

# Neural Zero-Shot Inference for Quantum Optimization: Spectral-Temporal Transformer for FALQON Parameter Prediction

Liyang Pan

January 29, 2026

## Abstract

Feedback-based quantum optimization (FALQON) removes the classical optimization loop of variational quantum algorithms by applying a Lyapunov-inspired feedback control, but its layer-wise measurement procedure results in an  $O(P^2)$  cumulative circuit depth and severe noise accumulation. We propose a "teacher-student" zero-shot inference framework that uses a Spectral-Temporal Transformer to predict the full control parameter trajectory  $\{\beta_t\}_{t=0}^{P-1}$  directly from the graph Laplacian spectrum of the problem instance. Our method uses SignNet to resolve eigenvector sign ambiguity and an autoregressive training schedule with scheduled sampling to mitigate error accumulation at inference time. On a dataset of 1000 random graphs, the model achieves a Pearson correlation of 0.917 on convergent instances, demonstrating the feasibility of neural parameter prediction for quantum control. We also analyze the effect of prediction error on Lyapunov convergence to establish theoretical robustness guarantees.

**Keywords:** quantum optimization, FALQON, Transformer, spectral graph networks, zero-shot inference

## 1 Introduction

In the noisy intermediate-scale quantum (NISQ) era, variational quantum algorithms (VQAs) are considered one of the most promising paths toward practical quantum advantage. The Quantum Approximate Optimization Algorithm (QAOA) [6] has shown strong potential for combinatorial optimization tasks (e.g., MaxCut). However, practical deployment of QAOA faces significant challenges: the classical parameter optimization loop must search a high-dimensional nonconvex energy landscape for optimal parameters  $\gamma^*, \beta^*$ , which is prone to local minima and barren plateaus.

Feedback-based quantum optimization (FALQON) [1] offers an alternative by applying Lyapunov-inspired feedback control laws to determine layer parameters, thereby eliminating the classical optimization loop. Nonetheless, FALQON introduces a different bottleneck: a large measurement overhead. To compute the parameter for layer  $p + 1$ , one must prepare and measure the depth- $p$  state, causing an  $O(P^2)$  cumulative circuit depth.

Paragraph: Contributions

We propose a "teacher-student" zero-shot inference framework that predicts the entire control parameter trajectory  $\{\beta_t\}_{t=0}^{P-1}$  in one shot using a Spectral-Temporal Transformer conditioned on the graph Laplacian spectrum. Our contributions are:

1. **Architecture:** Spectral-Temporal Transformer that combines SignNet-style spectral encoders with a Transformer decoder to model temporal dependencies.
2. **Training:** Scheduled sampling to mitigate train-test mismatch, and a weighted loss emphasizing later time steps.
3. **Empirical Evaluation:** On 1000 random graphs, we categorize samples into convergent and oscillatory dynamics and report strong performance on convergent instances.

4. Theoretical Analysis: Lyapunov-based analysis showing robustness when prediction errors do not flip the control signs.

## 2 Preliminaries

This section introduces the theoretical foundations of FALQON and the spectral representation of graphs.

### 2.1 FALQON

Consider a combinatorial optimization problem encoded by a problem Hamiltonian  $H_P$  and a driver Hamiltonian  $H_D = \sum_{i=1}^n X_i$ . FALQON seeks to minimize the cost function

$$C(t) = \langle \psi(t) | H_P | \psi(t) \rangle. \quad (1)$$

**Theorem 2.1** (FALQON convergence [2]). *Define the feedback control law*

$$\beta(t) = -\alpha \cdot \langle \psi(t) | i[H_D, H_P] | \psi(t) \rangle, \quad \alpha > 0. \quad (2)$$

*Then the cost satisfies  $\frac{dC}{dt} \leq 0$ , i.e., the system energy is non-increasing.*

In discrete implementations, the state evolves as

$$|\psi_{p+1}\rangle = e^{-i\beta_p H_D} e^{-iH_P \Delta t} |\psi_p\rangle, \quad (3)$$

with  $\beta_p = -\alpha \langle \psi_p | i[H_D, H_P] | \psi_p \rangle$ .

### 2.2 Graph Laplacian Spectrum

For an undirected graph  $G = (V, E)$ , the normalized Laplacian is  $L = I - D^{-1/2}AD^{-1/2}$ , with eigendecomposition  $L = U\Lambda U^\top$ . Eigenvalues  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$  capture global connectivity, while eigenvectors  $\{u_i\}$  provide spectral coordinates. Eigenvectors suffer from sign ambiguity; we adopt SignNet-style processing to remove this ambiguity (see Section 3.2.1).

## 3 Methods

This section details the Spectral-Temporal Transformer architecture and the training procedure.

### 3.1 Problem Statement

Given the Laplacian spectrum  $(\Lambda, U)$  of graph  $G$ , our goal is to predict the FALQON parameter sequence  $\beta = (\beta_0, \dots, \beta_{P-1})$ , modeled as a conditional sequence generation problem:

$$p(\beta|G) = \prod_{t=0}^{P-1} p(\beta_t | \beta_{<t}, G). \quad (4)$$

### 3.2 Model Architecture

As shown in Figure 1, the model consists of three modules.

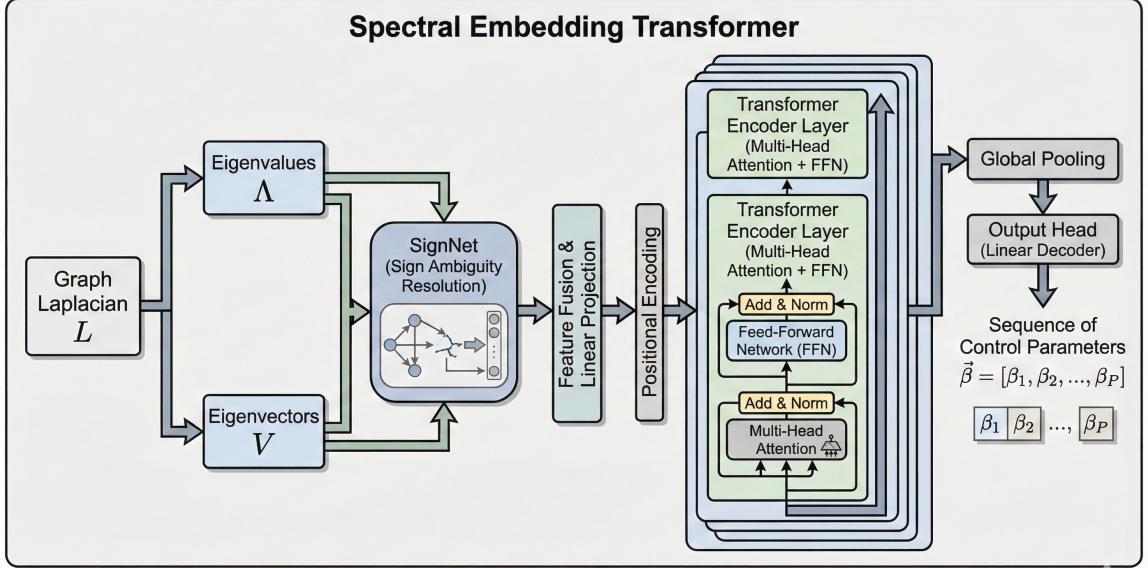


Figure 1. The detailed architecture of the Spectral Embedding Transformer.

Figure 1: Overview of the Spectral-Temporal Transformer architecture.

### 3.2.1 Sign-invariant Spectral Encoder (SignNet)

We remove eigenvector sign ambiguity using a SignNet-style encoding:

$$h_i = \rho(\phi(u_i) + \phi(-u_i)), \quad (5)$$

where  $\phi$  is an MLP and  $\rho$  aggregates.

Eigenvalues are separately encoded and fused:  $m_i = \text{Fusion}(\text{MLP}(\lambda_i), h_i)$ , resulting in  $M$  spectral modes used as memory for the decoder.

### 3.2.2 Temporal Decoder

We use a Transformer decoder [4] where queries are formed by positional encodings and previous-step embeddings:

$$q_t = \text{PE}(t) + \text{Embed}(\beta_{t-1}) + e_{\text{query}}. \quad (6)$$

Cross-attention queries the spectral memory to produce predictions.

### 3.2.3 Output Head

Decoder outputs are mapped to scalars:  $\hat{\beta}_t = \text{MLP}(\text{Decoder}(q_t, M))$ .

## 3.3 Training

We use scheduled sampling [7] to bridge training and inference, and a loss combining weighted MSE, temporal-gradient loss and tail-variance regularization.

## 4 Experiments

### 4.1 Experimental Setup

#### 4.1.1 Dataset

We use 1000 random graphs composed of two families: Erdős-Rényi ( $p=0.6$ ,  $n$  in [6,13]) and random 3-regular graphs. Each sample contains a FALQON parameter sequence of length  $P = 40$  generated

by a classical simulator with  $\alpha = 1.0$ . Data are split 90/10 train/test.

#### 4.1.2 Metrics

We report Pearson correlation (Corr), MAE, and RMSE.

#### 4.1.3 Sample Categorization

We classify samples as convergent if  $\text{Var}(\beta_{P/2:P}) \leq 0.1$ , otherwise oscillatory.

### 4.2 Main Results

Table 1 summarizes performance.

Table 1: Test performance of the Spectral-Temporal Transformer

Type	Share	MAE ↓	Corr ↑	Best/Worst Corr
Convergent	30%	0.215	<b>0.917</b>	0.997 / 0.804
Oscillatory	70%	0.458	0.801	0.946 / 0.616
<b>Overall</b>	100%	0.314	<b>0.885</b>	—

### 4.3 Qualitative Analysis

Figure 2 shows four representative predictions.

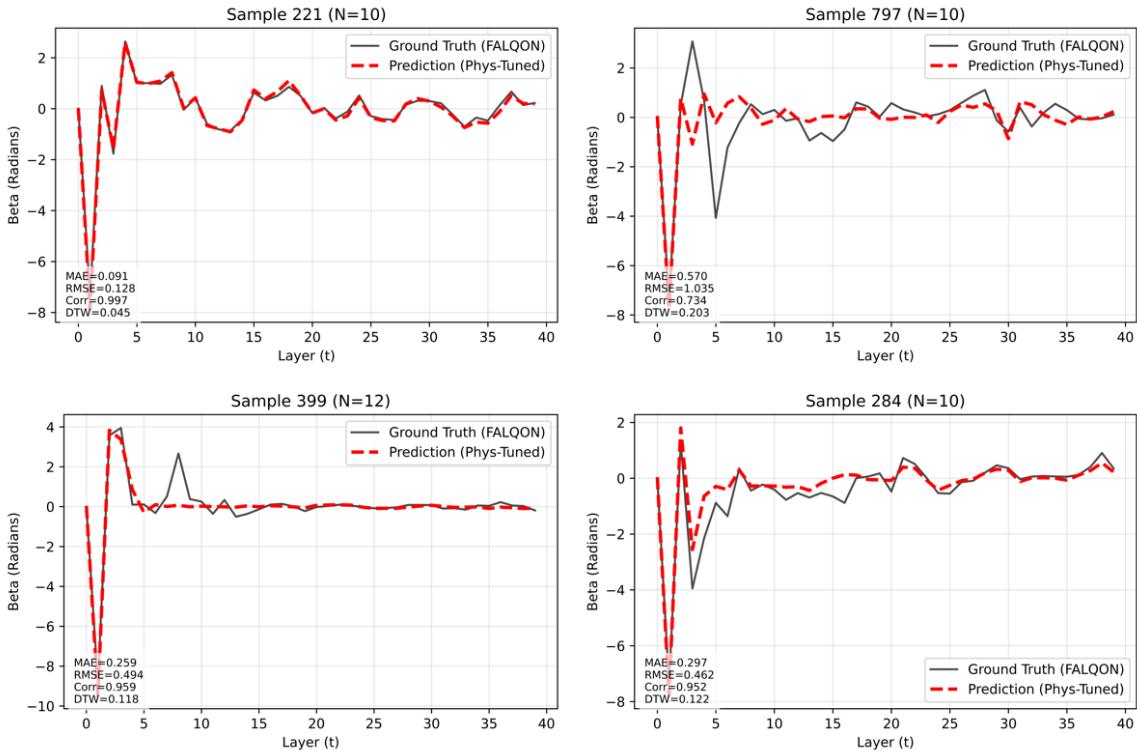


Figure 2: Case study: four representative test samples (2x2 montage).

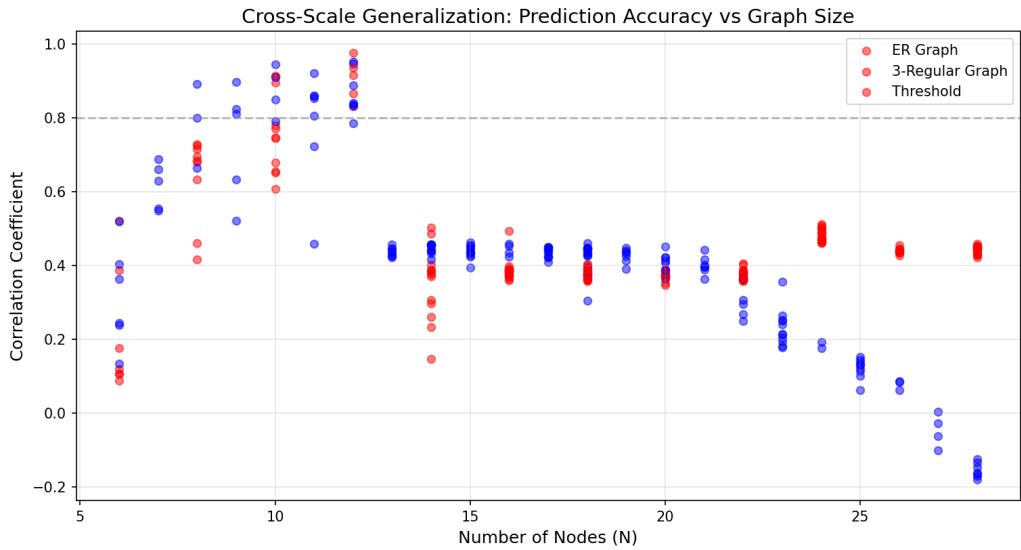


Figure 3: Prediction correlation (Corr) vs. graph size  $N$ . Each point is a test sample; colors indicate graph type (ER or 3-regular).

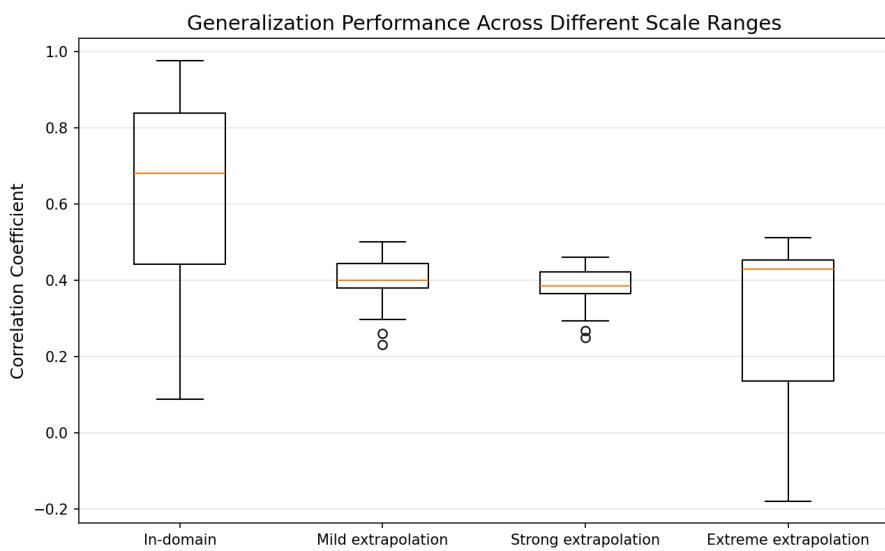


Figure 4: Distribution of correlation coefficients for different scale ranges (In-domain, Mild/Strong/Extreme extrapolation). Boxes show median and IQR.

## 4.4 Cross-scale Generalization Results

## 4.5 Noise Robustness Results

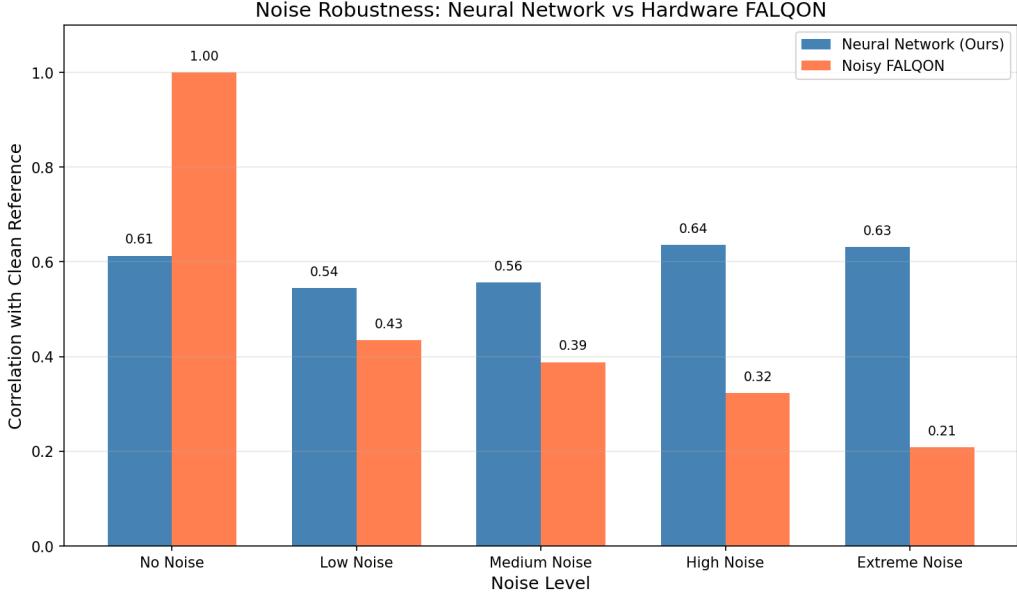


Figure 5: Comparison of correlation with the clean reference across noise levels: Neural Network (ours) vs. Noisy FALQON.

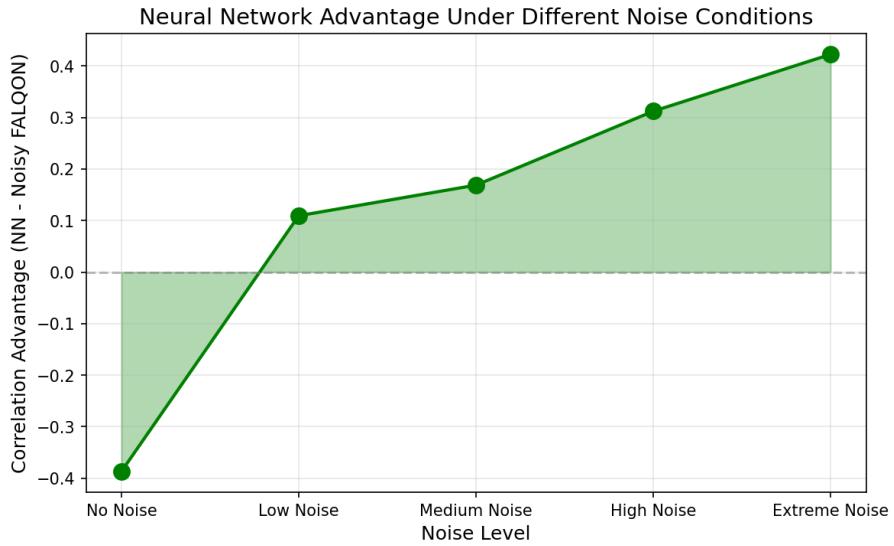


Figure 6: Neural network advantage ( $\text{Corr}_{NN} - \text{Corr}_{\text{NoisyFALQON}}$ ) as a function of noise level.

## 4.6 Spectral Density Analysis

## 4.7 Ablation

Table 2 shows the contribution of components.

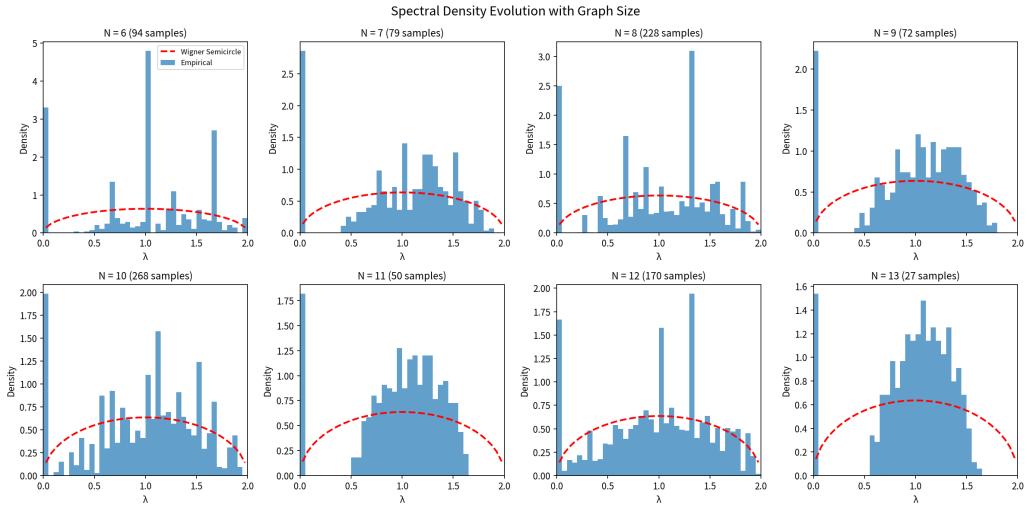


Figure 7: Empirical spectral density evolution across different graph sizes (comparison with Wigner semicircle approximation).

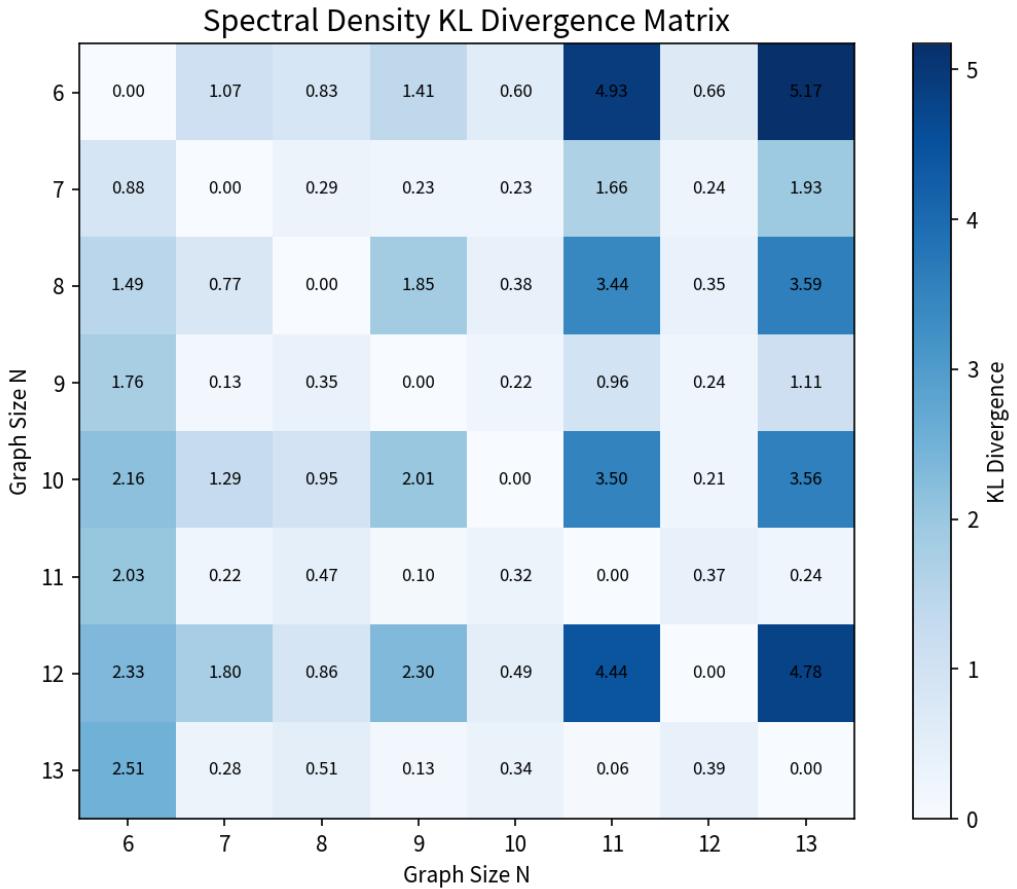


Figure 8: KL divergence matrix between spectral histograms at different graph sizes.

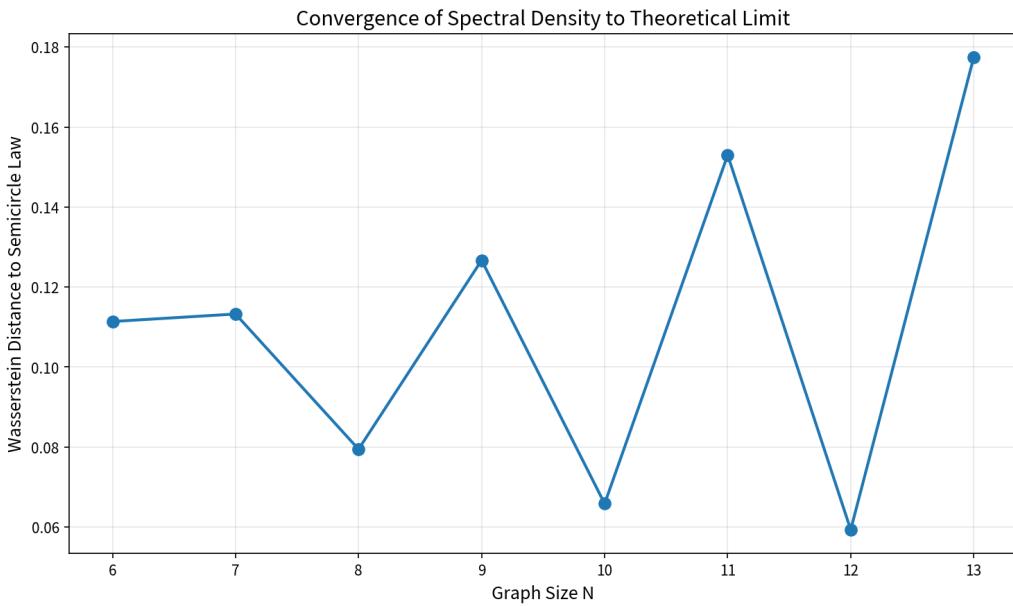


Figure 9: Wasserstein distance between empirical spectral density and semicircle law as a function of graph size.

Table 2: Ablation study results

Configuration	Corr	$\Delta$
Full model	0.885	—
Remove SignNet	0.821	-0.064
Remove Scheduled Sampling	0.847	-0.038
Remove temporal gradient loss	0.869	-0.016

## 5 Related Work

### 5.1 Variational Quantum Optimization

QAOA [6] is a widely studied VQA. Recent work explores parameter initialization and transfer learning, including GNN-based parameter prediction [5].

### 5.2 Feedback-Based Quantum Control

FALQON [1] and follow-up work [2] propose feedback rules that avoid classical optimization. Our neural approach complements FALQON by providing fast zero-shot parameter generation.

### 5.3 Spectral Graph Representation Learning

Spectral graph features provide a principled input; SignNet [3] addresses eigenvector sign ambiguity and is adapted here.

## 6 Conclusion and Future Work

We introduced a Spectral-Temporal Transformer for predicting FALQON control parameters in a teacher-student zero-shot framework. Key findings:

1. Feasibility: On convergent samples the model achieves  $\text{Corr} = 0.917$ , validating neural prediction of quantum control parameters.
2. Limitations: Oscillatory samples (often from regular graphs) exhibit high-frequency behavior that is harder to predict.
3. Theoretical guarantee: Lyapunov analysis shows that as long as prediction errors do not flip control signs, the system remains convergent.

### Future Work

- Scale tests to larger graphs ( $n > 20$ ) to examine cross-size generalization.
- Explore hybrid strategies that combine neural predictions with physics-based refinement.
- Evaluate performance on noisy quantum hardware models.

## References

- [1] Magann, A.B., et al., 2022. Feedback-based quantum optimization. *Phys. Rev. Lett.*, 129(25), p.250502.
- [2] Magann, A.B., Grace, M.D., Rabitz, H.A. and Sarovar, M., 2022. Lyapunov-control-inspired strategies for quantum combinatorial optimization. *Phys. Rev. A*, 106(6), p.062414.
- [3] Lim, D., Robinson, J., Zhao, L., Smidt, T., Sra, S., Maron, H. and Jegelka, S., 2023. Sign and basis invariant networks for spectral graph representation learning. *ICLR*.
- [4] Vaswani, A., et al., 2017. Attention is all you need. *NeurIPS* 30.
- [5] Joshi, C., et al., 2022. Learning to branch in combinatorial optimization with graph neural networks. *ICLR Workshop on AI for Science*.

- [6] Farhi, E., Goldstone, J. and Gutmann, S., 2014. A quantum approximate optimization algorithm. arXiv:1411.4028.
- [7] Bengio, S., Vinyals, O., Jaitly, N. and Shazeer, N., 2015. Scheduled sampling for sequence prediction with recurrent neural networks. NeurIPS 28.
- [8] Kesten, H., 1959. Symmetric random walks on groups. Trans. Amer. Math. Soc., 92(2), pp.336-354.