NeurIPS 2026

A Locality-Aware Surrogate for Natural-Gradient Descent in Quantum Optimization

Md Mobasshir Arshed Naved, Wenbo Xie, Wojciech Szpankowski, Ananth Grama

Department of Computer Science, Purdue University

Outline

01

Motivation

Why natural gradient, and why it’s expensive

02

Key idea & method

QFIM as a covariance; truncate, mask, estimate

03

Guarantees

Error decomposition, sample complexity, convergence

04

Results

VQE on spin chains, ablations, fidelity to the QFIM

Natural gradient follows the geometry of quantum states

Quantum natural gradient (QNG) preconditions with the **quantum Fisher information matrix** (QFIM), the Fubini–Study metric, to follow steepest descent on the state manifold:

$$\vec{\theta}^{\,t+1} = \vec{\theta}^{\,t} - \eta\, F\left(\vec{\theta}^{\,t}\right)^{-1} \nabla C\left(\vec{\theta}^{\,t}\right)$$

Often improves conditioning and convergence in practice.

THE BOTTLENECK

$O(p^2)$

QFIM entries to estimate per iteration, each from circuit shots: too many for realistic shot budgets.

Approximations differ in what they give up

Method	Geometry	Cross-parameter couplings	Metric cost
Block-diagonal QNG	QFIM	within blocks only	O(pq) entries
Random Natural Gradient	classical Fisher (lower bound)	measurement-induced	O(p) state preps
Shadow-Hybrid (ours)	QFIM surrogate	overlapping supports kept	one shadow dataset

Ours keeps the QFIM target and the couplings locality says matter, so it stays closest to the true QFIM.

Key idea: the QFIM is a covariance of generators

1 · Each derivative is generated by a propagated generator (via Pauli propagation):

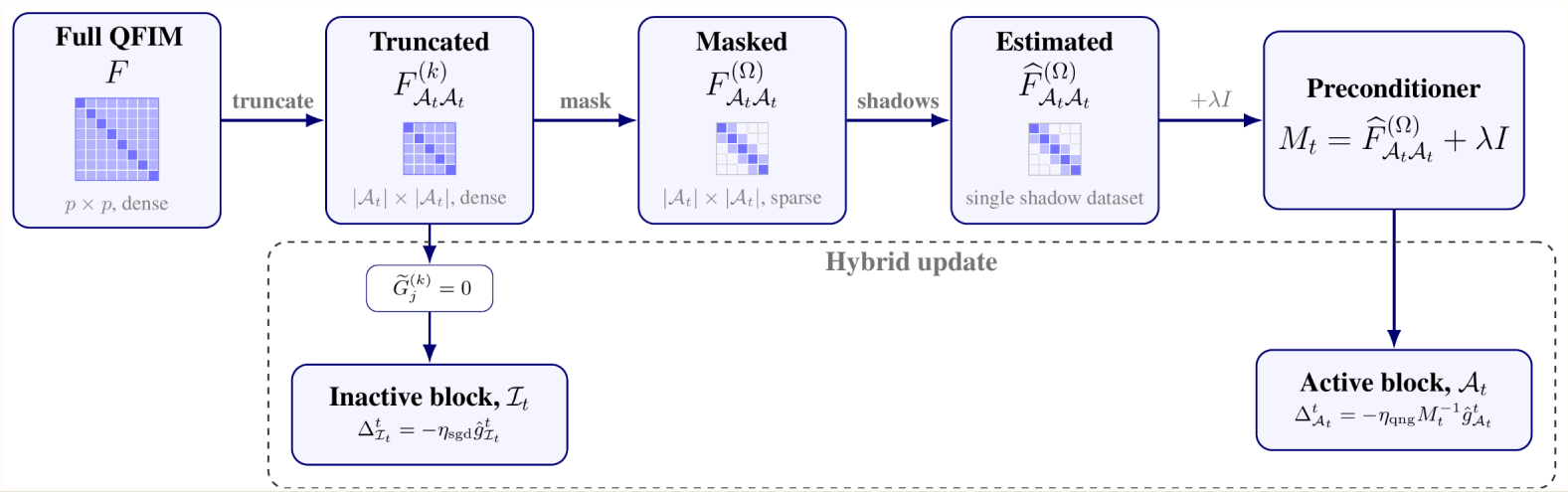
$$\frac{\partial}{\partial\theta_j}|\psi\rangle = i\,\widetilde{G}_j|\psi\rangle, \qquad \widetilde{G}_j = U_{>j}\,G_j\,U_{>j}^{\dagger}$$

2 · For pure states, every QFIM entry is a covariance, measurable from Pauli expectations:

$$F_{jl} = 4\,\Re\big[\mathrm{Cov}_\rho(\widetilde{G}_j,\widetilde{G}_l)\big] = 4\Big(\tfrac{1}{2}\big\langle\{\widetilde{G}_j,\widetilde{G}_l\}\big\rangle_\rho - \langle\widetilde{G}_j\rangle_\rho\langle\widetilde{G}_l\rangle_\rho\Big)$$

In local circuits, truncated generators stay localized, so circuit structure tells us where the QFIM can be sparsified.

Method: truncate, mask, estimate



1 · Truncate

Keep Pauli weight ≤ k; variance scales with k, not n.

2 · Mask

Keep only pairs whose supports overlap.

3 · Estimate

All retained entries, unbiased, from one shadow dataset.

4 · Update

Surrogate-QNG on active parameters, SGD on the rest.

Guarantees: every error source is accounted for

Error on the active block ≤ truncation bias + masking bias + finite-shot noise

$$\left\| \widehat{F}^{(\Omega)}_{\mathcal{A}_t\mathcal{A}_t} - F_{\mathcal{A}_t\mathcal{A}_t} \right\|_{\mathrm{op}} \leq \underbrace{E_{\mathrm{trunc}}}_{\text{weight-}k} + \underbrace{E_{\mathrm{mask}}}_{\text{support } \Omega_k} + \underbrace{E_{\mathrm{shot}}}_{\text{finite } N}$$

Sample complexity

Snapshots grow with the mask's row degree, only logarithmically with retained entries.

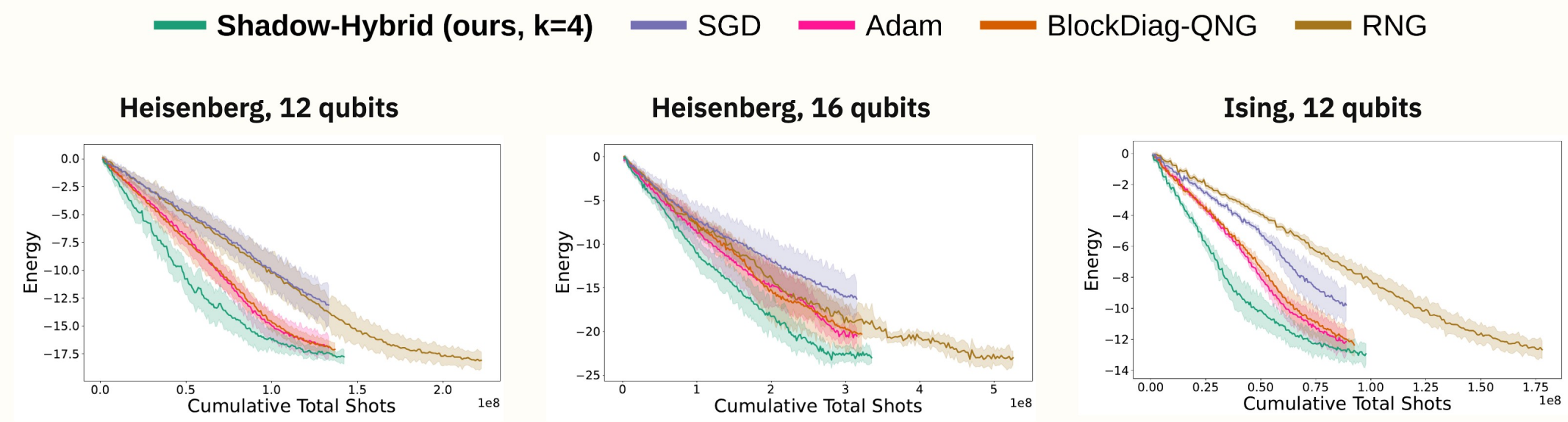
Convergence

To a stationary neighborhood set by gradient noise, under standard assumptions.

Choosing k

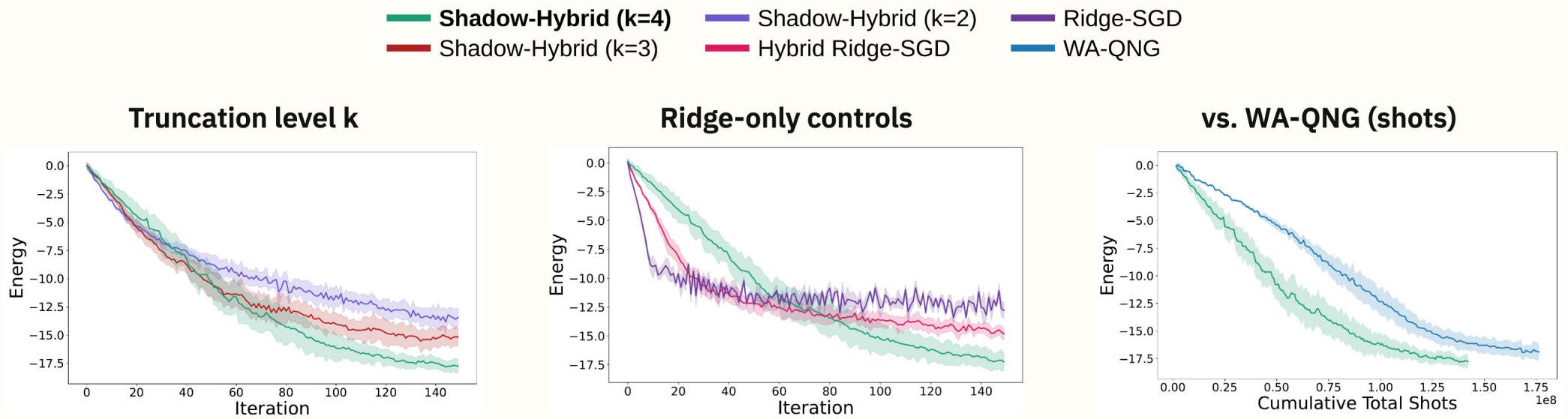
Larger k cuts truncation bias but raises shot variance; we use k = 4.

Lower energy per shot in every tested setting



8/8 seeds reach the target at 16 qubits, the only method to do so (RNG 7/8, others ≤ 3/8)

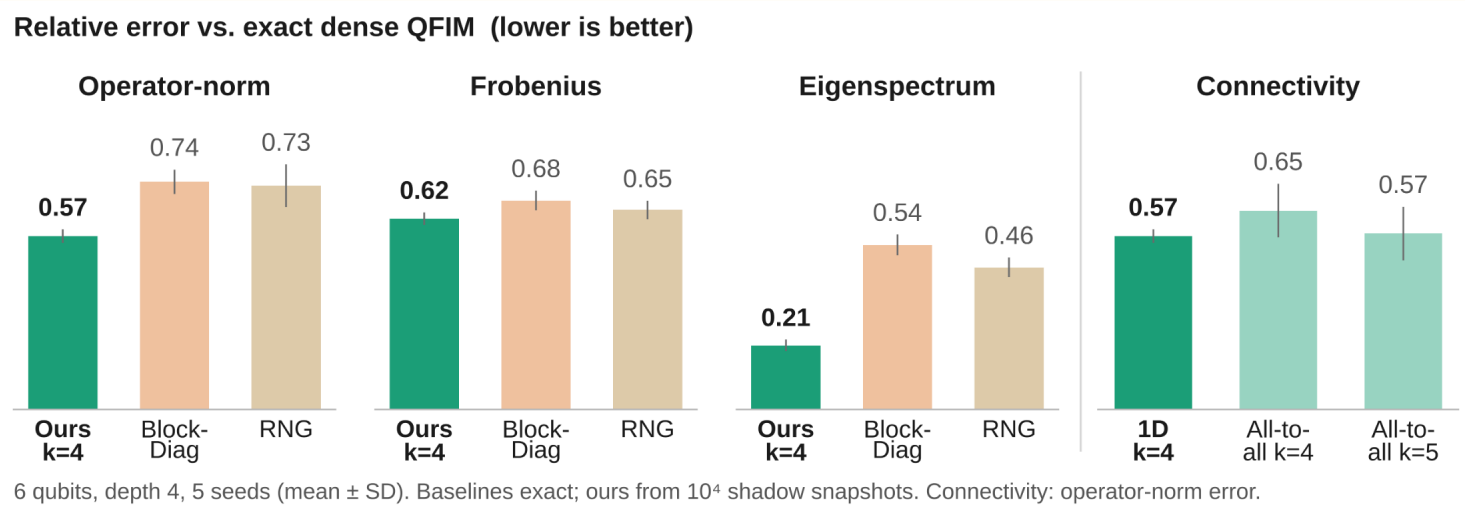
The gain comes from curvature



12-qubit Heisenberg: larger k helps; both ridge controls (global and hybrid) stall.

Beats locality-aware WA-QNG over nearly the whole shared shot budget.

Closer to the true QFIM, and sparse at scale



6 qubits, depth 4, 5 seeds (mean ± SD). Baselines exact; ours from 10^4 shadow snapshots. Connectivity: operator-norm error.

6 qubits, where the exact QFIM is computable; the all-to-all diagnostic benefits from a larger k.

≥2.2×
lower QFIM spectrum error than both baselines (6 qubits)

≈19%
of entries kept at 16 qubits, with 99% of Frobenius energy retained

≈25×
top surrogate eigenvalue over the ridge floor (12 qubits)

TAKEAWAY

Exploit circuit locality to build a sparse QFIM surrogate from a single classical-shadow dataset — natural-gradient updates under practical shot budgets.

Next

- Adaptive k
- Noisy (mixed-state) circuits
- Long-range connectivity

Thank you!

BACKUP

Experimental setup and shot budget

Setup

- VQE, 1D Heisenberg (12, 16 qubits) and transverse-field Ising (12 qubits)
- Hardware-efficient ansatz, nearest-neighbor CNOTs
- $k = 4$, ridge $\lambda = 0.1$ with PSD projection
- 8 matched seeds; learning rates tuned per method in $[0.01, 0.1]$
- Shared parameter-shift gradients, 1024 shots per measurement group
- Shadow snapshots N: 60k (12 qubits), 100k (16 qubits)

Metric-estimation shots per iteration (12 qubits, p = 144)

Method	Shots
BlockDiag-QNG (12 × 2048)	≈ 2.5×10^4
Ours (N snapshots)	6.0×10^4
RNG (289 × 2048)	≈ 5.9×10^5

About 2.4× BlockDiag-QNG and 9.9× cheaper than RNG, while keeping inter-block curvature.

BACKUP

Choosing k in practice

k trades truncation bias for shot variance. A fixed-budget **screening heuristic**, not an automatic rule:

1 · Screen

Check retained terms, active-set size, retained entries and max support-overlap degree.

Drop k values that exceed the budget.

2 · Test stability

Compare update directions from independent shadow batches at the same point.

Estimate shot variability for each k.

3 · Choose

Compare the change between adjacent k with that variability.

Prefer the smaller k unless the change exceeds the noise.

BACKUP

Measured error terms at 6 qubits

k	Truncation	Mask	Shot	Total
k = 2	0.761	0.150	0.078	0.783
k = 3	0.613	0.154	0.149	0.638
k = 4	0.540	0.159	0.249	0.566

Bias–variance in action: larger k cuts truncation error, shot error rises, mask error stays small; the total improves.

Relative operator-norm errors vs. the exact QFIM, 10^4 snapshots, 5 seeds. Components bound the total; they don’t add.

BACKUP

References

Quantum natural gradient. J. Stokes, J. Izaac, N. Killoran, G. Carleo. *Quantum* 4, 269 (2020).

Random natural gradient. I. Kolotouros, P. Wallden. *Quantum* 8, 1503 (2024).

Weighted approximate quantum natural gradient for VQE. C. Shi, V. Dunjko, H. Wang. *Quantum Sci. Technol.* 11, 015060 (2026).

Predicting many properties of a quantum system from very few measurements. H.-Y. Huang, R. Kueng, J. Preskill. *Nature Physics* 16, 1050 (2020).

Pauli propagation. M. S. Rudolph, T. Jones, Y. Teng, A. Angrisani, Z. Holmes. arXiv:2505.21606 (2025).