Scalable Quantum-Inspired Optimization through Dynamic Qubit Compression

Co Tran [1*] Quoc-Bao Tran [2*] Hy Truong Son [3] Thang N Dinh[2]

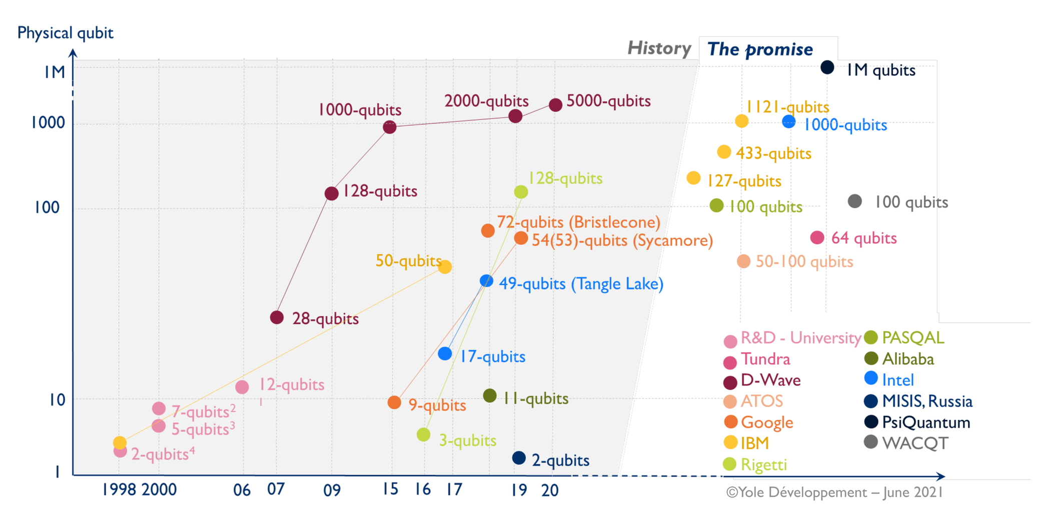
[1]University of Texas at Austin

[2]Department of Computer Science, Virginia Commonwealth University

[3]University of Alabama at Birmingham

*equal contributions

ABSTRACT



- A lack of current number of qubits for practical applications. For example: factoring RSA 2,048-bit numbers requires **20,000,000** physical qubits, far exceeding available hardware (maximum **5,000** qubits, promising **1,000,000** qubits).
- We present a GNN-based framework that **generalize well to Ising of different sizes**, automating the discovery of qubit reduction rules via **qubit alignments**.
- Our extensive testing reveals substantial multi-level size **reductions across various Ising topologies** while **preserving solution quality**.
- By significantly reducing qubit requirements, our approach expands the scope of tractable problems for current quantum devices.

BACKGROUND

Ising models. An Ising model consists of binary variables (spins) $s_i \in -1, +1$, with energy given by the Hamiltonian:

$$H = -\sum_{i,j} J_{ij} s_i s_j - \sum_i h_i s_i \qquad (1)$$

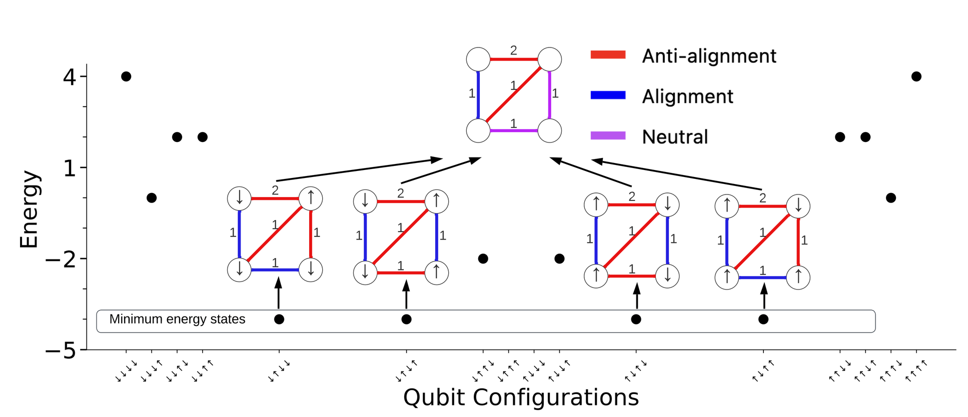
where J_{ij} is the coupling between spins i, j, and h_i is the external field at spin i.

Global minimum energy (ground states) can be found by using **Quantum annealing**:

$$H_{\text{system}}(s) = -\frac{A(s)}{2}\Big(\sum_i^n \sigma_i^x\Big) + \frac{B(s)}{2}\Big(H_{\text{problem}}\Big) \qquad (2)$$

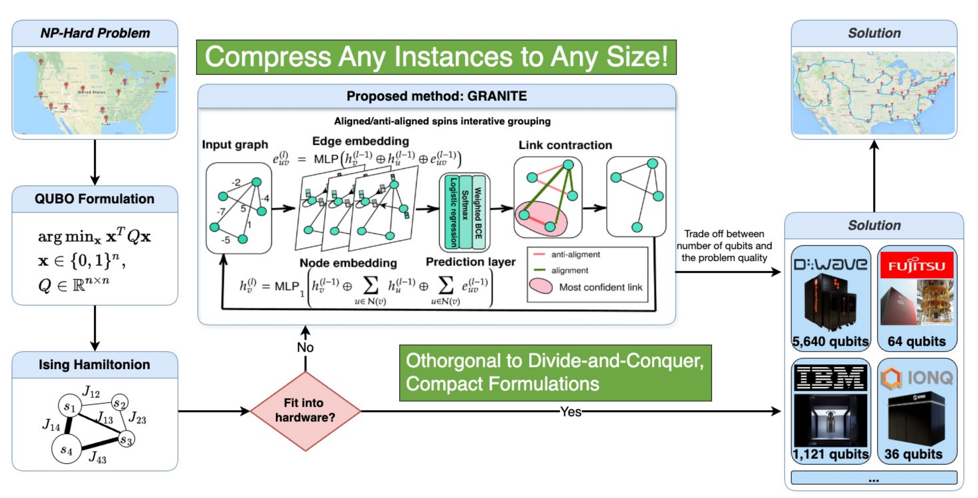
where $s \in [0, 1]$ is the anneal fraction, σ_i^x is the Pauli x-matrix for the i-th qubit, and H_{problem} is the problem Hamiltonian, equivalent to Hamiltonian (1). $A(s)$ and $B(s)$ define the anneal schedule.

Qubits alignment. We classify each edge (i, j) based on the behavior of connected spins across all ground states: alignment (s_i,s_j have same sign), anti-alignment (s_i, s_j have opposite signs), and neutral (s_i,s_j alignment varies).



Methodology

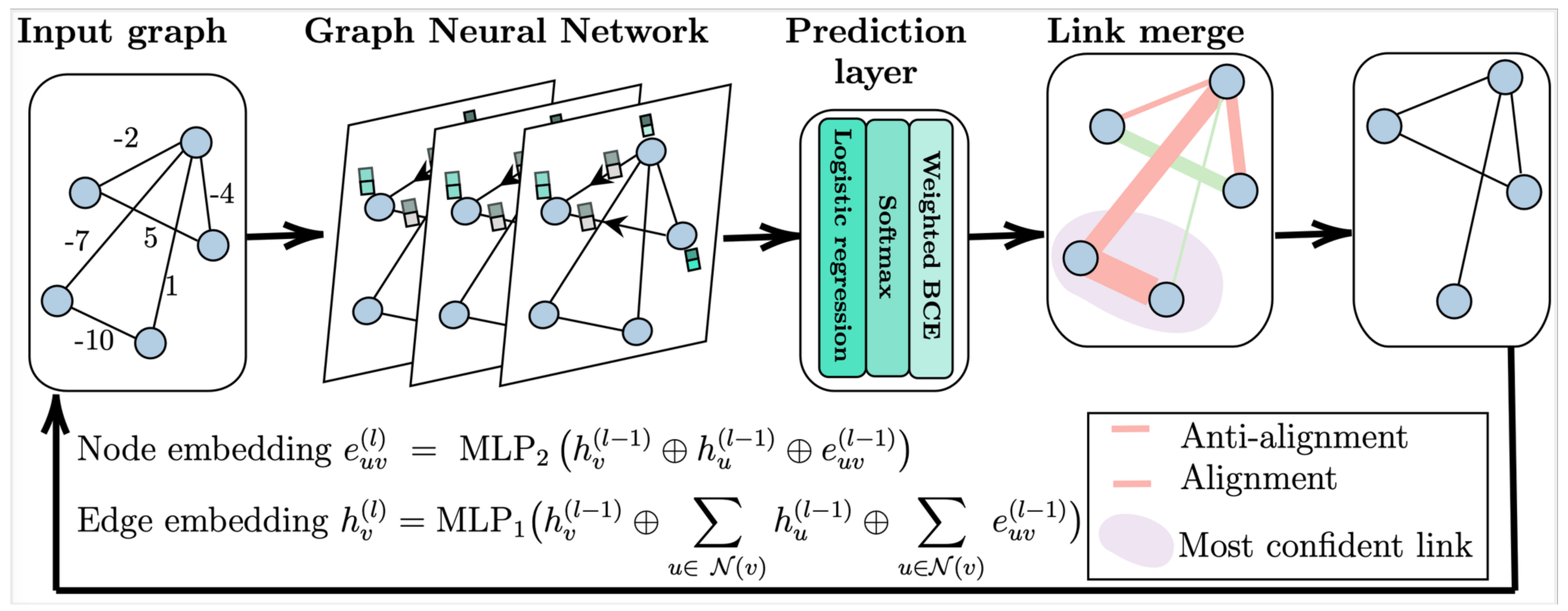
- **Input**: An Ising Hamiltonian H.
- **Output**: A compressed Ising Hamiltonian H' optimized for quantum hardware constraints.
- **Objective**: GNN architecture that generalizes across different Ising model sizes.



GNN Model. We respresent Ising Model with Hamiltonian graph $G_H = (V, E)$ where V is the set of nodes and $E \subseteq V \times V$ is the set of edges. The initial node and edge features are defined as:

$$\mathbf{H}^{(0)} = \{h_v^{(0)} \in \mathbb{R}^{d_h} \mid v \in V\},$$
$$\mathbf{E}^{(0)} = \{e_{uv}^{(0)} \in \mathbb{R}^{d_e} \mid (u, v) \in E\},$$

respectively, in which d_h and d_e are the corresponding numbers of input node and edge features.



At layer ℓ, the message passing scheme updates each node representation based on the neighboring nodes' representations at the previous layer $\ell - 1$; meanwhile, each edge representation is updated based on the edge's two corresponding nodes. Formally, we have:

$$h_v^{(\ell)} = \text{MLP}_1\left(h_v^{(\ell-1)} \oplus \sum_{u \in \mathcal{N}(v)} h_u^{(\ell-1)} \oplus \sum_{u \in \mathcal{N}(v)} e_{v,u}^{(\ell-1)}\right),$$
$$e_{uv}^{(\ell)} = \text{MLP}_2\left(h_u^{(\ell-1)} \oplus h_v^{(\ell-1)} \oplus e_{uv}^{(\ell-1)}\right),$$

Most confident link predict for if each edge (u, v) is aligned or anti-alignment. Let L denote the number of layers of message passing, a feature vector $z_{uv} = h_u^{(L)} \oplus h_v^{(L)} \oplus e_{uv}^{(L)}$.
The logistic regression model predicts the probability on each edge (i.e. edge confidence) as:

$$\hat{y}_{uv} = \sigma(\langle w, z_{uv} \rangle), \qquad (3)$$

We iteratively contract one edge per round to reduce the Ising Hamiltonian. Edges are classified as either alignment or anti-alignment via a weighted binary cross-entropy (BCE) loss, where the weight is the model's *confidence* in the prediction.

- **Negative Entropy (Confidence):**

$$c_i = Softmax\{-H(\hat{y}_i) = \hat{y}_i \log(\hat{y}_i) + (1 - \hat{y}_i)\log(1 - \hat{y}_i),$$

} where $\hat{y}_i \in [0, 1]$ is the probability that edge i is "alignment."

- **Weighted BCE Loss:**

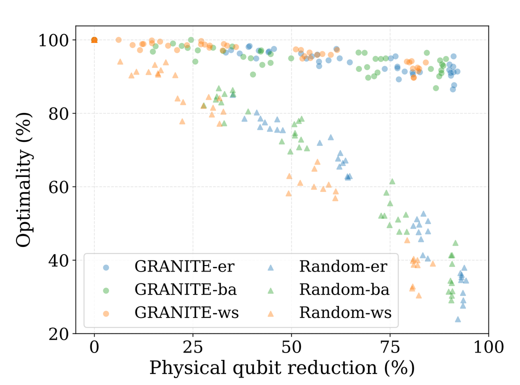
$$\mathcal{L} = \sum_{i=1}^{N} -c_i \Big[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\Big],$$

with $y_i \in \{0, 1\}$ denoting alignment or anti-alignment.

EXPERIMENTS RESULTS

Topology	Reduction (%)	n: 25	50	100	200	400
Erdős-Rényi	0.0% Original Ising	100.00 ± 0.00	100.00 ± 0.00	99.68 ± 0.19	96.77 ± 0.45	NaN
	12.5%	98.21 ± 1.09	98.67 ± 0.44	96.70 ± 0.55	94.87 ± 0.37	89.36 ± 0.37
	25.0%	96.03 ± 1.44	97.12 ± 0.76	94.26 ± 0.56	93.38 ± 0.54	88.58 ± 0.80
	50.0%	94.58 ± 1.64	94.19 ± 1.42	91.23 ± 1.00	91.56 ± 0.63	88.02 ± 0.63
	75.0%	93.20 ± 1.55	92.26 ± 1.40	89.60 ± 0.97	91.07 ± 0.82	88.17 ± 0.79
Barabási-Albert	0.0% Original Ising	100.00 ± 0.00	100.00 ± 0.00	99.93 ± 0.07	97.25 ± 0.25	89.11 ± 0.48
	12.5%	97.69 ± 1.22	97.54 ± 0.91	97.79 ± 0.65	94.71 ± 0.80	88.53 ± 1.11
	25.0%	97.12 ± 1.38	95.63 ± 1.24	97.09 ± 0.75	92.69 ± 1.08	87.84 ± 1.11
	50.0%	94.88 ± 1.59	93.00 ± 1.45	94.87 ± 1.14	92.66 ± 0.92	88.76 ± 0.60
	75.0%	92.43 ± 1.88	91.89 ± 1.21	93.37 ± 1.28	92.30 ± 0.87	88.72 ± 0.96
Watts-Strogatz	0.0% Original Ising	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	99.14 ± 0.14	96.57 ± 0.26
	12.5%	98.06 ± 1.13	98.85 ± 0.52	99.08 ± 0.40	96.73 ± 0.59	91.96 ± 0.72
	25.0%	95.95 ± 1.30	96.86 ± 0.90	97.40 ± 0.52	96.28 ± 0.56	90.73 ± 1.28
	50.0%	93.39 ± 1.08	96.26 ± 1.07	95.13 ± 0.83	94.69 ± 0.41	90.63 ± 0.94
	75.0%	91.69 ± 1.31	94.69 ± 1.22	92.74 ± 1.26	92.23 ± 0.53	91.03 ± 0.88

Table 1: Solution optimality on D-Wave quantum annealers after and before compressing Ising models with GRANITE.



ABLATION STUDY

# Layers	ER	BA	WS
1	76.64 ± 0.86	82.95 ± 0.72	85.73 ± 1.05
2	88.09 ± 0.45	86.47 ± 1.07	91.56 ± 1.03
3	$\mathbf{91.07 \pm 0.82}$	$\mathbf{92.30 \pm 0.87}$	$\mathbf{92.23 \pm 0.53}$
4	81.07 ± 1.27	86.47 ± 0.91	85.79 ± 1.39
5	84.48 ± 1.49	90.72 ± 0.85	90.94 ± 0.63

Table 3: Optimality with different number of GNN layers at n = 200 and reduction rate = 75%.

	BCE	MSE	Hybrid	Softmax
ER	86.90 ± 1.08	75.68 ± 3.85	$\mathbf{91.07 \pm 0.82}$	89.68 ± 1.07
BA	89.05 ± 1.35	92.07 ± 0.69	$\mathbf{92.30 \pm 0.87}$	78.89 ± 2.44
WS	92.58 ± 0.30	$\mathbf{94.02 \pm 0.60}$	92.23 ± 0.53	92.60 ± 0.51

Table 2: Optimality for different loss functions for n=200, edge reduction 75%.

CONCLUSION & FUTURE WORK

- A framework based on GNNs automates the process of finding reduction rules for Ising instances.
- Our framework enables solving larger optimization problems on current quantum devices.
- Extend GRANITE to handle real-world optimization problems from various domains.
- Adapt compression techniques for different quantum computing technologies and devices.

CONTACT

Quoc-Bao Tran, Ph.D. Student in CS, Virginia Commonwealth University, tranq3@vcu.edu
Co Tran, The University of Texas at Austin, cotran2@utexas.edu