotterization of the time evolution under the Hamiltonian in equation (4). That is, for a grid of times  $0 = t_0 < t_1 < \dots < t_M = t_{\text{sweep}}$  with uniform step size  $\Delta t = t_{\text{sweep}}/M$ , the time evolution operator can be approximated as [29]

$$\hat{\mathcal{T}} e^{-i t_{\text{sweep}} \int_0^1 ds' \hat{H}(s')/\hbar} = \prod_{m=1}^M e^{-i \Delta t (1-s_m) \hat{H}_{\text{drive}}/\hbar} e^{-i \Delta t s_m \hat{H}_{\text{cost}}/\hbar} + \mathcal{O}(\Delta t), \quad (9)$$

where  $\hat{\mathcal{T}}$  is the time-ordering operator and  $s_m = t_m/t_{\text{sweep}}$ . Note that the local error for each Trotter step is  $\mathcal{O}(\Delta t^2)$ , such that the accumulated global error scales as  $\mathcal{O}(\Delta t)$ , which may be improved through higher-order splitting methods [30]. Since all terms within the cost Hamiltonian commute with each other, the corresponding matrix exponential can be exactly decomposed into a series of single-qubit  $Z$  rotations and two-qubit  $ZZ$  rotations. Similarly, the matrix exponential of the driving Hamiltonian can be decomposed into distinct single-qubit  $X$  rotations. Each two-qubit rotation may then further be compiled into the native gate set of the respective architecture.

Compared to analog quantum annealing, this procedure incurs additional errors due to the discretization. Furthermore, since each time step  $\Delta t$  needs to be decomposed into the native gate set, there is in general a polynomial overhead in the number of gates. For sufficiently small Trotter steps, however, the limiting factors to the algorithm's performance are the same as those of the continuous protocol. In this situation, the adiabaticity time estimated from the minimum energy gap of the analog quantum annealing protocol acts as a proxy for the number of discrete time steps and thus the circuit depth required to reach the ground state of the cost Hamiltonian with the discrete protocol. Unless stated otherwise, we will generally consider analog quantum annealing throughout this work.

### 3. PUBO problems and their resource requirements

In this section, we discuss two instructive examples of optimization problems that can naturally be formulated as PUBO. We start in section 3.1 by illustrating the advantages of the PUBO formulation in terms of spatial resources, i.e. number of required qubits, for the simple toy model of minimizing a polynomial function. In section 3.2, we dive in some detail into the paradigmatic 3-SAT problem, which is the model we use for our numerical analysis in the rest of this work.

Appendix A briefly presents a collection of further examples (the  $p$ -spin model, hypergraph coloring, NAE-k-SAT, and the traveling salesperson problem with time windows), illustrating the ubiquity of PUBO problems.

#### 3.1. Toy model: minimization of a polynomial function

As a first simple example, we consider a continuous polynomial  $f: [a, b] \rightarrow \mathbb{R}$  as an objective function to be minimized. Without loss of generality, we choose the boundaries of the half-open interval as  $a = 0$  and  $b = 2^n$  for a convenient encoding of the continuous variable into bits. We can then represent any continuous variable  $x$  approximately as a bitstring  $(x_1, \dots, x_{n+m})$  via the mapping

$$x \approx \sum_{i=1}^n 2^{n-i} x_i + \sum_{i=1}^m 2^{-i} x_{n+i}, \quad (10)$$

where the first sum corresponds to the integer part and the second sum to the decimal part with  $m$ -digit resolution. Note that negative numbers can be handled in a similar way by introducing an additional Boolean variable  $x_0$ , encoding the sign of  $x$ , and multiplying equation (10) by  $(2x_0 - 1)$  [15].

As a concrete case, take the task to find the minimum of the cubic polynomial  $f(x) = x^3 + x$  in the domain  $[0, 2^2]$ . After inserting the encoding in equation (10) with a resolution of  $m = 1$ , we obtain

$$f(x_1, x_2, x_3) = 4x_1 + 2x_2 + \frac{5}{8}x_3 + 6x_1x_2 + \frac{15}{2}x_1x_3 + \frac{9}{4}x_2x_3 + 6x_1x_2x_3. \quad (11)$$

Due to the term  $\propto x_1x_2x_3$ , this optimization problem is of third-order PUBO form. The trivial solution in this case is given by  $(x_1, x_2, x_3) = (0, 0, 0)$ .

Equation (11) can be reduced to a QUBO problem by introducing an additional slack variable  $y = x_2x_3$ . An appropriate penalty term  $\propto (3y + x_2x_3 - 2x_2y - 2x_3y)$  needs to be added to the optimization problem in order to enforce consistency. While in this specific case the QUBO formulation requires only a single extra qubit compared to the equivalent PUBO formulation, the resource savings can get significant for larger and more complex problems (see below). Furthermore, the optimal introduction of slack variables (i.e. the assignment which requires the least amount of additional bits) is itself known to be an NP-hard problem [31]. Finally, the constraint that enforces consistency of the ancillary variables needs to be sufficiently strong such that other terms cannot overpower it. Since the maximal energy scales available in a given quantum machine are limited, this means in practice reducing the relative strength of the other terms, and thus incurring proportionally longer sweep times or deteriorating the adiabaticity of the sweep. Finding

the minimum penalty strength that enforces all constraints is in general NP-hard, but efficient approximate schemes with significant performance gains over default choices have been proposed [32].

### 3.2. Boolean satisfiability problem

Given a Boolean formula, the satisfiability problem (SAT) poses the decision problem whether there is an assignment of truth values such that the formula evaluates to *true*. In the case of  $k$ -SAT, the formula takes the form of a conjunction of  $M$  logical clauses  $C_m$ ,

$$C_1 \wedge C_2 \wedge \dots \wedge C_M, \quad (12)$$

with each clause being composed of a disjunction of  $k$  literals chosen from a set of  $N$  Boolean variables  $x_i$  and their negations  $\neg x_i$ . An example clause for  $k = 3$  is

$$C_m(x_i, x_j, x_\ell) = x_i \vee \neg x_j \vee \neg x_\ell. \quad (13)$$

Every SAT problem can be rewritten as a  $k$ -SAT problem using the rules of Boolean algebra. It is well known that  $k$ -SAT for  $k \leq 2$  belongs to the complexity class P, i.e. it is solvable in polynomial time on a classical computer [33]. In contrast, 3-SAT was the first decision problem proven to be NP-complete (Cook–Levin theorem [34, 35]). If  $P \neq NP$ , no classical algorithm could solve 3-SAT in polynomial time. Additionally, it is widely believed that problems in NP-complete are not contained in BQP [36], and could therefore not be solved in polynomial time on a quantum computer either. Nevertheless, even if an exponential quantum advantage may be out of reach, one can still hope for a practical speedup through quantum annealing [9, 37–40].

A phase transition for 3-SAT problems is known to occur asymptotically at a critical ratio of clauses to variables of  $(M/N)_{\text{crit}} \approx 4.24$  [41]. Around this critical point, one is likely to encounter hard-to-solve problems with unique solutions. Below the phase transition point, most problem instances are easy to solve with many solutions, while above the transition most instances become unsatisfiable. Note that although the probability of finding instances with unique solutions below the phase transition point decreases exponentially, it is well-known that those rare instances (with fewer clauses and unique solutions) tend to be the hardest ones to solve [42]. Our results in section 4 underline this fact by comparing the minimum energy gaps of instances constructed randomly by different generators with different ratios of clauses to variables.

#### 3.2.1. 3-SAT as PUBO

3-SAT can be reformulated in various ways as a PUBO or QUBO problem in order to make it suitable for quantum annealing. A straightforward encoding consists in a PUBO problem of degree three: each clause is cast into the form of an energy term penalizing any choice of truth values that does not satisfy the clause. For example, the clause  $C_m = x_i \vee \neg x_j \vee \neg x_\ell$  from above can be translated to  $(1 - x_i)x_jx_\ell$ , which only evaluates to zero if at least one variable satisfies the corresponding clause. Adding up these terms for each clause and translating the binary variables to quantum spin variables via  $x_i = 1/2 - \hat{S}_i^z$ , leads to the general three-body spin Hamiltonian in equation (3). Note that, in this specific PUBO formulation, the three-body and two-body interaction strengths  $J^{(3)}$  and  $J^{(2)}$  as well as the local biases  $\mathbf{h}^z$  are all of the same order of magnitude or zero. The ground state of the cost Hamiltonian then encodes the solution of the original 3-SAT problem (more precisely, it solves MAX-SAT, i.e. finding the solution that satisfies most clauses). This encoding requires  $N$  spins (qubits) for the  $N$  Boolean variables in the original problem.

#### 3.2.2. 3-SAT as QUBO via slack variables

Currently available quantum annealing hardware (e.g. the D-Wave machine [43]) natively only supports two-body interactions. Thus, one needs to find an equivalent encoding as a QUBO problem to solve 3-SAT on such hardware. A resource-efficient way to achieve this is to start with the PUBO formulation of 3-SAT given by the general Hamiltonian in equation (3) and to introduce slack variables that reduce all three-body interactions to quadratic and local terms (see section 3.1). Since there are at most  $M$  (number of clauses) unique three-body terms in the Hamiltonian, at most  $M$  additional qubits are required by this QUBO reduction. For hard problem instances around the phase transition, which we are most interested in here, this amounts to at most around  $4.24N$  additional qubits.

In appendix B, we discuss two other approaches to formulate 3-SAT as a QUBO problem. However, the direct approach of introducing slack variables is the most efficient in terms of qubit requirements and will be considered for the rest of this work.

## 4. Numerical results

In this section, we numerically test the quantum annealing performance of PUBO versus QUBO formulations for different instances of the 3-SAT problem (see section 3.2) using exact diagonalization. We focus our analysis on the behavior of the minimum energy gap  $\Delta E$  encountered during the annealing protocol, which is the key quantity that mainly determines the annealing performance, as explained in section 2.

### 4.1. Methods

In what follows, we briefly outline the framework and key algorithms employed in our numerical study. Throughout this section, we consider the annealing schedule given in equation (4), where the cost Hamiltonian is ramped up linearly, while the strength of a homogeneous transverse field is linearly decreased.

#### 4.1.1. Characteristic energy scales

In our numerics, we express the cost Hamiltonian  $\hat{H}_{\text{cost}}$  in units of the global energy scale  $J$ , which we identify with a characteristic energy of the physical device. In principle, it is desirable for  $J$  to be as large as possible in order to speed up the absolute computing time. In practice, however,  $J$  is limited by the maximum physical coupling rates the device can operate with. Consequently, a global rescaling of the Hamiltonian is required to ensure that none of the terms in equation (3) exceed these limits. On analog hardware, the physical couplings relevant for PUBO implementations up to third order include the maximum physical three-body and two-body interaction strengths as well as the maximum strength of local fields. On digital devices, the physical time scales associated with these interactions correspond, respectively, to the three-, two-, and single-qubit gate times. Since interactions of higher rank are generally slower compared to interactions of lower rank (see discussion in section 5.1), the former typically create a bottleneck: the global energy scale  $J$  is limited by the maximum physical three-body interaction strength if three-body interactions are present (third-order PUBO) and by the maximum physical two-body interaction strength if only two-body interactions are present (QUBO), where we assume that local fields can be scaled sufficiently high to pose no restriction. In particular, three-body interactions are the limiting factor for PUBO implementations of typical 3-SAT instances encountered in our small-scale numerical experiments, where the maximum ratio of three-body and two-body couplings is of order one (though this may be different for larger instances or problems with different PUBO structures). It is therefore natural in our case to choose the characteristic energy  $J = \max_{ijk} |J_{ijk}^{(3)}|$  as the global energy unit for third-order PUBO, while for QUBO we use  $J = \max_{ij} |J_{ij}^{(2)}|$ .

Importantly, for QUBO reductions of native PUBO problems,  $\hat{H}_{\text{cost}}$  incorporates both the problem Hamiltonian and penalty terms implementing consistency constraints. The two-body interactions can then be separated:  $J_{ij}^{(2)} = [J_{ij}^{(2)}]_{\text{problem}} + [J_{ij}^{(2)}]_{\text{constraints}}$ . Typically, enforcing the constraints requires the penalty terms to be much larger than the QUBO terms of the actual problem Hamiltonian. Since the characteristic energy scale  $J = \max_{ij} |J_{ij}^{(2)}|$  is limited by the maximum physical interaction rate supported by the hardware, we end up with  $[J_{ij}^{(2)}]_{\text{problem}}/J \ll 1$  and  $[J_{ij}^{(2)}]_{\text{constraints}}/J \sim 1$ , i.e. the part of the Hamiltonian encoding the actual problem is scaled down with respect to the penalty terms. In section 4.2.4, we investigate how the relative strength of the penalty terms influences the minimum energy gap and thus the adiabaticity time.

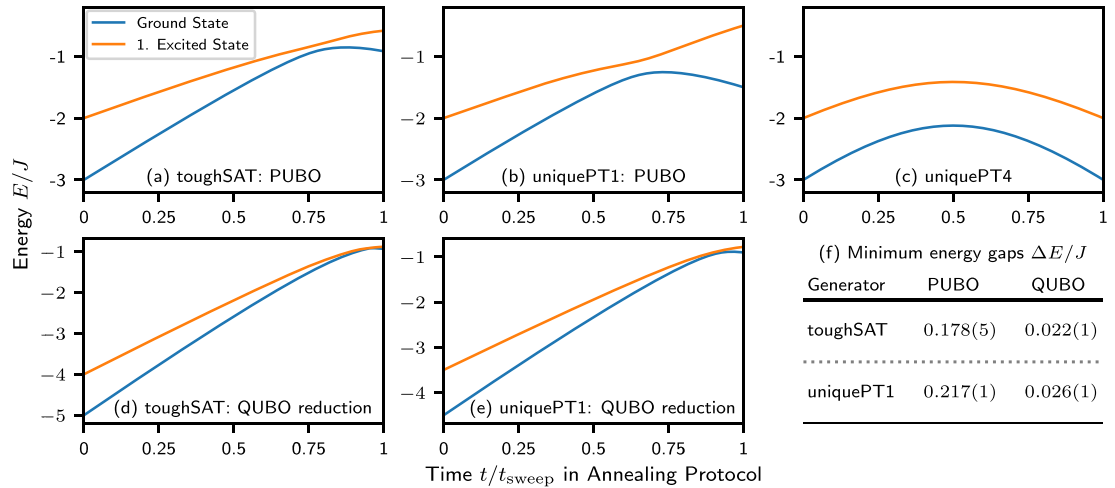
Unless stated otherwise, we choose the strength of the driving Hamiltonian in equation (4) to be the same as the characteristic energy scale of the cost Hamiltonian,  $h^x = J$ . In section 4.2.5, we justify this choice and discuss the effects of varying  $h^x$  on the annealing performance.

#### 4.1.2. 3-SAT problem generators

To generate instances of the 3-SAT problem, we adapt the three different generators toughSAT, uniquePT1, and uniquePT4 described in [16].

The toughSAT generator creates an arbitrary number  $M$  of 3-SAT clauses by randomly picking three out of  $N$  variables and then randomly applying negations to them. In our numerics, we set the ratio of clauses to variables close to the critical ratio  $(M/N)_{\text{crit}} \approx 4.24$ , as we expect to find hard-to-solve instances with unique ground states in the vicinity of this point (see section 3.2). To avoid ground state degeneracies, we postselect the randomly generated instances for those with unique solutions.

In contrast, the generators uniquePT1 and uniquePT4 always create instances with known and unique solutions (although not necessarily hard to solve in the latter case). Both generators work by picking a solution randomly and then constructing appropriate 3-SAT clauses to make said solution unique. While uniquePT4 creates  $M = 4N$  clauses to achieve that, uniquePT1 only needs  $M = N + 6$  clauses. However, we find that uniquePT4 constructs clauses in such a way that all third-order and second-order terms in the PUBO cost



**Figure 1.** Typical gap profiles of PUBO and QUBO formulations of randomly generated 3-SAT instances. The plots show the energies of the ground state (blue) and first excited state (orange) as a function of time during an annealing sweep according to equation (4) with driving strength  $h_x$  equal to the characteristic energy scale  $J$  of the cost Hamiltonian. The upper panels (a) and (b) depict the native PUBO formulations (cost Hamiltonian with up to three-body interactions,  $J = \max_{ijk} |J_{ijk}^{(3)}|$ ), while the lower panels (d) and (e) show the associated QUBO reductions (cost Hamiltonian with at most two-body interactions,  $J = \max_{ij} |J_{ij}^{(2)}|$ ). Note that two-body interactions are typically much stronger than three-body interactions. A direct decomposition of a 3-qubit unitary gate gives a ratio of  $J^{(3)}/J^{(2)} \approx 1/4$  (see section 5.1). The 3-SAT instances comprise  $N = 6$  classical variables and are generated by the toughSAT generator with  $M = \text{round}(4.24N) = 25$  clauses (panels (a) and (d)), the uniquePT1 generator with  $M = N + 6$  clauses (panels (b) and (e)), and the uniquePT4 generator with  $M/N = 4$  clauses (panel (c)). The QUBO reductions of the instances in (a) and (b) require four and three additional slack variables (ancillary qubits), respectively. For instances generated by uniquePT4 (c), no gap closing occurs since the generated instances are trivial (all three-body and two-body interactions cancel,  $J = \max_i |h_i^z|$ ). The table (f) shows the mean minimum energy gap and its standard error of the mean for an ensemble of 200 instances corresponding to the scenarios in (a), (b) and (d), (e).

function, i.e. three-body and two-body interactions in the cost Hamiltonian, cancel each other. What remains is a simple cost Hamiltonian with only local terms of the form  $\hat{H}_{\text{cost}} = -\sum_i h_i^z \hat{s}_i^z$ , whose solution can easily be read off. Thus, uniquePT4 creates trivial 3-SAT problems, where the time required to solve them scales linearly in the problem size.

For our numerical analysis, we are primarily interested in hard-to-solve problem instances generated by toughSAT and uniquePT1, which exhibit an exponential closing of the minimum energy gap with increasing problem size. For comparison, we also show the spectrum of a problem instance generated by uniquePT4, see figure 1(c).

#### 4.1.3. PUBO-to-QUBO reduction

To assess the annealing performance gains of PUBO over QUBO formulations when solving 3-SAT instances, we compare the minimum energy gap of the PUBO formulation with that of the equivalent QUBO reduction. Finding an optimal QUBO reduction for a given PUBO problem is in general NP-hard. For the small system sizes considered here, we employ the following brute-force algorithm: replace the pair of variables that appears most frequently in the third-order terms of the PUBO cost function by a slack variable (see section 3.1), then continue with the second-most-frequent pair of variables and so on, until all third-order terms have been replaced by quadratic ones with associated slack variables and constraints [44]. For each newly introduced ancilla variable  $y = x_i x_j$ , a penalty term  $\propto (3y + x_i x_j - 2x_i y - 2x_j y)$  needs to be added to the cost Hamiltonian (see section 3.1) and scaled by a suitable factor, thus ensuring that violating the penalty term in favor of the three-body interactions is never energetically favorable. This factor depends on the problem size and structure: in our numerics, we make the simple choice that for every three-body interaction term  $a_{ijk} x_i x_j x_k$  that is to be reduced by the ancilla variable  $x_{\text{ancilla}} = x_i x_j$ , we add  $|a_{ijk}| + 1$  to the strength of the corresponding penalty term. In section 4.2.5, we discuss how varying the strength of the penalty terms relative to the original cost Hamiltonian affects the overall annealing performance.

The QUBO formulation of toughSAT instances usually requires more slack variables than uniquePT1 instances of the same size. This is due to the fact that the number of ancilla variables is bounded from above by the number of clauses in the problem instance. For both generators, the increased resource requirements of the QUBO reductions generally lead to a smaller minimum gap and thus to a reduced annealing performance, as we will see below.

## 4.2. Analysis of the minimum gap

In what follows, we present our numerical results on the behavior of the minimum energy gap for randomly generated 3-SAT instances. To begin with, we illustrate the increase of the minimum gap in the PUBO formulation of the 3-SAT problem compared to the equivalent QUBO reduction and discuss the correlation of the minimum gap with the number of three-body interaction terms in the cost Hamiltonian. We then compare the scaling of the minimum gap with system size for the equivalent PUBO and QUBO formulations.

### 4.2.1. Gap profiles

Figure 1 illustrates typical gap profiles during an annealing sweep for specific instances with  $N = 6$  variables, generated by each of the three generators. The blue and yellow lines correspond, respectively, to the energies of the instantaneous ground state and first excited state during a sweep according to the annealing protocol in equation (4). While the toughSAT and uniquePT1 instances exhibit significant gap closings, the energy gap of the uniquePT4 instance stays consistently large over the entire sweep, indicating that the generated instances are trivial to solve. The gap profile of the depicted uniquePT4 instance is representative of every random instance created by this generator. For the toughSAT and uniquePT1 instances, we also plot the energies of the ground state and first excited state in the corresponding QUBO reduction.

In figure 1(f), we report the mean minimum energy gap and its standard error of the mean, obtained by averaging over a large number of samples. It can be seen that for this small system size, uniquePT1 and toughSAT produce on average comparable energy gaps (slightly larger for the former), even though uniquePT1 requires less clauses and thus less additional slack variables in the QUBO reduction. For larger systems, this is no longer true, as we will see in section 4.2.3.

Importantly, the gaps in the QUBO reductions are significantly smaller compared to those in the corresponding PUBO formulations. The larger relative variance in the mean energy gaps of the QUBO reductions can be explained by the fact that not every generated 3-SAT instance requires the same amount of slack variables for its reduction to a QUBO problem.

Recall that when characterizing the relative performance of PUBO versus QUBO via the adiabaticity time, it is necessary to take into account the different physical energy scales set by the three-body and two-body interaction rates  $J^{(3)}$  and  $J^{(2)}$ , respectively, see section 4.1.1. We will come back to this point in section 5.

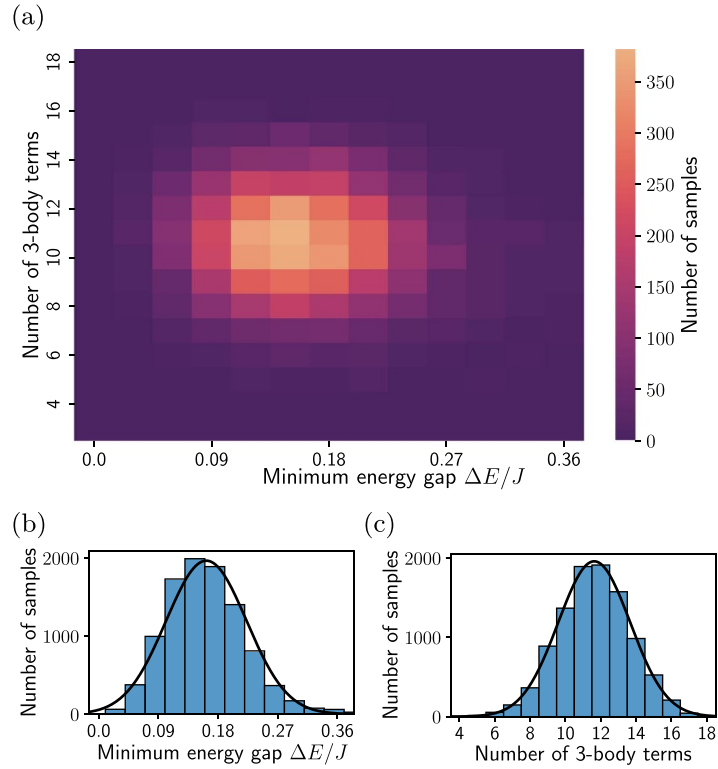
### 4.2.2. Absence of correlations between minimum gap and number of three-body terms

One may wonder whether the minimum gap (and thus the hardness to solve the PUBO problem with quantum annealing) depends on the number of nonzero three-body interaction coefficients present in the cost Hamiltonian. Figure 2 investigates the correlations between these two quantities for 10000 random instances generated via toughSAT with  $N = 6$  and  $M = \text{round}(4.24N) = 25$ . The samples for both the minimum gap and the number of three-body terms are broadly distributed around the mean, and are well approximated by a Gaussian (see marginal distributions in figure 2). The number of three-body terms spreads in the range 4–19. We do not find any significant correlation between the minimum gap and the number of three-body terms.

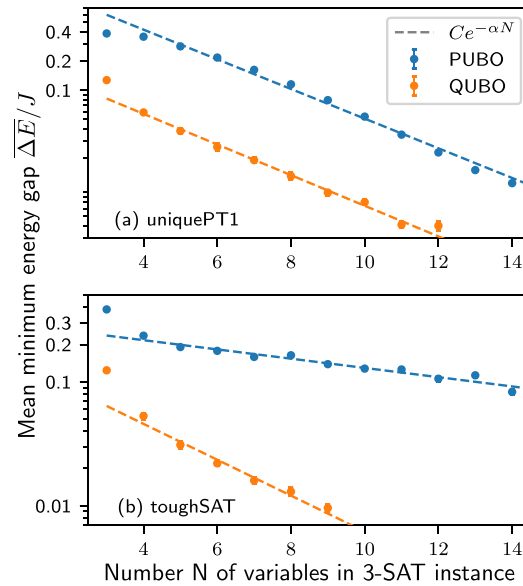
### 4.2.3. Scaling of the minimum gap versus system size

As explained in section 2, a key characteristics for predicting the performance of a quantum annealer is how the minimum energy gap encountered during the sweep scales with the problem size. In figure 3, we analyze the behavior of the minimum gap as a function of the number of variables for 3-SAT problems in their PUBO form as well as in their equivalent QUBO reduction (see section 4.1.3). The random 3-SAT instances are generated by the toughSAT and uniquePT1 generators and averaged over 200 realizations.

In the investigated range of problem sizes  $N$ , we find the mean minimum gap  $\overline{\Delta E}$  in all cases to close exponentially according to  $\overline{\Delta E} = \epsilon e^{-\alpha N}$ , as is to be expected for NP-complete problems like 3-SAT. The parameters of the best fits are given in table 1. toughSAT exhibits a slower scaling in the PUBO formulation than uniquePT1, indicating that the former generates on average easier (but still exponentially hard) instances. Notably, for ensembles generated by toughSAT (figure 3(b)), the gap closing in the PUBO formulation is significantly slower than in the QUBO formulation. Although the precise scaling may depend on the choice of the ratio of driving versus cost Hamiltonian as well as the strength of the penalty terms in the QUBO reduction, we argue in the following sections that the employed parameters are chosen close to optimal for the problem sizes investigated. This scaling thus suggests that for sufficiently large problems the PUBO encoding will always be advantageous. In the case of uniquePT1 (figure 3(a)), the gap closes approximately at the same rate for both PUBO and QUBO (slightly faster in the QUBO reduction). According to equation (8), a potential PUBO speedup then depends on the relative strength of the three-body with respect to the two-body interaction, which we analyze in more detail in section 5.



**Figure 2.** Distribution of the minimum energy gap and the number of third-order terms in the PUBO cost function (nonzero three-body interaction coefficients in the cost Hamiltonian). (a) Two-dimensional histogram, computed for 10 000 random 3-SAT instances generated via toughSAT with  $N = 6$  and  $M = 25$ . (b) and (c) Marginal histograms of the minimum energy gap and the number of third-order terms, respectively, with fitted Gaussian distributions. No significant correlations can be discerned.



**Figure 3.** Scaling of mean minimum energy gap  $\overline{\Delta E}$  with the number of Boolean variables  $N$  for ensembles of 200 randomly generated 3-SAT instances. The error bars show the statistical standard error of the mean, which is smaller than the marker size for most data points. The dashed lines show exponential fits to the data (ignoring  $N = 3$ ) according to  $\overline{\Delta E} = \epsilon e^{-\alpha N}$  with fit parameters  $\epsilon$  and  $\alpha$  given in table 1. For problems generated with uniquePT1 (a), the gap closes approximately at the same rate in both the PUBO and QUBO formulation, while for toughSAT (b), the QUBO formulation exhibits a faster gap closing with increasing  $N$ . In all cases, the minimum gap relative to the characteristic energy scales is larger for the PUBO formulation ( $J \sim J^{(3)}$ ) than for the corresponding QUBO reduction ( $J \sim J^{(2)}$ ).

**Table 1.** Parameters of the fits shown in figure 3, characterizing the exponential scaling of the mean minimum energy gap  $\overline{\Delta E} = \epsilon e^{-\alpha N}$  for randomly generated 3-SAT instances. The uncertainty of the fit parameters corresponds to the standard error of the mean for 200 realizations.

| generator | PUBO           |            | QUBO           |            |
|-----------|----------------|------------|----------------|------------|
|           | $\epsilon_P/J$ | $\alpha_P$ | $\epsilon_Q/J$ | $\alpha_Q$ |
| toughSAT  | 0.306(8)       | 0.086(4)   | 0.17(2)        | 0.33(2)    |
| uniquePT1 | 1.72(1)        | 0.352(1)   | 0.24(1)        | 0.364(8)   |

The numerically determined exponential scaling of the minimum energy gap also allows us to estimate the scaling of the adiabaticity time  $T$  as a figure of merit for the required time to solve the optimization problem with quantum annealing. Assuming that the matrix elements contributing to the quantity  $V$  in equation (6) scale at most polynomially with the system size  $N$ , the scaling of the adiabaticity time is dominated by the behavior of the mean minimum energy gap,  $T \propto \overline{\Delta E}^{-2}$ . For PUBO formulations of toughSAT instances, we then find that the adiabaticity time scales on average as  $T \propto 1.19^N$ . It is tempting to compare this number with classical solvers: the Unique 3-SAT problem, i.e. the promise version of 3-SAT, where each instance has either exactly one or no solution, can be solved classically in time  $\propto 1.307^N$  [45]. Remarkably, the adiabaticity time for toughSAT instances in PUBO form grows more slowly with the system size. Conversely, for the harder 3-SAT instances generated by uniquePT1, we find a scaling of the adiabaticity time as  $T \propto 2.02^N$  for the PUBO formulation, which is close to the  $2^N$  scaling of the classical brute-force approach. Although the adiabaticity time is only a crude approximation of the actual runtime required by quantum annealing and may not reflect accurately the true computational complexity, our findings raise the possibility that, for certain problem families, quantum annealing with a PUBO formulation of 3-SAT could be competitive with or even surpass classical performance.

#### 4.2.4. Dependence of the minimum energy gap on the QUBO penalty strength

Throughout our numerical analysis above, we have followed the heuristic described in section 4.1.3 to determine the strength of the energy penalty terms in the QUBO reduction of a PUBO problem such that the solution satisfies all constraints. In this section, we investigate how varying this strength affects the minimum energy gap.

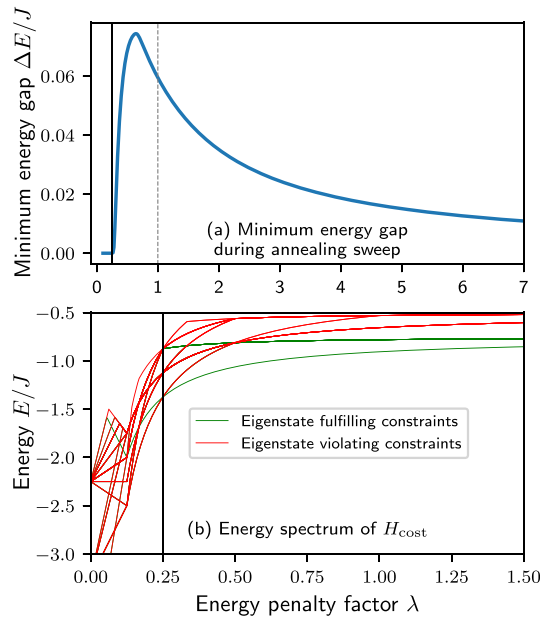
To this end, we separate the cost Hamiltonian as

$$\hat{H}_{\text{cost}} = \hat{H}_{\text{problem}} + \lambda \hat{H}_{\text{constraints}} \quad (14)$$

into the part encoding the actual problem,  $\hat{H}_{\text{problem}}$ , and the part ensuring the fulfillment of consistency constraints in the QUBO reduction,  $\hat{H}_{\text{constraints}}$ . Recall that the latter originates from penalty terms of the form  $3y + x_i x_j - 2x_i y - 2x_j y$ , one for each cubic term  $\propto x_i x_j x_k$  that is to be reduced by an ancilla variable  $y = x_i x_j$  in the PUBO cost function. To tune the strength of the energy penalty, we have introduced a factor  $\lambda \geq 0$  in equation (14), which globally scales the strength of  $\hat{H}_{\text{constraints}}$  relative to  $\hat{H}_{\text{problem}}$  (while preserving the relative weights of the individual terms in  $\hat{H}_{\text{constraints}}$ ). For each choice of  $\lambda$ , the cost Hamiltonian is normalized as before such that the characteristic energy scale is  $J = \max_{ij} |J_{ij}^{(2)}|$  with  $J_{ij}^{(2)}(\lambda) = [J_{ij}^{(2)}]_{\text{problem}} + \lambda [J_{ij}^{(2)}]_{\text{constraints}}$ , see section 4.1.1. The value  $\lambda = 1$  corresponds to the configuration used in the previous sections.

As a representative example, we consider a random 3-SAT instance generated by uniquePT1 with  $N = 6$  variables. Figure 4(a) shows the minimum energy gap as a function of the penalty strength  $\lambda$ . In this specific example, the largest value of the minimum energy gap,  $\Delta E/J \approx 0.074$ , is obtained for  $\lambda \approx 0.64$ . For comparison, our default choice,  $\lambda = 1$ , yields a gap that is about 19% smaller,  $\Delta E/J \approx 0.06$ . Higher values of  $\lambda$  further reduce the minimum gap since the relative strength of  $\hat{H}_{\text{problem}}$  with respect to  $\hat{H}_{\text{constraints}}$  decreases. Tuning  $\lambda$  in the opposite direction below the optimal value leads to a rapid decrease of the gap until it reaches zero where the ground state becomes degenerate and changes to a state that no longer fulfills the consistency constraints. In the example of figure 4(a), the point where the solution becomes unfeasible is reached at  $\lambda = 0.25$ , but for other problem instances, this can occur at larger values, e.g.  $\lambda = 0.5$ , in which case the optimum value lies even closer to our default choice  $\lambda = 1$ .

To gain a deeper understanding of this behavior, we study the energy spectrum of the cost Hamiltonian  $\hat{H}_{\text{cost}}$  as a function of  $\lambda$ , shown in figure 4(b) for the lowest 30 eigenvalues (all levels except for the unique solution are degenerate). For large enough values of the penalty parameter  $\lambda$ , all constraint-violating eigenstates (red color) are shifted up in energy, so that both the ground state and the first excited state fulfill all constraints (highlighted in green). Normalizing  $\hat{H}_{\text{cost}}$  to a fixed energy scale  $J$  reduces



**Figure 4.** Dependence of the minimum energy gap on the strength of the penalty terms in the QUBO reduction of a random 3-SAT instance generated by uniquePT1 with  $N = 6$  variables. (a) Minimum energy gap  $\Delta E/J$  ( $J = \max_{ij} |J_{ij}^{(2)}|$ ) encountered during the annealing sweep as a function of the penalty strength  $\lambda$  in the cost Hamiltonian  $\hat{H}_{\text{cost}} = \hat{H}_{\text{problem}} + \lambda \hat{H}_{\text{constraints}}$ . The gray dashed line shows the value  $\Delta E/J \approx 0.06$  corresponding to the default choice  $\lambda = 1$  used throughout this work. In this example, the optimum  $\Delta E/J \approx 0.074$  is reached at the slightly lower value  $\lambda \approx 0.64$ . (b) Energy spectrum of the cost Hamiltonian  $\hat{H}_{\text{cost}}$  for the lowest 30 eigenvalues, with most of them being degenerate. Levels whose associated eigenstates fulfill (violate any) of the constraints are colored in green (red). The solid black lines in the panels (a) and (b) mark the location where the ground state becomes degenerate and no longer fulfills the consistency constraints, causing the minimum gap to vanish ( $\lambda = 0.25$  in this example).

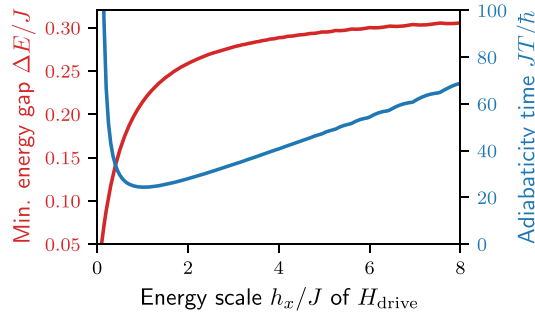
the energies (and thus the gap) of these two states by the factor  $\propto \lambda^{-1}$ . Consequently, it is counterproductive to choose the energy penalty too strong. Conversely,  $\lambda$  should not be chosen too small either: the largest minimum energy gap occurs close to the value of  $\lambda$  for which the first excited penalty-satisfying state and the first excited penalty-violating state become degenerate ( $\lambda = 0.5$  in figure 4(b)). Reducing  $\lambda$  further brings the first penalty-violating state closer to the ground state, resulting in a sharp decrease of the minimum gap.

To summarize, our example demonstrates that there is in general potential for optimization of the minimum energy gap in the QUBO reduction of a native PUBO problem by tuning the strength of the penalty terms. However, finding the optimal penalty strength generally requires solving the problem itself, which is prohibitively expensive for large exponentially hard problems. For our scaling analysis of the minimum gap, we have thus chosen the penalty strength in a problem-agnostic way as described in section 4.2.3. This heuristic choice is simple, incurs only moderate performance overhead, and reliably produces feasible solutions. Further optimizations may be achieved through more sophisticated schemes to approximate the optimal penalty strength [32], but such improvements would have only minor effects on the results of our PUBO versus QUBO analysis, as in our scenario the simple approach is already close to optimal.

#### 4.2.5. Influence of the driving strength on the annealing performance

So far, we have chosen the strength  $h^x$  of the driving Hamiltonian to be the same as the energy scale  $J$  of the cost Hamiltonian, see section 4.1.1. However,  $h^x$  is actually a free parameter that can be tuned to optimize the annealing schedule. In what follows, we discuss how the choice of the driving strength  $h^x$  influences the size of the minimum energy gap and the annealing performance.

To this end, we introduce the dimensionless cost Hamiltonian  $\hat{H}'_{\text{cost}} = \hat{H}_{\text{cost}}/J$  and the dimensionless driving Hamiltonian  $\hat{H}'_{\text{drive}} = \hat{H}_{\text{drive}}/h^x$ , expressed in units of the characteristic energy scale  $J$  and driving strength  $h^x$ , respectively. Consider now the dimensionless Hamiltonian  $\hat{H}'(g) = \hat{H}'_{\text{cost}} + g\hat{H}'_{\text{drive}}$  and let us assume that the energy gap between the ground state and the first excited state of  $\hat{H}'(g)$  becomes minimal at the (*a priori* unknown) critical value  $g_c$ . Writing the instantaneous Hamiltonian for the linear annealing schedule in equation (4) as  $\hat{H}(s) = sJ\hat{H}'_{\text{cost}} + (1-s)h^x\hat{H}'_{\text{drive}}$ , the relative strength  $g_c$  between  $\hat{H}_{\text{cost}}$  and  $\hat{H}_{\text{drive}}$ , and thus the minimum energy gap, is reached at the time  $s_c = (h^x/J)/(g_c + h^x/J)$  during the annealing sweep. The instantaneous Hamiltonian at this time reads  $\hat{H}(s_c) = s_c J \hat{H}'(g_c)$ . Consequently, the minimum energy gap



**Figure 5.** Influence of the driving strength on the minimum energy gap and the adiabaticity time for a linear annealing schedule  $(1-s)\hat{H}'_{\text{drive}} + s\hat{H}'_{\text{cost}}$ . The cost Hamiltonian  $\hat{H}_{\text{cost}} = J\hat{H}'_{\text{cost}}$  corresponds to the PUBO encoding of a random 3-SAT instance generated by uniquePT1 with  $N = 6$  variables and has characteristic energy scale  $J = \max_{ijk} |J_{ijk}^{(3)}|$ . The driving Hamiltonian  $\hat{H}_{\text{drive}} = h^x \hat{H}'_{\text{drive}} = -h^x \sum_i \hat{\sigma}_i^x$  represents a uniform transverse field of strength  $h^x$ . The minimum energy gap  $\Delta E$  increases monotonically with the driving strength (red curve), while the adiabaticity time  $T$ , corresponding to the right-hand side of equation (6), exhibits a minimum at  $h_x/J \approx 1.05$  (blue curve).

$\Delta E$  of the dimensionful Hamiltonian  $\hat{H}$  relates to the dimensionless gap  $\Delta E'_c$  of  $\hat{H}'(g_c)$  as

$$\Delta E/J = s_c \Delta E'_c = \frac{h^x/J}{g_c + h^x/J} \Delta E'_c. \quad (15)$$

This relation shows that the minimum energy gap increases monotonically with the driving strength  $h_x$  and saturates at  $J\Delta E'_c$  for large  $h_x/J$ , as illustrated in figure 5 for a single 3-SAT instance.

Despite the monotonic behavior of the minimum energy gap, a larger driving strength does not necessarily result in a better annealing performance. This is because a large value of  $h_x/J$  results in a steep ramp, making the schedule less adiabatic. Formally, this can be seen by examining the adiabatic condition in equation (6). As shown in figure 5 for a single 3-SAT instance, the adiabaticity time reaches a minimum where  $h_x$  and  $J$  become of comparable strength.

To understand this behavior, it is instructive to write the adiabatic condition in equation (6) as

$$t_{\text{sweep}} \gg T = \hbar \frac{V}{\Delta E^2} \approx \hbar \frac{J}{\Delta E^2} \left( V'_{\text{cost}} + \frac{h^x}{J} V'_{\text{drive}} \right). \quad (16)$$

Here, we have used the triangle inequality to separate the contributions of the cost and driving Hamiltonian to the quantity  $V$  into  $V'_{\text{cost}} = \max_s |\langle 1(s) | \hat{H}'_{\text{cost}} | 0(s) \rangle|$  and  $V'_{\text{drive}} = \max_s |\langle 1(s) | \hat{H}'_{\text{drive}} | 0(s) \rangle|$ , assuming a linear annealing schedule. This approximation becomes exact in the limit of either small or large  $h^x/J$ . For small  $h_x/J$ , the contribution of  $V'_{\text{drive}}$  can be neglected and the adiabaticity time is dominated by the behavior of the minimum energy gap according to equation (15), causing the adiabaticity time to diverge for  $h_x/J \rightarrow 0$ . For large  $h_x/J$ , the adiabaticity time is dominated by the linearly increasing second term in equation (16). The optimal driving strength corresponds to a tradeoff between these two scenarios, as illustrated in figure 5.

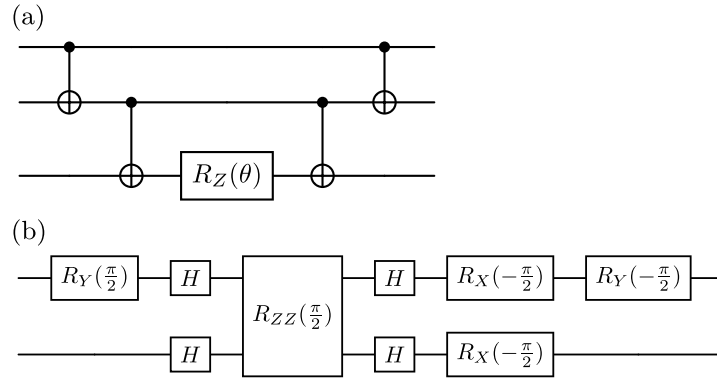
As this discussion shows, for a linear annealing schedule, the driving strength  $h^x$  should be chosen of comparable magnitude as the characteristic energy scale  $J$  of the cost Hamiltonian. The optimal value for  $h_x$  depends on (generally unknown) problem characteristics. For the specific instance used in figure 5, we find an optimal value of around  $h_x/J \approx 1.05$ . We checked that also for other problem instances the generic choice  $h_x = J$  used throughout this work is not far from the optimal condition for adiabaticity.

## 5. Implementing PUBO via multi-qubit interactions in quantum hardware

In this section, we discuss possibilities to implement three-body interactions in order to realize third-order PUBO formulations in quantum hardware. The exploration of such multi-body interaction terms and their role in enhancing quantum annealing protocols, for example by using them as non-stoquastic driving terms [46], has become an active area of research. As we will see, the constant overhead of directly implementing PUBO problems is more than compensated—already for small problem sizes—by the advantageous scaling of the minimal gap investigated in the previous section.

### 5.1. Native multi-qubit interactions and synthesis of three-body gates in quantum circuits

Several proposals exist to directly engineer native three-body spin interactions of the type  $\tilde{s}_i^z \tilde{s}_j^z \tilde{s}_k^z$  on the hardware level. For example, in trapped-ion systems, these include concurrent off-resonant driving of first



**Figure 6.** (a) The three-qubit gate  $R_{ZZZ}(\theta) = \exp(-i4\theta\tilde{s}_1^z\tilde{s}_2^z\tilde{s}_3^z)$  can be decomposed into four CNOT gates plus a single-qubit rotation. (b) Each CNOT gate can be further decomposed into one  $R_{ZZ}$  gate of fixed angle and multiple single-qubit rotations. The quantum gates in the decompositions are defined as  $R_\alpha(\theta) = \exp(-i\theta\tilde{s}^\alpha)$ , with  $\alpha \in \{X, Y, Z\}$ ,  $R_{ZZ}(\theta) = \exp(-2i\theta\tilde{s}^z\tilde{s}^z)$ , and  $H$  denoting the Hadamard gate.

and second sideband [47] (also relevant for other three-body interactions [48–50]) or a pulse sequence using squeezings and displacements of a phonon bus [51]. Another approach is via Floquet engineering of two-body interactions to obtain the three-body terms in a high-frequency expansion [52]. In addition, proposals for effective multi-body interactions on superconducting qubit architectures have been put forward [53]. Most available schemes share the common drawbacks that their implementation is somewhat complex and that the achievable higher-order interaction rates are considerably slower compared to those of two-body interactions. Moreover, the tunability of the relative strengths of three-body interactions between different triples of spins depends on the utilized scheme as well as hardware characteristics. Nonetheless, these techniques may be exploited in next-generation devices to realize analog quantum annealing with higher-order interactions that benefits from the advantages of direct PUBO encodings in terms of spatial and temporal resources.

While native higher-order interactions can also act as physical implementations of multi-qubit gates in digital quantum computers, due to the disadvantages mentioned above, most available platforms implement only single- and two-qubit gates natively. These can realize higher-order interaction gates, required for PUBO formulations, with only polynomial overhead. An example is given in figure 6: a three-qubit gate of the form  $R_{ZZZ}(\theta) = \exp(-i4\theta\tilde{s}_1^z\tilde{s}_2^z\tilde{s}_3^z)$  can be synthesized from four CNOT gates, plus a single-qubit rotation around the  $z$ -axis (see figure 6(a)). Each CNOT gate can be further decomposed into

$$\text{CNOT}_{1,2} = e^{-i\pi/4} R_{y_1}\left(-\frac{\pi}{2}\right) R_{x_1}\left(-\frac{\pi}{2}\right) R_{x_2}\left(-\frac{\pi}{2}\right) H_1 H_2 R_{zz}\left(\frac{\pi}{2}\right) H_1 H_2 R_{y_1}\left(\frac{\pi}{2}\right), \quad (17)$$

where we used a two-qubit controlled rotation gate  $R_{zz}$ , Hadamard gates  $H$  and single-qubit rotations around  $x$  and  $y$ -axis. We show the decomposition in figure 6(b). The overall circuit for the three-qubit gate then uses four two-body gates of type  $R_{zz}$  with fixed angle  $\pi/2$  and 29 single-qubit rotations. In this decomposition, two adjacent single-qubit gates could be further combined into only one single-qubit rotation. Similar decompositions are possible with other two-qubit gates like the Mølmer–Sørensen type gate  $R_{xx}$ . Note that in digital quantum annealing protocols, the three-body interaction strength in each Trotter step is controlled by the rotation angle  $\theta$ . Since  $\theta$  enters in figure 6(a) only via a single  $R_Z(\theta)$  gate, the overall three-qubit gate time is practically the same for different rotation angles.

As these examples show, the special structure of the  $ZZZ$  gate enables implementations with far less resources than necessary for a general three-body unitary operation (for example, state of the art numerical sequential optimization requires 15 CNOT gates to decompose a general three-body unitary operation [54, 55]). Importantly, while such an overhead may still be non-negligible in current noisy intermediate-scale quantum devices, it is a rather small constant factor. As discussed below, this overhead can be more than compensated by the higher resource efficiency of the PUBO encoding.

## 5.2. PUBO speedup

As suggested by our numerical results in section 4, the minimal gap decreases more slowly with increasing system size for the PUBO formulation of certain 3-SAT problems than for the corresponding QUBO reduction, indicating that there exists a system size above which the PUBO formulation will always be favorable. Specifically, this is the case for 3-SAT instances generated by toughSAT, see figure 3(b).

Even in scenarios where PUBO and QUBO scale approximately the same, as is the case for the 3-SAT problems generated by uniquePT1 shown in figure 3(a), a constant PUBO speedup is still possible. In the context of digital quantum annealing (see section 2.4), we can use the estimated overhead of the three-body interactions from section 5.1 together with the minimal gaps extracted numerically in section 4 in order to estimate the potential gain in computing time when using the direct PUBO formulation according to equation (8). In the example of figure 6, executing the three-qubit gate involves four two-qubit gates and therefore takes about four times as long as a single two-qubit gate (neglecting single-qubit gate times). In the context of quantum annealing, this corresponds to a ratio of three-body to two-body interaction strength of  $J^{(3)}/J^{(2)} \approx 1/4$ . Using equation (8) and the average values for the uniquePT1 generator listed in table 1, we obtain  $T_Q/T_P \approx 12.8 (\tilde{V}_Q/\tilde{V}_P) e^{0.024N}$ . In our numerics for small systems, we find a typical ratio of  $\tilde{V}_Q/\tilde{V}_P \approx 0.5$ , which only weakly depends on the system size. Our estimate therefore hints at a constant PUBO speedup in this regime; more precisely, an adiabatic sweep can be about six times faster if the PUBO instead of the QUBO formulation for this class of 3-SAT problems is used. If there is on top of that an (even small) exponential enhancement, as suggested by our numerical results, it will then dominate over a possible polynomial scaling of  $\tilde{V}_Q/\tilde{V}_P$  at larger system sizes and further improve the performance of PUBO over QUBO. Note that the precise value of the estimated speedup depends on the choice of the strength of the driving Hamiltonian as well as the strength of the constraint terms in the QUBO reduction. Using a better heuristic for the latter, e.g. the one proposed in [32], could possibly reduce the calculated PUBO improvement. From our analyses in sections 4.2.4 and 4.2.5, where we found the used parameters to be close to optimal, we expect such a change to be only minor.

For instances generated by toughSAT, we see numerical evidence for a much stronger exponential scaling advantage of the PUBO formulation compared to the corresponding QUBO reduction: under the same assumptions as above, we find  $T_Q/T_P \approx 0.81 (\tilde{V}_Q/\tilde{V}_P) e^{0.49N}$ . Already for small systems with  $N = 11$  classical variables, the adiabaticity time reduces by about two orders of magnitude in the PUBO formulation. That is, solving this family of exponentially hard 3-SAT problems via quantum annealing is expected to be two orders of magnitude faster when using the PUBO formulation over the conventional approach of applying a QUBO reduction, and this estimate increases by another order of magnitude for every  $\approx 4.7$  additional classical variables in the problem.

## 6. Conclusion

In this work, we have presented a variety of combinatorial problems that are naturally phrased in terms of higher-order PUBO. Using the PUBO formulation directly instead of the corresponding QUBO reduction—currently the standard choice in quantum annealing—can have a number of advantages: (1) PUBO formulations avoid additional ancillary qubits, thus reducing the spatial resources required. (2) The ancillary qubits required by QUBO reductions are typically subject to consistency constraints, which require large energy scales—a central limiting factor in existing quantum annealing hardware. (3) As we have illustrated numerically at the example of the paradigmatic 3-SAT problem, PUBO formulations can have significantly larger minimum energy gaps, allowing for faster annealing sweeps on both analog and digital platforms. This gain can outweigh the overhead incurred by higher-order interactions, as we have exemplified for digital quantum annealing with three-qubit gates synthesized from universal sets of single- and two-qubit gates. In particular, we find an exponential advantage of PUBO over QUBO in the scaling of the minimum energy gap with the system size for certain classes of problems (in our example, already for small toughSAT instances with  $N = 11$  variables, this suggests a potential speedup of about two orders of magnitude). Altogether, implementing optimization problems in quantum annealers as native PUBO rather than QUBO can thus lead to savings in terms of both spatial and temporal resources.

On top of that, formulating certain problems as PUBO can reveal underlying structures and give insights into the complexity of certain problems. As a case in point, rewriting 3-SAT instances generated by uniquePT4 (a generator constructing benchmark problems with unique solutions [16]) in PUBO form leads to a cost Hamiltonian described by only local fields that can be trivially solved. In addition, our numerical analysis of the minimum energy gap reveals characteristics of other 3-SAT generators found in [16] (e.g. the fact that the generator uniquePT1 creates on average harder instances than toughSAT), which are relevant for future studies of both classical and quantum algorithms where these generators can be used for benchmark purposes.

Although in this work we have focused on the paradigmatic 3-SAT problem, it is plausible that a similar exponential scaling advantage holds for other problems which are naturally described as a PUBO, such as those exemplified in appendix A. In the future, it will be interesting to quantitatively study potential speedups also for these and other problems. Another interesting question is whether the presence of multi-qubit operators modifies the behavior of quantum resources during the annealing sweep [56–59]. Finally, the favorable scaling of the adiabaticity time for certain families of 3-SAT problems, as suggested by

our analysis of the minimum energy gap in section 4.2.3, encourages more systematic studies comparing the time complexity of quantum annealing to state-of-the-art classical algorithms [17–19].

### Data availability statement

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

### Acknowledgments

We acknowledge useful discussions with Marcel Seelbach Benkner, Ivan Boldin, Michael Möller, Junichi Okamoto, Sebastian Rubbert, Theerapot Sriarunothai and Christof Wunderlich. The work reported in this publication is based on a project that was funded by the German Federal Ministry for Education and Research under the funding Reference No. 13N16437. The authors are solely responsible for the content of this publication. This work has benefited from Q@TN, the joint lab between University of Trento, FBK—Fondazione Bruno Kessler, INFN—National Institute for Nuclear Physics, and CNR—National Research Council. We acknowledge support by Provincia Autonoma di Trento.

### Appendix A. Selected PUBO problems

There exists a plethora of optimization problems that are naturally posed as PUBO. In this appendix, we mention a few relevant examples.

#### A.1. $p$ -spin model

The cost Hamiltonian of the  $p$ -spin model ( $p \geq 2$ ) reads

$$\hat{H}_{\text{cost}} = -N \left( \frac{1}{N} \sum_{i=1}^N \hat{\sigma}_i^z \right)^p. \quad (\text{A1})$$

For  $p = 2$ , this Hamiltonian is known as the Lipkin–Meshkov–Glick model [60–62], while for  $p \rightarrow \infty$  it recovers the Grover problem [63]. The solution of this model is trivial (all spin up for  $p$  odd; all spin down or all spin up for  $p$  even), and the large structure by which it is characterized enables efficient classical simulations even of non-stoquastic annealing protocols [64]. Nevertheless, its infinite range all-to-all interactions make it a useful toy model for testing improvements of quantum-optimization algorithms. For example, while standard annealing protocols with a transverse field as the driving Hamiltonian encounter a first-order phase transition in the  $p$ -spin model [63], non-stoquastic protocols soften it to a second-order phase transition. This example illustrates the power of non-stoquastic protocols for improving adiabaticity and thus saving computing time [65, 66].

#### A.2. Hypergraph coloring

Consider an undirected graph  $G = (V, E)$ , consisting of a finite set of vertices  $V$  and a set of edges  $E$ , the latter being a set of subsets of  $V$ . For a regular undirected graph, all edges are unordered pairs of vertices. For a hypergraph on the other hand, each edge can contain any number of vertices.

Given  $G$  and a number of  $k$  colors, the hypergraph vertex coloring problem asks whether there is an assignment of exactly one colour to each vertex such that no edge contains vertices all of the same colour [67]. A QUBO formulation of the regular graph coloring is well-known [68], which can be easily generalized to hypergraphs: let  $x_n^i = 1$  if vertex  $n$  has color  $i$ , else 0. The cost function of the problem can then be written as

$$f(\mathbf{x}) = \sum_{n \in V} \left( 1 - \sum_{i=1}^k x_n^i \right)^2 + \sum_{e \in E} \sum_{i=1}^k \prod_{n \in e} x_n^i, \quad (\text{A2})$$

where the first term ensures that each vertex has exactly one color assigned to it and the second term penalizes each edge whose vertices are all of the same color. Thus, any solution with zero energy is a proper vertex coloring of the hypergraph.

To implement this PUBO problem encoding on a quantum computer,  $k|V|$  qubits are required (with  $|V|$  being the number of vertices in the graph). To reduce equation (A2) to a QUBO problem, one needs to introduce ancilla variables for every edge in the graph containing more than two vertices (the exact amount needed will depend on the degree of the edges and their overlap with each other).

Other generalizations of graph problems to hypergraphs similarly yield PUBO problems with higher-order interaction terms, e.g. the vertex cover problem or the graph partitioning problem [69].

### A.3. NAE- $k$ -SAT

Apart from  $k$ -SAT, which we have discussed in detail in section 3.2, there are many other variations and constraints on the general satisfiability problem. One of those is the NAE- $k$ -SAT (Not-All-Equal- $k$ -SAT) problem [70], which can also be naturally formulated as a PUBO. Like  $k$ -SAT, NAE- $k$ -SAT consists of a conjunction of clauses, where each clause contains  $k$  Boolean variables and is only *true*, if at least one variable is *true* and at least one variable is *false*. Only for  $k \geq 4$  is the problem naturally formulated as a PUBO with higher-order interactions [70]. For example, for  $k = 4$ , a clause  $C(x_i, x_j, x_k, x_l)$  can be encoded into one four-body interaction  $\sim \tilde{s}_i^z \tilde{s}_j^z \tilde{s}_k^z \tilde{s}_l^z$  and six two-body interaction penalty terms of the cost Hamiltonian.

As for  $k$ -SAT, this PUBO encoding requires one qubit for each Boolean variable. The equivalent reduction to a QUBO, on the other hand, requires (for  $k = 4$ ) in the worst case two additional ancilla qubits for each clause in the problem.

### A.4. Traveling salesperson problem (TSP) with time windows

A variation on the famous TSP is the TSP with time windows. Here, like in the TSP, we ask for a Hamiltonian path with minimum weight on a graph with weights assigned to all edges. In addition, each vertex has to be visited in a specific time window. Both QUBO and PUBO formulations of this problem are known. In [71], the authors propose two different QUBO and one PUBO formulation of degree four for the problem. In the PUBO case, the number of required qubits scales as  $\mathcal{O}(n^2 + n\delta)$  in the number of vertices  $n$  (here,  $\delta$  is a number scaling at most linearly in  $n$ ). In contrast, the QUBO formulations scale as  $\mathcal{O}(n^3 + n\delta)$  and  $\mathcal{O}(n^2 + n^2\delta)$ , respectively, for the edge-based and integer linear programming formulations.

## Appendix B. QUBO formulations of 3-SAT

The most resource-efficient way (in terms of qubit requirements) to formulate a 3-SAT instance as a QUBO problem is the direct reduction of the PUBO formulation to QUBO via the introduction of slack variables (see section 3.2.2). Here, we also list two more resource-intensive methods.

### B.1. 3-SAT as QUBO via the maximum independent set problem

A well-known example of a QUBO formulation of 3-SAT uses the equivalent maximal independent set problem (MIS) [72]: Draw a graph with three vertices for each clause (corresponding to the three involved literals). All vertices belonging to the same clause as well as pairs of vertices corresponding to mutually exclusive literals (i.e.  $x_i$  and  $\neg x_i$ ) are connected by edges, the set of which is denoted with  $E$ . The MIS problem then asks whether it is possible to color  $M$  vertices without two colored vertices being connected by an edge. If the answer is yes, the original 3-SAT problem is satisfiable. This MIS problem can be easily translated to a QUBO problem requiring  $3M$  qubits (or about  $\text{round}(3 \times 4.24N) = \text{round}(12.72N)$  for hard problems):

$$f_{\text{MIS}}(\mathbf{x}) = \sum_{ij \in E} x_i x_j. \quad (\text{B1})$$

Although generically more resources are required than for the reduction with slack variables, a potential advantage of the MIS formulation, especially regarding its use on physical quantum devices, lies in the fact that all interactions have the same energy scale, whereas slack variables for a direct QUBO reduction are introduced with constraints that are imposed through high energy penalties.

### B.2. 3-SAT as QUBO via linear inequality constraints

Finally, we mention also a linear formulation of 3-SAT as a QUBO problem [73], where one introduces one ancillary Boolean variable  $z_m$  for each clause  $C_m$  that indicates whether said clause is fulfilled or not. One then asks to maximize the number of fulfilled clauses

$$\max \sum_m z_m \quad (\text{B2})$$

under the inequality constraints

$$\sum_{x_i \in \text{TRUE}_m} x_i + \sum_{x_i \in \text{FALSE}_m} (1 - x_i) \geq z_m, \quad (\text{B3})$$

where  $\text{TRUE}_m$  ( $\text{FALSE}_m$ ) is the set of literals contained in  $C_m$  satisfying the clause if chosen as *true* (*false*). One can then introduce two slack variables per inequality constraint to express them as equalities and write the problem in QUBO form.

Overall, this approach needs  $3M$  ancilla qubits in addition to the  $N$  qubits for the literals, as well as an energy penalty to enforce the constraints. This scaling has to be compared to the one using the mapping to the MIS problem ( $3M$  qubits, no different energy scales), the one decomposing the PUBO problem using slack variables ( $N$  qubits plus at most  $M$  ancillas, plus energy penalties), and the direct PUBO implementation ( $N$  qubits, interactions of degree three, no different energy scales).

## ORCID iDs

Sebastian Nagies  <https://orcid.org/0009-0009-8269-3237>

Kevin T Geier  <https://orcid.org/0000-0001-9411-3898>

Javed Akram  <https://orcid.org/0000-0003-3115-7601>

Michael Johanning  <https://orcid.org/0000-0003-4746-7559>

Philipp Hauke  <https://orcid.org/0000-0002-0414-1754>

## References

- [1] Santoro G E, Martoňák R, Tosatti E and Car R 2002 Theory of quantum annealing of an Ising spin glass *Science* **295** 2427
- [2] Martoňák R, Santoro G E and Tosatti E 2004 Quantum annealing of the traveling-salesman problem *Phys. Rev. E* **70** 057701
- [3] Hauke P, Katzgraber H G, Lechner W, Nishimori H and Oliver W D 2020 Perspectives of quantum annealing: methods and implementations *Rep. Progr. Phys.* **83** 054401
- [4] Rajak A, Suzuki S, Dutta A and Chakrabarti B K 2022 Quantum annealing: an overview *Philos. Trans. R. Soc.* **381** 20210417
- [5] Yarkoni S, Raponi E, Bäck T and Schmitt S 2022 Quantum annealing for industry applications: introduction and review *Rep. Progr. Phys.* **85** 104001
- [6] King A D et al 2024 Computational supremacy in quantum simulation (arXiv:2403.00910 [quant-ph])
- [7] Weinberg S J, Sanches F, Ide T, Kamiya K and Correll R 2023 Supply chain logistics with quantum and classical annealing algorithms *Sci. Rep.* **13** 4770
- [8] Micheletti C, Hauke P and Faccioli P 2021 Polymer physics by quantum computing *Phys. Rev. Lett.* **127** 080501
- [9] Slongo F, Hauke P, Faccioli P and Micheletti C 2023 Quantum-inspired encoding enhances stochastic sampling of soft matter systems *Sci. Adv.* **9** eadi0204
- [10] Outeiral C, Morris G M, Shi J, Strahm M, Benjamin S C and Deane C M 2021 Investigating the potential for a limited quantum speedup on protein lattice problems *New J. Phys.* **23** 103030
- [11] Chermoshentsev D A, Malyshev A O, Esencan M, Tiunov E S, Mendoza D, Aspuru-Guzik A, Fedorov A K and Lvovsky A I 2022 Polynomial unconstrained binary optimisation inspired by optical simulation (arXiv:2106.13167 [quant-ph])
- [12] Schmidbauer L, Wintersperger K, Lobe E and Maurer W 2024 Polynomial reduction methods and their impact on QAOA circuits (arXiv:2406.08889 [quant-ph])
- [13] Biamonte J D 2008 Nonperturbative  $k$ -body to two-body commuting conversion Hamiltonians and embedding problem instances into Ising spins *Phys. Rev. A* **77** 052331
- [14] Babbush R, O’Gorman B and Aspuru-Guzik A 2013 Resource efficient gadgets for compiling adiabatic quantum optimization problems *Ann. Phys. (Berl.)* **525** 877
- [15] Stein J, Chamanian F, Zorn M, Nülein J, Zielinski S, Kölle M and Linnhoff-Popien C 2023 Evidence that PUBO outperforms QUBO when solving continuous optimization problems with the QAOA *Proc. Companion Conf. on Genetic and Evolutionary Computation (GECCO’23 Companion)* (Association for Computing Machinery) pp 2254–62
- [16] Hsieh C-Y, Amador R and Chiang C-F 2021 Exploration of hard to solve 3-SAT problems *J. Inf. Technol. Ind.* **7** 23
- [17] Punnen A P 2022 *The Quadratic Unconstrained Binary Optimization Problem: Theory, Algorithms and Applications* 1st edn (Springer)
- [18] Glover F, Kochenberger G, Hennig R and Du Y 2022 Quantum bridge analytics I: a tutorial on formulating and using QUBO models *Ann. Oper. Res.* **314** 141
- [19] Du Y, Wang H, Hennig R, Hulandageri A, Kochenberger G and Glover F 2025 New advances for quantum-inspired optimization *Intl. Trans. in Op. Res.* **32** 6
- [20] Jansen S, Ruskai M-B and Seiler R 2007 Bounds for the adiabatic approximation with applications to quantum computation *J. Math. Phys.* **48** 102111
- [21] Lidar D A, Rezakhanlou A T and Hama A 2009 Adiabatic approximation with exponential accuracy for many-body systems and quantum computation *J. Math. Phys.* **50** 102106
- [22] Amin M H S 2009 Consistency of the adiabatic theorem *Phys. Rev. Lett.* **102** 220401
- [23] Cheung D, Høyer P and Wiebe N 2011 Improved error bounds for the adiabatic approximation *J. Phys. A: Math. Theor.* **44** 415302
- [24] Čepaitė I, Polkovnikov A, Daley A J and Duncan C W 2023 Counterdiabatic optimized local driving *PRX Quantum* **4** 010312
- [25] García-Pintos L P, Sahasrabudhe M and Arenz C 2024 Tighter lower bounds on quantum annealing times (arXiv:2410.14779 [quant-ph])
- [26] Bottarelli A, de Andoin M G, Chandarana P, Paul K, Chen X, Sanz M and Hauke P 2024 Symmetry-enhanced counterdiabatic quantum algorithm for qudits (arXiv:2410.06710 [quant-ph])
- [27] Altshuler B, Krovi H and Roland J 2010 Anderson localization makes adiabatic quantum optimization fail *Proc. Natl Acad. Sci.* **107** 12446
- [28] Farhi E, Goldstone J and Gutmann S 2014 A quantum approximate optimization algorithm (arXiv:1411.4028 [quant-ph])
- [29] Sack S H and Serbyn M 2021 Quantum annealing initialization of the quantum approximate optimization algorithm *Quantum* **5** 491
- [30] Blanes S, Casas F and Murua A 2024 Splitting methods for differential equations *Acta Numer.* **33** 1–161
- [31] Boros E and Hammer P L 2002 Pseudo-boolean optimization *Discrete Appl. Math.* **123** 155
- [32] Alessandrini E, Ramos-Calderer S, Roth I, Traversi E and Aolita L Alleviating the quantum Big- $M$  problem 2023 (arXiv:2307.10379 [quant-ph])

- [33] Aspöqvall B, Plass M F and Tarjan R E 1982 A linear-time algorithm for testing the truth of certain quantified boolean formulas *Inform. Process. Lett.* **14** 195
- [34] Cook S A 1971 The complexity of theorem-proving procedures *Proc. Third Annual ACM Symp. on Theory of Computing (STOC'71)* (Association for Computing Machinery) pp 151–8
- [35] Karp R M 1972 Reducibility among combinatorial problems *Complexity of Computer Computations* (Springer US) pp 85–103
- [36] Aaronson S 2010 BQP and the polynomial hierarchy *Proc. 42nd ACM Symp. on Theory of Computing (STOC'10)* (New York, NY, USA) (Association for Computing Machinery) pp 141–50
- [37] Heim B, Rønnow T F, Isakov S V and Troyer M 2015 Quantum versus classical annealing of Ising spin glasses *Science* **348** 215
- [38] Albash T and Lidar D A 2018 Adiabatic quantum computation *Rev. Mod. Phys.* **90** 015002
- [39] Willsch D, Willsch M, Gonzalez Calaza C D, Jin F, De Raedt H, Svensson M and Michielsen K 2022 Benchmarking advantage and D-Wave 2000Q quantum annealers with exact cover problems *Quantum Inf. Process.* **21** 141
- [40] Bernaschi M, González-Adalid Pemartín I, Martín-Mayor V and Parisi G 2024 The quantum transition of the two-dimensional Ising spin glass *Nature* **631** 749
- [41] Mézard M, Parisi G and Zecchina R 2002 Analytic and algorithmic solution of random satisfiability problems *Science* **297** 812
- [42] Žnidarič M 2005 Scaling of the running time of the quantum adiabatic algorithm for propositional satisfiability *Phys. Rev. A* **71** 062305
- [43] D-Wave Systems (available at: [www.dwavesys.com/](http://www.dwavesys.com/))
- [44] Verma A, Lewis M and Kochenberger G 2021 Efficient QUBO transformation for higher degree pseudo boolean functions (arXiv:2107.11695 [math.OC])
- [45] Hansen T D, Kaplan H, Zamir O and Zwick U 2019 Faster k-SAT algorithms using biased-PPSZ *Proc. 51st Annual ACM SIGACT Symp. on Theory of Computing (STOC'19)* (ACM) pp 578–89
- [46] Ghosh R, Nutricati L A, Feinstein N, Warburton P A and Bose S 2024 Exponential speed-up of quantum annealing via n-local catalysts (arXiv:2409.13029 [quant-ph])
- [47] Bermudez A, Porras D and Martin-Delgado M A 2009 Competing many-body interactions in systems of trapped ions *Phys. Rev. A* **79** 060303
- [48] Yang D, Giri G S, Johanning M, Wunderlich C, Zoller P and Hauke P 2016 Analog quantum simulation of  $(1 + 1)$ -dimensional lattice QED with trapped ions *Phys. Rev. A* **94** 052321
- [49] Andrade B, Davoudi Z, Graß T, Hafezi M, Pagano G and Seif A 2022 Engineering an effective three-spin Hamiltonian in trapped-ion systems for applications in quantum simulation *Quantum Sci. Technol.* **7** 034001
- [50] Nagies S, Geier K T, Akram J, Okamoto J, Bantounas D, Wunderlich C, Johanning M and Hauke P 2025 The role of higher-order terms in trapped-ion quantum computing with magnetic gradient induced coupling *Quantum Sci. Technol.* **10** 025051
- [51] Katz O, Cetina M and Monroe C 2022  $n$ -body interactions between trapped ion qubits via spin-dependent squeezing *Phys. Rev. Lett.* **129** 063603
- [52] Decker K, Karrasch C, Eisert J and Kennes D 2020 Floquet engineering topological many-body localized systems *Phys. Rev. Lett.* **124** 190601
- [53] Chancellor N, Zohren S and Warburton P A 2017 Circuit design for multi-body interactions in superconducting quantum annealing systems with applications to a scalable architecture *npj Quantum Inf.* **3** 21
- [54] Rakyta P and Zimborás Z 2022 Approaching the theoretical limit in quantum gate decomposition *Quantum* **6** 710
- [55] Madden L and Simonetto A 2022 Best approximate quantum compiling problems *ACM Trans. Quantum Comput.* **3** 1–29
- [56] Orús R and Latorre J I 2004 Universality of entanglement and quantum-computation complexity *Phys. Rev. A* **69** 052308
- [57] Lanting T et al 2014 Entanglement in a quantum annealing processor *Phys. Rev. X* **4** 021041
- [58] Hauke P, Bonnes L, Heyl M and Lechner W 2015 Probing entanglement in adiabatic quantum optimization with trapped ions *Front. Phys.* **3** 21
- [59] Santra G C, Roy S S, Egger D J and Hauke P 2024 Genuine multipartite entanglement in quantum optimization (arXiv:2411.08119 [quant-ph])
- [60] Lipkin H, Meshkov N and Glick A 1965 Validity of many-body approximation methods for a solvable model *Nucl. Phys. B* **62** 188
- [61] Meshkov N, Glick A and Lipkin H 1965 Validity of many-body approximation methods for a solvable model *Nucl. Phys. B* **62** 199
- [62] Glick A, Lipkin H and Meshkov N 1965 Validity of many-body approximation methods for a solvable model *Nucl. Phys. B* **62** 211
- [63] Jörg T, Krzakala F, Kurchan J, Maggs A C and Pujos J 2010 Energy gaps in quantum first-order mean-field-like transitions: The problems that quantum annealing cannot solve *Europhys. Lett.* **89** 40004
- [64] Ohzeki M 2017 Quantum Monte Carlo simulation of a particular class of non-stoquastic Hamiltonians in quantum annealing *Sci. Rep.* **7** 41186
- [65] Seki Y and Nishimori H 2012 Quantum annealing with antiferromagnetic fluctuations *Phys. Rev. E* **85** 051112
- [66] Durkin G A 2019 Quantum speedup at zero temperature via coherent catalysis *Phys. Rev. A* **99** 032315
- [67] Kang D Y, Kelly T, Kühn D, Methuku A and Osthus D 2023 Graph and hypergraph colouring via nibble methods: A survey *European Congress of Mathematics* (EMS Press) pp 771–823
- [68] Lucas A 2014 Ising formulations of many NP problems *Front. Phys.* **2** 5
- [69] Dominguez F, Unger J, Traube M, Mant B, Ertler C and Lechner W 2023 Encoding-independent optimization problem formulation for quantum computing *Front. Quantum Sci. Technol.* **2** 1229471
- [70] Bashar M K and Shukla N 2023 Designing Ising machines with higher order spin interactions and their application in solving combinatorial optimization *Sci. Rep.* **13** 9558
- [71] Salehi O, Glos A and Miszczak J A 2022 Unconstrained binary models of the travelling salesman problem variants for quantum optimization *Quantum Inf. Process.* **21** 67
- [72] Choi V 2010 Adiabatic quantum algorithms for the np-complete maximum-weight independent set, exact cover and 3SAT problems (arXiv:1004.2226 [quant-ph])
- [73] Herrman R, Treffert L, Ostrowski J, Lotshaw P C, Humble T S and Siopsis G 2021 Globally optimizing QAOA circuit depth for constrained optimization problems *Lect. Notes. Comput. Sci.* **14** 294