Quantum Science and Technology

ACCEPTED MANUSCRIPT

Unified Sparse Optimization via Quantum Architectures and Hybrid Techniques

To cite this article before publication: Wen-Xin Li *et al* 2025 *Quantum Sci. Technol.* in press https://doi.org/10.1088/2058-9565/adbcd1

Unified Sparse Optimization via Quantum Architectures and Hybrid Techniques

Wenxin Li[1], Chuan Wang [2*], Hai Wei[1], Shuai Hou[3], Chongyu Cao[1], Chengkang Pan[3], Yin Ma[1], Kai Wen[1*]

[1] Beijing QBoson Quantum Technology Co., Ltd., Beijing 100015, China
[2] School of Artificial Intelligence, Beijing Normal University, Beijing, 100875, China
[3] Future Mobile Technology Lab, China Mobile Communications Research Institute, Beijing, 100032, China
[*] Corresponding Authors

E-mail: wangchuan@bnu.edu.cn, wenk@boseq.com

Abstract. In an era of rapid technological advancements and unprecedented data inundation, sparsity has emerged as a key property with profound implications in various fields. One important application of sparsity is sparse signal recovery, which involves reconstructing signals from limited observations and is of great importance in medical imaging, communication systems, and data compression. However, traditional sparse signal recovery methods often require computationally intensive algorithms, especially for large-scale problems. In this paper, we investigate the application of the coherent Ising machine (CIM), a hybrid quantum computing paradigm, as a novel approach to efficiently solve several sparsity-related optimization problems, presenting significant contributions in terms of model development and experimental validation. Our proposed models surpass existing approaches by reducing the computational resource requirements and enhancing problem-solving capabilities. Additionally, we also provide theoretical analysis on the performance guarantees of the proposed models, offering insights into their reliability and robustness. To further enhance the scalability and efficiency of the proposed model, we incorporate Benders Decomposition to decompose large-scale problems into smaller subproblems that can be solved more effectively. In addition, the efficiency and accuracy of the CIM-based sparse optimization approach are demonstrated through the experiments on the CIM platform, which highlights its potential to solve complex combinatorial optimization problems in practical scenarios.

Keywords: Sparse Optimization, Coherent Ising Machine, Combinatorial Optimization.

1. Introduction

In today's rapidly evolving technological landscape, characterized by an unprecedented deluge of data, sparsity has emerged as a precious property, particularly in high-impact applications. Sparsity, the property that a significant fraction of elements are zero or negligible in a given representation, is a cornerstone in several fields, ranging from signal processing [1], image processing [2] to machine learning [3]. The

concept of sparsity has garnered immense attention for its inherent capacity to represent complex data structures succinctly. This essential attribute has played a key role in simplifying computations, reducing memory requirements, and improving the efficiency of algorithms in a wide variety of fields.

Sparse optimization is a branch of optimization that focuses on finding solutions with a sparse representation. These problems typically involve ℓ_0 norm regularization or sparsity constraints, aiming to find solutions with a small number of non-zero elements. Developing techniques that are computationally tractable to solve these optimization problems are critical. As a special case, sparse signal recovery, the process of reconstructing a signal from a limited set of measurements or observations, has attracted considerable interest due to its applicability in fields such as medical imaging [4] and communication systems [5]. In the standard compressed sensing (CS) theory, the fundamental objective is to find the sparsest approximation, *i.e.*, minimize the ℓ_0 norm, within the constraints of an under-determined problem, as shown in Figure 1. However, this is an NP-hard problem in general [6]. One prominent approach in CS involves minimizing the ℓ_1-norm of the signal, which serves as a convex surrogate for the non-convex ℓ_0-norm. While the direct application of ℓ_0 regularization often necessitates intricate and complex techniques, ℓ_1 regularization offers a more tractable solution. It can be efficiently addressed utilizing methods such as first-order convex optimization and Orthogonal Matching Pursuit (OMP) [7], thereby simplifying the optimization process significantly. However, it is proven that ℓ_0-norm requires fewer measurements for accurate signal reconstruction compared to other norms like the ℓ_1-norm or ℓ_2-norm [8]. This makes it particularly advantageous in scenarios with stringent resource constraints, such as limited storage or bandwidth. In fact, the ℓ_0-norm achieves the theoretically optimal reconstruction performance. Specifically, it guarantees successful signal reconstruction when the compression rate $\alpha = m/n$ exceeds the sparsity density $\rho = k/n$, *i.e.*, $\alpha > \rho$ [8]. This represents the best possible performance boundary for sparse signal recovery.

Traditional approaches for ℓ_0 minimization often rely on iterative algorithms that can be computationally demanding and time-consuming, particularly when dealing with large-scale problems. However, recent advancements in quantum computing, with its extraordinary processing capabilities, have opened up a promising avenue for harnessing the potential of sparsity in intricate problem-solving endeavors. The integration of quantum computing principles presents an exciting opportunity to expedite and enhance the speed and accuracy of sparse signal recovery processes, thereby revolutionizing the landscape of signal processing and data analysis. Quantum computers boast the capability to explore an exponentially larger solution space within a fraction of the time compared to their classical counterparts. This paper investigates the application of coherent Ising machines (CIM), a type of hybrid quantum device, as a novel approach to efficiently recovering sparse signals, thus addressing a fundamental challenge in various fields.

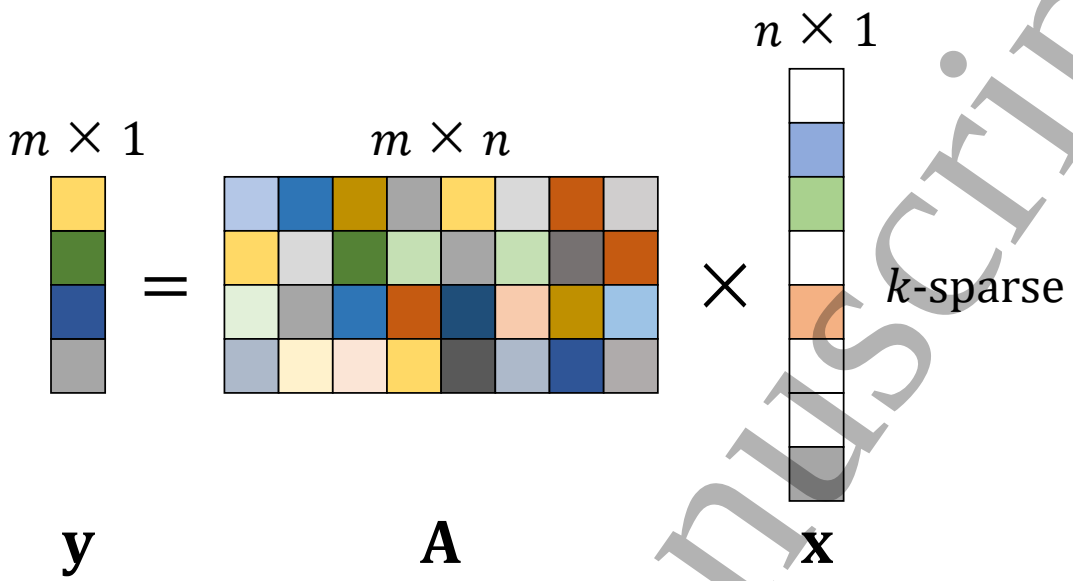


Figure 1: Schematic representation of the compressed sensing problem, a specific instance within the realm of sparse optimization. The figure demonstrates the transformation of a high-dimensional sparse signal into a lower-dimensional measurement space.

CIM represents a notable advancement in the realm of quantum computing, which harnesses the quantum properties of degenerate optical parametric oscillators (DOPOs) to tackle complex optimization problems [9, 10]. Unlike conventional computers, CIMs leverage the distinctive dynamics inherent in DOPO networks, which can be interpreted as "artificial spins" with two states. This approach allows CIMs to efficiently find the optimal solutions to combinatorial optimization problems, such as sparse optimization that are computationally intensive for conventional algorithms. By utilizing a measurement and feedback scheme to simulate arbitrary spins coupling [11], CIMs can tackle large-scale problems invloving thousands of spins [12, 13], overcoming the limitations of previous physical Ising machines.

Note that when the system of linear equations $A\mathbf{x} = b$ is reformulated into a QUBO model by encoding $\mathbf{x}$ as binary variables, the resulting QUBO matrix has $O(n^2K^2)$ non-zero entries, leading to a densely connected network. CIM is well-suited for solving sparse optimization problems due to its ability to directly handle densely connected networks. Unlike other quantum hardware devices, such as quantum annealers, which typically rely on hardware-restricted local graphs like Chimera, CIM allows pairwise coupling between any variables without requiring a minor embedding scheme. This eliminates the need for additional resources to map densely connected networks onto a limited graph structure. In addition, according to [14], CIM shows advantages in handling high-density matrices. CIM also demonstrates advantages in terms of speed,

scalability, and energy efficiency as it consumes lower energy in optical systems. For instance, it has been shown to achieve solutions three orders of magnitude faster than Simulated Annealing (SA) [13], and its scalability has been proven by solving Ising problems involving up to 100,000 variables.

In this paper, we delve into the optimization landscape of ℓ_0-norm minimization, a formidable challenge owing to its inherent combinatorial complexity. We explore the adaptation and utilization of CIM for the following three ℓ_0-related optimization problems, capitalizing on their ability to traverse solution spaces with reduced computational effort:

- Problem 1: Minimize the least-squares regularized with ℓ_0 norm, specified in Definition 1.
- Problem 2: k-sparsity constrained minimization, specified in Definition 2.
- Problem 3: Least-squares constrained by fixed sparsity, specified in Definition 3.

In many practical applications, it is sufficient to achieve a certain sparsity level rather than seeking the absolute sparsest solution. With prior experience on the upper bound of sparsity, we can obtain the optimal solutions by addressing the aforementioned sparse constraint optimization problems 2 and 3. The difficulty of solving these models lies in the sparsity constraint. Note that the sparsity-constrained problem aligns closely with the statistical concept of feature selection [15]. In the realm of statistics, the process of feature selection involves choosing a subset of relevant features from a larger set to build a predictive model. The process aims to identify the most informative and influential features while discarding those that are less significant.

In addition to the development of quantum hardware devices, the integration of quantum and classical computational methods has proven to be a beneficial step toward practical implementation [16–18]. Hybrid quantum-classical algorithms capitalize on the strengths of both paradigms, combining the flexibility of classical optimization with the power of quantum computing. Among these hybrid approaches, the quantum Benders decomposition [19–21] algorithm has emerged as a powerful method for tackling challenging optimization problems that involve both continuous and discrete variables. Aonishi et al. [22] exploit the innovative divide-and-conquer strategy that is conceptually similar to Benders' decomposition, where the optimization of each type of variable can be handled separately while the other type of variables are fixed. The ℓ_0-norm problem in [22] involves a quadratic cost function rather than a linear one, which deviates from classical Benders' decomposition that typically handles linear optimization problems. It remains an interesting question to explore the compatibility of Benders decomposition with sparse optimization problems and its effectiveness.

1.1. Main Results

The integration of quantum principles in solving sparsity-related challenges signifies a progressive step toward harnessing quantum advantages for practical applications. As summarized in Table 1, our contributions can be concluded in the following parts:

	Problem 1	Problem 2	Problem 3	# of Ising spins
Ide and Ohzeki [23]	✓	✓	✓	$2n(K-1)$
Romano et al. [24]	✓	✗	✗	$n(K+1)$
Section 3.1 (Model 1)	✓	✓	✓	$n(K+2)$+slack variables*
Section 3.3 (Model 2.1)	✓	✓	✗	$n(K+1)$
Section 4.2 (Model 2.2)	✓	✓	✓	$n(K+\lceil \log K \rceil + 1)$*

Table 1: Comparison of current work with previous works. *The slack variables and $\lceil \log K \rceil$ term can be avoided using the quantum conditional gradient method in [17].

- In Section 3.1, we initially introduce a unified model, referred to as model 1, that addresses three distinct problems concurrently. This model, where the signal itself comprises real-valued variables while the ℓ_0 norm introduces binary variables, exemplifies the complexity inherent in such mixed-variable systems. We also present the use of the quantum Benders decomposition algorithm to efficiently partition the optimization process into subproblems, addressing real-valued and binary variable components iteratively. This approach further reduces computational consumption while maintaining the solution's fidelity.
- In addition to the mixed-variable model, two further models leveraging fixed-point representations of the signal are proposed. In Section 3.3, we present model 2.1 as an extension of model 1, which exhibits a reduction in the number of required Ising spins compared to model 1, while retaining the capability to address problems 1 and 2. Building upon model 2.1, we further refine our approach and present model 2.2 in Section 4.2, which effectively addresses problem 3 with only $\lceil \log K \rceil$ cost of additional Ising spins per signal. Compared to the approach detailed in [24], our proposed models exhibit a superior problem-solving capacity concerning problem 2 and problem 3, as discussed in Section 4.1. Notably, our models significantly reduce the bit requirements in [23], enabling existing devices to tackle larger-scale problems efficiently. It is worth mentioning that the slack variables in Model 1 and the $\lceil \log K \rceil$ additional Ising spins per signal in Model 2.2 can be eliminated if we employ the quantum conditional gradient method in [17].
- We also present a theoretical analysis of the performance guarantees of our proposed model, providing a solid foundation for its reliability and effectiveness. We then apply our model to the sparse signal recovery problem using the CIM physical platform. The experimental results conducted on the CIM physical platform showcase its remarkable capability to accurately recover the ground truth signal within a millisecond-level time frame. This outcome underscores the efficiency and efficacy of the CIM approach in addressing complex combinatorial optimization problems within real-world scenarios.

1.2. Related Work

In the field of compressed sensing, there are three primary classes of algorithms for sparse signal recovery: convex optimization, greedy iterative, and Bayesian inference methods [25]. Convex optimization approaches frame the recovery as a convex problem, solving it within a convex framework [26]. Greedy iterative algorithms iteratively detect the non-zero elements' positions and recover the signal based on the identified support [27]. Bayesian inference methods, or probabilistic approaches, utilize Bayesian reasoning to generate sparse signals with a given prior [28]. In essence, these techniques recover sparse signals through optimization, iteration, and probabilistic inference.

In the realm of quantum algorithms, recent literature has presented innovative approaches to address the ℓ_0-regularized compressed sensing problem under the quadratic unconstrained binary optimization (QUBO) framework. Aonishi et al. [22] proposed a quantum-classical hybrid system comprising quantum machinery and classical digital processors. To theoretically assess the performance of the hybrid CIM system, [22] introduces truncated Wigner stochastic differential equations (W-SDEs) as a model for degenerate optical parametric oscillator networks. The study employs statistical mechanics on the W-SDEs to derive macroscopic equations. Notably, the findings suggest that in practical scenarios like magnetic resonance imaging data analysis, the estimation accuracy of this hybrid system may surpass that of ℓ_1-RBCS. Gunathilaka et al. [29] extended this work by introducing a chaotic amplitude-controlled closed-loop CIM (CAC-CIM), leveraging chaotic dynamics to escape local minima and achieve enhanced accuracy in magnetic resonance imaging data. Building on this, Aonishi et al. [30] developed a highly versatile Cyber CIM implemented on FPGA, enabling both open-loop and closed-loop architectures with FP32 precision for interaction terms, and achieving computation speeds over ten times faster than GPU implementations for applications including L0RBCS. To address the computational challenges of CIM's stochastic differential equations, Gunathilaka et al. [31] introduced the mean-field CIM (MF-CIM) model, a physics-inspired heuristic solver that simplifies differential equations, enabling efficient FPGA-based implementations while preserving performance. These advancements collectively illustrate the evolution of CIM technology and its practical relevance in compressed sensing tasks.

In a related vein, [32] introduces a hybrid methodology that alternates between discrete and continuous optimization steps. Within the discrete optimization phase, the Ising machine is employed to minimize the objective function described by the QUBO form. The study demonstrates that even when substituting the Ising machine with simulated annealing (SA), the method retains its functionality. In the reported numerical experiments, SA is utilized for the discrete optimization stage. [33] investigates compressed sensing for binary signal scenarios. This work formulates an equivalent QUBO problem and leverages classical and hybrid quantum computing techniques based on quantum annealing to tackle the challenge. The experimental results intriguingly highlight a substantial enhancement in the reconstructability of problem instances

compared to traditional ℓ_1 optimization methods.

The CIM [9–13, 34–36] represents a type of Ising-like hardware solver designed for specialized optimization of Ising models. The Ising model describes spin interaction behavior and serves as a theoretical model for addressing combinatorial optimization problems. Finding the optimal combination from a set of numerous choices can be defined as discovering the ground state of an Ising model with appropriate parameters. The precursor to Ising computers involves quantum annealers (QA), which employ superconducting qubits to emulate artificial spins. Inagaki et al. [12] introduced a large-scale artificial spin network based on degenerate optical parametric oscillators (DOPO), paving the way for photon-based Ising machines capable of tackling challenging combinatorial optimization problems. Inagaki et al. [12] demonstrated that for 2000-spin problems, encompassing random graphs, scale-free graphs, and complete graphs, CIM could attain highly accurate approximate solutions within 5 ms, surpassing the precision of classical computers using the GW-SDP algorithm. Notably, in complete graphs, CIM achieved more accurate solutions compared to classical computers utilizing simulated annealing (SA) while improving the speed by approximately 20 to 50 times. CIM's prowess and advantage in solving QUBO problems are evident. CIM has been demonstrated to solve problems such as Maximum Likelihood Detection (ML-MIMO) [37–40], error control decoding [41], MIMO beam selection [42], computing power scheduling [43], and molecular docking [44].

2. Problem Definitions and Formulations

In sparse signal recovery, a fundamental challenge arises in accurately estimating a sparse signal based solely on a limited set of observations. This scenario frequently involves addressing the linear equation $\mathbf{y} = \mathbf{A}\mathbf{x}$, where the vector $\mathbf{y} \in \mathbb{R}^m$ denotes the measurement data, the matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ represents the observation matrix (with $m \ll n$), $\mathbf{x}$ signifies the sparse signal to be recovered. In [45], the authors proved a crucial condition for signal recovery. Specifically, considering an $\mathbf{x}$ that adheres to a k-sparse property (indicating that it has at most k nonzero components), it turns out that if the restricted isometry constant δ_{2k} satisfies the condition $\delta_{2k} < 1$, an intriguing result unfolds. Namely, the minimization of the ℓ_0 norm,

$$\min ||\mathbf{x}||_0 \tag{1}$$

$$\text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{y} \tag{2}$$

which seeks to find the sparsest solution, yields a unique outcome in the form of $\mathbf{x}$. The problem is known to be NP-hard [6]. In addition, in many practical scenarios, the measured signal $\mathbf{y}$ is perturbed by a small amount of noise $\mathbf{n}$, *i.e.*, $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$, our models in this paper can be extended to accommodate this situation.

By formulating Eq. (1)-(2) into the QUBO form, penalty methods are introduced to attack the NP-hard unconstrained problem as $\min_{\mathbf{x}} ||\mathbf{x}||_0 + \lambda \cdot ||\mathbf{A}\mathbf{x} - \mathbf{y}||_2^2$.

Alternatively, we can consider the equivalent ℓ_0-regularized problem that appears more frequently in the literature.

Definition 1 (Minimize the least-squares regularized with ℓ_0 norm)**.** *Given measurement matrix* $\mathbf{A} \in \mathbb{R}^{m\times n}(m \ll n)$*, measurements* $\mathbf{y} \in \mathbb{R}^m$*, we aim to recover the sparse signal* $\mathbf{x} \in \mathbb{R}^n$ *via the following objective function:*

$$\min_{\mathbf{x}} ||\mathbf{A}\mathbf{x} - \mathbf{y}||_2^2 + \lambda \cdot ||\mathbf{x}||_0 \tag{3}$$

The function above is to find an estimate of $\mathbf{x}$ that minimizes the least-squares error, given by $||\mathbf{A}\mathbf{x} - \mathbf{y}||_2^2$. To encourage sparsity in the solution, a regularization term is introduced using the ℓ_0 norm. There are several linear models in various contexts, including inverse problems like compressed sensing and sparse regression [46]. The distinctive aspect here lies in the primary focus of each domain. In statistical contexts, the focus remains on predictive accuracy, whereas in solving inverse problems, the emphasis shifts towards unraveling the coefficients with their inherent physical meanings intact.

Definition 2 (k-sparsity constrained minimization problem)**.**

$$\min ||\mathbf{A}\mathbf{x} - \mathbf{y}||_2 \tag{4}$$

$$s.t.\ ||\mathbf{x}||_0 \leq k \tag{5}$$

In the context of sparse signal recovery, we further examine a scenario in which an additional constraint is imposed during the solution of the least-squares problem. The goal is to find a sparse signal $\mathbf{x}$ that best fits the measurement data while adhering to the sparsity constraint. The constraint requires that the sparsity of the solution vector $\mathbf{x}$ does not exceed a given threshold k.

The k-sparsity constrained minimization problem, as depicted in equation (4)-(5), closely relates to the concept of feature selection in statistics [15], which is formally described as following,

$$\min_{\mathbf{w}\in\mathbb{R}^p} \sum_{i=1}^{n} \ell(y_i, \mathbf{w}^{\mathrm{T}}\mathbf{x}_i) \text{ s.t. } \|\mathbf{w}\|_0 \leq k, \tag{6}$$

where ℓ is an appropriate convex loss function, such as the ones of ordinary least square, *i.e.*, $\ell(y,u) = \frac{1}{2}(y-u)^2$, considered in this paper. In this problem, the objective is to minimize the sum of the loss terms, given by the function $\ell(y_i, \mathbf{w}^{\mathrm{T}}x_i)$, where y_i is the target variable associated with the i-th data point, $\mathbf{x}_i$ is the corresponding feature vector and $\mathbf{w}$ is a parameter vector to be optimized. The optimization goal is subject to the constraint that the ℓ_0 norm of the parameter vector $\mathbf{w}$ is less than or equal to a predefined value k, allowing us to select only a limited number of features for our model.

Definition 3 (Least-squares constrained by fixed sparsity)**.**

$$\min ||\mathbf{A}\mathbf{x} - \mathbf{y}||_2 \tag{7}$$

$$s.t.\ ||\mathbf{x}||_0 = k \tag{8}$$

The third problem is a specific variant of the second problem, where the sparsity level is fixed at exactly k nonzero elements in the signal $\mathbf{x}$. This problem is particularly relevant when the exact sparsity level is known or when enforcing strict sparsity is crucial for interpretability or resource-efficient signal representation.

3. QUBO Model for ℓ_0 regularization

The QUBO problem belongs to a distinctive category of combinatorial optimization challenges. Its primary objective is to find a binary vector that minimizes a quadratic function. This specific problem class can be directly tackled using QA, positioning it as a pivotal domain within quantum computing research. It serves as a bridge connecting traditional optimization problems with the quantum realm. As the field of quantum computing continues to flourish, investigating and comprehending QUBO problems remains a cornerstone for realizing practical quantum applications in optimization.

The QUBO problem aims to identify a binary vector that can minimize a quadratic function. The standard formulation for QUBO problems can be represented using the equation:

$$\min_{\mathbf{x}\in\{0,1\}^n} \mathbf{x}^{\mathrm{T}}\mathbf{Q}\mathbf{x}$$

where $\mathbf{Q}$ is the coefficient matrix and $\mathbf{x}$ is the binary vector. This mathematical framework finds broad applications in various practical scenarios, including computer vision [47–52]; machine learning tasks such as feature selection [53, 54] and clustering [55, 56]; complex financial models like portfolio optimization [57]. Taking into account the compatibility of QUBO problems with coherent Ising machines, it is noteworthy that coherent Ising machines are highly suited for addressing QUBO issues, making them an excellent approach for optimization and sampling tasks.

3.1. A Unified Model

Consider a signal $\mathbf{x}$ composed of n components, where the i-th component is defined as $x_i = g(\boldsymbol{\delta}_i)$, where $\boldsymbol{\delta}_i$ is a binary vector of length K, representing the encoding of the i-th component x_i, *i.e.*,

$$\boldsymbol{\delta}_i = [\delta_{i1}, \delta_{i2}, \ldots, \delta_{iK}]^{\mathrm{T}} \tag{9}$$

and $\delta_{ij} \in \{0, 1\}$. Here, $g(\cdot) : 2^K \to \mathbb{R}$ is a function that maps binary vectors of length K to real numbers. We assume that the range of x_i is known, specifically, $x_i \in [\underline{x}, \bar{x}]$.

Let $\varepsilon = \min_{x_i \neq 0} |x_i|$, which represents the minimum absolute value of non-zero signal components. We introduce binary variables z_i to indicate whether the i-th component is zero, *i.e.*, $z_i = 1$ if x_i is zero. Similarly, we define binary variables z_i^+ and z_i^- to indicate whether x_i is positive or negative, respectively. To ascertain the non-zero nature of each signal, we introduce the following constraints:

- Bounds on signal components:

$$\underline{x} z_i^- + \varepsilon z_i^+ \le x_i \le \bar{x} z_i^+ - \varepsilon z_i^- \tag{10}$$

for all $1 \le i \le n$. Note that $x_i \in [\varepsilon, \bar{x}]$ when $x_i > 0$, and $x_i \in [\underline{x}, -\varepsilon]$ when $x_i < 0$. This constraint ensure that the variables z_i, z_i^+, z_i^- correctly assume values to restrict the signal x_i within a specified range.

- Binary constraints:

$$z_i^- + z_i + z_i^+ = 1 \tag{11}$$

for all $1 \le i \le n$. These constraints ensure that exactly one out of the three binary variables z_i^-, z_i, and z_i^+ are active at a time, capturing the sign of x_i. We can replace z_i^- using the substitution $z_i = 1 - z_i^- - z_i^+$, but it is important to note that the values of z_i^- and z_i^+ must ensure that z_i^- remains within the range $\{0, 1\}$. Thus, we need to add the following constraint:

$$z_i^- + z_i^+ \le 1$$

This can be achieved by adding a penalty term $z_i^- \cdot z_i^+$ to the objective function.

Now, consider the QUBO model:

$$\begin{aligned}
&\min \mathcal{H} \\
&= \min \sum_{i=1}^{n} (z_i^- + z_i^+) \qquad (12) \\
&\quad + \lambda \cdot \sum_{\ell=1}^{m} \Big(\sum_{i=1}^{n} A_{\ell,i} \cdot g(\boldsymbol{\delta}_i) - y_\ell \Big)^2 \\
&\quad + \lambda \cdot \sum_{i=1}^{n} z_i^- z_i^+ \\
&\quad + \lambda \cdot \sum_{i=1}^{n} [\underline{x} z_i^- + \varepsilon z_i^+ + s_i^{(1)} - g(\boldsymbol{\delta}_i)]^2 \\
&\quad + \lambda \cdot \sum_{i=1}^{n} [\bar{x} z_i^+ - \varepsilon z_i^- - s_i^{(2)} - g(\boldsymbol{\delta}_i)]^2 \qquad (13)
\end{aligned}$$

The third and fourth terms enforce the lower and upper bounds on signal components by introducing slack variables, for example, $s_i^{(h)} = \sum_j s_{ij}^{(h)} \cdot 2^j$ $(h = 1, 2)$.

3.2. Quantum Benders Decomposition Algorithm

The problem can be reformulated as follows:

$$\min_{\mathbf{x},\mathbf{z}} \ \mathbf{p}^{\mathrm{T}} \mathbf{z} \tag{14}$$

$$s.t. \quad \mathbf{A}\mathbf{x} = \mathbf{b} \tag{15}$$

$$\mathbf{B}\mathbf{x} + \mathbf{C}\mathbf{z} \le \mathbf{h} \tag{16}$$

$$\mathbf{x} \in \mathbb{R}^n, \mathbf{z} \in \{0, 1\}^{2n} \tag{17}$$

where $\mathbf{z} = [z_1^-, z_2^-, \ldots, z_n^-, z_1^+, z_2^+, \ldots, z_n^+]$, $\mathbf{p} = \mathbf{1}_{2n\times 1}$, and

$$\mathbf{B} = \begin{bmatrix} -\text{diag}(\mathbf{1}_n) \\ \text{diag}(\mathbf{1}_n) \\ \mathbf{0}^{n\times n} \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} \text{diag}(\underline{x}\cdot\mathbf{1}_n) & \text{diag}(\varepsilon\cdot\mathbf{1}_n) \\ \text{diag}(\varepsilon\cdot\mathbf{1}_n) & -\text{diag}(\bar{x}\cdot\mathbf{1}_n) \\ \text{diag}(\mathbf{1}_n) & \text{diag}(\mathbf{1}_n) \end{bmatrix}, \quad \mathbf{h} = \begin{bmatrix} \mathbf{0}^{2n\times 1} \\ \mathbf{1}^{n\times 1} \end{bmatrix}$$

The ℓ_0-norm optimization problem is particularly suitable for Benders decomposition due to its two-level structure, where the upper level involves binary decisions to select a subset of variables, and the lower level involves solving a continuous optimization problem based on the selected subset. This separation allows the binary selection to be handled in the master problem, as the dependent continuous optimization is addressed in the subproblem. The dual information obtained from solving the subproblem can then be used to iteratively refine the master problem with Benders cuts, ensuring convergence. By decomposing the problem, Benders' decomposition efficiently manages the combinatorial complexity of ℓ_0-norm constraints while leveraging the scalability of solving smaller, more focused subproblems.

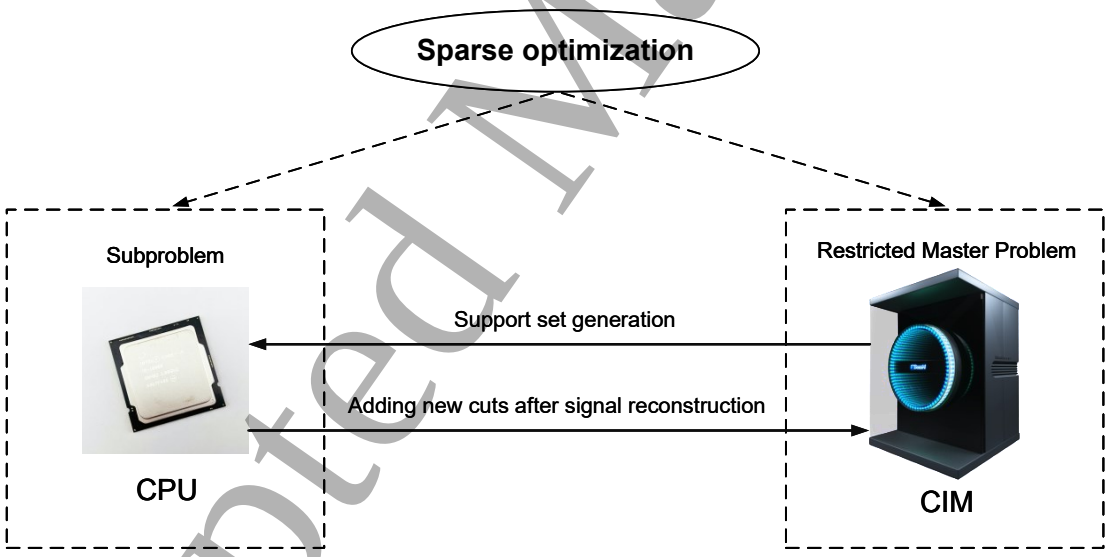


Figure 2: The workflow of Quantum Benders Decomposition for sparse optimization, where the Restricted Master Problem (RMP) is solved using CIM to generate support sets. The Subproblem, handled on a classical CPU, introduces new cuts based on reconstructed signals to iteratively refine the solution.

We rewrite the optimization problem as

$$\min_{\mathbf{z}} \mathbf{p}^{\mathrm{T}}\mathbf{z} + t$$

where t is the optimal objective value of the primal subproblem under a given value of $\mathbf{z}$, *i.e.*, when the support vector of the solution $\mathbf{x}$ that represents the locations of the

non-zero entries in $\mathbf{x}$ is fixed:

$$\begin{aligned} \text{(Primal Subproblem)} \quad t = \min_{\mathbf{x}} \; & 0 \\ s.t. \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{B}\mathbf{x} \leq \mathbf{h} - \mathbf{C}\mathbf{z} \end{aligned}$$

The value of t can be interpreted as:

$$t = \begin{cases} 0, & \text{If } \exists\ \mathbf{x} \text{ with given non-zero entries,} \\ \infty, & \text{Otherwise.} \end{cases}$$

In this context, $t = 0$ indicates that the second-stage optimization problem is feasible and has a solution for the given $\mathbf{z}$, whereas $t = \infty$ means that no feasible solution satisfying the linear equations $\mathbf{A}\mathbf{x} = \mathbf{b}$ exists under the given support set.

According to the duality theory, t can be determined by solving the following dual problem:

$$\begin{aligned} \text{(Dual Subproblem)} \quad t = \max_{\lambda,\nu} \; & -\lambda^{\mathrm{T}}\mathbf{b} - \nu^{\mathrm{T}}(\mathbf{h} - \mathbf{C}\mathbf{z}) && (18) \\ \text{s.t.} \quad & \mathbf{A}^{\mathrm{T}}\lambda + \mathbf{B}^{\mathrm{T}}\nu = \mathbf{0}, && (19) \\ & \nu \geq \mathbf{0}. && (20) \end{aligned}$$

λ and ν are the Lagrange multipliers associated with the equality and inequality constraints, respectively. The feasible region of the dual problem is fixed and determined by the structure of the primal constraints. When addressing the dual subproblem under the given parameter $\hat{\mathbf{z}}$, the algorithm will lead to one of two distinct outcomes:

- t **is unbounded above.** The algorithm will identify this condition and return an extreme ray $[\lambda_i^{\dagger\mathrm{T}}, \nu_i^{\dagger\mathrm{T}}]$ that satisfies:

  $$\lambda_i^{\dagger\mathrm{T}}\mathbf{b} + \nu_i^{\dagger\mathrm{T}}(\mathbf{h} - \mathbf{C}\hat{\mathbf{z}}) < 0.$$

  In this scenario, the objective function value $f(\mathbf{z})$ is unbounded and approaches $+\infty$. To ensure feasibility, a new *feasibility cut* is added to the master problem:

  $$\lambda_i^{\dagger\mathrm{T}}\mathbf{b} + \nu_i^{\dagger\mathrm{T}}(\mathbf{h} - \mathbf{C}\mathbf{z}) \geq 0. \tag{21}$$

  This ensures that the objective function forms an obtuse angle with each extreme ray of the dual feasible region.

- **Existence of an optimal solution.** The algorithm will return the corresponding extreme point $[\lambda_j^{\ddagger\mathrm{T}}, \nu_j^{\ddagger\mathrm{T}}]$ along with the corresponding optimal objective value of

  $$\lambda_j^{\ddagger\mathrm{T}}\mathbf{b} + \nu_j^{\ddagger\mathrm{T}}(\mathbf{h} - \mathbf{C}\hat{\mathbf{z}}). \tag{22}$$

  This expression forms the basis for generating the following *optimality cuts*

  $$t \geq \lambda_j^{\ddagger\mathrm{T}}\mathbf{b} + \nu_j^{\ddagger\mathrm{T}}(\mathbf{h} - \mathbf{C}\mathbf{z}),$$

  to tighten the master problem's representation.

We use $\mathcal{I}$ and $\mathcal{J}$ to denote the whole set of extreme points and extreme rays of the fixed dual feasible region. The value of t is the dual objective at the best extreme point in $\mathcal{I}$ while ruling out all extreme rays in $\mathcal{J}$, directions of unbounded improvement in the dual subproblem. Benders decomposition adopts a delayed constraint generation strategy, by using only a subset of constraints in $\mathcal{I}$ and $\mathcal{J}$. To this end, we introduce the Restricted Master Problem (RMP$_k$), a relaxed version of the original ℓ_0 optimization problem. RMP$_k$ initially includes a subset of the total constraints, which are progressively added as the algorithm identifies violations through the subproblem. By solving RMP$_k$, we obtain an approximate solution, which is then refined in subsequent iterations.

$$\text{(Restricted Master Problem, RMP}_k\text{)} \quad \min_{\mathbf{z},t} \quad \mathbf{p}^{\mathrm{T}}\mathbf{z} + t \tag{23}$$

$$s.t. \quad \lambda_i^{\dagger\mathrm{T}}\mathbf{b} + \nu_i^{\dagger\mathrm{T}}(\mathbf{h} - \mathbf{C}\mathbf{z}) \geq 0, \quad i \in \mathcal{I}_k, \tag{24}$$

$$t \geq \lambda_j^{\ddagger\mathrm{T}}\mathbf{b} + \nu_j^{\ddagger\mathrm{T}}(\mathbf{h} - \mathbf{C}\mathbf{z}), \quad j \in \mathcal{J}_k. \tag{25}$$

The restricted master problem can be transformed to the following QUBO form:

$$\mathbf{p}^{\mathrm{T}}\mathbf{z} + \sum_i 2^i \cdot t_i + \sum_{i\in\mathcal{I}_k} \Big(\lambda_i^{\dagger\mathrm{T}}\mathbf{b} + \nu_i^{\dagger\mathrm{T}}(\mathbf{h} - \mathbf{C}\mathbf{z}) - \sum_\ell s_{i,\ell}^{(1)} \cdot 2^\ell\Big)^2 + \sum_{j\in\mathcal{J}_k} \Big(t + \sum_\ell s_{j,\ell}^{(2)} \cdot 2^\ell - \lambda_j^{\ddagger\mathrm{T}}\mathbf{b} - \nu_j^{\ddagger\mathrm{T}}(\mathbf{h} - \mathbf{C}\mathbf{z})\Big)^2,$$

where t_i, $s_{j,\ell}^{(1)}$, $s_{j,\ell}^{(2)}$ are binary variables for representing the slack variables.

Algorithm 1 Quantum Benders Decomposition Algorithm

Initialize LB $\leftarrow -\infty$, UB $\leftarrow +\infty$.
Define relaxed master problem RMP_k of the ℓ_0 norm minimization problem.
while UB $-$ LB $> \epsilon$ **do**
  Solve RMP_k to obtain optimal solution $(\bar{\mathbf{z}}, \bar{t})$.
  Update LB $\leftarrow \text{VAL}_k$.
  Solve the subproblem under $\bar{\mathbf{z}}$
  if Subproblem has an optimal solution $(\bar{\mathbf{x}}, \bar{\lambda}, \bar{\nu})$ **then**
    if $\bar{\lambda}_j^{\mathrm{T}}\mathbf{b} + \bar{\nu}_j^{\mathrm{T}}(\mathbf{h} - \mathbf{C}\bar{\mathbf{z}}) \leq \bar{t}$ **then**
      $(\bar{x}, \bar{z})$ is optimal.
      Terminate.
    else
      Add constraint $\bar{\lambda}_j^{\mathrm{T}}\mathbf{b} + \bar{\nu}_j^{\mathrm{T}}(\mathbf{h} - \mathbf{C}\mathbf{z}) \leq t$ to RMP_k.
      Update UB $\leftarrow \min\{\text{UB}, \mathbf{p}^{\mathrm{T}}\bar{\mathbf{z}}\}$.
    end if
  else
    Let $[\lambda_i^{\dagger\mathrm{T}}, \nu_i^{\dagger\mathrm{T}}]$ be the extreme ray of the Dual Subproblem found under $\bar{\mathbf{z}}$.
    Add constraint $\lambda_i^{\dagger\mathrm{T}}\mathbf{b} + \nu_i^{\dagger\mathrm{T}}(\mathbf{h} - \mathbf{C}\mathbf{z}) \geq 0$ to RMP_k.
  end if
end while

The Quantum Benders Decomposition Algorithm combines classical optimization with quantum computing to solve sparse optimization problems efficiently. By decomposing the optimization problem into a master problem and a sub-problem, this hybrid approach uses quantum computing to solve RMP_k, accelerating the optimization process. The relaxed master problem involves fewer decision variables and constraints and is often simpler to solve than the original full problem. By solving this relaxed version using quantum methods, the algorithm benefits from the parallelism and efficiency of quantum computers, which can provide faster solutions for large optimization problems.

3.3. A Simplified Model

In this section, we introduce a simpler model that requires fewer Ising spins for representation. Similarly, we make the assumption that each x_i is represented by a set of K binary variables δ_{ij}, where $x_i = g(\boldsymbol{\delta}_i)$ and $g(\cdot) : \{0,1\}^K \to \mathbb{R}$. We assume that function $g(\cdot)$ possesses a unique zero point and denote the zero of $g(\cdot)$ as

$$\boldsymbol{\delta}_0 = [\delta_1^{(0)}, \ldots, \delta_K^{(0)}]^{\mathrm{T}},$$

i.e., $g(\boldsymbol{\delta}_0) = 0$.

Let $\mathcal{I}_0 = \{j \in [K] | \delta_j^{(0)} = 0\}$ represent the indices of zero values in the vector $\boldsymbol{\delta}^{(0)}$, and let $\mathcal{I}_1 = \{j \in [K] | \delta_j^{(0)} = 1\}$ represent the indices of one values. We introduce the

following constraints to characterize whether x_i is equal to zero:

$$z_i \geq \delta_{ij}, \quad j \in \mathcal{I}_0 \tag{26}$$

$$z_i \geq 1 - \delta_{ij}, \quad j \in \mathcal{I}_1 \tag{27}$$

The constraints above are pivotal in determining the value of z_i. It is important to observe that for the indices j within the set $\mathcal{I}_0$, the condition for x_i does not coincide with the zero point of the function $g(\cdot)$ is met if there exists at least one index j for which $\delta_{ij} = 1$. Similarly, for the indices j in the set $\mathcal{I}_1$, x_i deviates from the zero point if there exists at least one x_i such that $\delta_{ij} = 0$. In both scenarios, the inequalities in question necessitate that z_i must be assigned a value of 1. However, when the right-hand sides of the inequalities evaluate to 0, it indicates that x_i aligns with the zeros of the function $g(\cdot)$. In this particular case, given that the objective function is designed to minimize the cumulative sum of z_i values, it is optimal to set z_i to 0. This ensures that the solution aligns with the minimization objective while adhering to the logical structure of the constraints.

Let $\mathbf{z} = [z_1, z_2, \ldots, z_n]^{\mathrm{T}}$, $\boldsymbol{\delta} = [\boldsymbol{\delta}_1^{\mathrm{T}}, \boldsymbol{\delta}_2^{\mathrm{T}}, \ldots, \boldsymbol{\delta}_n^{\mathrm{T}}]^{\mathrm{T}}$ and $\mathbf{p} = [\mathbf{z}^{\mathrm{T}}, \boldsymbol{\delta}^{\mathrm{T}}]^{\mathrm{T}} \in \{0,1\}^{n(K+1)}$. The quadratic unconstrained binary optimization formulation of this model is given by:

$$\begin{aligned}\min_{\mathbf{p}} \mathcal{H} = \min_{\mathbf{p}} & \sum_{\ell=1}^{m} \Big(\sum_{i=1}^{n} A_{\ell,i} \cdot g(\boldsymbol{\delta}_i) - y_\ell \Big)^2 + \lambda \cdot \sum_{i=1}^{n} z_i \\ & + \lambda \cdot \sum_{i=1}^{n} \Big(\sum_{j \in \mathcal{I}_0} (\delta_{ij} - z_i \delta_{ij}) + \sum_{j \in \mathcal{I}_1} (1 - z_i)(1 - \delta_{ij}) \Big).\end{aligned} \tag{28}$$

For notational convenience, we let

$$\mathcal{H}_0 = \sum_{i=1}^{n} \Big(\sum_{j \in \mathcal{I}_0} (\delta_{ij} - z_i \delta_{ij}) + \sum_{j \in \mathcal{I}_1} (1 - z_i)(1 - \delta_{ij}) \Big), \tag{29}$$

the term $(\delta_{ij} - z_i\delta_{ij})$ and $(1 - z_i)(1 - \delta_{ij})$ in $\mathcal{H}_0$ encodes inequalities (26) and (27) respectively.

Consider when

$$g(\mathbf{t}) = \mathbf{w}^{\mathrm{T}}\mathbf{t} + w_0 \ (\mathbf{w} \in \mathbb{R}^K, w_0 \in \mathbb{R}),$$

then the i-th signal can be expressed as $x_i = \mathbf{w}^{\mathrm{T}}\boldsymbol{\delta}_i + w_0$ and consequently

$$\mathbf{x}^{\mathrm{T}} = \boldsymbol{\delta}^{\mathrm{T}}(\mathbf{I}_n \otimes \mathbf{w}) + w_0 \mathbf{1}_n^{\mathrm{T}}, \tag{30}$$

where $\otimes$ is the Kronecker product. The Hamiltonian of this model is represented as follows,

$$\mathcal{H} = \mathbf{p}^{\mathrm{T}} \mathbf{Q}_2 \mathbf{p}, \tag{31}$$

where matrix $\mathbf{Q}_2$ along with the detailed derivation, are provided in Appendix Appendix A.

Example 1 (Real valued signal)**.** *In scenarios involving signal components with both positive and negative values, let us consider the case where we define* $\bar{x} = -\underline{x} \geq 0$*. The signal* x_i *can be expressed as:*

$$x_i = 2 \times \bar{x} \times \Big(\sum_{j=1}^{K} 2^{-j} \times \delta_{ij} - 0.5 \Big) \in [-\bar{x}, \bar{x}). \tag{32}$$

For inequalities related to the L_0 *norm, they take the form:*

$$z_i \geq 1 - \delta_{i1} \tag{33}$$

$$z_i \geq \delta_{ij}, \quad j \geq 2 \tag{34}$$

In the context of quantum optimization, the QUBO expression is provided below:

$$\sum_{\ell=1}^{m} \Big(2 \sum_{i=1}^{n} A_{\ell,i} \cdot \bar{x} \cdot \Big(\sum_{j=1}^{K} 2^{-j} \cdot \delta_{ij} - 0.5 \Big) - y_\ell \Big)^2 + \lambda \cdot \sum_{i=1}^{n} \Big((1 - z_i)(1 - \delta_{i1}) + \sum_{j=2}^{K} (\delta_{ij} - z_i \delta_{ij}) \Big) + \lambda \cdot \sum_{i=1}^{n} z_i \tag{35}$$

Example 2 (Non-negative signal)**.** *Consider a scenario where the signal components are non-negative,* i.e., $\underline{x} = 0$*. We let*

$$x_i = \bar{x} \times \sum_{j=1}^{K} 2^{-j} \times \delta_{ij} \in [0, \bar{x}). \tag{36}$$

With the constraint that $z_i \geq \delta_{ij}$ $(\forall i, j)$*, the corresponding QUBO model is presented as following:*

$$\sum_{\ell=1}^{m} \Big(\sum_{i=1}^{n} A_{\ell,i} \cdot \bar{x} \cdot \sum_{j=1}^{K} 2^{-j} \cdot \delta_{ij} - y_\ell \Big)^2 + \lambda \cdot \sum_{i=1}^{n} \sum_{j=1}^{K} (\delta_{ij} - z_i \delta_{ij}) + \lambda \cdot \sum_{i=1}^{n} z_i \tag{37}$$

3.4. Hyperparameter selection in the QUBO model

Within the QUBO model outlined earlier, an aspect that requires careful consideration is related to the selection of the parameter λ. The impact of λ on the optimization process is noteworthy. Larger values of λ enforce a convergence of the solution vectors to a state of zero, while considerably diminutive values steer the model towards resembling a least-squares estimate. Optimal results are typically achieved for intermediary λ values, which facilitates the emergence of parsimonious models capable of faithfully recovering the latent ground.

However, when adapting these considerations to the quantum computing domain, the precision limitations intrinsic to quantum systems necessitate a cautious calibration of the penalization coefficient λ. Imposing excessively stringent penalties is counterproductive due to these precision constraints. This prompts a pivotal inquiry: how can the minimum threshold, denoted as $\lambda_{\min}$, be determined? The overarching strategy hinges upon the principle that solution vectors should not violate the constraints to an extent that leads to a reduction in the objective value. Thus, the motivation behind selecting a sufficiently substantial penalty coefficient lies in compelling solution vectors to adhere within the feasible domain. In essence, this is to ensure that the penalty incurred due to constraint violations surpasses the corresponding reduction in the objective function value.

In the context of an Ising model of the objective function, defined as $\min_{\sigma_i\in\{1,-1\}} \sigma^{\mathrm{T}}\mathbf{J}\sigma$, the energy variation ensuing from the flipping of the i-th element of σ can be succinctly formulated as:

$$-4\cdot\sigma_i\cdot\langle\sigma,\mathbf{J}_i\rangle, \tag{38}$$

where $\mathbf{J}_i$ denotes the i-th row of matrix $\mathbf{J}$ and $\langle,\rangle$ represents the inner product. Note that after normalization (*e.g.*, after multiplying a factor of $1/\min_{A_{ij}\neq 0} A_{ij}^2$), the variation in the quadratic formation of the constraint is at least one, hence one possible lower bound for λ we can set is

$$\begin{aligned}\lambda_{\min} &= \max_{i,\sigma} |-4\cdot\sigma_i\cdot\langle\sigma,\mathbf{J}_i\rangle| \\ &= 4\max_{i,\sigma} |\langle\sigma,\mathbf{J}_i\rangle| \\ &= 4\max_i \sum_j |J_{ij}|\end{aligned} \tag{39}$$

4. Sparsity constrained optimization

Upon a quick examination, it becomes evident that our model in Section 3.1 can be adapted to tackle both Problem 2 and Problem 3. Due to the constraint of brevity in this context, we will skip the detailed discussion of these modifications here. In this section, we will explore the task of addressing both problem 2 and problem 3 within the framework of our established model 2.1. To effectively tackle Problem 3, modifications and enhancements were introduced to our Model 2.1, resulting in the development of Model 2.2.

4.1. Inequality constraint

For model 2.1 presented in Section 3.3, the variable z_i does not entirely capture the true nature of the signal x_i being exactly zero. This is because when x_i equals zero, z_i can take on either the value 1 or 0. As a result, this model falls short of effectively addressing problem 3. Despite this limitation, our model demonstrates notable efficacy without

necessitating any modifications when the constraints are transformed into inequality constraints. The rationale behind this is that when x_i is non-zero, z_i is necessarily set to 1. Consequently, the quantity $\sum_{i=1}^{n} z_i$ serves as an achievable upper bound for $||\mathbf{x}||_0$, as demonstrated by the following problem formulation:

$$\min ||\mathbf{A}\mathbf{x} - \mathbf{y}||_2^2 \tag{40}$$

$$\text{s.t.} \sum_{i=1}^{n} z_i = k \tag{41}$$

$$x_i = g(\boldsymbol{\delta}_i) \tag{42}$$

$$z_i \geq \delta_{ij}, \quad j \in \mathcal{I}_0 \tag{43}$$

$$z_i \geq 1 - \delta_{ij}, \quad j \in \mathcal{I}_1 \tag{44}$$

which is equivalent to problem 2.

We provide an analysis of the limitations and shortcomings of the model presented in [24] when applied to sparsity-constrained optimization. The approach in [24] is known as *termwise quadratization* [58] in the literature, which converts a monomial into a quadratic polynomial using exactly one auxiliary variable. More specifically, there are several negative monomials in the form of $-\prod_{i=1}^{d} b_i$ in the QUBO objective function in [24], the domain of each b_i is $\{0, 1\}$. The following two polynomials are equivalent:

$$-\prod_{i=1}^{d} b_i \text{ and } \min_{a \in \{0,1\}} \Big\{ (d-1)a - \sum_{i=1}^{d} a b_i \Big\}. \tag{45}$$

This involves representing the ℓ_0-norm using quadratic terms, resulting in a formulation in the form of higher-order unconstrained binary optimization (HUBO) when dealing with problem 2. Consequently, an order reduction step becomes necessary, which adds complexity to the model and increases the number of binary variables. More importantly, the model described in [24] consistently requires the inclusion of the quadratic polynomial containing auxiliary variables, as outlined in equation (45), within the objective function. This aspect of the model inclines it towards seeking solutions with smaller ℓ_0-norms, which is not equivalent to addressing problem 2.

4.2. Equality constraint

In this section, we focus on augmenting model 2.1 to effectively tackle the challenges posed by problem 3. The primary objective is to elucidate the key alterations made to model 2.1 and the resulting QUBO formulation. To achieve this, we have incorporated specific constraints within Model 2.1, affecting each z_i as follows:

$$z_i \leq \sum_{j \in \mathcal{I}_0} \delta_{ij} - \sum_{j \in \mathcal{I}_1} \delta_{ij} + |\mathcal{I}_1|, \forall 1 \leq i \leq n. \tag{46}$$

This constraint ensures that if the signal x_i is 0, z_i must also be constrained to 0. Correspondingly, we only need to augment the QUBO expression outlined in (A.1) with

the following term:

$$\sum_{i=1}^{n}\Big(\sum_{j\in\mathcal{I}_0}\delta_{ij}-\sum_{j\in\mathcal{I}_1}\delta_{ij}+|\mathcal{I}_1|-z_i-s_i\Big)^2+\Big(\sum_{i=1}^{n}z_i-k\Big)^2, \tag{47}$$

where $s_i=\sum_{j=0}^{\lceil\log K\rceil-1}2^j\cdot s_{ij}$ serves as the slack variable.

By introducing these constraints and their respective QUBO representation, we can tailor Model 2.1 to effectively address Problem 3, thus enhancing its applicability and performance in this specific context.

4.3. Further Applications

Our model can be extended to support vector machines (SVM), a powerful class of supervised learning algorithms used for classification and regression tasks. The optimization problem for SVM can be formulated as follows:

$$\lambda\cdot||\mathbf{w}||_0+\frac{1}{m}\sum_{i=1}^{m}\max(0,1-y_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i-b)) \tag{48}$$

where $\mathbf{w}$ is the weight vector, b is the bias term, y_i are the labels of the training data $\mathbf{x}_i$, and m is the number of training examples. The max operator in the objective function can be transformed into the following quadratic objective function:

$$\lambda\cdot||\mathbf{w}||_0+\frac{1}{m}\sum_{i=1}^{m}r_i(1-y_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i-b)), \tag{49}$$

by introducing variable $r_i\in\{0,1\}$ and the following inequality:

$$M(r_i-1)\le 1-y_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i-b)\le Mr_i \tag{50}$$

where M is a sufficiently large constant.

According to the previous methods, we can transform this problem into a QUBO model, enabling us to leverage quantum computing for SVM training and inference. This extension demonstrates the versatility of our model and its potential to address a wide range of optimization problems beyond those discussed in earlier sections.

5. Theoretical Guarantee of Models under Fixed-Point Representation

In our proposed models, the discretization of signals using Ising spins introduces certain limitations compared to their continuous counterparts. This quantization process can be mathematically modeled as mapping the signals onto a finite alphabet through a quantizer function. It inevitably leads to errors due to the reduction in the signal's value domain. In this section, we analyze the errors introduced by signal quantization in these ℓ_0-based models.

Existing literature has explored the challenges in compressed sensing associated with quantized measurements [59–61]. In conventional models, it is generally assumed that continuous-valued measurements of an unknown signal are available. In real-world applications, these measurements need to be converted into a limited bit representation for purposes such as transmission, storage, and processing. In extreme cases, such as one-bit sign measurements [62–64], only the sign of the measurement is retained, discarding all magnitude information. Despite this, several studies have investigated the error characteristics of ℓ_1-based models under quantized measurements and analyzed the performance of associated algorithms [59, 65, 66].

Our analysis in this section focuses specifically on quantization in ℓ_0-based models and evaluating the resulting reconstruction errors. This investigation aims to deepen our understanding of the theoretical guarantee imposed by discrete representations. We first introduce quantizer $Q(\cdot) : \mathbb{R}^n \to \mathbb{R}_Q^n$, where $\mathbb{R}_Q^n$ denotes the domain of the quantizer $Q(\mathbf{x})$, *i.e.*, the set of all possible quantized representations of signals, and its worst case distortion ϵ_Q, which is defined as the supremum of the squared Euclidean distance between $\mathbf{x}$ and its quantized counterpart $Q(\mathbf{x})$ over all possible signals in $\mathbb{R}^n$:

$$\epsilon_Q = \sup_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{x} - Q(\mathbf{x})\|_2^2. \tag{51}$$

For example, when considering binary encoding in Example 1 and 2,

$$\epsilon_Q = K \cdot \bar{x}^2 \cdot 2^{-2K},$$

where K is the number of bits used in the encoding.

We can choose $Q(\cdot)$ to be a sparsity-preserving quantizer, which ensures nonzero elements remain nonzero after quantization, while zero elements remain zero. In other words, the quantizer only alters nonzero elements in a way that its nonzero nature is retained. For example,

$$Q(\mathbf{x})_i = \begin{cases} \frac{\lceil 2^K \cdot x_i \rceil}{2^K}, x \geq 0 \\ \frac{\lfloor 2^K \cdot x_i \rfloor}{2^K}, x < 0 \end{cases}, \forall 1 \leq i \leq n.$$

When the signal $\hat{\mathbf{x}}$ is reconstructed by solving the least square problem among sparse quantized vectors, *i.e.*, under a sparsity constraint $\mathcal{S}$:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}\in\mathbb{R}_Q^n \cap \mathcal{S}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2, \tag{52}$$

we have the following conclusion.

Theorem 1. *For sparsity constrained problems (4)-(5) and (7)-(8), or more generally problem (52),*

$$\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2^2 \leq \frac{1 + \delta_k(\mathbf{A})}{1 - \delta_{2k}(\mathbf{A})} \epsilon_Q. \tag{53}$$

Proof. As $Q(\cdot)$ preserves sparsity, $Q(\mathbf{x}^*)$ is also a feasible solution to (52). Using the optimality of $\hat{\mathbf{x}}$, we can establish the following inequality:

$$\|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}\|_2^2 \leq \|\mathbf{y} - \mathbf{A}Q(\mathbf{x}^*)\|_2^2, \tag{54}$$

Using the fact that $\mathbf{y} = \mathbf{A}\mathbf{x}^*$, the left-hand side of (54) satisfies that

$$\|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}\|_2^2 = \|\mathbf{A}(\mathbf{x}^* - \hat{\mathbf{x}})\|_2^2 \geq (1 - \delta_{2k}(\mathbf{A}))\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2^2. \tag{55}$$

The inequality above follows from the following two facts:

- measurement matrix $\mathbf{A}$ satisfies the restricted isometry property (RIP), with constant $\delta_k(\mathbf{A}) \in [0, 1]$, *i.e.*, for all $\mathbf{v}$ that are k-sparse,

$$(1 - \delta_k(\mathbf{A}))\|\mathbf{v}\|_2^2 \leq \|\mathbf{A}\mathbf{v}\|_2^2 \leq (1 + \delta_k(\mathbf{A}))\|\mathbf{v}\|_2^2.$$

- $||\hat{\mathbf{x}} - \mathbf{x}^*||_0 \leq 2k$.

For the right-hand side of (54), we have:

$$\begin{aligned} \|\mathbf{y} - \mathbf{A}Q(\mathbf{x}^*)\|_2^2 &\leq \|\mathbf{A}(\mathbf{x}^* - Q(\mathbf{x}^*))\|_2^2 && (56)\\ &\leq (1 + \delta_{\mathbf{A}})\|\mathbf{x}^* - Q(\mathbf{x}^*)\|_2^2 && (57)\\ &\leq (1 + \delta_{\mathbf{A}})\epsilon_Q. && (58) \end{aligned}$$

Inequality (57) is based on the fact that the quantization error $\mathbf{x} - Q(\mathbf{x})$ is k-sparse for any k-sparse vector $\mathbf{x}$ and sparsity preserving $Q(\cdot)$.

Combining these inequalities, we obtain the following bound on the error between the reconstructed signal and the optimal signal:

$$\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2^2 \leq \frac{1 + \delta_k(\mathbf{A})}{1 - \delta_{2k}(\mathbf{A})}\epsilon_Q. \tag{59}$$

□

Theorem 2. *By setting* $\lambda \leq \frac{(1-\delta_k(\mathbf{A}))\epsilon_Q}{k}$ *in the QUBO formulation (3), or equivalently, setting a sufficiently large penalty coefficient for reconstruction error term,*

$$||\hat{\mathbf{x}} - \mathbf{x}^*||_2^2 \leq \frac{(3 + \delta_k(\mathbf{A})) \cdot \epsilon_Q}{1 - \delta_{2k}(\mathbf{A})} + \epsilon_Q.$$

Proof. Let $\hat{\mathbf{x}}$ be the optimal quantized solution to the ℓ_0 regularized least square problem, *i.e.*,

$$\hat{\mathbf{x}} = \underset{\mathbf{x} \in \mathbb{R}_Q^n}{\operatorname{argmin}} f(\mathbf{x}) = ||\mathbf{A}\mathbf{x} - \mathbf{y}||_2^2 + \lambda||\mathbf{x}||_0. \tag{60}$$

Note that

$$||\mathbf{A}Q(\mathbf{x}^*) - \mathbf{y}||_2^2 = ||\mathbf{A}(Q(\mathbf{x}^*) - \mathbf{x}^*)||_2^2 \leq (1 + \delta_k(\mathbf{A}))||Q(\mathbf{x}^*) - \mathbf{x}^*||_2^2 \leq (1 + \delta_k(\mathbf{A}))\epsilon,$$

and thus $f(Q(\mathbf{x}^*)) \leq (1+\delta_k(\mathbf{A}))\epsilon_Q + \lambda k \leq 2\epsilon_Q$. We next claim that $||\mathbf{A}\hat{\mathbf{x}} - \mathbf{y}||_2^2 \leq 2\epsilon_Q$, otherwise

$$f(\hat{\mathbf{x}}) \geq ||\mathbf{A}\hat{\mathbf{x}} - \mathbf{y}||_2^2 > 2\epsilon_Q \geq f(Q(\mathbf{x}^*)),$$

which contradicts the fact that $\hat{\mathbf{x}}$ is the optimal solution to (60).

Furthermore, we have

$$||\hat{\mathbf{x}} - Q(\mathbf{x}^*)||_2^2 \leq \frac{||\mathbf{A}(\hat{\mathbf{x}} - Q(\mathbf{x}^*))||_2^2}{1-\delta_{2k}(\mathbf{A})} = \frac{||(\mathbf{A}\hat{\mathbf{x}} - \mathbf{y}) - (\mathbf{A}Q(\mathbf{x}^*) - \mathbf{y})||_2^2}{1-\delta_{2k}(\mathbf{A})} \leq \frac{(3+\delta_k(\mathbf{A}))\cdot\epsilon_Q}{1-\delta_{2k}(\mathbf{A})},$$

and consequently

$$||\hat{\mathbf{x}} - \mathbf{x}^*||_2^2 \leq ||Q(\mathbf{x}^*) - \mathbf{x}^*||_2^2 + ||\hat{\mathbf{x}} - Q(\mathbf{x}^*)||_2^2 = \frac{(3+\delta_k(\mathbf{A}))\cdot\epsilon_Q}{1-\delta_{2k}(\mathbf{A})} + \epsilon_Q.$$

□

For all the three problems considered in this paper, the error induced by the quantization, which represents the distance to the optimal solution, i.e., $\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2^2$, is in the order of $O(\epsilon_Q)$. For the quantization scheme in Example 1 and 2, the number of bits K, required for coding the real-valued signal, scales as $O\left(\log 1/\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2^2\right)$.

Our worst-case results above also apply for randomly distributed signals $\mathbf{x}$. In addition, when random signals are quantized into binary representations, the quantization process introduces a form of observation noise that is inherently dependent on the signal itself, where measurements $\mathbf{y} = \mathbf{A}(\mathbf{x} + \mathbf{n}_{ob})$, $\mathbf{n}_{ob}$ denotes the observation noise introduced before sampling. This is different from measurement noise that is introduced after sampling [67]. In this context, the recovery of sparse signals from quantized signals can be viewed as a compressive sensing problem with signal-dependent observation noise.

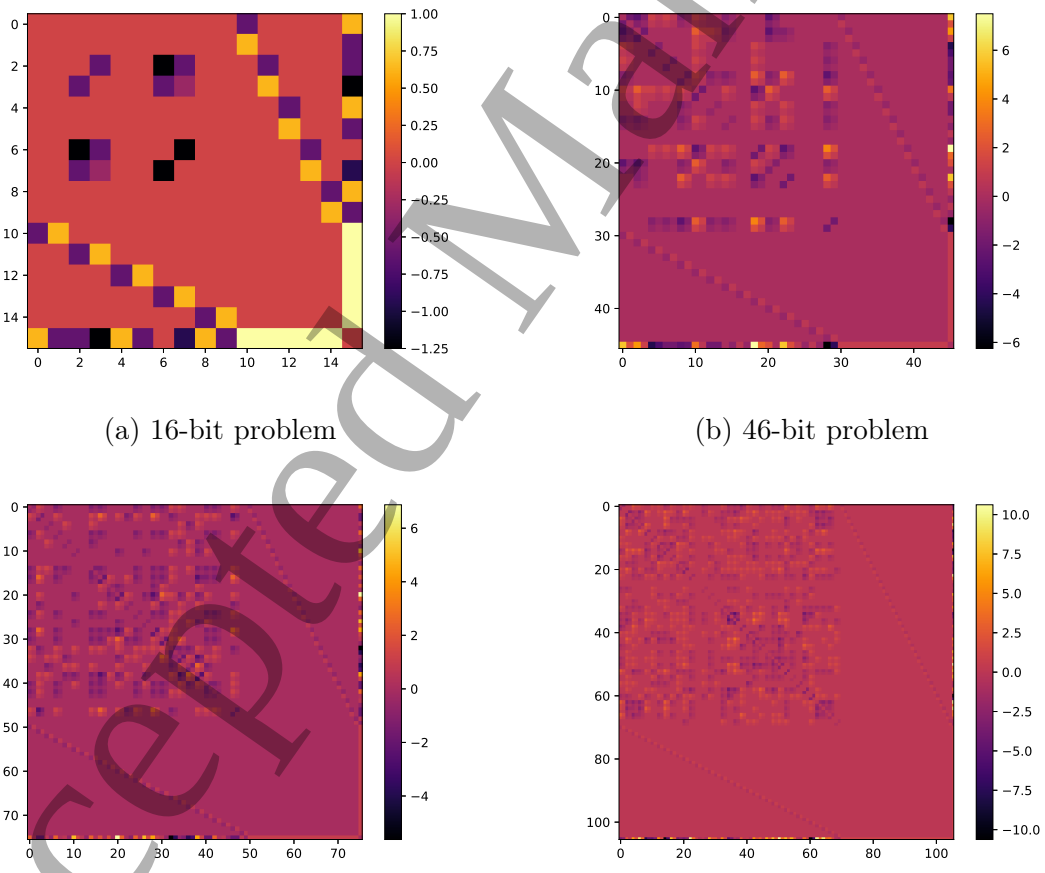
Existing literature has studied several aspects of compressive sensing under noisy observations and measurements. Reeves et al. [67] shows that observation noise has less impact on signal reconstruction compared to sampling noise or using low sampling rates. Kipnis et al. [68] examined the relationship in distribution between the compressed form of a high-dimensional signal $\mathbf{x}$ obtained via a random spherical code and the observation of $\mathbf{x}$ affected by additive white Gaussian noise (AWGN). Understanding the fundamental limits [69] and theoretical guarantees [70] of random signals quantization in compressed sensing, with the additional presence of observation and measurement noise, is a potential future research direction.

6. Experiment Results

The experiment is designed based on model 2.1 in Section 3.3. We performed four experiments with varying bit lengths of 16, 46, 76, and 106 Ising spins, respectively.

6.1. Data generation procedures

Firstly, we generate the measurement matrices with 0 and ±1 entries randomly and independently for each experiment, following a Bernoulli distribution. The choice of a Bernoulli distribution [71] demonstrates that with exponentially high probability, matrices generated in this manner satisfy the restricted isometry property (RIP) property with the number of measurements nearly linear in the sparsity level. To assess the effectiveness of our model, we need a ground truth. Hence, we randomly generate a set of true signal values with a predetermined number of non-zero entries, which are set to 1, 2, 3, and 4, respectively. In the presence of both the measurement matrices and the true signal values, we proceed to compute the measurement values. The key step entails the utilization of these measurement values in conjunction with the measurement matrix to reconstruct the original true signal values via CIM.



(a) 16-bit problem

(b) 46-bit problem

(c) 76-bit problem

(d) 106-bit problem

Figure 3: Heatmaps of Ising matrices for the four problems. The Heatmaps exhibit a spectrum of hues, signifying a varied landscape of spin interactions. With the augmentation of the bit count, the heatmap patterns evolve to a higher density, thereby unveiling intricate motifs that are less discernible in the context of smaller problem instances.

Based on the data and transformations, we can derive the corresponding QUBO and Ising models for the optimization problem. The Ising model is another representation of the same optimization problem and is related to the QUBO model through a simple transformation. The heatmaps shown in Fig. 3 represent the Ising matrices, *i.e.*, the coefficient matrix of the Ising model, for the four different problem instances with varying numbers of Ising spins. The intensity of the colors in the heatmaps reflects the magnitude of the coefficients in the Ising matrices, with darker colors indicating lower values. The visualization of Ising matrices provides insights into the complexity and interconnections of the qubits in optimization problems.

6.2. Methodology

In our experimental setup, we conducted a comparative analysis involving CIM quantum hardware and five classical algorithms: Orthogonal Matching Pursuit (OMP) [72], Basis Pursuit De-Noising (BPDN) [73], and Least Absolute Shrinkage and Selection Operator (LASSO) [74], which are classic algorithms within the field of compressed sensing, as well as SA and Tabu Search, two classical algorithms employed for solving QUBO problems. In the following, we give a brief introduction to the SA and Tabu algorithms.

- Simulated Annealing [75] (SA): The simulated annealing algorithm is a stochastic optimization technique inspired by the physical process of annealing in metallurgy. It aims to solve combinatorial optimization problems by iteratively exploring the solution space while gradually decreasing the exploration intensity. At each iteration, the algorithm considers a neighboring solution and probabilistically accepts it, allowing for the exploration of both local and global optima. The acceptance probability is determined by the difference in the objective function values and a "temperature" parameter, which controls the degree of randomness in the exploration process. As the algorithm progresses, the temperature is gradually reduced, leading to a more deterministic search towards the optimal solution. Simulated annealing can escape from local optima and its convergence properties make it a valuable tool for tackling complex optimization problems in various domains.
- Tabu Search [76]: Tabu search is a metaheuristic optimization algorithm employed to solve combinatorial and discrete optimization problems. Operating within a guided search framework, it iteratively explores the solution space by maintaining a tabu list that records recently visited solutions and associated moves, preventing immediate revisits to encourage diversification. Additionally, the search incorporates intensification by leveraging aspiration criteria to override tabu status for promising solutions. These dynamics enhance the algorithm's ability to escape local optima and converge towards near-optimal solutions. Tabu search offers a robust and effective strategy for addressing complex optimization challenges through its adaptive and memory-driven approach.

Our coherent Ising machine (CIM) [77] is a quantum computing system that employs laser pulses in optical fibers as its fundamental computational units, referred to as qubits. As shown in Figure 4, we utilize a 1560-nm mode-locked fiber laser with a 100 MHz repetition rate and a 100 fs pulse width. This laser generates 780-nm pulses through second harmonic generation (SHG), which are then used to synchronously pump a periodically poled lithium niobate (PPLN) bulk crystal within a fiber-ring cavity. The fiber loop includes several key components: a coupler for injecting feedback signal pulses, another coupler for outputting degenerate optical parametric oscillator (DOPO) pulses for homodyne measurement, and a fiber stretcher to mitigate interference from the external environment. The system also includes EDFA (Erbium-Doped Fiber Amplifier) for signal amplification, BHD (Balanced Homodyne Detection) for precise pulse measurement, AD (Analog-to-Digital) conversion for signal digitization, DA (Data Acquisition) for data collection, PM (Phase Modulator) for phase control, and LO (Local Oscillator) for stable reference signals.

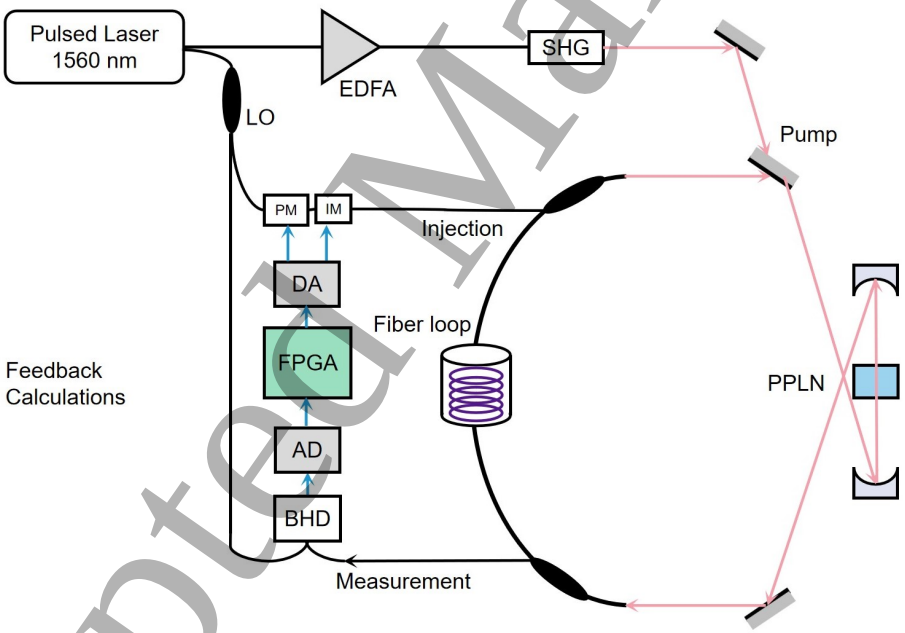


Figure 4: Schematic diagram of a coherent Ising machine, illustrating the integrated system for simulating quantum dynamics and its components including the pulsed laser input, EDFA, FPGA-based feedback computation, and measurement units.

The DOPO network comprises 211 optical pulses, each representing a spin of the coherent Ising machine. Following the analog-to-digital conversion, these pulses are sent into a field-programmable gate array (FPGA). Given the Ising matrix, the FPGA can calculate the feedback injection amplitude of each pulse in each round.

The measurement and feedback scheme begins with a homodyne detection stage, where the in-phase component of each OPO pulse is measured. This measurement captures the amplitude and phase information of the light pulses, which are critical for

determining the state of each simulated spin. The measured data is then processed by FPGA, which computes the feedback signal based on the predefined coupling matrix. The computed feedback signal is subsequently used to modulate the phase and intensity of local oscillator pulses. These modulated pulses are injected back into the optical cavity, effectively coupling the OPO pulses according to the desired interactions. This closed-loop feedback system ensures that the optical pulses are dynamically adjusted to reflect the evolving state of the simulated spin system.

In the developed CIM, our approach involves transforming the problem into a corresponding Maximum Cut (Max-Cut) problem for solving. It has been demonstrated in the literature that finding the ground state of the Ising model without an applied magnetic field can be reformulated as a Max-Cut problem. Indeed, for Ising formulation

$$\sigma^* = \underset{\sigma}{\operatorname{argmin}} -\sum_{i,j} J_{i,j}\sigma_i\sigma_j - \sum_i h_i\sigma_i, \tag{61}$$

we can add an auxiliary spin variable s_a to convert the linear terms in (61) to quadratic terms as following,

$$(\bar{\sigma}^*, \sigma_a^*) = \underset{\sigma,\sigma_a}{\operatorname{argmin}} -\sum_{i,j} J_{i,j}\sigma_i\sigma_j - \sum_i h_i\sigma_i\sigma_a, \tag{62}$$

and we can obtain the original optimal σ^* via equation $\sigma^* = \bar{\sigma}^* \cdot \sigma_a^*$. Take the 76-bit problem as an example, as shown in Fig 8, we perform the solving of the 76-bit max-cut problem in CIM. The numerous edges connecting the nodes in the graph illustrate the intricate coupling relationships among the variables, highlighting the complexity of our problem.

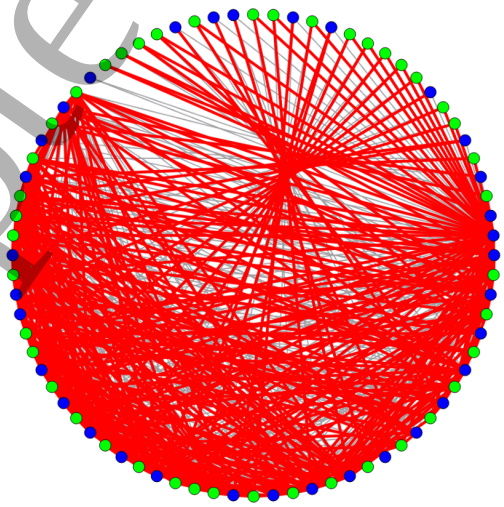


Figure 5: Solution for 76-bit Max-cut problem. The nodes are divided into two categories: one set of nodes is colored with blue, representing the spin values of $+1$, while the other set is colored with green, representing spin values of -1.

6.3. Analysis of the results

In Figure 6, we present a single Hamiltonian evolution of solutions obtained using CIM. From these figures, it is clearly discernible that within a fleeting time frame of several milliseconds, specifically, not surpassing 3.39 milliseconds, CIM converges to the ground energy state.

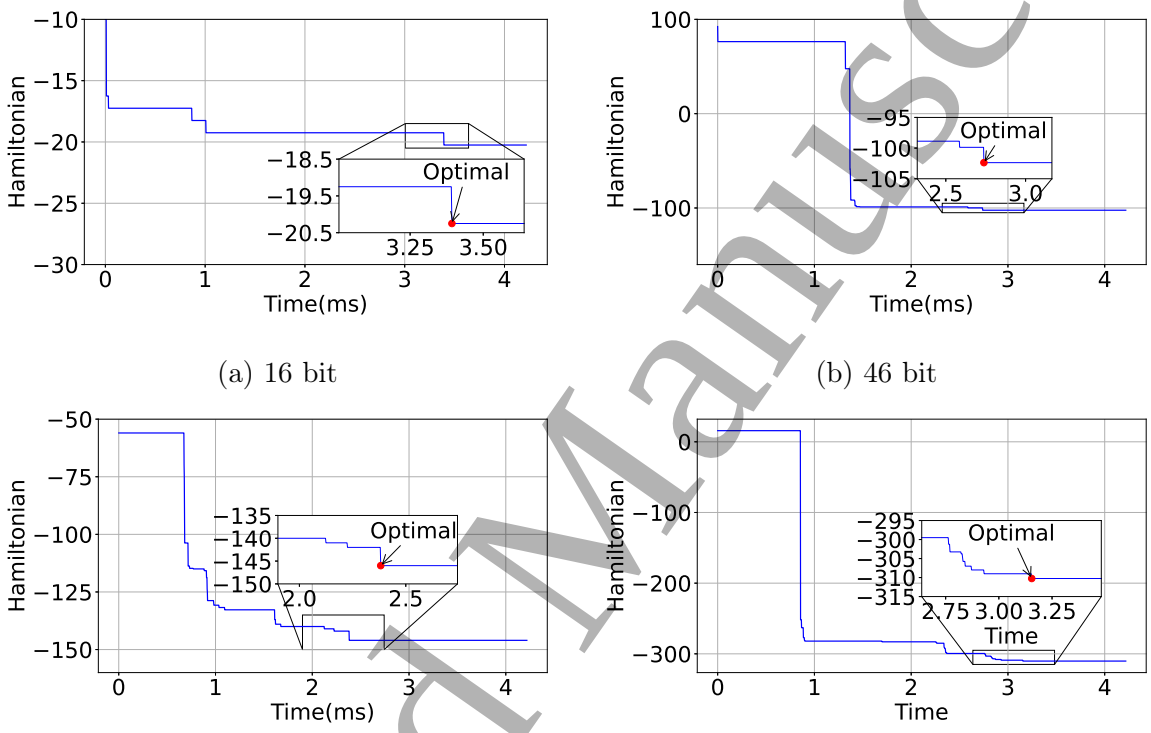


Figure 6: A single evolution of Hamiltonian values over time for different system sizes, based on the best single shot. The main panels show the dynamics of the Hamiltonian (y-axis) as a function of time in milliseconds (x-axis) for four distinct bit lengths: (a) 16-bit, (b) 46-bit, (c) 76-bit, and (d) 106-bit systems. The red dots indicate instances of optimal energy states. The inset in the panel is a zoomed-in view of a specific region of interest, highlighting the fluctuations in detail in the Hamiltonian's trajectory. The graphs illustrate the varying complexity and convergence to optimal states with increasing system size.

In the CIM's evolution curve, the time interval between adjacent data points is 2.11 microseconds. Indeed, within the fiber loop, there were 211 oscillating pulses spaced 10 nanoseconds apart between each pair of pulses. This precise arrangement results in a transmission time of 2.11 microseconds for the optical pulses to traverse the loop. As we observe, the Hamiltonian undergoes a gradual descent over time, and as the pump power increases to the critical threshold for oscillation, a profound phase transition unfolds.

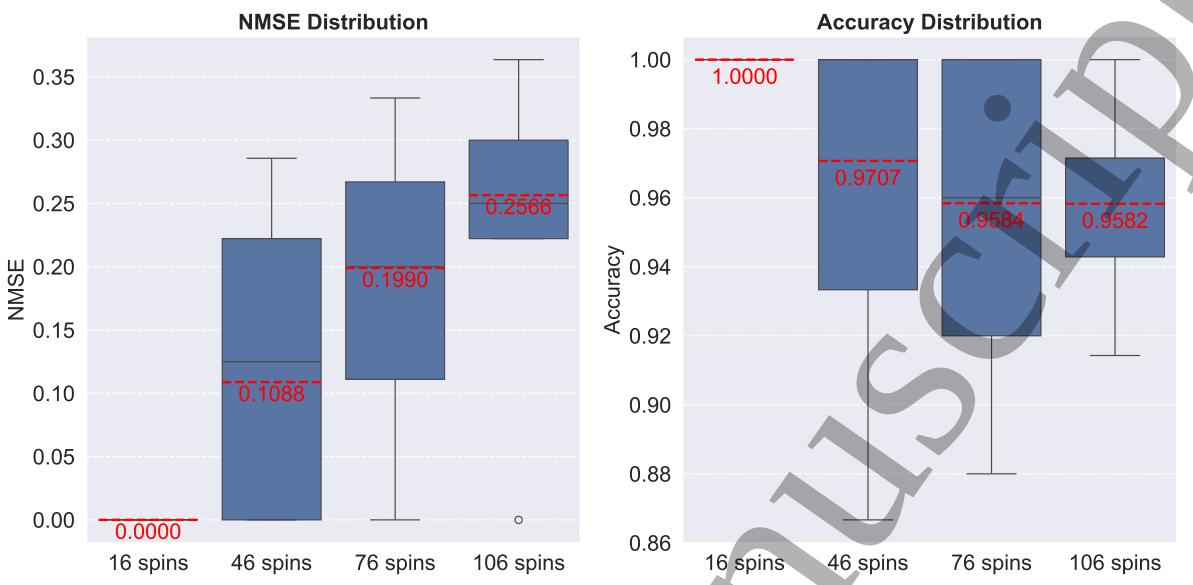


Figure 7: Boxplot distributions of Normalized Mean Squared Error (NMSE) and Accuracy for CIM across different problem sizes. The red dashed lines indicate the mean values for each group. CIM demonstrates excellent performance with perfect accuracy (100%) and zero NMSE for 16 spins, while maintaining high accuracy and low NMSE as the problem size increases, showcasing its scalability and robustness in solving sparse optimization problems.

This set of results demonstrates the efficiency of the optimization process in approaching the optimal Hamiltonian value within a short time frame for different bit sizes.

Table 2 presents the Normalized Mean Squared Error (NMSE) and accuracy of solutions obtained by various algorithms. The definitions for NMSE and accuracy are as follows:

$$\text{NMSE} = \frac{||\mathbf{x}_{\text{true}} - \mathbf{x}||^2}{||\mathbf{x}||^2}, \tag{63}$$

$$\text{Accuracy} = \frac{|\{i|x_{\text{true},i} = 0, x_i = 0\}| + |\{i|x_{\text{true},i} \neq 0, x_i \neq 0\}|}{n} \tag{64}$$

As shown in the table, it is evident that CIM can accurately recover the signal values with an NMSE of 0 and 100% accuracy over different problem sizes. We also validate the superior performance of CIM through statistical analysis of multiple experimental runs. Figure 7 presents the boxplot distributions of NMSE and accuracy for CIM with different problem sizes.

Experiments	**Algorithm**	**NMSE**	**Accuracy**
$m = 2, n = 5, k = 1$	OMP	0.0000	100.00%
	LASSO	0.0008	100.00%
	BPDN	0.0064	80.00%
	SA_{short}	0.3320	80.00%
	SA_{long}	0.1700	87.20%
	$\text{Tabu}_{\text{short}}$	0.5879	75.60%
	$\text{Tabu}_{\text{long}}$	0.2930	76.80%
	CIM	0.0000	100.00%
$m = 5, n = 15, k = 2$	OMP	1.6000	86.67%
	LASSO	1.3525	86.67%
	BPDN	0.4009	33.33%
	SA_{short}	0.8630	60.40%
	SA_{long}	0.4109	90.53%
	$\text{Tabu}_{\text{short}}$	0.8958	65.73%
	$\text{Tabu}_{\text{long}}$	0.8184	68.53%
	CIM	0.0000	100.00%
$m = 5, n = 25, k = 3$	OMP	0.0000	100.00%
	LASSO	1.1137	84.00%
	BPDN	0.0043	88.00%
	SA_{short}	1.0475	50.08%
	SA_{long}	1.0726	74.72%
	$\text{Tabu}_{\text{short}}$	1.0858	60.64%
	$\text{Tabu}_{\text{long}}$	0.9912	61.44%
	CIM	0.0000	100.00%
$m = 7, n = 35, k = 4$	OMP	0.2737	94.29%
	LASSO	0.7348	94.29%
	BPDN	0.2705	14.29%
	SA_{short}	1.0111	44.80%
	SA_{long}	0.9856	68.51%
	$\text{Tabu}_{\text{short}}$	1.0100	59.20%
	$\text{Tabu}_{\text{long}}$	0.9858	58.06%
	CIM	0.0000	100.00%

Table 2: Performance comparison between CIM and classical algorithms. For both SA and Tabu, we tested two different runtimes. The runtime of SA_{short} is approximately 0.16s, with the four experiments taking 0.1531s, 0.1570s, 0.1616s, 0.1611s, respectively. As for SA_{long}, its runtime is around 15s, and the individual experiment times are 14.96s, 15.09s, 15.78s, 15.72s. $\text{Tabu}_{\text{short}}$ demonstrates a runtime close to 0.1s, with corresponding times of 0.0756s, 0.0813s, 0.0917s, 0.1276s. Lastly, $\text{Tabu}_{\text{long}}$ runs for about 0.8s, and the four running time are 0.5717s, 0.7129s, 0.8149s, 1.0518s. It can be observed that increasing the runtime improves the performance of both SA and Tabu.

The results demonstrate the CIM's robust performance in solving combinatorial optimization problems. For smaller problem sizes, such as 16 spins, the CIM achieves perfect accuracy (100%) with an NMSE of 0, indicating highly precise solutions. Through the increases of the size, the NMSE exhibits a slight upward trend, with means of 0.1088, 0.1990, and 0.2566 for 46, 76, and 106 spins, respectively, while maintaining a high mean accuracy of 97.07%, 95.84%, and 95.82%. These results highlight the scalability and competitive advantage of the current work over classical algorithms, as further evidenced in Table 2, where CIM almost consistently outperforms traditional methods.

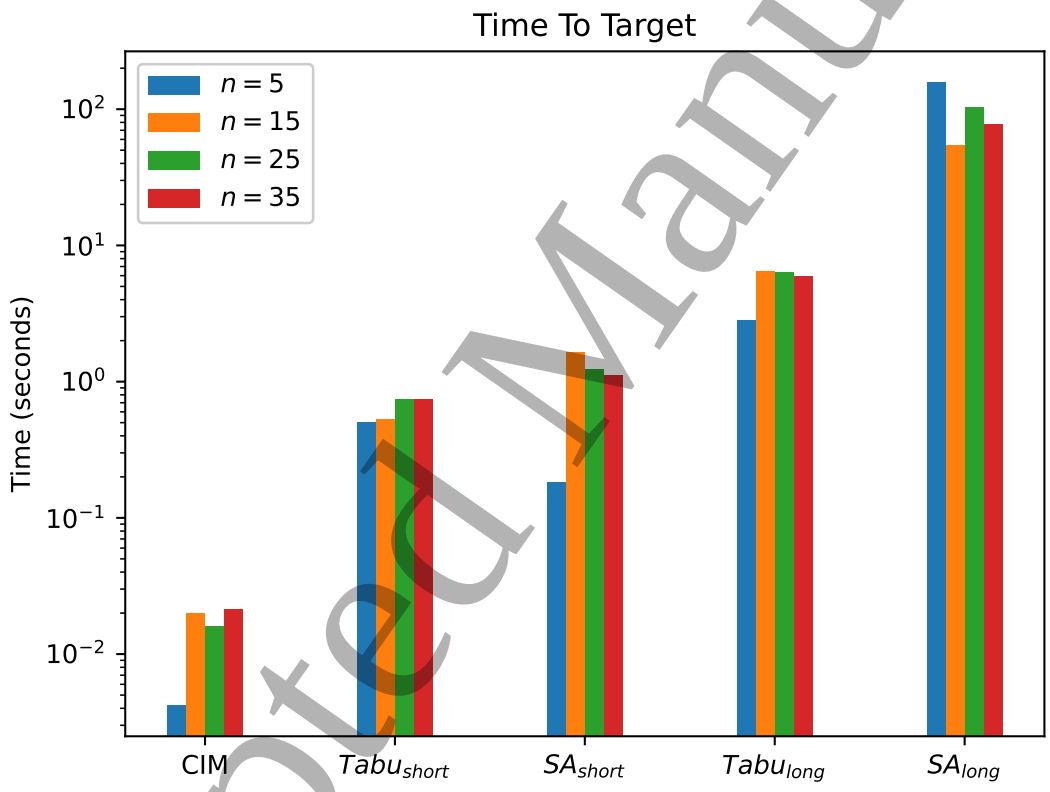


Figure 8: Time To Target (TTT) comparison of several approaches. The performance of CIM, Tabu, and Simulated Annealing (SA) is evaluated under different problem sizes, represented by the colors: blue ($n = 5$), orange ($n = 15$), green ($n = 25$), and red ($n = 35$).

For those approaches with random output solutions, including CIM, SA, and Tabu search, we evaluate the time required to achieve a solution with sufficient quality at a high probability. Let p_0 be the constant of desired success probability, here we calculate

the following metric of *time to target (TTT)* [78, 79],

$$\tau \cdot \frac{\log(1-p_0)}{\log(1-p_\tau)},$$

where τ represents the duration of a single shot, and p_τ denotes the probability of achieving the target in a single shot.

In our experiments, we set the expected accuracy of the three approaches as its target value, and p_τ is calculated accordingly. For CIM, each shot consists of 2000 loops, and the time per loop is 2.11 microseconds, so the total time τ is 4.22 milliseconds. To be rigorous, we set τ as the longest running time among the calls that exceed the specified target for both SA and Tabu. Since the parameters for SA and Tabu are fixed in each experiment, the runtime for these two algorithms remains relatively stable, with minimal variation. Thus, using this method to define τ for SA and Tabu will not exaggerate their TTT values. In the experiment, we set the desired probability p_0 to 0.99.

From Table 2, CIM consistently outperforms both SA and Tabu in terms of average accuracy and average NMSE, achieving accuracy levels above 95.8%. Despite this, as shown in Figure 8, among the three approaches, CIM demonstrates a lower Time To Target (TTT) compared to both SA and Tabu, indicating that CIM can reach superior accuracy in a shorter amount of time, with high probability.

6.4. Hybrid Algorithm

In this experiment, we compare the complexity of the quantum Benders decomposition method with direct QUBO modeling for solving sparse optimization problems. In direct QUBO modeling, the total number of binary variables grows as least $n(K+1)$, making the model increasingly complex and difficult to solve as K increases. By iteratively refining upper and lower bounds, the Benders decomposition simulated effectively narrows the solution space, as depicted in the figures. The gray regions in the plots represent the gap between the bounds, which consistently decreases over iterations, highlighting the algorithm's convergence properties. Note that there are $2n$ binary variables and one real variable, and the number of constraints for the RMP problem is no more than the number of iterations. This leads to fewer qubits being required when solving the RMP using quantum hardware devices, especially when using the quantum conditional gradient method [17] for constrained programming problem variables, compared to directly encoding the original QUBO model with $n(K+1)$ binary variables.

7. Discussions

A limitation shared by hardware-based solvers, including CIM, is their constrained parameter precision [80]. While the QUBO framework theoretically operates with continuous parameters, practical implementations on physical devices are inherently restricted by finite numerical representations. As a result, the precision of the encoded

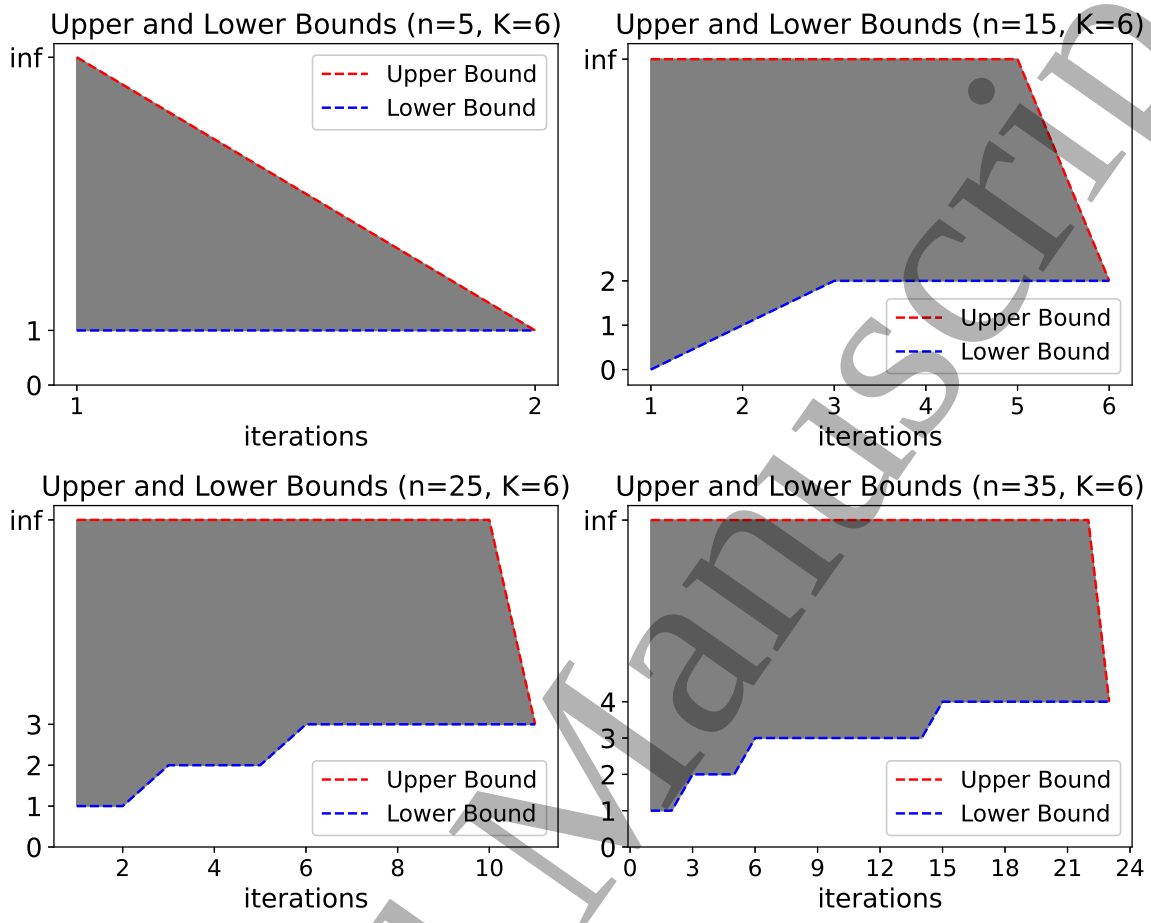


Figure 9: Convergence of upper and lower bounds using the Benders decomposition method for sparse optimization. Subplots correspond to the problem with the signal length of $n = 5, 15, 25$ and 35, with $K = 6$ denoting the number of Ising spins required if we encode each real variable in binary variables.

parameters is inherently limited, potentially leading to approximations that may affect the accuracy of the solutions.

Despite these practical limitations, recent theoretical advancements have provided a deeper understanding of CIMs. Cheng et al. [81] analyzed the stability and convergence properties of oscillator-based Ising systems, and established conditions for the stability of equilibrium points. These results offer an understanding of how CIMs can reliably solve combinatorial optimization problems. Pramanik et al. [82] provides the first rigorous theoretical analysis of the convergence properties of the optoelectronic oscillator (OEO)-based CIMs, demonstrating that they are not merely heuristic but can be analytically understood. By establishing bounds on the expected difference between the objective value and the optimal solution, OEO-CIM can converge to a region around the optimum under reasonable assumptions. Although the OEO-CIM analyzed in [82] is not identical to the CIM used in our study, their theoretical analysis offers valuable insights into

the convergence properties of CIMs in general. Further exploration of these theoretical guarantees in the context of our specific CIM implementation could be an interesting direction for future research.

8. Conclusion

In this paper, we explore the interplay between the concept of sparsity and the cutting-edge field of quantum computing using CIM. Our contributions focus on the development of efficient models capable of addressing multiple sparsity-related optimization problems. We have also introduced a hybrid quantum-classical approach to further enhance computational efficiency and bridge the gap between quantum and classical paradigms. In addition, we have demonstrated the practical applicability of CIM in sparse signal recovery through real-world experiments. We believe that the fusion of these two domains opens doors to more efficient and powerful solutions with far-reaching implications in various fields of information processing.

Data availability statement

All data that support the findings of this study are included in the article.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grants No. 62371050 and No.62461160263).

Conflict of interest

The authors declare that they have no conflict of interest.

Appendix A. Derivation of the QUBO model 2.1

The Hamiltonian of the model is

$$\begin{aligned}\mathcal{H} =& ||\mathbf{Ax}-\mathbf{y}||_2^2+\lambda\cdot||\mathbf{x}||_0+\mathcal{H}_0\\ =& \underbrace{\mathbf{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{Ax}-2\mathbf{y}^{\mathrm{T}}\mathbf{Ax}}_{\mathcal{H}_1}+\underbrace{\lambda\cdot||\mathbf{x}||_0+\mathcal{H}_0}_{\mathcal{H}_2}+||\mathbf{y}||_2^2\end{aligned}$$

By substituting equation (30) into this, we have

$$\begin{aligned}\mathcal{H}_1 =&\mathbf{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{x} - 2\mathbf{y}^{\mathrm{T}}\mathbf{A}\mathbf{x}\\ =&(\boldsymbol{\delta}^{\mathrm{T}}(\mathbf{I}_n\otimes\mathbf{w}) + w_0\mathbf{1}_n^{\mathrm{T}})\mathbf{A}^{\mathrm{T}}\mathbf{A}((\mathbf{I}_n\otimes\mathbf{w}^{\mathrm{T}})\boldsymbol{\delta} + w_0\mathbf{1}_n)\\ &- 2\mathbf{y}^{\mathrm{T}}\mathbf{A}((\mathbf{I}_n\otimes\mathbf{w}^{\mathrm{T}})\boldsymbol{\delta} + w_0\mathbf{1}_n)\\ =&\boldsymbol{\delta}^{\mathrm{T}}(\mathbf{I}_n\otimes\mathbf{w})\mathbf{A}^{\mathrm{T}}\mathbf{A}(\mathbf{I}_n\otimes\mathbf{w}^{\mathrm{T}})\boldsymbol{\delta}\\ &+ w_0\boldsymbol{\delta}^{\mathrm{T}}(\mathbf{I}_n\otimes\mathbf{w})\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{1}_n + w_0\mathbf{1}_n^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}(\mathbf{I}_n\otimes\mathbf{w}^{\mathrm{T}})\boldsymbol{\delta}\\ &- 2\mathbf{y}^{\mathrm{T}}\mathbf{A}(\mathbf{I}_n\otimes\mathbf{w}^{\mathrm{T}})\boldsymbol{\delta} + w_0^2\mathbf{1}_n^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{1}_n - 2w_0\mathbf{y}^{\mathrm{T}}\mathbf{A}\mathbf{1}_n\\ =&\boldsymbol{\delta}^{\mathrm{T}}(\mathbf{A}^{\mathrm{T}}\mathbf{A}\otimes\mathbf{W})\boldsymbol{\delta} + 2[w_0\mathbf{1}_n^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}(\mathbf{I}_n\otimes\mathbf{w}^{\mathrm{T}})\\ &- \mathbf{y}^{\mathrm{T}}\mathbf{A}(\mathbf{I}_n\otimes\mathbf{w}^{\mathrm{T}})]\boldsymbol{\delta} + w_0^2\mathbf{1}_n^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{1}_n - 2w_0\mathbf{y}^{\mathrm{T}}\mathbf{A}\mathbf{1}_n.\end{aligned}$$

In the last equation, we use the fact that

$$(\mathbf{I}_n\otimes\mathbf{w})\mathbf{Q}(\mathbf{I}_n\otimes\mathbf{w}^{\mathrm{T}}) = \mathbf{Q}\otimes\mathbf{W},$$

where $\mathbf{W} = \mathbf{w}\mathbf{w}^{\mathrm{T}}$. Neglecting constant terms, the Hamiltonian $\mathcal{H}_1$ can be equivalently expressed by minimizing the following term, utilizing the fact that

$$\begin{aligned}&[w_0\mathbf{1}_n^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}(\mathbf{I}_n\otimes\mathbf{w}^{\mathrm{T}}) - \mathbf{y}^{\mathrm{T}}\mathbf{A}(\mathbf{I}_n\otimes\mathbf{w}^{\mathrm{T}})]\boldsymbol{\delta}\\ &= \boldsymbol{\delta}^{\mathrm{T}}\mathrm{Diag}(w_0\mathbf{1}_n^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}(\mathbf{I}_n\otimes\mathbf{w}^{\mathrm{T}}) - \mathbf{y}^{\mathrm{T}}\mathbf{A}(\mathbf{I}_n\otimes\mathbf{w}^{\mathrm{T}}))]\boldsymbol{\delta}\end{aligned}$$

holds for binary vector $\boldsymbol{\delta}$:

$$\mathcal{H}_1 \equiv\boldsymbol{\delta}^{\mathrm{T}}[\mathbf{A}^{\mathrm{T}}\mathbf{A}\otimes\mathbf{W} + 2\mathrm{Diag}(w_0\mathbf{1}_n^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}(\mathbf{I}_n\otimes\mathbf{w}^{\mathrm{T}}) - \mathbf{y}^{\mathrm{T}}\mathbf{A}(\mathbf{I}_n\otimes\mathbf{w}^{\mathrm{T}}))]\boldsymbol{\delta}.$$

Let

$$\mathbf{Q}_1 =\mathbf{A}^{\mathrm{T}}\mathbf{A}\otimes\mathbf{W} + 2\mathrm{Diag}(w_0\mathbf{1}_n^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}(\mathbf{I}_n\otimes\mathbf{w}^{\mathrm{T}}) - \mathbf{y}^{\mathrm{T}}\mathbf{A}(\mathbf{I}_n\otimes\mathbf{w}^{\mathrm{T}})),$$

then

$$\mathcal{H}_1 = \mathbf{p}^{\mathrm{T}}\begin{bmatrix}\mathbf{Q}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{p}.$$

For the Hamiltonian associated with sparsity, we have the following result,

$$\begin{aligned}\mathcal{H}_2 =&\lambda\cdot\sum_{i=1}^{n} z_i + \lambda\cdot\sum_{i=1}^{n}\Big(\sum_{j\in\mathcal{I}_0}(\delta_{ij} - z_i\delta_{ij}) + \sum_{j\in\mathcal{I}_1}(1-z_i)(1-\delta_{ij})\Big)\\ =&\lambda\cdot\Big(\sum_{i\in\mathcal{I}_0} z_i + \sum_{i=1}^{n}\sum_{j\in\mathcal{I}_0}\delta_{ij} - \sum_{i=1}^{n}\sum_{j\in\mathcal{I}_1}\delta_{ij}\Big)\\ &+\lambda\cdot\Big(\sum_{i=1}^{n}\sum_{j\in\mathcal{I}_1} z_i\delta_{ij} - \sum_{i=1}^{n}\sum_{j\in\mathcal{I}_0} z_i\delta_{ij}\Big) + \lambda\cdot|\mathcal{I}_1|\\ =&\mathbf{L}_0^{\mathrm{T}}\mathbf{p} + \mathbf{p}^{\mathrm{T}}\mathbf{Q}_0\mathbf{p}\\ =&\mathbf{p}^{\mathrm{T}}(\mathbf{Q}_0 + \mathrm{Diag}(\mathbf{L}_0))\mathbf{p}\end{aligned}$$

where $\mathbf{L}_0$ and $\mathbf{Q}_0$ consist of λ, $-\lambda$ and 0. The last equation holds because $\mathbf{p}$ is a binary vector. Hence

$$\mathcal{H} = \mathbf{p}^{\mathrm{T}}\mathbf{Q}_2\mathbf{p}, \tag{A.1}$$

where

$$\mathbf{Q}_2 = \mathbf{Q}_0 + \mathrm{Diag}(\mathbf{L}_0) + \begin{bmatrix} \mathbf{Q}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

References

[1] L. Stanković, Ervin Sejdić, Srdjan S. Stanković, Miloš Daković, and Irena Orović. A tutorial on sparse signal reconstruction and its applications in signal processing. *Circuits, Systems, and Signal Processing*, 38:1206–1263, 2018.

[2] Joaquin Zepeda, Christine M. Guillemot, and Ewa Kijak. Image compression using sparse representations and the iteration-tuned and aligned dictionary. *IEEE Journal of Selected Topics in Signal Processing*, 5:1061–1073, 2011.

[3] Michael E Tipping. Sparse bayesian learning and the relevance vector machine. *Journal of machine learning research*, 1(Jun):211–244, 2001.

[4] Hyekyoung Lee, Dong Soo Lee, Hyejin Kang, Boong-Nyun Kim, and Moo K. Chung. Sparse brain network recovery under compressed sensing. *IEEE Transactions on Medical Imaging*, 30:1154–1165, 2011.

[5] Zhen Gao, Linglong Dai, Shuangfeng Han, Chih lin I, Zhaocheng Wang, and Lajos Hanzo. Compressive sensing techniques for next-generation wireless communications. *IEEE Wireless Communications*, 25:144–153, 2017.

[6] Balas Kausik Natarajan. Sparse approximate solutions to linear systems. *SIAM journal on computing*, 24(2):227–234, 1995.

[7] Yagyensh C. Pati, Ramin Rezaiifar, and Perinkulam S. Krishnaprasad. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, pages 40–44 vol.1, 1993.

[8] Yoshiyuki Kabashima, Tadashi Wadayama, and Toshiyuki Tanaka. A typical reconstruction limit for compressed sensing based on lp-norm minimization. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(09):L09003, 2009.

[9] Zhe Wang, Alireza Marandi, Kai Wen, Robert L Byer, and Yoshihisa Yamamoto. Coherent ising machine based on degenerate optical parametric oscillators. *Physical Review A*, 88(6):063853, 2013.

[10] Alireza Marandi, Zhe Wang, Kenta Takata, Robert L Byer, and Yoshihisa Yamamoto. Network of time-multiplexed optical parametric oscillators as a coherent ising machine. *Nature Photonics*, 8(12):937–942, 2014.

[11] Peter L McMahon et al. A fully programmable 100-spin coherent ising machine with all-to-all connections. *Science*, 354(6312):614–617, 2016.

[12] Takahiro Inagaki, Yoshitaka Haribara, Koji Igarashi, Tomohiro Sonobe, Shuhei Tamate, Toshimori Honjo, Alireza Marandi, Peter L McMahon, Takeshi Umeki, Koji Enbutsu, et al. A coherent ising machine for 2000-node optimization problems. *Science*, 354(6312):603–606, 2016.

[13] Toshimori Honjo, Tomohiro Sonobe, Kensuke Inaba, Takahiro Inagaki, Takuya Ikuta, Yasuhiro Yamada, Takushi Kazama, Koji Enbutsu, Takeshi Umeki, Ryoichi Kasahara, et al. 100,000-spin coherent ising machine. *Science advances*, 7(40), 2021.

[14] Ryan Hamerly, Takahiro Inagaki, Peter L McMahon, Davide Venturelli, Alireza Marandi, Tatsuhiro Onodera, Edwin Ng, Carsten Langrock, Kensuke Inaba, Toshimori Honjo, et al. Experimental

investigation of performance differences between coherent ising machines and a quantum annealer. *Science advances*, 5(5):eaau0823, 2019.

[15] Dimitris Bertsimas, Jean Pauphilet, and Bart Van Parys. Sparse regression: Scalable algorithms and empirical performance. 2020.

[16] Peter J Karalekas, Nikolas A Tezak, Eric C Peterson, Colm A Ryan, Marcus P Da Silva, and Robert S Smith. A quantum-classical cloud platform optimized for variational hybrid algorithms. *Quantum Science and Technology*, 5(2):024003, 2020.

[17] Alp Yurtsever, Tolga Birdal, and Vladislav Golyanik. Q-fw: A hybrid classical-quantum frank-wolfe for quadratic binary optimization. In *European Conference on Computer Vision (ECCV)*, pages 352–369, 2022.

[18] Anh-Dzung Doan, Michele Sasdelli, David Suter, and Tat-Jun Chin. A hybrid quantum-classical algorithm for robust fitting. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 417–427, 2022.

[19] Zhongqi Zhao, Lei Fan, and Zhu Han. Hybrid quantum benders' decomposition for mixed-integer linear programming. In *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 2536–2540, 2022.

[20] Xinliang Wei, Lei Fan, Yuanxiong Guo, Yanmin Gong, Zhu Han, and Yu Wang. Hybrid quantum-classical benders' decomposition for federated learning scheduling in distributed networks. *IEEE Transactions on Network Science and Engineering*, 2024.

[21] M Yassine Naghmouchi and Wesley da Silva Coelho. Mixed-integer linear programming solver using benders decomposition assisted by a neutral-atom quantum processor. *Physical Review A*, 110(1):012434, 2024.

[22] Toru Aonishi, Kazushi Mimura, Masato Okada, and Yoshihisa Yamamoto. L0 regularization-based compressed sensing with quantum–classical hybrid approach. *Quantum Science and Technology*, 7(3):035013, 2022.

[23] Naoki Ide and Masayuki Ohzeki. Sparse signal reconstruction with qubo formulation in l0-regularized linear regression. *arXiv preprint arXiv:2202.00452*, 2022.

[24] Yaniv Romano, Harel Primack, Talya Vaknin, Idan Meirzada, Ilan Karpas, Dov Furman, Chene Tradonsky, and Ruti Ben Shlomi. Quantum sparse coding. *Quantum Machine Intelligence*, 6(1):4, 2024.

[25] Justin A Ziniel. *Message passing approaches to compressive inference under structured signal priors*. The Ohio State University, 2014.

[26] David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

[27] Joel A Tropp and Anna C Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on information theory*, 53(12):4655–4666, 2007.

[28] David L Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences (PNAS)*, 106(45):18914–18919, 2009.

[29] Mastiyage Don Sudeera Hasaranga Gunathilaka, Satoshi Kako, Yoshitaka Inui, Kazushi Mimura, Masato Okada, Yoshihisa Yamamoto, and Toru Aonishi. Effective implementation of l 0-regularised compressed sensing with chaotic-amplitude-controlled coherent ising machines. *Scientific Reports*, 13(1):16140, 2023.

[30] Toru Aonishi, Tatsuya Nagasawa, Toshiyuki Koizumi, Mastiyage Don Sudeera Hasaranga Gunathilaka, Kazushi Mimura, Masato Okada, Satoshi Kako, and Yoshihisa Yamamoto. Highly versatile fpga-implemented cyber coherent ising machine. *arXiv:2406.05377*.

[31] Mastiyage Don Sudeera Hasaranga Gunathilaka, Yoshitaka Inui, Satoshi Kako, Kazushi Mimura, Masato Okada, Yoshihisa Yamamoto, and Toru Aonishi. L0-regularized compressed sensing with mean-field coherent ising machines. *arXiv:2405.00366*.

[32] Kazue Kudo. Compressed sensing based on qubo formulation. In *Journal of Physics: Conference Series*, volume 2207, page 012033. IOP Publishing, 2022.

[33] Robert S Wezeman, Irina Chiscop, Laura Anitori, and Wim van Rossum. Quantum-classical solution methods for binary compressive sensing problems. In *International Conference on Computational Science*, pages 107–121, 2022.

[34] Bo Lu, Yong-Pan Gao, Kai Wen, and Chuan Wang. Combinatorial optimization solving by coherent ising machines based on spiking neural networks. *Quantum*, 7:1151, 2023.

[35] Bo Lu, Chen-Rui Fan, Lu Liu, Kai Wen, and Chuan Wang. Speed-up coherent ising machine with a spiking neural network. *Optics Express*, 31(3):3676–3684, 2023.

[36] Bo Lu, Lu Liu, Jun-Yang Song, Kai Wen, and Chuan Wang. Recent progress on coherent computation based on quantum squeezing. *AAPPS Bulletin*, 33(1):7, 2023.

[37] Minsung Kim, Davide Venturelli, and Kyle Jamieson. Leveraging quantum annealing for large mimo processing in centralized radio access networks. In *Proceedings of the ACM special interest group on data communication*, pages 241–255. 2019.

[38] Minsung Kim, Salvatore Mandrà, Davide Venturelli, and Kyle Jamieson. Physics-inspired heuristics for soft mimo detection in 5g new radio and beyond. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking (MobiCom)*, pages 42–55, 2021.

[39] Abhishek Kumar Singh, Kyle Jamieson, Peter L McMahon, and Davide Venturelli. Ising machines' dynamics and regularization for near-optimal mimo detection. *IEEE Transactions on Wireless Communications*, 21(12):11080–11094, 2022.

[40] Abhishek Kumar Singh, Davide Venturelli, and Kyle Jamieson. Perturbation-based formulation of maximum likelihood mimo detection for coherent ising machines. In *2022 IEEE Global Communications Conference (GLOBECOM 2022)*, pages 2523–2528, 2022.

[41] Srikar Kasi, John Kaewelh, and Kyle Jamieson. A quantum annealer-enabled decoder and hardware topology for nextg wireless polar codes. *IEEE Transactions on Wireless Communications*, 2023.

[42] Yuhong Huang, Wenxin Li, Chengkang Pan, Shuai Hou, Xian Lu, Chunfeng Cui, Jingwei Wen, Jiaqi Xu, Chongyu Cao, Yin Ma, et al. Quantum computing for mimo beam selection problem: Model and optical experimental solution. In *2023 IEEE Global Communications Conference (GLOBECOM 2023)*, pages 5463–5468, 2023.

[43] Jingwei Wen, Zhenming Wang, Zhiguo Huang, Dunbo Cai, Bingjie Jia, Chongyu Cao, Yin Ma, Hai Wei, Kai Wen, and Ling Qian. Optical experimental solution for the multiway number partitioning problem and its application to computing power scheduling. *Science China Physics, Mechanics & Astronomy*, 66(9):290313, 2023.

[44] Jinyin Zha, Jiaqi Su, Tiange Li, Chongyu Cao, Yin Ma, Hai Wei, Zhiguo Huang, Ling Qian, Kai Wen, and Jian Zhang. Encoding molecular docking for quantum computers. *Journal of Chemical Theory and Computation*, 19(24):9018–9024, 2023.

[45] Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2):489–509, 2006.

[46] Sohail Bahmani, Bhiksha Raj, and Petros T Boufounos. Greedy sparsity-constrained optimization. *The Journal of Machine Learning Research*, 14(1):807–841, 2013.

[47] Junde Li and Swaroop Ghosh. Quantum-soft qubo suppression for accurate object detection. In *European Conference on Computer Vision (ECCV)*, pages 158–173, 2020.

[48] Tolga Birdal, Vladislav Golyanik, Christian Theobalt, and Leonidas J Guibas. Quantum permutation synchronization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13122–13133, 2021.

[49] Marcel Seelbach Benkner, Zorah Lähner, Vladislav Golyanik, Christof Wunderlich, Christian Theobalt, and Michael Moeller. Q-match: Iterative shape matching via quantum annealing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7586–7596, 2021.

[50] Jan-Nico Zaech, Alexander Liniger, Martin Danelljan, Dengxin Dai, and Luc Van Gool. Adiabatic

quantum computing for multi object tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8811–8822, 2022.

[51] Federica Arrigoni, Willi Menapace, Marcel Seelbach Benkner, Elisa Ricci, and Vladislav Golyanik. Quantum motion segmentation. In *European Conference on Computer Vision (ECCV)*, pages 506–523, 2022.

[52] Harshil Bhatia, Edith Tretschk, Zorah Lähner, Marcel Seelbach Benkner, Michael Moeller, Christian Theobalt, and Vladislav Golyanik. Ccuantumm: Cycle-consistent quantum-hybrid matching of multiple shapes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1296–1305, 2023.

[53] Riccardo Nembrini, Maurizio Ferrari Dacrema, and Paolo Cremonesi. Feature selection for recommender systems with quantum computing. *Entropy*, 23(8):970, 2021.

[54] Maurizio Ferrari Dacrema, Fabio Moroni, Riccardo Nembrini, Nicola Ferro, Guglielmo Faggioli, and Paolo Cremonesi. Towards feature selection for ranking and classification exploiting quantum annealers. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2814–2824, 2022.

[55] Eldan Cohen, Hayato Ushijima-Mwesigwa, Avradip Mandal, and Arnab Roy. Unified clustering and outlier detection on specialized hardware. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3770–3774, 2021.

[56] Prasanna Date, Davis Arthur, and Lauren Pusey-Nazzaro. Qubo formulations for training machine learning models. *Scientific reports*, 11(1):10029, 2021.

[57] Davide Venturelli and Alexei Kondratyev. Reverse quantum annealing approach to portfolio optimization problems. *Quantum Machine Intelligence*, 1(1-2):17–30, 2019.

[58] Avradip Mandal, Arnab Roy, Sarvagya Upadhyay, and Hayato Ushijima-Mwesigwa. Compressed quadratization of higher order binary optimization problems. In *2020 Data Compression Conference (DCC)*, pages 383–383, 2020.

[59] Argyrios Zymnis, Stephen Boyd, and Emmanuel Candes. Compressed sensing with quantized measurements. *IEEE Signal Processing Letters*, 17(2):149–152, 2009.

[60] Hans Christian Jung, Johannes Maly, Lars Palzer, and Alexander Stollenwerk. Quantized compressed sensing by rectified linear units. *IEEE transactions on information theory*, 67(6):4125–4149, 2021.

[61] Xiangming Meng and Yoshiyuki Kabashima. Quantized compressed sensing with score-based generative models. In *2023 The Eleventh International Conference on Learning Representations*.

[62] Yaniv Plan and Roman Vershynin. One-bit compressed sensing by linear programming. *Communications on pure and Applied Mathematics*, 66(8):1275–1297, 2013.

[63] Albert Ai, Alex Lapanowski, Yaniv Plan, and Roman Vershynin. One-bit compressed sensing with non-gaussian measurements. *Linear Algebra and its Applications*, 441:222–239, 2014.

[64] Pranjal Awasthi, Maria-Florina Balcan, Nika Haghtalab, and Hongyang Zhang. Learning and 1-bit compressed sensing under asymmetric noise. In *Conference on Learning Theory*, pages 152–192, 2016.

[65] Laurent Jacques, David K Hammond, and Jalal M Fadili. Dequantizing compressed sensing: When oversampling and non-gaussian constraints combine. *IEEE Transactions on Information Theory*, 57(1):559–571, 2010.

[66] Wei Dai and Olgica Milenkovic. Information theoretical and algorithmic approaches to quantized compressive sensing. *IEEE transactions on communications*, 59(7):1857–1866, 2011.

[67] Galen Reeves and Michael Gastpar. Differences between observation and sampling error in sparse signal reconstruction. In *2007 IEEE/SP 14th Workshop on Statistical Signal Processing*, pages 690–694, 2007.

[68] Alon Kipnis and Galen Reeves. Gaussian approximation of quantization error for estimation from compressed data. *IEEE Transactions on Information Theory*, 67(8):5562–5579, 2021.

[69] Galen Reeves. The fundamental limits of stable recovery in compressed sensing. In *2014 IEEE International Symposium on Information Theory*, pages 3017–3021, 2014.

[70] Alon Kipnis, Galen Reeves, Yonina C Eldar, and Andrea J Goldsmith. Compressed sensing under optimal quantization. In *2017 IEEE international symposium on information theory (ISIT)*, pages 2148–2152, 2017.

[71] Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive approximation*, 28:253–263, 2008.

[72] Joel A. Tropp and Anna C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 53:4655–4666, 2007.

[73] Scott Saobing Chen, David L. Donoho, and Michael A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20:33–61, 1998.

[74] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the royal statistical society series b-methodological*, 58:267–288, 1996.

[75] Dimitris Bertsimas and John Tsitsiklis. Simulated annealing. *Statistical science*, 8(1):10–15, 1993.

[76] Fred Glover and Manuel Laguna. *Tabu search*. Springer, 1998.

[77] Qboson. `https://qboson.com`.

[78] James King, Sheir Yarkoni, Mayssam M Nevisi, Jeremy P Hilton, and Catherine C McGeoch. Benchmarking a quantum annealing processor with the time-to-target metric. *arXiv preprint arXiv:1508.05087*.

[79] Hiroki Takesue, Kensuke Inaba, Toshimori Honjo, Yasuhiro Yamada, Takuya Ikuta, Yuya Yonezu, Takahiro Inagaki, Takeshi Umeki, and Ryoichi Kasahara. Finding independent sets in large-scale graphs with a coherent ising machine. *Science Advances*, 11(7):eads7223, 2025.

[80] Sascha Mücke, Thore Gerlach, and Nico Piatkowski. Optimum-preserving qubo parameter compression. *Quantum Machine Intelligence*, 7(1):1–18, 2025.

[81] Yi Cheng, Mohammad Khairul Bashar, Nikhil Shukla, and Zongli Lin. A control theoretic analysis of oscillator ising machines. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 34(7), 2024.

[82] Sayantan Pramanik, Sourav Chatterjee, and Harshkumar Oza. Convergence analysis of opto-electronic oscillator based coherent ising machines. In *2024 16th International Conference on COMmunication Systems & NETworkS (COMSNETS)*, pages 1076–1081, 2024.