SymQNet: VQE Hamiltonian Estimation for Molecular Optimization with Quantum Computing
"Significantly Faster Drug Candidate Screening"

Yash V. Tomar
West Lafayette, Indiana

Drug Discovery is Broken

$2.6B	12yr	90%
Average Drug Cost[1]	Time to Market[2]	Failure Rate[3]

Delayed treatment has been proven to **increase mortality rates**[4].

Faster molecular optimization (MO) can significantly speed up drug candidate screening.

MO's Flawed Role

Molecular Optimization identifies most **stable configuration** to accurately evaluate a candidate[5,6,7].

- Stable structures are most commonly found configuration of candidate in nature[5].
- Important because it **accurately** determines how a potential candidate **behaves** in biological systems[5,6]

Born-Oppenheimer approximation implies most stable structure lies at lowest ground-state energy[5].

Models such as DFT attempt to **minimize ground-state energy** to arrive at the most stable structure[5].

MO methods have limits in accuracy[5,6]

Struggle with weak *intermolecular forces* (e.g., London-Dispersion)[7]

Solution: Quantum MO (QMO)

Uses **Variational Quantum Eigensolver** (VQE) to approximate wavefunction better.[8,9]

VQE is a Potential Solution

VQE reduces synthesis by 30-50%, saving **6-12 months**[10]

- **Thousand-year** long problems *potentially* solved in hours[10]

Hamiltonian is operator that encodes molecule's energy[11]

Calculating Hamiltonian Energy is Impossible

Mathematically impossible to diagonalize & directly calculate the energy of a structure[11]

Estimating Energy is "Next Best Bet"

VQE[8,9] uses various methods to try estimate the energy on Hamiltonian

Choosing measurements becomes important

Optimally choosing what parameter to measure could speed up MO

Estimation Requires Slow Measurements

To be accurate, methods must continuously measure parameters (potentially **millions of times**) to reduce variation[11,12]

Hypothesis based on RL

Molecule → *Resources Wasted* → Optimized

Choose Measurement Usually chosen poorly, wasting shots later — **Pain Point**

Conduct Measurement ...on chosen qubit. expensive but essential.

Hypothesis: If canonical **information** in each measurement shot can be **maximized** using reinforcement learning, then the number of estimation shots for a Hamiltonian's energy can be **significantly reduced** because maximizing information reduces uncertainty in the energy estimate more efficiently, allowing convergence with **fewer measurements**.

Key Challenge: Engineer a way around measurements' noisy and probalistic nature. Such conditions are **unfavorable** for learning.

Methods: Problem Framed as POMDP and Architecture Iteratively Engineered

Reinforcement Learning Offers Reward Based Strategy

Reinforcement learning trains an "agent" to take **optimal actions** based on observations of its environment guided by a reward signal[14,15]
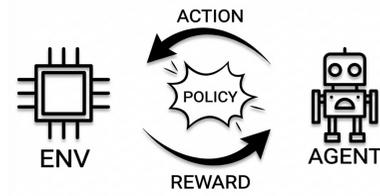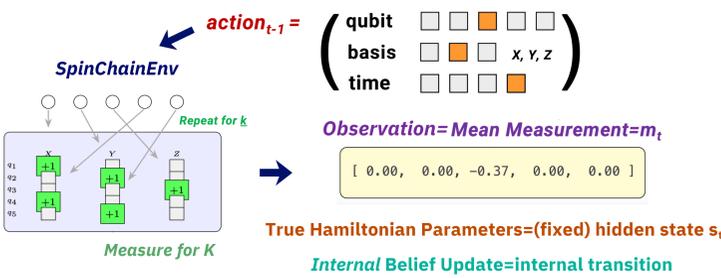
Infogain Reward

$$r_t = \max(0,\ H_{t-1} - H_t)$$

where H is *entropy* of belief and t represents the time index

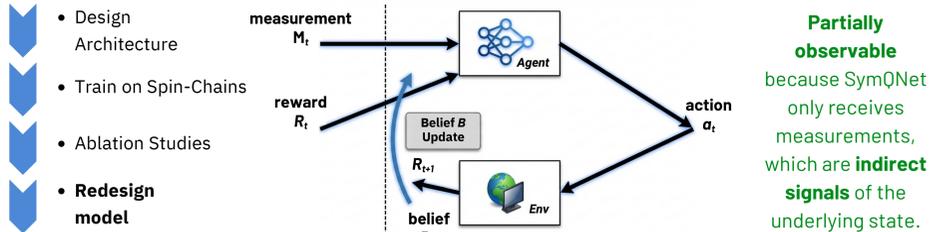ACTION — POLICY — REWARD / ENV — AGENT

Core Idea: Learn a **strategy** to choose which measurements give the **most information** to estimate energy **faster**.

Action Defined as Selecting Measurement (with Qubit and Basis)

$action_{t-1}$ =
- qubit
- basis — X, Y, Z
- time

SpinChainEnv

Repeat for k

Measure for K

Observation= Mean Measurement=m_t

[0.00, 0.00, -0.37, 0.00, 0.00]

True Hamiltonian Parameters=(fixed) hidden state s_t
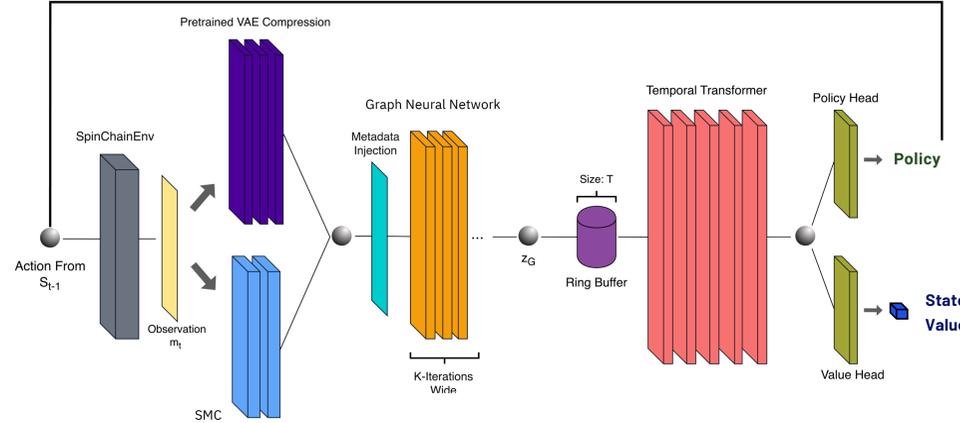
Internal Belief Update=internal transition

Bayesian Sequential Monte Carlo (SMC) is our **estimator**. It's what internally estimates energies and tells us how uncertain it is about parameters[16]. **SymQNet's job is to accurately guide SMC.**

Engineering Cycle & Partially Observable Markov-Decision Process (POMDP) Framing

- Design Architecture
- Train on Spin-Chains
- Ablation Studies
- **Redesign model**

measurement M_t → Agent
reward R_t → Belief B Update → R_{t+1}
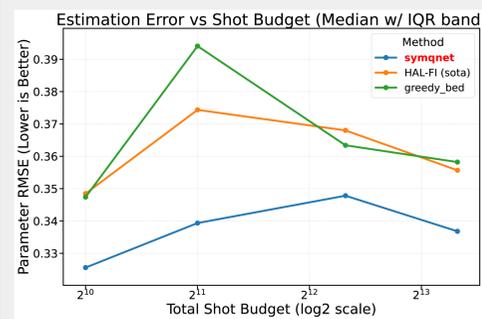action a_t
belief B_{t-1} → Env

Partially observable because SymQNet only receives measurements, which are **indirect signals** of the underlying state.

Architecture Iteratively Designed to Account for Tough Learning Conditions

Pretrained VAE Compression — Graph Neural Network — Temporal Transformer — Policy Head

SpinChainEnv — Metadata Injection — Size: T — Ring Buffer

Action From S_{t-1} — Observation m_t — z_G — K-Iterations Wide — SMC

Policy Head → **Policy**
→ **State Value**
Value Head

Results: Evaluation of Performance, Scaling Behavior, VQE Integration, and Reproducibility

SymQNet Offers Decisive Win over State-of-the-art Baselines

Estimation Error vs Shot Budget (Median w/ IQR band)

Method: symqnet, HAL-FI (sota), greedy_bed

X-axis: Total Shot Budget (log2 scale) 2^{10} – 2^{13}
Y-axis: Parameter RMSE (Lower is Better)

Decisive Win Over SOTA

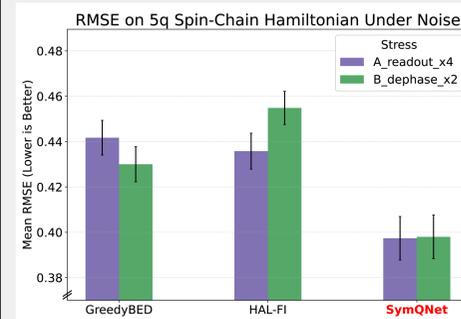Achieves **lower Parameter RMSE** than HAL-FI (SOTA) and GreedyBED across all shot budgets (2^{10}-2^{13}).

Statistically significant: **p< 0.05** via paired-Wilcoxon rank test (Holm-corrected)

- Softens exponential error growth in downstream Hamiltonian use.

Same budget. Better Results.

Shot Budget Benchmarking versus SOTA

Noise Stress-Test Shows Support for Hypothesis

RMSE on 5q Spin-Chain Hamiltonian Under Noise

Stress: A_readout_x4, B_dephase_x2
X-axis: GreedyBED, HAL-FI, SymQNet
Y-axis: Mean RMSE (Lower is Better)

Robust Under Realistic Noise

SymQNet outperforms both GreedyBED and HAL-FI under **readout noise (x4)** and **dephasing noise (x2)**.

- Statistically significant (±2SEM).

p < 0.05 via paired-Wilcoxon test (Holm-corrected)

Gains maintained under realistic conditions: **ready for deployment.**

Deployment: `pip install symqnet-molopt`

Noise Stress-Test (Readout & Dephase) for Parameter Estimation

Molecular Application Shows Significant Speedup and Accuracy Gains

46.7% Shot Reduction — Fewer shots to reach success threshold

4/4 Molecules Won — Better solution on Pyridine, Imidazole, Urea & Ising (p < 0.05)

21.2% MAE Reduction — Better mean energy error reduction vs. baseline
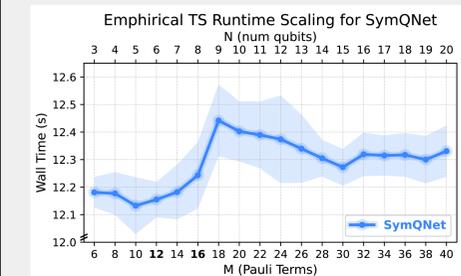
1.88x Drug Discovery Speedup — More molecules screened under the same budget

Energy Estimation with SymQNet on VQE Job versus Standard Pauli-Term Estimation

Molecular Optimization of **Pyridine Drug** (SymQNet Assisted VQE (frac=0.3), SPSA Downstream VQE (frac=0.3)) — 270k 740k
Molecular Optimization of **Imidazole Drug** — 420k 670k
Molecular Optimization of **Ising Hamiltonian** — 510k 950k
Molecular Optimization of **Urea** — 490k 810k
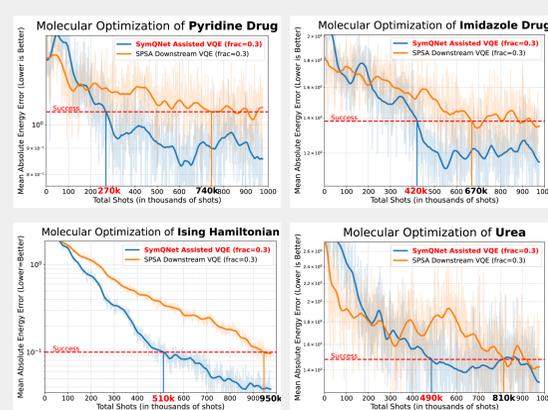
VQE MO Jobs with Energy Estimation on SymQNet vs Regular

Emphirical Runtime Scaling Tests Show O(1) Scaling

Emphirical TS Runtime Scaling for SymQNet
N (num qubits): 3 4 7 10 13 16 19 20
X-axis: M (Pauli Terms) — 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40
Y-axis: Wall Time (s)

How can SymQNet be accessible?

Accessible in One Line

`pip install symqnet-molopt`

Open-source Repository

symqnet-molopt Public

SymQNet is currently **patent-pending**

Why Should I Care?

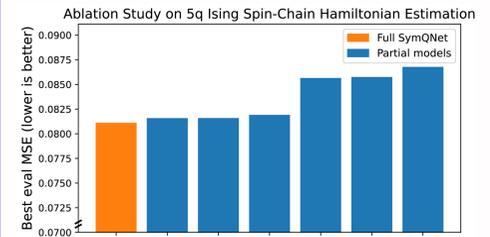Cost doesn't explode as molecules get larger. *Significant* applicibility potential.

Reference: Baseline methods typically scale as O(N[4]) and O(M).[17]

Emphirically Flat Runtime Scaling (up to 20 qubits) and Reproducibility

Ablation Studies on Prototype

Ablation Studies show Maximal Component Utilization

*Provides experimental **evidence** SymQNet **learned.***

Ablation Study on 5q Ising Spin-Chain Hamiltonian Estimation

Full SymQNet / Partial models
X-axis: SymQNet, no_shots_metadata, no_theta_cov_features, vae_sampled_z, roughen_off, no_graph, particles_16
Y-axis: Best eval MSE (lower is better)

- Paired t-test reveals p < 0.05 (Holm-corrected).

No Metadata — Removing metadata injection degrades performance (context matters!).

No Graph — Removing qubit graph structure hurts spatial relationships are informative.

Reduced Particles — Fewer SMC particles (1.6×) reduces Bayesian estimation quality.

Ablation studies systematically remove or degrade each component of SymQNet to verify its contribution. Across 30 seeds, **SymQNet (Full) achieves the lowest MSE**, confirming that every design choice contributes meaningfully to performance.

Conclusions & Project Recap

From results, hypothesis appears to be supported!

Maximizing information from each shot with RL can lead to significant speedup of QMO

🚀 **Drug Discovery Speedup**
1.88× more molecules screened per budget. 46.7% shot reduction. 21.2% better energy error.

🏆 **Beats state-of-the-art**
Lower RMSE than HAL-FI & GreedyBED under same shot budget and under quantum noise (p < 0.05).

⚡ **Empirical O(1) Scaling**
Improves from O(N[4]), O(M) to O(1) (empirically).

Deployed: `pip install symqnet-molopt`
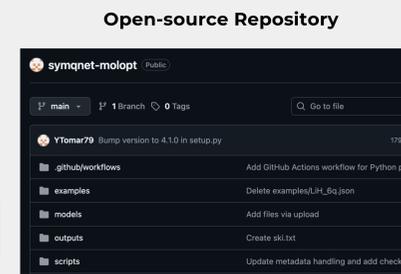
Next Steps and My Vision

Next Steps for SymQNet

1 Higher-Dimensional Hamiltonians
Train on larger, more complex molecular systems

2 Real-Life Deployment
Trade off reward with device cost on actual quantum hardware

My Vision of Accelerating Medicine

- **Faster drug development** rate with lower costs
- **Better drugs** with lower clinical failure rates