ILP-based resource optimization realized by quantum annealing for optical wide-area communication networks—A framework for solving combinatorial problems of a real-world application by quantum annealing

Arthur Witt [1]*, Jangho Kim [2]*, Christopher Körber [3] and Thomas Luu [2]

[1]Institute of Communication Networks and Computer Engineering, University of Stuttgart, Stuttgart, Germany, [2]Institute for Advanced Simulation (IAS-4) & Jülich Aachen Research Alliance High Performance Computing, Forschungszentrum Jülich, Jülich, Germany, [3]Fraunhofer Research Institution for Energy Infrastructures and Geothermal Systems IEG, Fraunhofer IEG, Bochum, Germany

Resource allocation of wide-area internet networks is inherently a combinatorial optimization problem that if solved quickly, could provide near real-time adaptive control of internet-protocol traffic ensuring increased network efficacy and robustness, while minimizing energy requirements coming from power-hungry transceivers. In recent works we demonstrated how such a problem could be cast as a quadratic unconstrained binary optimization (QUBO) problem that can be embedded onto the D-Wave Advantage^{TM} quantum annealer system, demonstrating proof of principle. Our initial studies left open the possibility for improvement of D-Wave solutions via judicious choices of system run parameters. Here we report on our investigations for optimizing these system parameters, and how we incorporate machine learning (ML) techniques to further improve on the quality of solutions. In particular, we use the Hamming distance to investigate correlations between various system-run parameters and solution vectors. We then apply a decision tree neural network (NN) to learn these correlations, with the goal of using the neural network to provide further guesses to solution vectors. We successfully implement this NN in a simple integer linear programming (ILP) example, demonstrating how the NN can fully map out the solution space that was not captured by D-Wave. We find, however, for the 3-node network problem the NN is not able to enhance the quality of space of solutions.

KEYWORDS

discrete optimization, integer linear program, machine learning, quantum annealing, quantum computing, resource allocation, wide-area networks

1 Introduction

Quantum computing is a cutting-edge technology that has gained significant relevance during the last decades. Algorithms for searching and optimization are currently studied intensively on quantum computers as they hold the potential for solving problems with non-polynomial (NP) complexity very efficiently. Nowadays, quantum computers have reached a scale that allows for the solution of non-trivial problems which have real-world applications.

One example is the energy-aware resource allocation of wide-area networks (Chiaraviglio et al., 2009). In these cases, one can consider the resource allocation as an optimization problem and introduce it as a relevant application of quantum computing, and in particular quantum annealing (QA), as it inherently provides a certain failure tolerance with self-healing capability. Further, the problem has NP complexity and requires frequent solutions for just-in-time adaptation of the network. If solutions are generated quickly, say on the order of seconds, a revolution in network operation with increased network efficiency might be possible since current solutions obtained from classical and/or heuristic algorithms require 15 minutes or more for time-to-solution as shown in Feller (2012) and Tornatore et al. (2002). In previous studies, (Witt et al., 2023), we have demonstrated how this resource allocation problem can be formulated, based on an integer linear program (ILP) model, as a quadratic unconstrained binary problem (QUBO) which can then be embedded onto a quantum annealer.

In our initial studies we used the D-Wave Advantage™ system (JUPSI) at the Forschungszentrum Jülich to perform the quantum annealing. As part of the solution process, the QUBO problem was embedded onto the quantum qubits prior to performing the quantum annealing. This entailed mapping the problem onto a network of logical qubits, whereby each logical qubit consists of a constellation, or "chain", of physical qubits. This mapping ensures the requisite "connectivity" of the logical qubits as dictated by the QUBO problem. We discovered that the network optimizing approach was greatly limited by this embedding process. For example, the optimization problem of a network with three nodes can be described as a QUBO with ∼100 binary variables (logical qubits). Even though this 3-node problem is quite small, the required amount of physical qubits was in the range of 500 qubits, representing already roughly 10% of the physical qubits available in the D-Wave Advantage™ system. With the current embedding process that we employed at the time, a simulation of a 15-node problem, corresponding to a real-world network, would require a quantum annealer with ∼50,000 physical qubits, which is an order of magnitude larger then current systems. We note that the embedding process is not unique.

Our initial studies also had limited scope in system run parameters, such as annealing time and profile of the annealing process, penalty factor of the QUBO matrix, and chain strength between physical qubits constituting logical qubits. Our choice of run parameters were constrained mainly to system default values, with little exploration on the dependence of quality of solutions on these run parameters. Therefore there is potential room for increasing the efficiency of the quantum annealing process (which would result in better quality and more *feasible* solutions) by judicious choice of optimized run parameters.

In this paper we address some of these issues by studying the process of annealing with the aim to optimize the parameters for the quantum annealing procedure. We introduce solution quality metrics for evaluation purposes. Of particular import is the Hamming distance metric, which rates the distance between the ideal and obtained solution vector in binary space. By using D-Wave solutions in conjunction with the Hamming distance to optimal solution, we empirically determine correlations between various run parameters and the quality of solution. These correlations guide us in determining optimized run parameters for the system in question, with the hope that the same optimized run parameters can be applied to similar, but larger, systems. Furthermore, we apply a decision tree neural network (NN) to learn these correlations, after which we use the NN to "guess" improved solutions. This NN represents a machine learning (ML) approach that we couple with D-Wave generated solutions that aims at providing better quality solutions, and represents an example of a hybrid classical (ML)/quantum (QA) procedure for solving the combinatorial optimization problem.

Our paper is organized as follows. In the following section we give a cursory description of ILPs in general, the used method to solve ILPs on quantum annealer, and two ILPs that we examined in our study. We then introduce in Subsection 2.4 a Hamming distance metric, and demonstrate how it is used to derive correlations between quality of solutions and various system run parameters. Such correlations will be "learned" by our decision tree NN, which we describe in detail in Subsection 2.5. In Section 3 we present our findings. We first concentrate on a simple ILP problem, demonstrating that our hybrid classical/quantum procedure does indeed result in new feasible solutions while at the same time providing guidance on optimized run parameters. We then apply the formalism to the 3-node network problem mentioned above, where here we see limited improvement in solutions, all of which unfortunately are nowhere near the optimal solution. In Section 4 we discuss our findings and recapitulate. We comment on possible future directions of investigation.

2 Materials and methods

2.1 The concept of integer linear programs

The investigations performed in our work fall under the class of discrete optimization problems, meaning variables x to be optimized take on only discrete values. Such problems can be cast succinctly as an integer linear program (ILP), where certain constraints, given as a set of linear (in-)equation, have to be satisfied while minimizing a linear function. An ILP can be defined in its canonical form by (Equations 1–4):

$$\text{objective} \quad x_0 = \underset{x}{\operatorname{argmin}}\{c^{\mathrm{T}} x\} \qquad (1)$$

$$\text{constraints} \quad Ax + b \leq 0 \qquad (2)$$

$$\text{variables} \quad x \in \mathbb{N}^n, \, x_i \geq 0 \qquad (3)$$

$$\text{constants} \quad c \in \mathbb{R}^n, b \in \mathbb{R}^m, A \in \mathbb{R}^{m \times n}. \qquad (4)$$

The ILP's objective function can be seen as a loss function and is defined in Equation (1) with a vector of cost terms c weighting

the variable vector $\boldsymbol{x}$. Matrix $\boldsymbol{A}$ and vector $\boldsymbol{b}$ parameterize the linear equations that represent the inequality constraints (Equation 2). They can be reshaped to equality constraints, $\boldsymbol{Ax} + \boldsymbol{b} + \boldsymbol{s} = \boldsymbol{0}$, by introducing *slack* variables $\boldsymbol{s} \in \mathbb{R}, s_j \geq 0$. This is a typical step within the classical ILP-solving algorithm *simplex*, see Nash (2000). We use the convention, that $\mathbb{R}$ are real-valued numbers, $\mathbb{N}$ natural numbers inclusive zero, and $\mathbb{B}$ binary numbers.

Such problems are well-known to be non-polynomial (NP)-hard in general. According to Karp (1972) and Adler et al. (2014), linear programs are a rare class of problems in NP that resists the classification as NP-complete or polynomial-solvable problems. Lenstra (1983) argued, that mixed-integer linear programs with fixed number of variables are solvable in polynomial-time. In contrary, Nguyen and Pak (2017) present integer programs that are NP-complete, even for fixed number of variables. Their work further shows, that some integer programs are solvable in polynomial-time. We can conclude, that bounded integer linear programs are NP-complete and are solvable within polynomial time in few cases.

2.2 Solving integer linear programs on quantum annealer

Quantum annealers are well-suited for investigating ILP problems. However, an additional modification to the ILP problem is required prior to embedding the problem on the quantum qubits. Here the constraints are included into the cost function (to be minimized) by introduction of penalty weight p. In so doing, the original ILP problem with constraints is recast into quadratic form without constraints,

$$\left.\begin{array}{c} \boldsymbol{x}_0 = \underset{\boldsymbol{x}}{\arg\min} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \right\} \\ \boldsymbol{Ax} + \boldsymbol{b} \leq \boldsymbol{0} \\ \boldsymbol{x} \in \mathbb{N}^n \geq 0 \end{array}\right\} \longleftrightarrow \left\{\begin{array}{c} \boldsymbol{q}_0 = \underset{\boldsymbol{q}}{\arg\min} \left\{ \boldsymbol{q}^\top \boldsymbol{Q} \boldsymbol{q} + C \right\} \\ \boldsymbol{q} \in \{0,1\}^k \end{array}\right. .$$

Here $\boldsymbol{A}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ are problem specific parameters as introduced in the previous section. It is useful to classify solution vectors $\boldsymbol{x}$ into two categories: **feasible** solutions which fulfill the constraints and **unfeasible** solutions which violate the constraints. While a feasible solution to the ILP is not necessarily an optimal solution, an unfeasible solution hypothetically can have a smaller objective value than the optimal feasible solution.

The problem is mapped to the quadratic unconstrained binary optimization (QUBO) by definition of matrix $\boldsymbol{Q}$ that includes a penalty factor p and a constant C. The inequality can be expressed by an equation and another minimization over a slack variable $\boldsymbol{s}$ incorporated into the bit vector $\boldsymbol{q}$. The QUBO objective function minimizes both the ILP objective function plus another objective function representing the constraints. The penalty term expresses the relative weight between both (ILP objective and constraint) objective functions, see Chang et al. (2020) and Witt et al. (2023) for a more detailed description. Finding the solution set $\boldsymbol{q}_0$ that provides the absolute minimum of $\boldsymbol{q}^\top \boldsymbol{Q} \boldsymbol{q}$ is equivalent to solving the original ILP problem with solution vector $\boldsymbol{x}_0$.

The D-Wave Advantage$^{\text{TM}}$ system is adapted to solving the Ising spin system that represents an array of binary spins with interactions between spins $\boldsymbol{\sigma}$ giving by some connectivity matrix $\boldsymbol{J}$ and external magnetic field $\boldsymbol{h}$. Our QUBO matrix can easily be rewritten using $\boldsymbol{J}$ and $\boldsymbol{h}$ without any loss of generality,

$$\boldsymbol{q}^\top \boldsymbol{Q} \boldsymbol{q} + C \Leftrightarrow \boldsymbol{\sigma}^\top \boldsymbol{J} \boldsymbol{\sigma} + \boldsymbol{h}^\top \boldsymbol{\sigma} + g$$

$$\text{with} \quad \begin{cases} \boldsymbol{J} &= \frac{1}{4} \boldsymbol{Q}_0 \\ \boldsymbol{h} &= \frac{1}{2} \hat{\boldsymbol{q}} + \frac{1}{2} \boldsymbol{Q}_0 \mathbf{1} \\ g &= \frac{1}{4} \mathbf{1}^\top \boldsymbol{Q}_0 \mathbf{1} + \frac{1}{2} \mathbf{1}^\top \hat{\boldsymbol{q}} + C \end{cases},$$

with $\boldsymbol{Q}_0 = \boldsymbol{Q} - \text{diag}\{\hat{\boldsymbol{q}}\}$, $\hat{\boldsymbol{q}} = \text{diag}^{-1}\{\boldsymbol{Q}\}$, and g some constant. The problem is now well-suited for the D-Wave machine. In Chang et al. (2020) and Witt et al. (2023), we demonstrated proof of principle that such a problem can be solved on a quantum annealer.

2.3 Investigated integer linear programs

In this work we have investigated multiple ILPs, two of which we define explicitly here. The details of the remaining ILPs we considered are described in our accompanying Supplementary material. The first ILP optimizes the selection of two integer variables under some constraints. It provides a test case where all possible solutions can be studied with the approach of brute force sampling, i.e., it provides a well-suited setup for benchmarking. The second ILP describes a realistic network resource optimization as studied in Witt et al. (2023). As possible solutions are representable as binary vectors with more than 60 variables, a brute force sampling is not applicable within reasonable time for this case.

2.3.1 Trivial ILP

Based on expressions (1)–(4), we can define a particular ILP problem by

$$\boldsymbol{A} = \begin{bmatrix} -1/3 & -1 \\ -3 & -1 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 2 \\ 6 \\ -2 \end{bmatrix}, \quad \boldsymbol{c} = \begin{bmatrix} 1 \\ 3 \end{bmatrix},$$
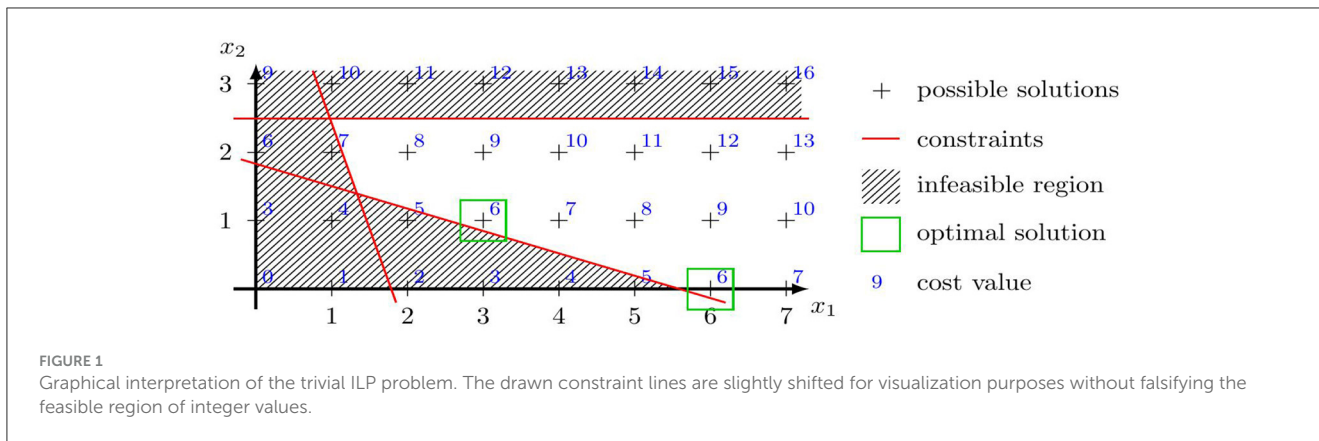
$$\boldsymbol{x} \in \mathbb{N}^2, \quad \boldsymbol{s} \in \mathbb{N}^3.$$

A graphical interpretation of this ILP is depicted in Figure 1. We can easily obtain the optimal solution vectors,

$$\boldsymbol{x}_0 = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \text{ or } \begin{bmatrix} 6 \\ 0 \end{bmatrix},$$

and the optimal cost value $\boldsymbol{c}^\top \boldsymbol{x}_0 = 6$ from this graph. This problem is an explicit example of an ILP that can have more than one optimal solution, which in turn can cause misleading results in benchmarking experiments. In general, ILPs can have zero, one or more solutions. We restrict the values in $\boldsymbol{x}$ to $x_i \in \{0,1,2,3\}, \forall i \in \{1,2\}$, such that the ILP is uniquely solvable.

For binary representation of integer values, we will use 2 bits for each variable in $\boldsymbol{x}$ and 3 bits for each element in $\boldsymbol{s}$. Then $\boldsymbol{q}$ is the binary search vector to be optimized. According to Chang

FIGURE 1
Graphical interpretation of the trivial ILP problem. The drawn constraint lines are slightly shifted for visualization purposes without falsifying the feasible region of integer values.

et al. (2020), this mapping with integer mapping matrix Z can be described by

$$
\begin{bmatrix} x \\ -- \\ s \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ -- \\ s_1 \\ s_2 \\ s_3 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 0 & 0 & | & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & | & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & | & 4 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & | & 0 & 0 & 0 & 4 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & | & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 2 & 1 \end{bmatrix} \underbrace{\begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_{13} \end{bmatrix}}_{q} .
$$

In Supplementary Table S1 we enumerate other trivial ILPs that we have investigated. These ILPs encompass a range of optimal solutions, parameters, and dimensions.

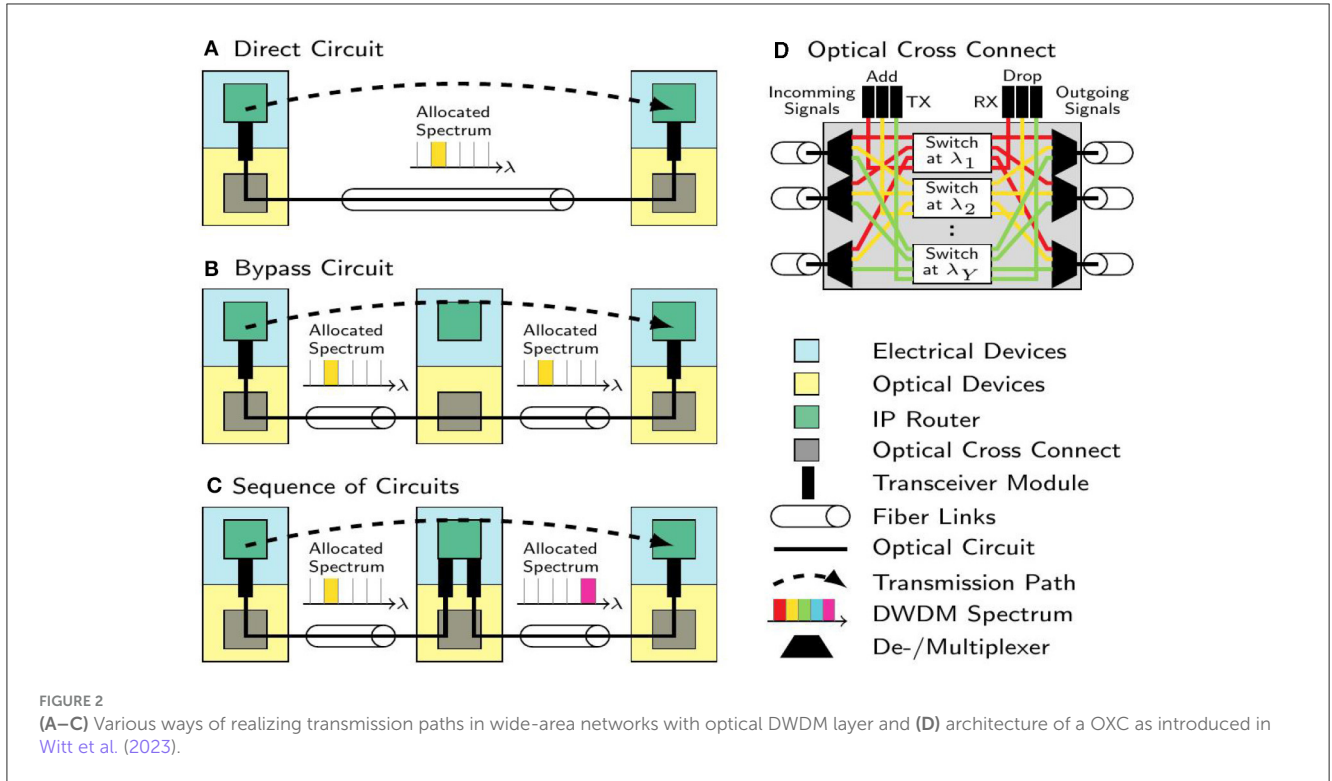### 2.3.2 Network resource allocation problem

Optical wide-area networks consist of nodes that are linked by optical fiber systems in form of a meshed topology. Nodes $v \in V$ are two layered. They are equipped with electrical IP routers in the upper layer and optical cross connects (OXCs) in the lower layer. Traffic from connected networks that traverses the wide area network is "handed over" at the IP layer. Signal transitions between layers inside the WAN are performed with optical bidirectional transceivers, that are configured for unidirectional use as required. Optical transceivers generate optical signals with a bandwidth of 50 GHz at various center frequencies. A finite number of signals can be combined in a dense wavelength division multiplexing (DWDM) scheme on a particular optical fiber link. This schemes are specified according to ITU-T (2020). Thus, usable frequency bands in the optical region, typically referred as wavelengths, are uniquely defined. Optical cross connects enable wavelength-selective forwarding and redirection of optical signals between connected fibers. Fiber links are realized by a sequence of fiber spans and fiber amplifiers and provide a hardware-wise connection between nodes according to the networks topology. The maximal reach of optical signals depends on the signal configuration (specified by modulation schemes, and used forward error correction, etc.) and the transceiver type itself. As an example, a tunable

coherent transceiver[1] achieves an optical reach of 1,000 km at a rate of 100 GBit/s. Typically, optical transmission paths are organized as a sequence of transmission sections c with at least one section to enable a end-to-end data transfer. Transmission sections are abstract links in the lower layer that provide optical transparent transmission on multiple wavelength. Their spanning distance is limited by the optical reach of the driving transceivers. Figure 2 illustrates how transmission paths in wide-area networks can be realized.

Energy-aware traffic engineering can be seen as a major task for economic network operation. Therefore, network resources like transceivers and wavelengths on fiber links have to be allocated to assign the required capacity to a transmission section c. Assuming that the network is operated as single rate system, i.e., all transceivers have the same signal rate, e.g., $\xi = 100$ GBit/s, capacities at a transmission section c can be scaled if multiple transceivers, enumerated by w_c, are activated. Thus, the transmission section's capacity is $w_c \xi$. A unidirectional traffic demand d represents a connectivity request between two network nodes. We assume, that a demand exist for all disjunct node pairs (u, v) with $u \neq v$ and $u, v \in V$. The network has to provide appropriate transmission paths, i.e., routes through the network topology along a sequence of transmission sections, to enable the transport of the demand's traffic with volume h_d. We prepare a network-specific collection T of possible transmission path realization prior the optimization, whereas possible transmission path realizations per demand d are defined as $t_d \in T_d \subset T$, see Witt et al. (2023), Section II-C. The economic resource allocation within WDM networks is a discrete and combinatorial optimization problem.

Witt et al. (2023) devised an integer linear program (ILP) based on Enderle et al. (2020) to address the energy-aware resource allocation problem within wide-area networks and prepared the ILP according to the ILP-to-QUBO mapping formalism as presented in Chang et al. (2020) and delineated in Subsection 2.2. They further studied the solvability of this ILP, prepared in QUBO form, on the D-Wave Advantage[TM], a state-of-the-art quantum

---

1 100/200 G Tunable Coherent CFP2-DCO Transceiver: https://www.fs.com/de-en/products/120128.html?attribute=5320&id= 297112.

FIGURE 2
(A–C) Various ways of realizing transmission paths in wide-area networks with optical DWDM layer and **(D)** architecture of a OXC as introduced in Witt et al. (2023).

annealer with over 5,000 qubits. Since the current work focuses on improvement methods within the algorithmic part and not on the application itself, we refer to Witt et al. (2023) for a more in-depth explanation and interpretation of the ILP. In the following, we recapitulate the ILP briefly. Parameters and variables are given in Table 1. Traffic volumes h_d per demand d, that are varying over time, are held constant during the optimization and will be updated frequently in a real scenario. The equality constraint (Equation 5) enforces that a demand is routed on exactly one transmission path. Constraint (Equation 6) combines traffic flows per transmission section as selected in Equation (5) and reserves the required capacity in terms of optical channels w_c. Constraint (Equation 7) activates installed transceivers to drive the transmission sections. Minimizing the number of optical channels w_c as defined in objective (Equation 8), reduces the total amount of active transceivers as well. This enables a energy-aware network operation.

Constraints:

$$\sum_{t_d \in T_d} g_{t_d} = 1 \quad \forall d \in D \tag{5}$$

$$-w_c + \sum_{d \in D} \sum_{t_d \in T_d} \rho_{c,t_d} \cdot \frac{h_d}{\xi} \cdot g_{t_d} \leq 0 \quad \forall c \in C \tag{6}$$

$$\sum_{c \in C} w_c \cdot \varphi_{v,c} \leq \eta_v \quad \forall v \in V \tag{7}$$

Objective:

$$\sum_{c \in C} w_c \rightarrow \min_{.} \tag{8}$$

The network under test is a fully-connected 3 node network, e.g., the topology has a triangular shape with two short edges of 300 km length and a long edge spanning a 424 km distance. Each network edge is realized by two fiber links to realize bidirectional transmission. Traffic demand values h_d are taken from a normal distribution with mean 75 Gbit/s and standard deviation 20 Gbit/s. As they represent floating numbers, we discretize them with an accuracy of a = 1 (acc. to Witt et al., 2023, Section III-C), i.e., fractions are rounded to "x.0" or "x.5". We set the number of installed transceivers per node to $\eta_v = 15$ and the maximal number of parallel optical signals per transmission path to $\omega_{c,\max} = 3$. The parameter $\omega_{c,\max}$ influences the QUBO's matrix sizes as described in Witt et al. (2023), Section III-C. The parameters ρ_{c,t_d} and $\varphi_{v,c}$ represent the connectivity described by the topology. They are predefined together with the transmission path realization sets T_d. The boolean selector variable g_{t_d}, indicating the selection of a predefined transmission path realization t_d for demand d, and the number of parallel optical signals per transmission path ω_c are determined during the optimization.

2.4 Correlations between solution metrics and system run parameters

With the intent to minimize the objective function, D-Wave provides a distribution of solutions, all of which are not equally important nor of equal quality. The setup of the ILP scenario (penalty term, float variable solution, integer sizes) and the QUBO (sparsity-affecting transformations, embedding, chain strength) parameterize the problem. The annealing procedure (annealing schedule, spin transformations, thermalization/decorrelation

TABLE 1 List of parameters used in the ILP for network optimization.

Parameter		Interpretation
Constants	$\xi \in \mathbb{R}$	Data rate of a single transceiver
	$\eta_v \in \mathbb{N}$	Amount of transceivers installed at node v
	$\rho_{c,t_d} \in \mathbb{B}$	Indicates whether the transmission section c is part of the demand-specific transmission path realization t_d
	$\varphi_{v,c} \in \mathbb{B}$	Indicates whether transmission section c is connected to node v
	$h_d \in \mathbb{R}$	Traffic volume of demand d
Variables	$g_{t_d} \in \mathbb{B}$	Path selector is 1 if a transmission path for demand d is realized by circuit configuration $t_d \in T_d$
	$\omega_c \in \mathbb{N}$	Amount of active transceivers, driving a transmission section c

pauses) can also have significant influence on the obtained distribution of solutions. Studies like Willsch et al. (2022) show that a proper parameter selection in terms of annealing schedule and embedding variants can change the situation significantly. Furthermore, the effect of thermalization in the context of quantum annealing processes can have an impact, as was shown in Dickson et al. (2013). Ideally, the solutions to the problem should not be affected by the choices for these meta parameters. Still, we selected the range of parameters to be tested using our experience garnered from our previous study, see Witt et al. (2023).

However, as we show in later sections, different combinations of parameters significantly affect the likelihood of obtaining feasible solutions. Choices for such meta parameters can be highly correlated. For example, longer (slower) annealing profiles can provide higher probabilities for finding a feasible solution, yet at the cost of generating fewer total number of solutions.

Our first studies, Witt et al. (2023), found that probabilities for finding a minimal feasible solution for the three-node network problem were at the order of 10^{-4}% and below. This presented a non-trivial task to evaluate the quality of the distribution of solutions when only having a sample sizes of $< 10^6$. To address this issue, we formulate statistical measures based on the distribution of samples to quantify the quality of our D-Wave setup. Since the optimal solution is, by definition, a feasible solution, we are interested in the rate in which feasible solutions are produced. We thus consider the feasibility ratio,

$$r_{\text{feasible}} = \frac{N_{\text{feasible}}}{N_{\text{samples}}} , \tag{9}$$

that rates the success of finding N_{feasible} feasible solutions within a solution set with N_{samples} samples.

Another metric of choice for solutions in the binary search space that we use in our research here is the Hamming distance

$$\text{dist}\{\boldsymbol{x}, \boldsymbol{y}\} = \sum_i XOR(x_i, y_i) .$$

This metric gives the number of flipped bits between an ideal solution $\boldsymbol{x}$ obtained by a classical ILP solver like CPLEX or GLPK and a non-ideal solution $\boldsymbol{y}$ obtained by the quantum annealer. For binary solution vectors the Hamming distance is equivalent to the L_2-norm of the difference between $\boldsymbol{x}$ and $\boldsymbol{y}$. This metric provides a sense of "distance" between two solution vectors, essentially telling us how many "bit-flips" are required to bring one solution into

another. Ultimately it allows us to perform a direct comparison between particular D-Wave solution vectors and a known desired solution vector.

Finally, we can train a neural network (NN) on these correlations, with the goal that once trained, we can use the NN to make further guesses on optimal solutions vectors. We describe our NN in the following section.

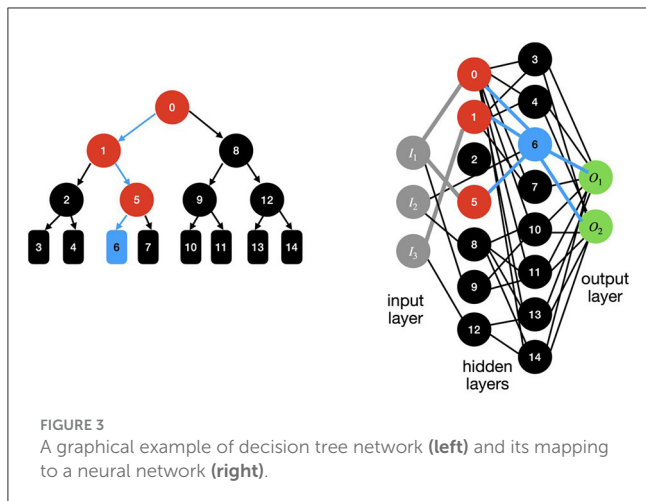## 2.5 Machine-learning approach

We employ a decision tree (DT) neural network in our investigations. This NN is a type of supervised machine learning (ML) algorithm that is used typically for regression and classification analysis. It is a model that represents a series of decisions and their possible consequences in the form of a tree-like structure (Breiman et al., 1984). Each node in the tree represents a decision, and each branch represents a possible outcome or path that can be taken based on that decision. In Figure 3 we provide a graphical example of a decision tree and its mapping to a neural network.

A major advantage of decision tree NNs is their ease of use, understandability, and interpretability. This makes their implementation simple and their application efficient. Another advantage comes from their inherent robustness to data outliers. They can even handle missing values in the data. The data itself can be both categorical and numerical in nature.

However, a potential drawback of DTs is that they can easily overfit the data. This ultimately means that, though they may be sufficiently *expressive* to explain the trained data, they fail when extrapolating to new, or unseen data. Thus, the NN is limited in its *generalizability*. This issue can to a certain degree be mitigated by pruning the tree or using other techniques to reduce the complexity of the model. In our studies we did not employ such mitigation techniques, and leave such potential studies for later investigations. We used the Scikit-learn python module (Pedregosa et al., 2011) and its functionalities to implement our DT networks.

2.6 Construction of Sherrington-Kirkpatrick graph

The Sherrington-Kirkpatrick (SK) graph encloses the coupling strength and external fields of a Ising Hamiltonian. As mentioned

FIGURE 3
A graphical example of decision tree network (**left**) and its mapping
to a neural network (**right**).

in Thai et al. (2022), finding the weighted minimal cut in this graph
is equivalent to finding the ground state in the Ising Hamiltonian.
Further, the Hamiltonian's energy landscape can be explored by
exploration of the SK graph's cut space.

The corresponding SK graph of the Ising Hamiltonian $\mathcal{H}(x) =$
$h^\top x + x^\top J x$ with n variables x_i can be denoted as $G_H^{SK} = (V, E, w)$
with node set V, undirected edges $(i, j) \in E$ and their weights W.
The first n nodes of V correspond to the variables x_i. A further
node is added to V to capture the external fields h. Set E contains
only edges with non-zero weights according to $w_{ij} = J_{ij} + J_{ji}$ for
$1 \leq i, j \leq n$ and weights $w_{i,n+1} = h_i$. Then, the weighted adjacency
matrix J' of graph G_H^{SK} with $J'_{ij} = J_{ij} + J_{ji}, J'_{ji} = 0$ and $J'_{i,n+1} = h_i$ can
be used together with $y \in \mathbb{S}^{n+1}$ to define the SK Hamiltonian as

$$\mathcal{H}^{SK}(y) = y^\top J' y.$$

To apply a weighted minimal-cut approach on the SK graph
for graph reduction, a cut is defined by a subset $S \subseteq V$, such that
$\langle S, V \backslash S \rangle$ contains a set of edges that needs to be cut for separation of
S and $V \backslash X$. With $c(S) = \sum_{(u,v) \in \langle S, V \backslash S \rangle} w_{uv}$, the capacity of the cut,
a minimal cut is defined as $S^* = mc(G^{SK}) = \operatorname*{argmin}_{S \subseteq V} c(S)$ with the
minimal capacity of $MC(G^{SK}) = \min_{S \subseteq V} c(S)$.

3 Results

3.1 Trivial ILP problem

We now provide our findings for our simple ILP problem that
we described in Subsubsection 2.3.1. Similar results for the other
trivial ILPs we considered are found in the Supplementary material.
Note that this problem is sufficiently small that we can
determine all possible feasible solutions via brute force, which
includes the optimal solution. In this case this corresponds
to a total number of $N_{\text{feasible}} = 1,536$ feasible solutions.
The whole solution space contains $N_{\text{samples}} = 2^{13} =$
$8,192$ possible vectors as our binary search vector q has a
dimension of 13, see Subsubsection 2.3.1. Thus, the feasibility
ratio (Equation 9) for brute force sampling is $r_{\text{feasible}} =$
18.75%.

3.1.1 Observations and correlations

In Figure 4 we show results for brute force sampling (left
side) and a run on the D-Wave Advantage™ (right side) using
a penalty of $p = 2$ and default run parameters. In the D-
Wave case, just 200 samples are taken, which is a relative
small portion (∼2.4%) compared to the complete solution space.
We have to remark, that the optimal solution can be found
even if the sample set is small. Figure 4 shows the distribution
of solutions over energy (upper row) and Hamming distance
(middle row) obtained by the mentioned sampling methods
and classified by their feasibility demarcated by feasible (green),
infeasible (red), and all (blue) solutions. We can observe, that
solutions obtained with D-Wave show low energies and only
Hamming distances of up to 8. This indicates, that the aimed
optimization takes place and only solutions with mostly good
qualities are found by D-Wave quantum annealer. But, we still
have to sort solutions by feasibility after sampling as minimizing
the energy can not entirely sort out infeasible solutions. The lower
row of Figure 4 shows the correlations between the solution's
energy and their Hamming distance in relation to the best
feasible solution. Solutions with small Hamming distances tend
to have smaller energy values as observable and indicated by the
best fitting curves. We identified, that higher-energy solutions
are correlated with increasing Hamming distances to optimal
solution as the slope of the fitting curves are non-zero. The
energy range for solutions at same Hamming distances is spread
widely if the whole search space is considered. Feasible solutions
could be found only at the lower energy range. Within the D-
Wave sample set, solutions with small Hamming distances are
over-proportionally feasible solutions which is indicted by the
regression curves.

In the brute force sampling case, we can describe the
distribution of solutions upon the Hamming distance
(Figure 4 left, middle row) by a cumulative distribution
function,

$$CDF(d) = \frac{1}{2^{N_q}} \sum_d \binom{N_q}{d} \quad \forall \quad d \in \{0, 1, \ldots, N_q\},$$

with d representing possible Hamming distances
for binary search vectors q of length N_q. This
relation can be used for benchmarking as it forms
a fundamental boundary that only depends on the
vector size.

In Figure 5 (upper left panel) we show dependence of
the feasibility ratio (9) as a function of anneal time and
penalty factor, as well as the average hamming distance
(upper right panel) as a function of the same parameters.
There is seemingly little correlation between p and the
anneal time as long as $p \gtrsim 10$. However, these results
suggest that increasing beyond $p \gtrsim 100$ is beneficial since
in this region solutions with lower Hamming distance are
more likely. It is remarkable, that all feasibility ratios that are
shown in Figure 5 are significantly larger than the theoretical
value of 18.75% for the case that all possible solutions
are considered.

We also encountered a number of individual feasible solution
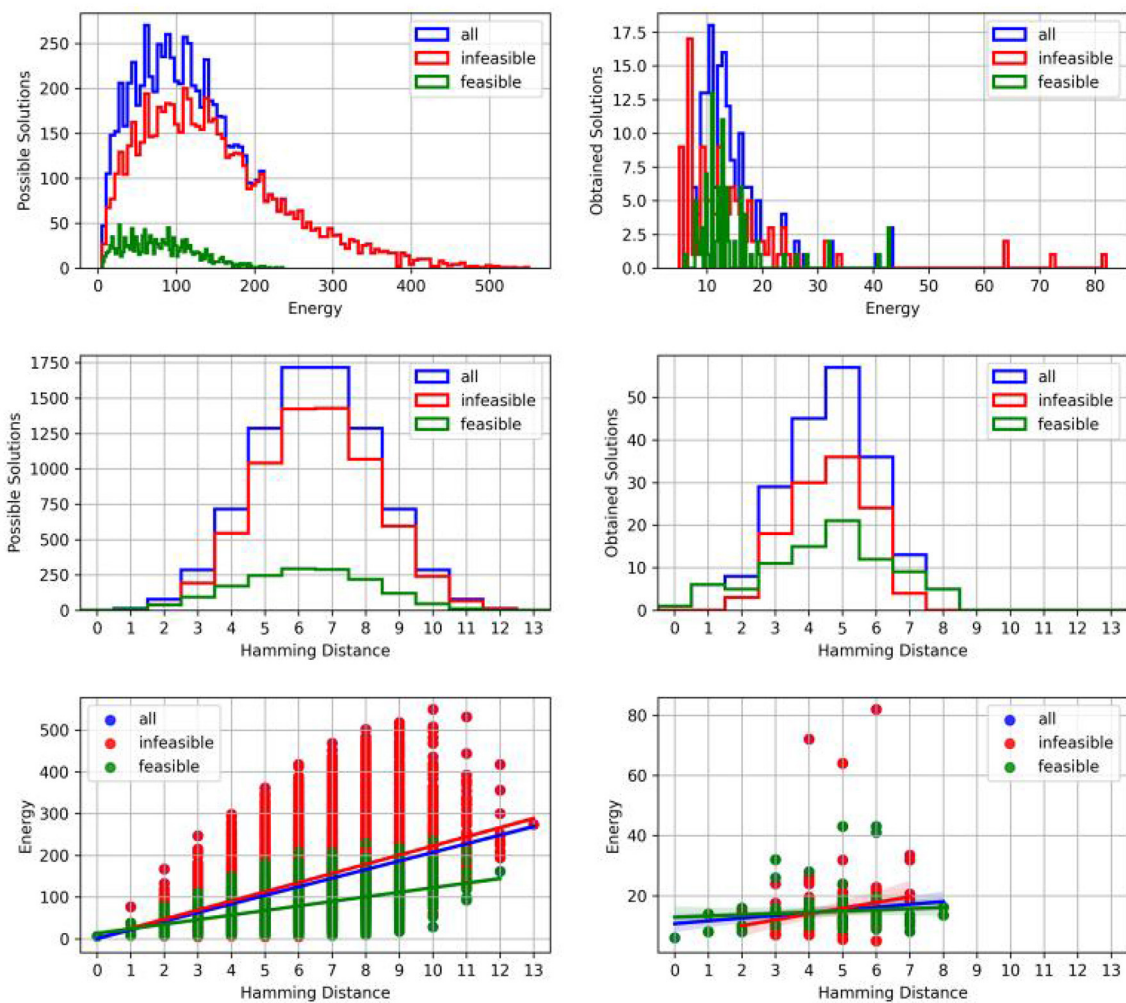samples obtained in a single run whereby the solution vectors

FIGURE 4
(Left) Brute force sampling to investigate the entire solution space. **(Right)** D-Wave sampling with penalty $p = 2$ and a set of 200 samples. **(Upper row)** Histogram of solutions sorted by energy values. As solutions gathered by D-Wave's quantum annealer have only energy values in the lower area compared to the brute force case, x-axis are scaled differently. **(Middle row)** Histogram of solutions over Hamming distance with respect to the best feasible solution vector. **(Lower row)** Scatter plot of solutions with reference to their energy values and the Hamming distance with respect to the best feasible solution vector. (Blue) solution set under investigation. (Red) infeasible solutions. (Green) feasible solutions.

fulfill the ILP's constraints and differ from each other in at least one of its components, but are not necessarily optimal solutions. It can happen that some of these individual feasible solutions can share the same cost value. The parameter dependence of the number of individual feasible solutions is presented in lower left panel of Figure 5. We find that short annealing times generate more individual solutions, however at the expense of reducing the low-energy solutions. So the D-Wave quantum annealer finds more solutions with higher energies if shorter annealing profiles are applied. To no surprise, these correlations suggest that optimizing to longer anneal times will provide lower energy solutions. Similar findings are found for the other trivial ILPs listed in Supplementary material. The relevant figures in this case are Supplementary Figures S1–S3.

We point out that we find no correlations between the parameters `chain_strength` and `annealing_time`, suggesting that further optimization of the `chain_strength` parameter is not possible.

3.1.2 Improvements obtained by machine learning approach

Within a sample set, generated by D-Wave Advantage™, we have 110 independent solutions for our trivial ILP problem when using $p = 2$ and an annealing time of 20 μs. This represents ∼10% of the possible feasible solutions. To improve upon this, we train a NN on the correlations described above and then use the NN to generate more solutions.

In particular, we train our NN using the solution vector vs. the energy and feasibility correlations obtained from D-Wave data. With input of energy and feasibility, the decision tree regression predicts a new solution which has the corresponding input energy and feasibility. We note that our NN does not always provide new solutions whose output energy coincides with the same input energy. This is readily seen in the left plot in Figure 6, where the output energy E_{out} is plotted as a function of input energy E_{in}. A one-to-one correspondence would provide a straight line with slope of unity, which is clearly not seen. However, the correlation
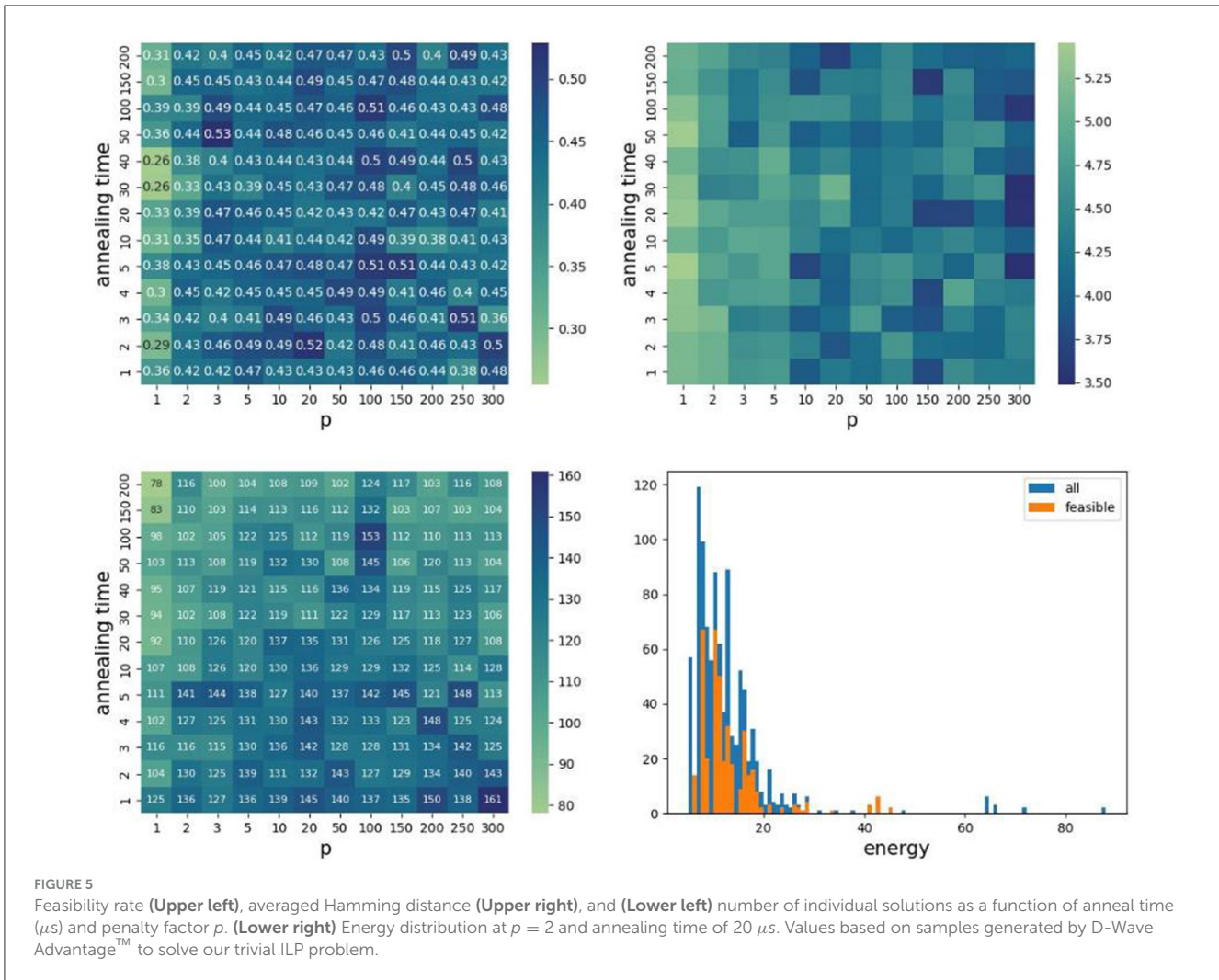
FIGURE 5
Feasibility rate **(Upper left)**, averaged Hamming distance **(Upper right)**, and **(Lower left)** number of individual solutions as a function of anneal time (μs) and penalty factor p. **(Lower right)** Energy distribution at $p = 2$ and annealing time of 20 μs. Values based on samples generated by D-Wave Advantage™ to solve our trivial ILP problem.

between input and output energy as captured by our NN is still positive. We find that the slope of this correlation depends on the p value, whereby larger p values provide a slope closer to unity. Qualitatively similar behavior is found for the ILPs listed in Supplementary material, as can be seen in Supplementary Figure S5 of this document.

We expect the decision tree to recognize the feasibility condition, but predicted solutions of the NN are not always feasible. As mentioned above, the NN predicts solution vectors whose energy ranges have some correlation with the input energies. This feature provides, in principle, an advantage over brute-force sampling since we can target solutions within a specific energy range using our NN, whereas such control via brute force sampling is not possible. However, there isn't a complete one-to-one correspondence between input and output energy since ∼20% of the predicted solution vectors have components that are not binary but contain fractional numbers. In these cases we round the fraction to zero if the fractional number is smaller than 0.2, and to one if larger than 0.8. Between 0.2 and 0.8, we enumerate all possible combinations of 0 and 1, generating in these case new proposed solution vectors. We then perform another feasibility test on these NN solutions to filter

out infeasible solutions. The energy distribution of feasible vs. infeasible solutions after this treatment is shown in Figure 6 right.

3.2 Three-node network

We now turn our attention to the 3-node problem, which represents the smallest, non-trivial system of wide-area networks. Here we use CPLEX to obtain the optimal solution vector, from which we make comparisons with D-Wave solution vectors. The distribution of D-Wave solutions as a function of Hamming distance to the optimal solution is given in Figure 7. Note that in this case the optimal solution is not captured by D-Wave. In fact, D-Wave cannot find any feasible solutions within a set of 600,000+ samples. As remark, the entire search space for this case is 2^{63}.

When we investigate inter-parameter correlations, we find little to no correlations between the Hamming distance, `chain_strength`, `anneal_time`, and penalty factor p. This is demonstrated by the nearly flat dependence of the data in Figure 8. This lack of correlation prevents us from obtaining optimized run parameters for this system, and unfortunately
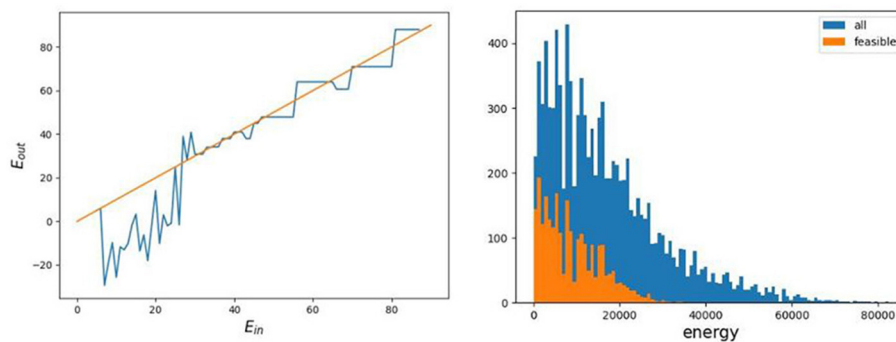
FIGURE 6
Decision tree method to find more feasible solutions based on D-Wave data at $p = 2$ and annealing time = 20 μs.
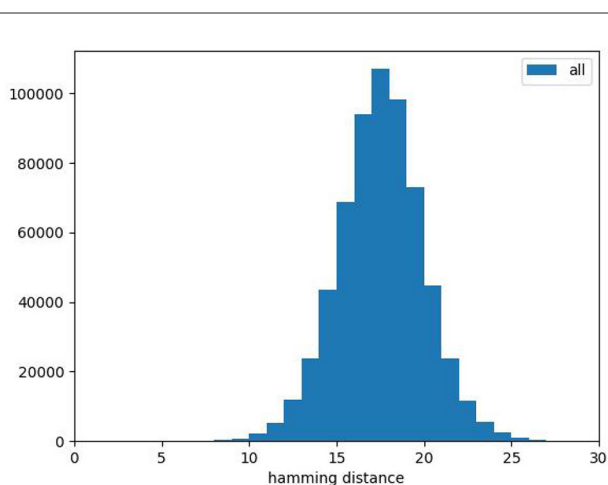

FIGURE 7
Histogram of Hamming distance of the all solutions for 3-node problem. There are no feasible solutions.

suggests that larger node problems will become just as difficult, if not more difficult, to optimize.

These findings already hint at the difficulties we encounter when applying an NN to this system, as we describe in the following section. But we nonetheless train a DT network on the energy and Hamming distance to optimal solution, exactly as described in Subsubsection 3.1.2.

4 Discussion

4.1 Interpretation of findings

The total number of feasible solutions of the trivial ILP problem is 1,536. As previously mentioned, D-Wave finds a little <10% of these solutions, but with our NN we can fully ascertain the full solution space distribution (compare the lower right panel of Figure 5 with that of Figure 6 and see also Supplementary Figures S2, S5). More concretely, we provide the exact number of addition feasible solutions found with our NN as a function of input parameters `annealing time` and penalty factor p in Figure 9 (see Supplementary Figure S4 for our other

trivial ILPs). This means that, for our simple ILP problem, the decision tree after round off treatment provided 1,426 new independent feasible solutions. The distribution of new solutions as a function of Hamming distance provided by our ML technique is given in Figure 10. So combining our NN results with D-Wave's, all possible 1,536 solutions were found. Thus, our hybrid classical (ML)/quantum (D-Wave) method allowed us to fully map out the full solution space. We note that our NN is not generalizable to all trivial ILPs, but is unique for each ILP. This is because the solution vector space generated by D-wave is specific for each ILP, and so each NN is trained with this specific solution vector layout. Within our formalism a "master" NN for all trivial ILPs is not possible.

We now discuss our 3-node problem. Note that in this case, D-Wave could not find the optimal solution provided by CPLEX, despite our system parameter investigations mentioned in the previous section. It is not viable to assume that feasible solutions can be found by luck or random guessing. The probability to find the optimal (minimal and feasible) solution is $1/2^{63} \sim 10^{-19}$ in our case. The fact that we could not find feasible solutions within a set of 600.000 samples indicates that feasible solutions are very rare. This was already observed in our previous study (Witt et al., 2023). There, we were not able to find any feasible solutions for some of the test sets and in other cases around 0.2–11 per million samples. There are some possible hints for why this is the case here. First of all, the entire solution set contains only a small portion of feasible solutions that fulfill the ILP. Further, the annealer minimizes the energy of the QUBO Hamiltonian. As it is possible that the lowest energy state can be obtained with various solution vectors someone could find also a energy-wise optimized vector that does not fulfill the ILP. Furthermore, hardware imperfections like noise or limited detection resolution can cause this undesirable behavior.

At this point, critical voices could rate the annealer as an expensive random sampler. But this is not the case as we were able to show that trivial ILP problems are definitely solvable with D-Wave. In these cases, we explicitly used less samples than the solution space's size to avoid an oversampling—somebody could also solve small problems by oversampling even if the sampler is neither a random guess sampler where each solution is equal probable or a minimizing sampler like the quantum annealer. Thus, D-Wave performs better than a random sampler. Clearly the solvability is not the same for the network problem case, and this
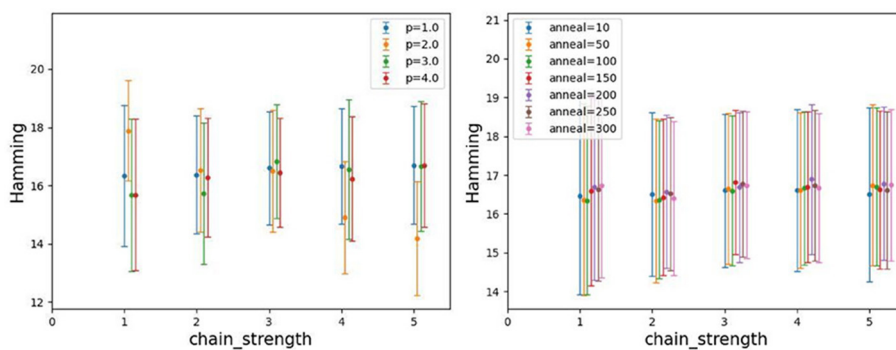
FIGURE 8
Hamming distance in dependence of penalty p and annealing parameters chain strength and annealing time. The Hamming distance is obtained by comparing with the best solution obtained by CPLEX.
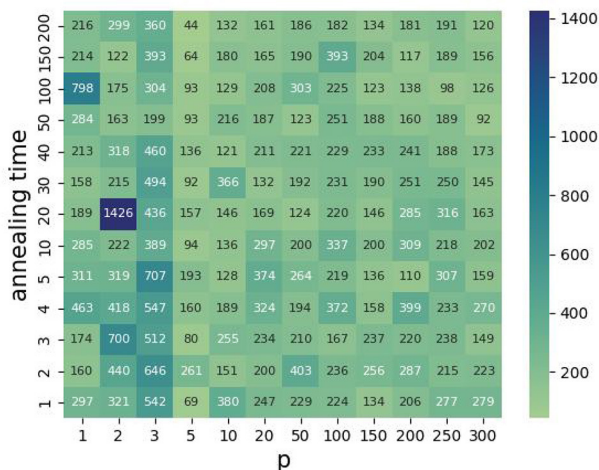

FIGURE 9
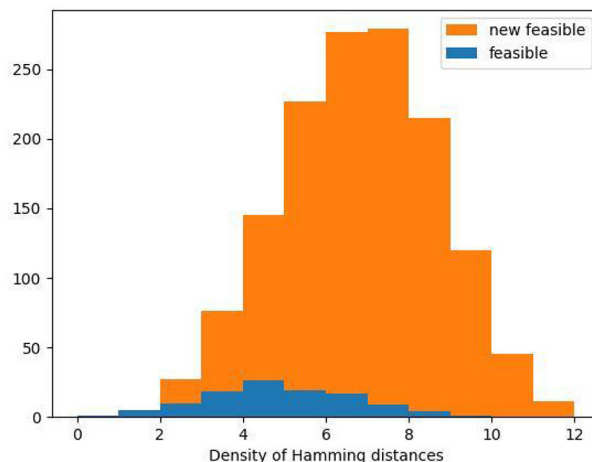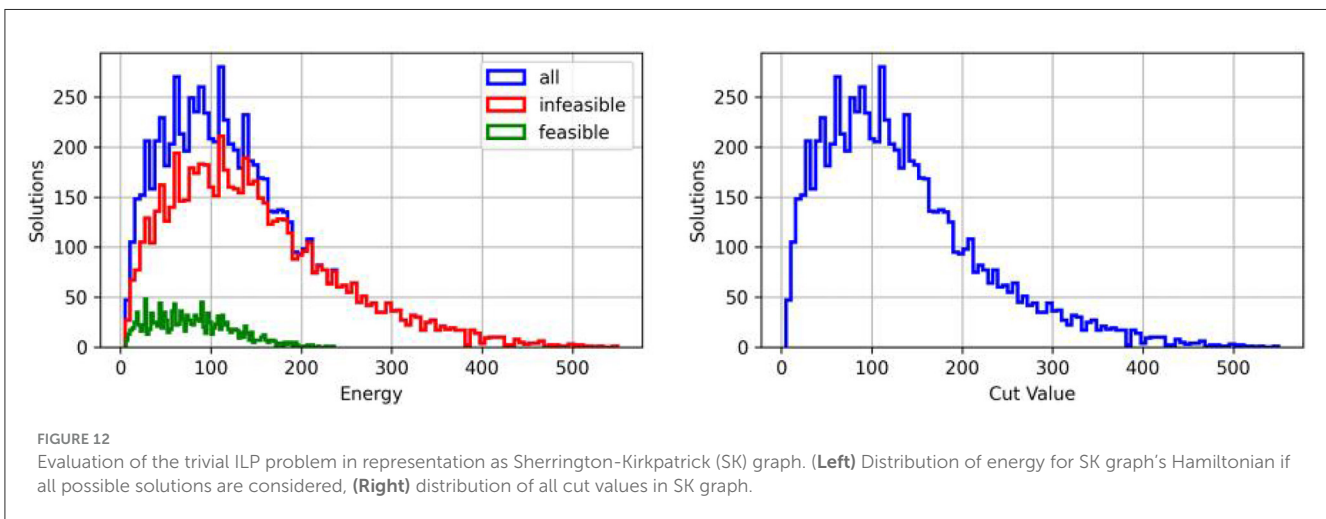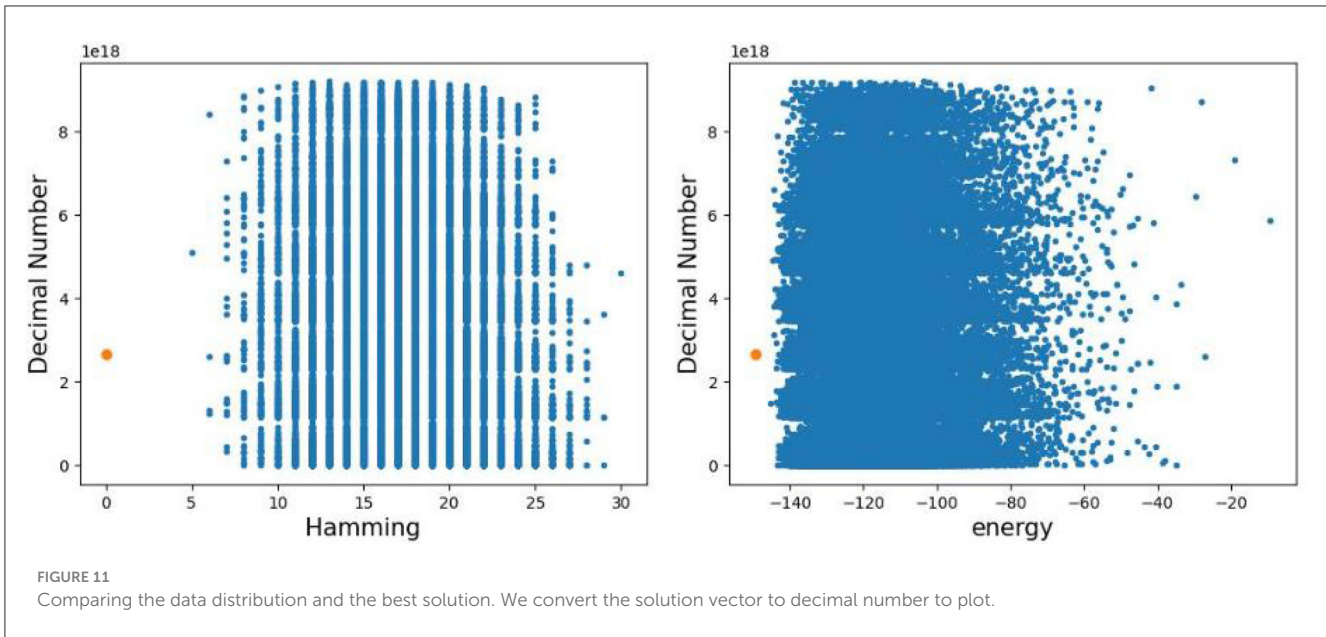The number of new independent feasible solutions found by decision tree for the ILP problem.

FIGURE 10
Histogram of hamming distance of the feasible solution obtained by D-Wave (blue) and the new feasible solutions obtained by decision tree method for the ILP problem.

may be due to (a) a higher connected QUBO and longer chains of qubits that represent a logical qubit, which cause chain breaks in the quantum annealing hardware to be more likely, (b) numbers in the QUBO matrix have a higher differing range that may be not represented in the hardware well-enough, and of course (c) other issues that are beyond our knowledge.

As part of an approach for improvements, we trained our NN for the 3-node problem with the distributions that we generated from our correlation studies in a comparable way as it was done in the simple ILP problem. Once trained, we found, however, that the NN was unsuccessful in finding any new feasible solutions, let alone the optimal solution. We attribute this to the fact that our D-Wave data distribution of energies (which is used to train the NN) does not cover the energy region of the optimal solution. In fact, as shown in Figure 11, the distribution of D-Wave solutions is far from the optimal solution. Our NN could therefore not generalize sufficiently to lower energy solutions. Compounding the issue is the fact that the distribution of D-Wave solutions contained no feasible

solutions, and this in turn limited what the NN could "learn". Thus, our hybrid (ML)/quantum (D-Wave) method failed to produce any new solutions for our 3-node problem.

An obvious question to raise is whether another choice of NN is better suited for our 3-node problem. As we discussed in Subsection 2.5, one of the main advantages that motivated our choice of the decision tree NN is admittedly its ease of use, interpretability, and implementation. However, because of its potential lack of expressivity, one could argue that another choice of NN, e.g., convolutional or recurrent, might lead to better results. This indeed may be the case, and at the least warrants further research. We point out, however, that regardless of the NN architecture, our formalism requires that there exist correlations between hyper-parameters and the resulting D-Wave solutions vectors. It is these correlations that are "learned" by the NN. Since we found no such correlations in our 3-node problem, we suspect that any other type of NN will have similar difficulties as those encountered by our decision tree NN.

FIGURE 11
Comparing the data distribution and the best solution. We convert the solution vector to decimal number to plot.

FIGURE 12
Evaluation of the trivial ILP problem in representation as Sherrington-Kirkpatrick (SK) graph. (**Left**) Distribution of energy for SK graph's Hamiltonian if all possible solutions are considered, (**Right**) distribution of all cut values in SK graph.

4.2 Outlook on further improvements

Still there may be ways to improve the situation. Our studies to date have only varied the annealing profile. Instead, one may perform reverse annealing, where the annealing is run "backwards" from a starting classical solution, allowing for exploration of the energy landscape around the classical solution. We are actively investigating this procedure. Reverse annealing may be also applicable to set initial states as shown in Pelofske et al. (2023). Thus, expected solutions or solutions that are close to an expected solution can be set as start value for the annealing process. If the optimizer is applied frequently—a typical situation in network optimization—the last obtained solution can be used for the initialization of the next run as new optimal network configurations might be close to the last configuration.

Annealing parameter like annealing schedules and various embeddings for our problems can be studied more detailed like in the study of Willsch et al. (2022). The authors of Willsch et al. (2022)

discovered an increase in the success rate for proper settings in the annealing schedule. In our case we observed a more or less constant success rate, especially for the 3-node network problem. Apart from that it may be valuable to study our approach on a larger set of similar problems to get a more general perspective. Unfortunately, we had to restrict our study on a single problem instance as the amount of feasible solutions for our problem is very rare and large sampling sets are required for the analyzes. Besides, thermalization within the annealing process can be studied as well, see Dickson et al. (2013).

Furthermore, since the size of the problem that is embedded on the quantum annealer plays a crucial role for its solvability, methods for efficient embedding or problem reduction should be incorporated within future studies. We point out that the work of Thai et al. (2022) seems promising in reducing the demands on the number of physical qubits. Here the authors introduced a fast Hamiltonian reduction algorithm (FastHare) that defines non-separable groups of qubits, i.e., qubits that obtain the

same value in optimal solutions, and performed a reduction by merging non-separable groups into single qubits. This could be done within a worst case time complexity of $\mathcal{O}(\alpha n^2)$ with a user-defined parameter α. The authors of Thai et al. (2022) showed in a benchmark that their algorithm is capable of saving 62% of physical qubits on average within a processing time of 0.3 s, outperforming the roof duality–the reduction used within the D-Wave's software development kit SDK. We reviewed parts of their work. In particular, we mapped our trivial ILP problem to a so-called Sherrington-Kirkpatrick (SK) graph. We further evaluated all cut values within this graph. The results (Figure 12) show that the cut values in the SK graph correspond to the energy values of QUBO or Ising problem solution vectors. As the Hamiltonian reduction is based on graph compression on basis of minimal cuts, we expect that the proposed algorithm (Thai et al., 2022) can improve the situation, as a reduced Hamiltonian might be better solvable on the D-Wave quantum annealer.

Unfortunately, we were not able to fully implement and apply this sophisticated algorithm as we struggled at the following point. The algorithm applies a min-cut algorithm on the SK graph to detect non-separable qubit groups. Originally, we though that a standard min-cut algorithm could be applied here. Unfortunately, the for us available min-cut algorithms can be only applied in graphs with positive-weighted edges. But, due to the nature of QUBO, Ising or SK Hamiltonians, the edges in a SK graph may have negative-valued edge weights. This issue was not addressed in their work (Thai et al., 2022). However, it remains unsure, if the fast Hamiltonian Reduction (FastHare) algorithm can improve the solvability of our ILPs with D-Wave's annealer as the authors used randomly generated graph structures in their evaluation, i.e., the graphs are weakly connected and as such well-suited for graph compression.

Beside the ILP to QUBO mapping formalism that was described in Chang et al. (2020) and Witt et al. (2023), someone could model the problem in a differing way. One possibility is the introduction of constraint-specific penalty factors, that create new degrees of freedom usable for problem-specific optimization of the algorithm. It can be achieved by the use of a penalty vector $\boldsymbol{p}^{\top} = [p_1, p_2, \ldots, p_m]$ and a corresponding penalty matrix $\boldsymbol{P} = \boldsymbol{Ip}$ inside the formulations. The QUBO Hamiltonian and thus the objective to be optimized is then

$$\mathcal{H}(q) = \boldsymbol{q}^{\top} \boldsymbol{Q} \boldsymbol{q} + C \rightarrow \min,$$
$$\text{with} \quad \boldsymbol{Q} = \begin{bmatrix} \boldsymbol{Q}_{xx} & \boldsymbol{Q}_{xs} \\ \boldsymbol{Q}_{sx} & \boldsymbol{Q}_{ss} \end{bmatrix}, \quad C = \boldsymbol{b}^{\top} \boldsymbol{Pb} \text{ and}$$

$$\boldsymbol{Q}_{xx} = \boldsymbol{Z}_x^{\top} \boldsymbol{A}^{\top} \boldsymbol{PAZ}_x + \text{diag}\left\{ \left(2\boldsymbol{b}^{\top}\boldsymbol{PA} + \boldsymbol{c}^{\top}\right)\boldsymbol{Z}_x \right\},$$
$$\boldsymbol{Q}_{xs} = \boldsymbol{Q}_{sx}^{\top} = \boldsymbol{Z}_x^{\top} \boldsymbol{A}^{\top} \boldsymbol{PZ}_s,$$
$$\boldsymbol{Q}_{ss} = \boldsymbol{Z}_s^{\top} \boldsymbol{PZ}_s + 2\text{diag}\left\{ \boldsymbol{Z}_s^{\top} \boldsymbol{Pb} \right\}.$$

This extends the ILP to QUBO mapping formalism to a generalized form. Required details could be found in Witt et al. (2023), Section III-D.

4.3 Outcome

Our work can be summarized as follows. The approach aims to solve ILPs with a quantum annealing attempt. We tried to find optimal annealing parameters and discovered weak correlations between annealing parameters and success rates in the 3-node network case. Further, a decision tree ML approach was applied to increase the rate of feasible ILP solutions. We realized that further improvements are needed to overcome remaining hurdles and discussed some attempts therefore. Even as the results for the 3-node problem are not fully satisfying, we are able to show with less complicated ILP problems that the approach works in principle. Thus, we expect that the approach can be extended in a way that larger problem instances are also solvable.

Finally, fast ILP-solving methods can have a significant impact on systems that should be optimized in real time. As an example, a novel mode of real-time network operation in wide-area networks is studied in Witt (2024). Here, similar ILPs are used to define a frequently applied network optimization.

Data availability statement

The datasets presented in this article are not readily available because it is not intended to share the data publicly. Requests to access the datasets should be directed to AW, arthur.witt@ieee.org.

Author contributions

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomp.2024.1356983/full#supplementary-material

References

Adler, I., Papadimitriou, C., and Rubinstein, A. (2014). "On simplex pivoting rules and complexity theory," in *Integer Programming and Combinatorial Optimization*, eds. J. Lee, and J. Vygen (Cham: Springer International Publishing), 13–24.

Breiman, L., Friedman, J., Stone, C., and Olshen, R. (1984). *Classification and Regression Trees*. New York, NY: Taylor & Francis.

Chang, C. C., Chen, C.-C., Körber, C., Humble, T. S., and Ostrowski, J. (2020). *Integer Programming from Quantum Annealing and Open Quantum Systems*. Ithaca, NY.

Chiaraviglio, L., Mellia, M., and Neri, F. (2009). "Energy-aware backbone networks: a case study," in *2009 IEEE International Conference on Communications Workshops* (Dresden), 1–5.

Dickson, N. G., Johnson, M. W., Amin, M. H., Harris, R., Altomare, F., Berkley, A. J., et al. (2013). Thermally assisted quantum annealing of a 16-qubit problem. *Nat. Commun.* 4:2920. doi: 10.1038/ncomms2920

Enderle, T., Witt, A., and Christou, F. (2020). "Delay-differentiated routing in meshed backbone networks," in *Photonic Networks; 21th ITG-Symposium* (Berlin: VDE Verlag), 20–27.

Feller, F. (2012). "An optimization-heuristic approach to dynamic optical bypassing," in *13th ITG Fachtagung on Photonic Networks* (Leipzig: VDE Verlag).

ITU-T (2020). *Recommendation ITU-T G.694.1, SERIES G: Transmission Systems and Media, Digital Systems and Networks Transmission Media and Optical Systems Characteristics—Characteristics of Optical Systems; Spectral Grids for WDM Applications: DWDM Frequency Grid*. Geneva: International Telecommunication Union - Telecommunication Standardization Sector.

Karp, R. M. (1972). *Reducibility among Combinatorial Problems*. Boston, MA: Springer US, 85–103.

Lenstra, H. W. (1983). Integer programming with a fixed number of variables. *Math. Operat. Res.* 8, 538–548. doi: 10.1287/moor.8.4.538

Nash, J. (2000). The (Dantzig) simplex method for linear programming. *Comp. Sci. Eng.* 2, 29–31. doi: 10.1109/5992.814654

Nguyen, D., and Pak, I. (2017). "The computational complexity of integer programming with alternations," in *32nd Computational Complexity Conference (CCC 2017), number 6, 6:1-6:18. Leibniz International Proceedings in Informatics*, ed. R. O'Donnell (Dagstuhl).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830. doi: 10.48550/arXiv.1201.0490

Pelofske, E., Hahn, G., and Djidjev, H. (2023). Initial state encoding via reverse quantum annealing and h-gain features. *IEEE Transact. Quant. Eng.* 4, 1, 21. doi: 10.1109/TQE.2023.33 19586

Thai, P., Thai, M. T., Vu, T., and Dinh, T. N. (2022). "FastHare: fast hamiltonian reduction for large-scale quantum annealing," in *2022 IEEE International Conference on Quantum Computing and Engineering (QCE)* (Los Alamitos, CA: IEEE Computer Society), 114–124.

Tornatore, M., Maier, G., and Pattavina, A. (2002). "WDM network optimization by ILP based on source formulation," in *Proceedings. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies, Volume 3* (New York, NY), 1813–1821.

Willsch, D., Willsch, M., Gonzalez Calaza, C. D., Jin, F., De Raedt, H., Svensson, M., et al. (2022). Benchmarking advantage and D-wave 2000Q quantum annealers with exact cover problems. *Quant. Inf. Process.* 21, 1–22. doi: 10.1007/s11128-022-0 3476-y

Witt, A. (2024). "Queue-aware network control algorithm with a high quantum computing readiness-evaluated in discrete-time flow simulator for fat-pipe networks," in *IEEE 25th International Conference on High Performance Switching and Routing (HPSR) 2024* (Pisa).

Witt, A., Körber, C., Kirstädter, A., and Luu, T. (2023). "Tactile network resource allocation enabled by quantum annealing based on ILP modeling," in *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)* (Bellevue, WA), 670–680.