Optimal Ansatz-free Hamiltonian Learning In Situ

Taiqi Zhou

The Chinese University of Hong Kong

Weiyuan Gong

Harvard University

NeurIPS 2026

Hamiltonian learning in situ

Unknown interactions

The Pauli labels and coefficients are unknown.
We assume a norm bound, with no locality restriction or predefined interaction graph.

$$H = \sum_{u \neq 0} \lambda_u P_u , \qquad \|H\| \leq \Lambda$$

u = (x | z) labels an n-qubit Pauli string.

Coefficientwise accuracy

$$\max_{u \neq 0} |\hat{\lambda}_u - \lambda_u| \leq \varepsilon$$

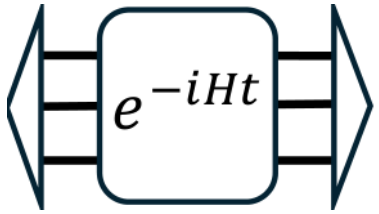
Omitted coefficients are assigned zero. The identity term is excluded: it contributes only a global phase.

$$\sum_{u \neq 0} {\lambda_u}^2 = 2^{-n} \operatorname{Tr}(H^2) \leq \Lambda^2$$

$$S_\varepsilon = \{u \neq 0 : |\lambda_u| \geq \varepsilon\}, \qquad |S_\varepsilon| \leq \frac{\Lambda^2}{\varepsilon^2}$$

Experimental access

Prepare a Pauli product state.
Apply one forward evolution.
Measure in local Pauli bases.



No ancillas, interleaved control, or inverse evolution.
Classical choices may depend on earlier outcomes.

$$T_{\text{tot}} = \sum_{\text{shots}} t_{\text{shot}}$$

This counts evolution time across all shots. Preparation, readout, and classical processing are separate costs.

Measured correlation

Prepare a uniformly random product eigenstate of Q, with eigenvalue s. After time t, measure P to obtain m. Their product estimates a normalized trace correlation:

$$\mathbb{E}[sm \mid t] = F_{P,Q}(t) = 2^{-n} \operatorname{Tr}\left(e^{iHt} P e^{-iHt} Q\right)$$

Zhou and Gong, §2.2–2.3, Problem 2.1 and Proposition 2.3. Circuit: Fig. 1(a).

Product measurements attain the SQL scaling

Theorem

For $0 < \varepsilon \le \Lambda$ and fixed success probability, every coefficient of magnitude at least ε is detected.
Retained labels have magnitude above $\varepsilon/3$; the interval between these thresholds allows a transition region.

$$S_\varepsilon \subseteq \hat{S}_\varepsilon \subseteq \left\{u \neq 0 : |\lambda_u| > \frac{\varepsilon}{3}\right\}$$

$$\max_{u \in S_\varepsilon} |\hat{\lambda}_u - \lambda_u| \le \frac{\varepsilon}{6}$$

$$N_{\text{shots}} = O\left(\frac{\Lambda^2}{\varepsilon^2} \log \frac{4\Lambda}{\varepsilon}\right)$$

$$T_{\text{tot}} = O\left(\frac{\Lambda}{\varepsilon^2} \log \frac{4\Lambda}{\varepsilon}\right)$$

Precision dependence

Setting	Total evolution time
Earlier: no control or ancillas	$O(\varepsilon^{-4})$
This protocol	$O(\varepsilon^{-2} \log(1/\varepsilon))$
Worst-case lower bound	$\Omega(\varepsilon^{-2} \log(1/\varepsilon))$

Λ and success probability are fixed. The lower bound applies in the dimension and precision regime stated on Slide 6.

Estimate candidates to $\varepsilon/6$; retain $|\tilde{\lambda}u| > \varepsilon/2$.

Zhou and Gong, Theorem 4.1, Algorithm 3 and §4.2.

Probe times and classical cost

$$\mathbb{E}[t] = O\left(\frac{1}{\Lambda}\right)$$

$$t \in \{0, t_0, 2t_0, \ldots\}, \qquad t_0 < \frac{\pi}{6\Lambda}$$

The displayed grid condition is sufficient. Its spacing does not shrink as the target precision improves.

Classical processing is polynomial in n and Λ/ε.
The evolution-time bound has no explicit n.

Bit flips locate the candidate interactions

Same-basis displacement sieve

Prepare a uniformly random string a, evolve, and measure b in the same basis. A displacement collects the squared coefficients of all Pauli terms with that projection.

$$d = a \oplus b, \qquad W_x(d) = \sum_z \lambda_{(d \mid z)}{}^2$$

Preparation and readout	Displacement identifies
Z basis	x-projection of u = (x | z)
X basis	z-projection of u = (x | z)

A coefficient of size ε contributes mass ε². Direct sampling finds heavy projections without estimating every mass.

$$N_{\text{proj}} = O\left(\frac{\Lambda^2}{\varepsilon^2} \log \frac{4\Lambda}{\varepsilon}\right)$$

Finite-time derivative kernel

$C_P(t) = F_{P,P}(t)$ is the same-basis signal.

Its frequencies lie in [−2Λ, 2Λ].

$$C_P''(0) = -4 \sum_{u:\{P_u,P\}=0} \lambda_u{}^2$$

$$C_P''(0) = 2\int_0^\infty K_\Lambda(t) C_P(t)\, \mathrm{d}t$$

$$q_\Lambda(t) = \frac{2|K_\Lambda(t)|}{\|K_\Lambda\|_1}, \qquad \|K_\Lambda\|_1 = O(\Lambda^2)$$

$$\Pr[D = d] \geq \frac{2W_x(d)}{\|K_\Lambda\|_1} = \Omega\left(\frac{W_x(d)}{\Lambda^2}\right)$$

The probability bound holds for every nonzero d.

Candidate set

$$S_\varepsilon \subseteq \hat{\mathcal{U}}_{\text{proj}}, \qquad |\hat{\mathcal{U}}_{\text{proj}}| = O\left(\left(\frac{\Lambda}{\varepsilon}\right)^4\right)$$

$$\hat{\mathcal{U}}_{\text{proj}} = \left(\left(\hat{\Pi}_x \cup \{0\}\right) \times \left(\hat{\Pi}_z \cup \{0\}\right)\right) \setminus \{(0,0)\}$$

Add zero projections explicitly. The Cartesian product may include extra labels; coefficient estimates filter them.

Zhou and Gong, §3.1–3.2, Propositions 3.1–3.2 and Lemma 3.3.

The same records estimate many coefficients

Cross-Pauli signal

Choose a uniformly random frame C in $\{X, Y, Z\}^n$.
At each qubit, prepare and read out along the other two Pauli directions, with their order randomly swapped.
For compatible Pauli strings P and Q,

$$F'_{P,Q}(0) = 2\sigma\lambda_u, \qquad \sigma \in \{-1, +1\}$$

σ comes from Pauli multiplication. The signed parity Z uses the input, output, and orientation signs.
Sample T from a density proportional to $|L\Lambda(T)|$, where $L\Lambda$ is the first-derivative kernel. For a visible label,

$$Y_u^{(C)} = \frac{\sigma}{2}\|L_\Lambda\|_1 \operatorname{sign}\big(L_\Lambda(T)\big)\, Z$$

Swapping preparation and readout supplies the antisymmetric signal. Both orientations evolve forward.

Visibility correction and data reuse

A label is visible if it matches C on an odd number of its nonidentity positions. If its weight is w(u), the visibility probability remains bounded below:

$$q(u) = \frac{1 - 3^{-w(u)}}{2} \geq \frac{1}{3}$$

$$\tilde{Y}_u = \frac{1[u \in \mathcal{V}(C)]}{q(u)} Y_u^{(C)}$$

$$\mathbb{E}[\tilde{Y}_u] = \lambda_u, \qquad \left|\tilde{Y}_u\right| = O(\Lambda)$$

Keep each complete input and output sign string. Within this second-stage data set, every record supplies parities for all visible candidates. A union bound adds only log |U| to the simultaneous estimation cost.

For nonempty U: uniform error α at fixed confidence. The full algorithm uses α = ε/6.

$$N_{\text{coef}} = O\left(\frac{\Lambda^2}{\alpha^2}\log(4|\mathcal{U}|)\right) \qquad T_{\text{coef}} = O\left(\frac{\Lambda}{\alpha^2}\log(4|\mathcal{U}|)\right)$$

Zhou and Gong, §3.3, Proposition 3.4, Lemma 3.5 and Algorithm 2.

Unknown support adds a logarithmic cost

Hard family

Choose one hidden label x_a in each of M blocks and add a random sign. Pauli strings from distinct blocks anticommute, keeping the norm exactly Λ.

$$H_{x,\sigma} = \frac{\Lambda}{\sqrt{M}} \sum_{a=1}^{M} (-1)^{\sigma_a} Q_{a,x_a}$$

$$H_{x,\sigma}{}^2 = \Lambda^2 I, \qquad M = \left\lfloor \left(\frac{\Lambda}{4\varepsilon}\right)^2 \right\rfloor$$

Each active coefficient has magnitude at least 4ε. The construction uses $M + \lceil \log_2 M \rceil$ qubits, so an ε-accurate learner must identify every active label.

Information in one experiment

After averaging signs and erasing them from the transcript, a shot has a latent index J. It is uninformative for J = 0; otherwise it depends on one uniformly random block.

$$\Pr[J = 0] = \cos^2(\Lambda t), \qquad \Pr[J = a] = \frac{\sin^2(\Lambda t)}{M}$$

Theorem

Any control-free protocol with success probability at least 2/3 requires, for a worst-case Hamiltonian,

$$T_{\mathrm{tot}} = \Omega\left(\frac{\Lambda}{\varepsilon^2} \log \frac{\Lambda}{\varepsilon}\right)$$

$$0 < \varepsilon \le \frac{\Lambda}{16}, \qquad n \ge C\frac{\Lambda^2}{\varepsilon^2}$$

C is a sufficiently large absolute constant. The qubit requirement allows the hard family to fit.

Coupon-collector cost

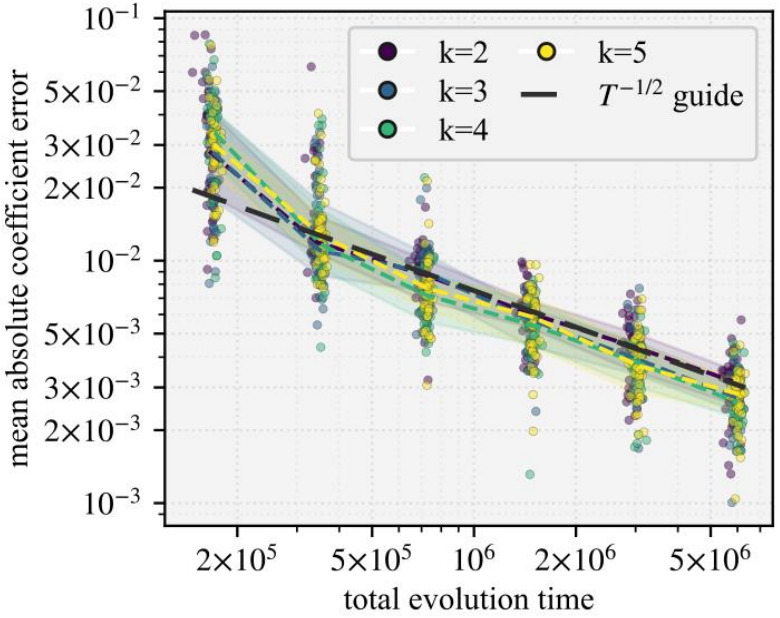
$$\mathbb{E}[K] = \Omega(M \log M), \qquad \mathbb{E}[K] \le \Lambda T_{\mathrm{tot}}$$

K counts informative shots; $\sin^2(\Lambda t) \le \Lambda t$. Even a learner given the latent indices must encounter every block. This matches the upper bound in the stated regime.

Zhou and Gong, Theorem 5.2 and Appendix G.3–G.4 (especially Theorem G.8).

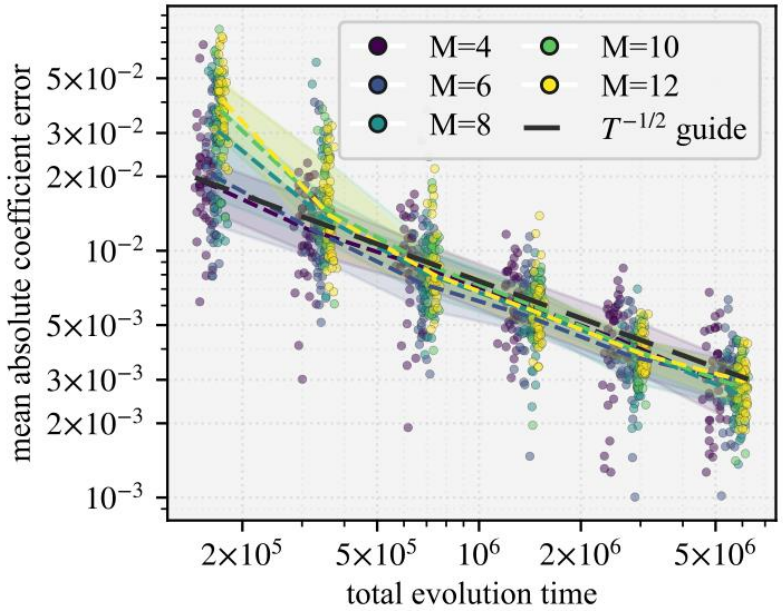
Numerical results

Local Hamiltonians



n = 6, M = 8; locality k = 2, 3, 4, 5.

Unrestricted sparse Hamiltonians



n = 5; sparsity M = 4, 6, 8, 10, 12.

Error over the true support S

$$\mathrm{MAE}_S = \frac{1}{|S|}\sum_{u \in S} |\hat{\lambda}_u - \lambda_u|$$

A missed true term contributes with $\hat{\lambda}_u = 0$.

Simulation settings and interpretation

Λ = 1; six budgets; 40 trials for each setting. The dashed line is a $T^{-1/2}$ guide. Mean error follows this trend; the theorem instead gives a simultaneous coefficient guarantee.

Zhou and Gong, §6 and Fig. 2.