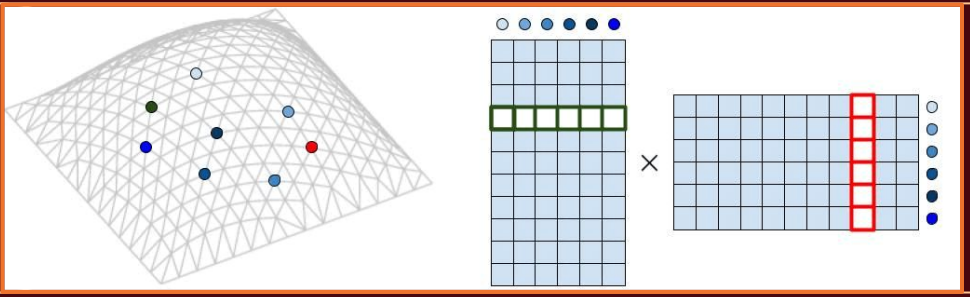
Neurips | Sydney, Australia

Walking the Hypercube: Unbiased Quantum Partition Functions Without the Matrix

Google Research, Kensho Technologies,
Google DeepMind, Columbia University



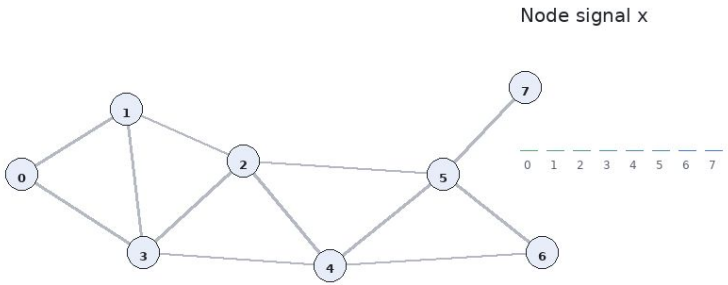
Avinava Dubey*, Arijit Sehanobish*,
Krzysztof Choromanski*

* equal contribution

Graph Random Features (GRFs) in a nutshell

1. Start with a graph

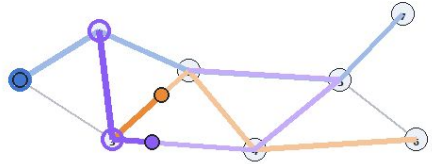
Information lives on nodes and flows along edges.



Goal: replace expensive graph kernels
or message passing by cheap random features.

2. Sample random graph probes

Random walks can act as simple stochastic probes of graph structure.



Random walks as probes

Walk 1
0 → 1 → 2 → 5 → 7

Walk 2
3 → 2 → 4 → 6 → 4

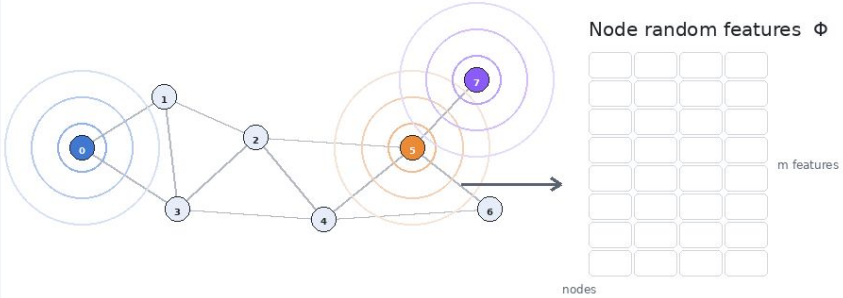
Walk 3
1 → 3 → 4 → 5 → 2

Random probing can be visualized by sampling walks on the graph
and using their trajectories to build node-dependent features.

Graph Random Features (GRFs) in a nutshell

3. Build a feature map Φ

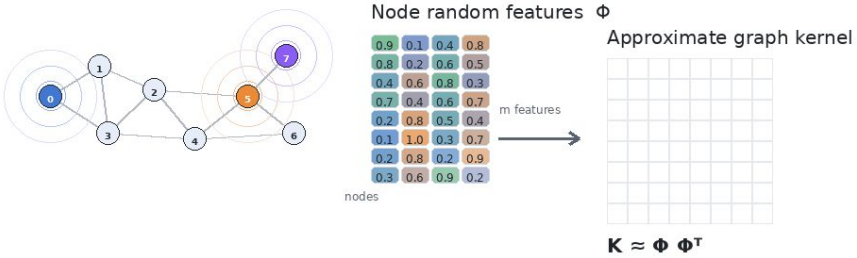
Each node becomes a compact vector of random graph measurements.



A node v is embedded as: $\Phi(v) = [\varphi_1(v), \varphi_2(v), \ldots, \varphi_m(v)]$

4. Use Φ to approximate graph computation

Inner products of random features approximate graph kernels and fast attention.



Instead of forming dense K exactly, compute with Φ:

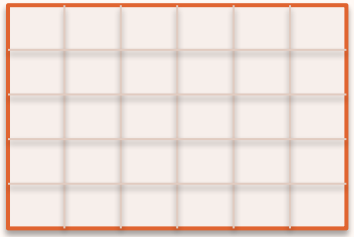
- **$Kx \approx \Phi(\Phi^T x)$** **cost: O(nm) rather than $O(n^2)$**

The bottleneck: tracing an operator over an exponential state space

Quantum partition functions are traces of exponentiated Hamiltonians

$$Z(\beta) = \mathrm{Tr}\left[\exp(-\beta H)\right]$$

n spins → 2^n basis states

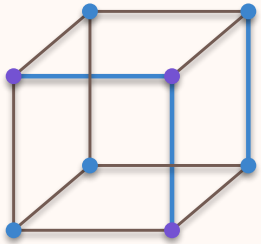


$2^n \times 2^n$

4^n entries
if stored densely

Matrix-based trace estimation becomes memory-limited.

But the model has local structure



Local transitions between basis states

Exploit walks on the hypercube instead of materializing the operator.

Goal: estimate $\mathrm{Tr}(\exp(-\beta H))$ without constructing the full matrix as GRFs still materialize H.

From General Graphs to Quantum Ising Models

The analysis of several quantum models heavily rely on understanding spectral properties of the exponentiated adjacency matrices:

- **Transverse Field Ising Model (TFIM)**: this is the canonical model of quantum phase transitions. The off-diagonal part induces quantum effects.
- **Sherrington-Kirkpatrick (SK) Model**: the canonical spin glass model.
- **XXZ Model**: The off-diagonal part now generates two-spin swap moves that preserve total magnetization, giving each configuration $O(n^2)$ neighbors. Special instantiations include XX model and Heisenberg model.

Hypercube Walk Estimator: main idea

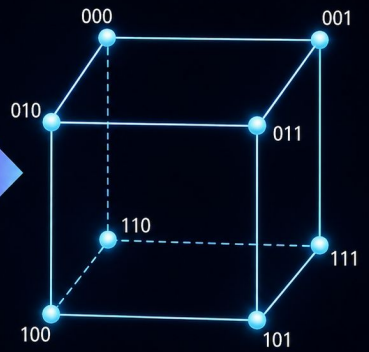
1. From a huge matrix to a walk on basis states

$$Z = \mathrm{Tr}(e^{-\beta H})$$

$$\begin{bmatrix} H_{00,00} & H_{00,01} & H_{00,10} & H_{00,11} & H_{00,100} & \cdots & H_{00,2^n-1} \\ H_{01,00} & H_{01,01} & H_{01,10} & H_{01,11} & H_{01,100} & \cdots & H_{01,2^n-1} \\ H_{10,00} & H_{10,01} & H_{10,10} & H_{10,11} & H_{10,100} & \cdots & H_{10,2^n-1} \\ H_{11,00} & H_{11,01} & H_{11,10} & H_{11,11} & H_{11,100} & \cdots & H_{11,2^n-1} \\ H_{100,00} & H_{100,01} & H_{100,10} & H_{100,11} & H_{100,100} & \cdots & H_{100,2^n-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ H_{2^n-1,00} & H_{2^n-1,01} & H_{2^n-1,10} & H_{2^n-1,11} & H_{2^n-1,100} & \cdots & H_{2^n-1,2^n-1} \end{bmatrix}$$

Size: $2^n \times 2^n$ (exponential)

Hypercube of computational basis states ($n = 3$ qubits)



No full matrix

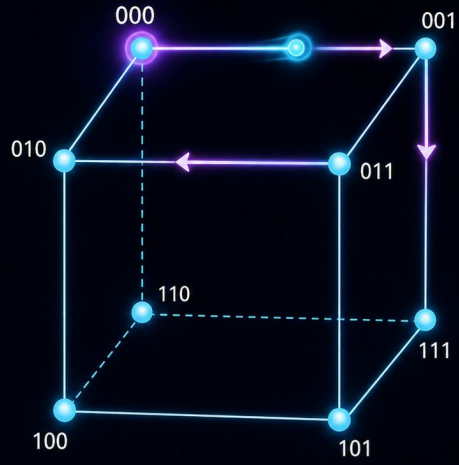
Building or storing H is infeasible.

Idea: estimate the partition function without materializing H.

2. Basis states become vertices, transitions become edges

Computational basis states

000
001
010
011
100
101
110
111
⋮



Path example:

000 → 001
↓
011
↓
010

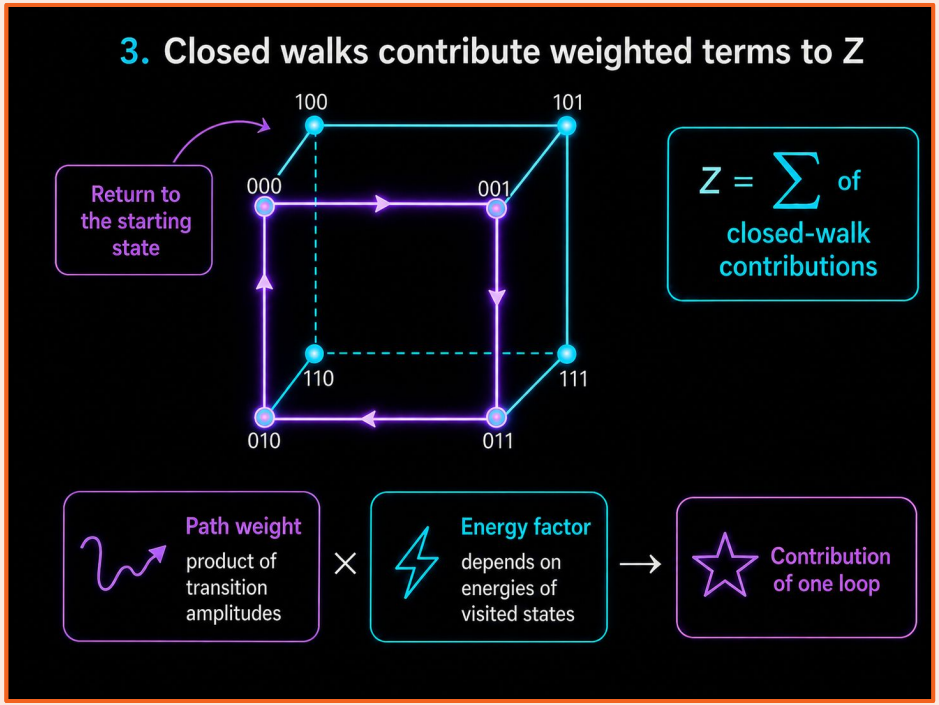
Off-diagonal term

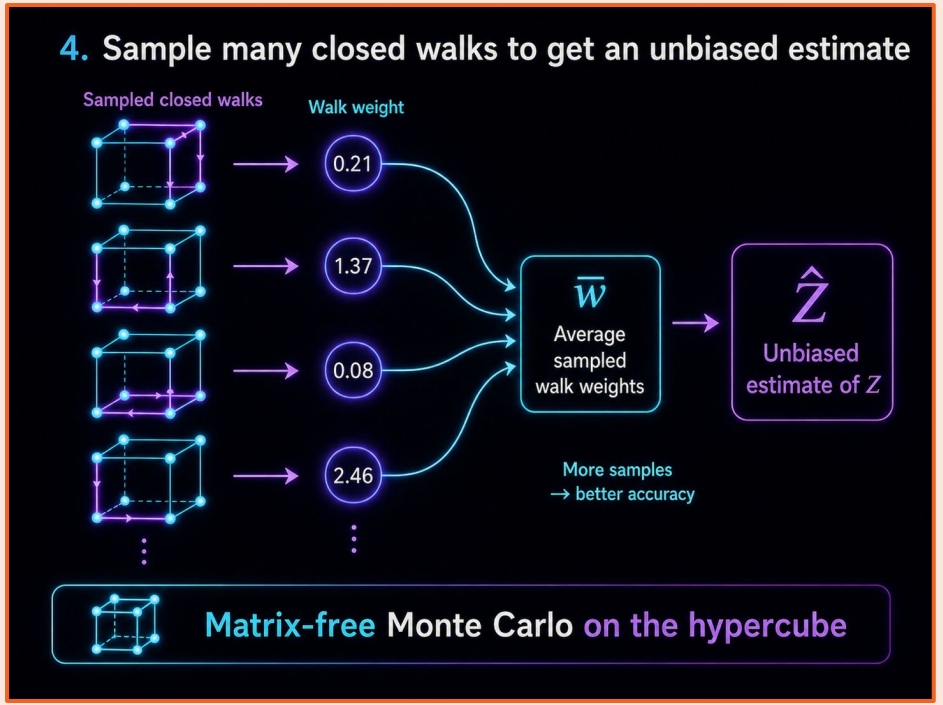
→

move to a neighboring basis state.

We work with **local moves,** not the full matrix.

Hypercube Walk Estimator: main idea





HWE estimates the two matrix-function factors using independent GRF walks, projects them onto Rademacher probes, and averages their products.

Why HWE Scales

Local computation replaces exponential matrix operations; independent samples expose parallelism.

O(n)

Memory

No explicit Hamiltonian or matrix exponential is stored.

O(n + d)

Per-walk step

Local transitions and incremental energy updates.

Three independent sampling axes

Hutchinson probes

Independent Rademacher vectors

Average over probes

Source nodes

Independent starting configurations

Sample different roots

Walk pairs

Independent trajectories per source

Generate pairs in parallel

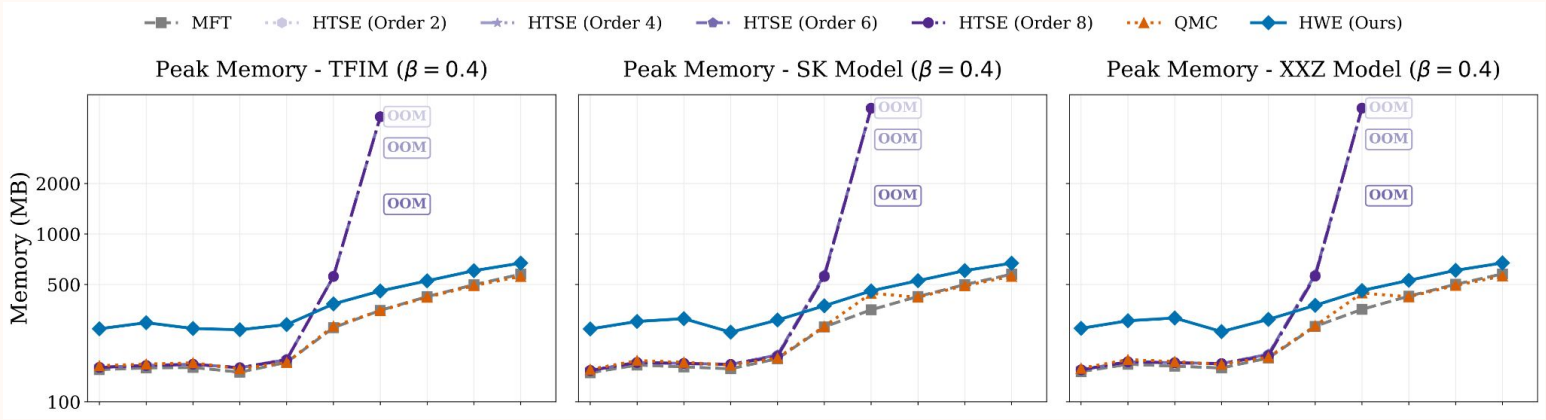
Aggregate independent estimates

Key distinction: memory and per-step cost avoid dependence on $N = 2^n$; total sampling effort may still grow.

Matrix-free estimation changes the memory bottleneck

The estimator trades matrix materialization for controllable sampling

Peak memory across system sizes



Matrix-based methods hit the 32 GB limit before n = 16;
No Matrix Materialization. Random walks are generated on the fly.

Unbiased trace estimate

Averaging sampled closed-walk contributions gives an unbiased estimate of Z.

Controllable error bounds

Statistical variance drops predictably with sample size, allowing targeted precision limits.

Hypercube Walk Estimator: Results

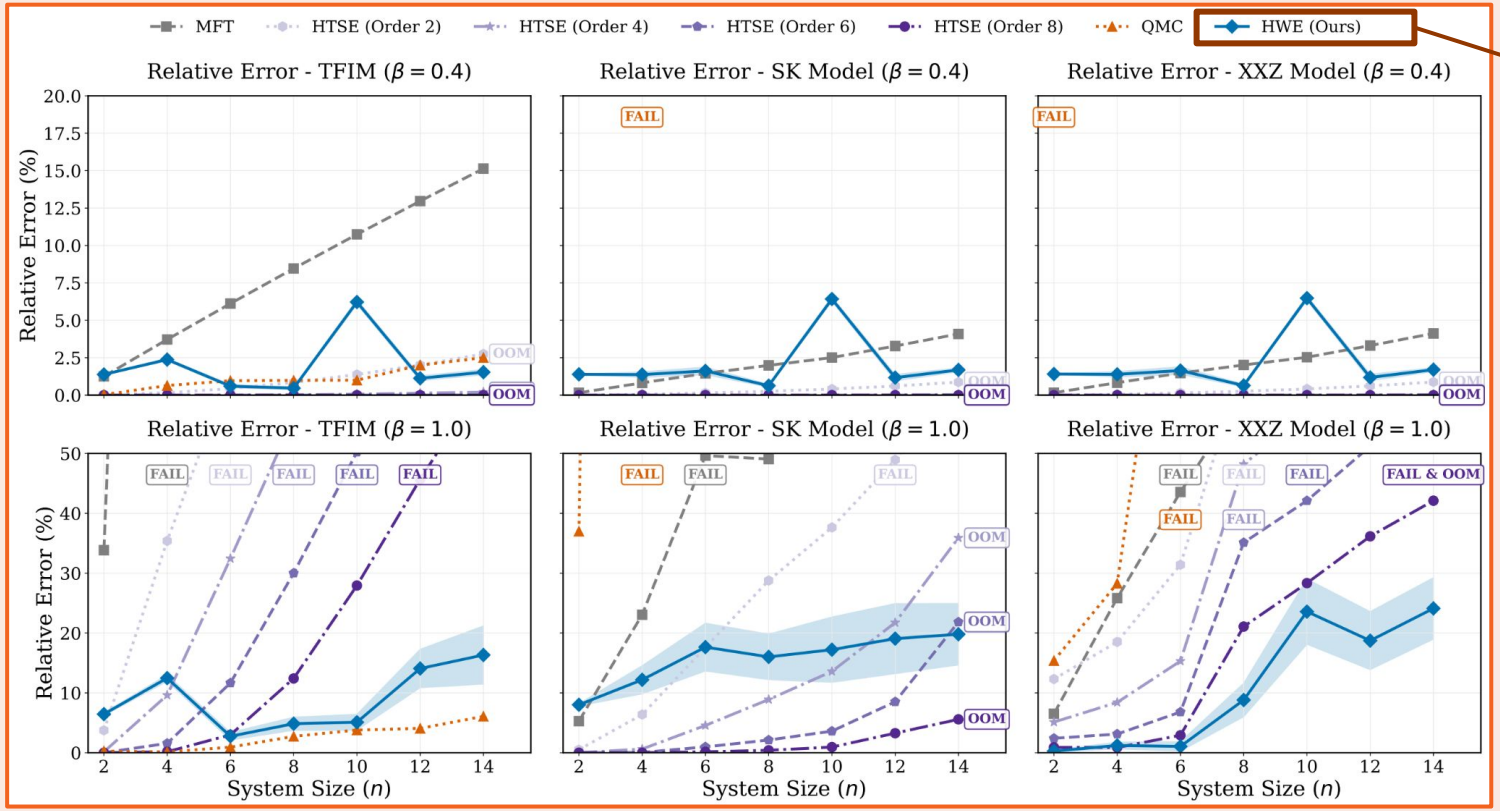
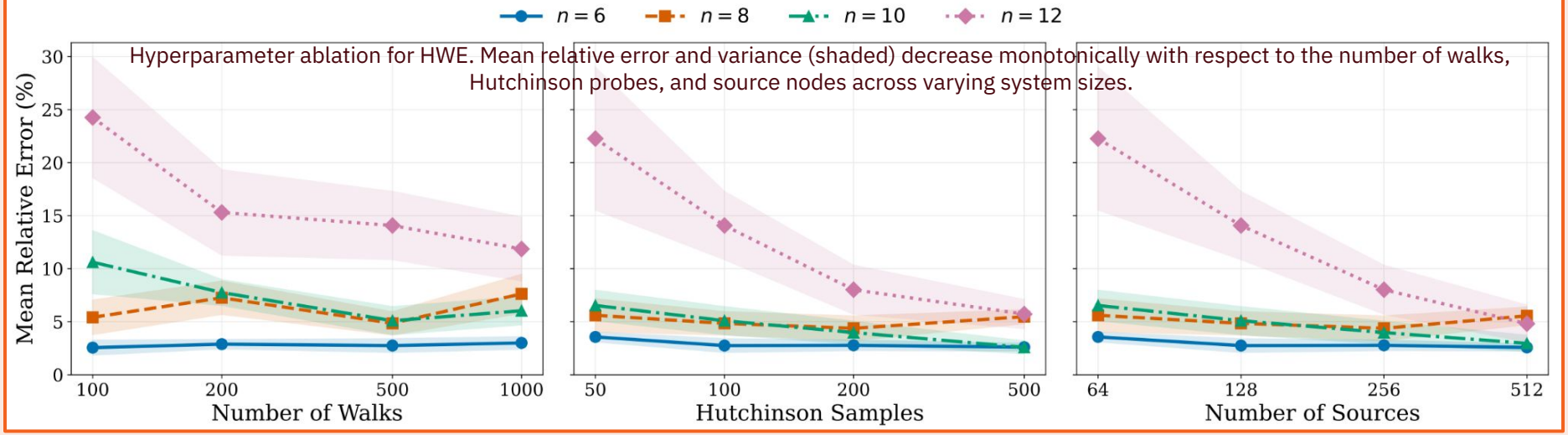


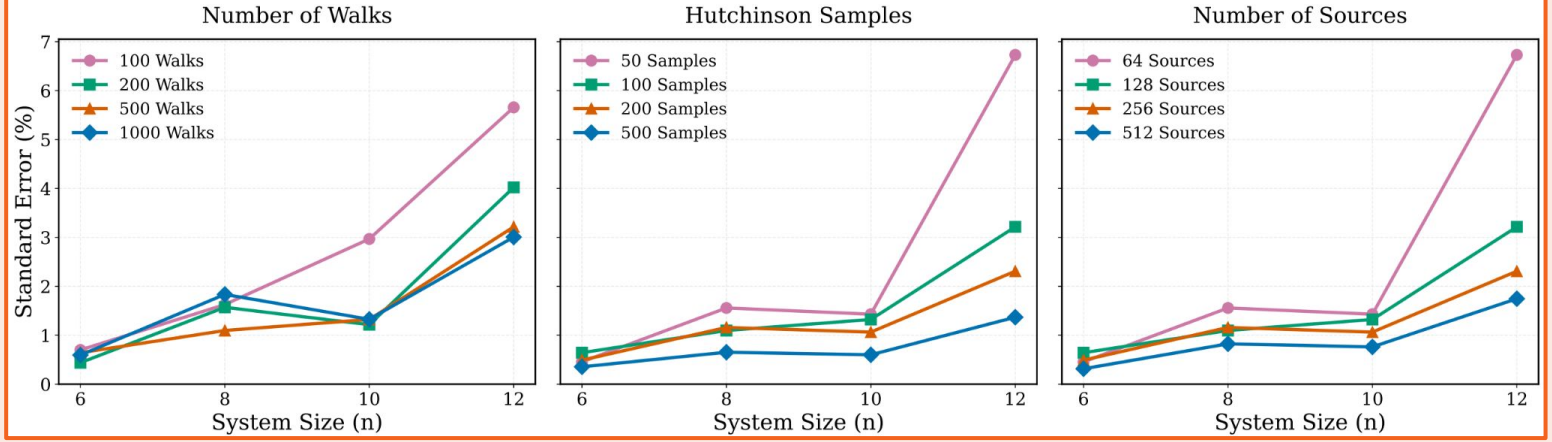
Fig.: Relative Error across System Sizes. A comparison of trace estimators on the TFIM, SK, and XXZ models at high (β = 0.4, top) and low (β = 1.0, bottom) temperatures. Methods exceeding a 40% error threshold are strictly labeled as **FAIL**, while those exceeding available memory (32 GB) at n = 16 are labeled **OOM**. At low temperatures, MFT fails to capture critical quantum fluctuations, QMC performs well for TFIM but poorly for rest of the evaluations, and HTSE mathematically diverge with increasing n even though they are the most accurate at smaller values of n.

Hypercube Walk Estimator: Results



Hyperparameter ablation for HWE. Mean relative error and variance (shaded) decrease monotonically with respect to the number of walks, Hutchinson probes, and source nodes across varying system sizes.

Standard Error scaling of the HWE estimator as a function of system size (n) across varying numbers of Walks, Hutchinson Samples, and Source Nodes.



Thank You