

Learning k -body Hamiltonians via sublinear-time sparse recovery

Franz J. Schreiber and Jens Eisert

Learning an unknown Hamiltonian from dynamical data is a central problem in quantum learning theory [1–4] and a key primitive for the benchmarking and verification of quantum devices [5–7]. Hamiltonians realized in nature and in engineered devices are typically composed of few-body interactions, reflecting short ranged interactions, of which often only a small number carry significant weight. This observation has motivated a line of work importing ideas from compressed sensing into notions of Hamiltonian learning [8–10].

In this work, we focus on the task of learning an unknown k -body Pauli Hamiltonian, described by a coefficient vector μ over the non-identity Pauli strings of weight at most k . Hamiltonian learning built on the standard compressed-sensing toolkit has several favorable features in this context. First, no prior knowledge of which interaction terms are present is required. Second, such schemes remain sample-efficient for k -body Hamiltonians, whose ambient space of $O(n^k)$ candidate Pauli terms, with n the number of qubits, renders exhaustive strategies impractical. Third, they offer recovery guarantees that improve with the compressibility of the signal. In compressed sensing one observes noisy linear measurements $y = A\mu + e$ of the unknown coefficient vector, with far fewer measurements than μ has entries, and aims at a reconstruction $\hat{\mu}$ obeying a guarantee of the form

$$\|\mu - \hat{\mu}\|_2 \leq \frac{C}{\sqrt{M}} \|\mu - \mu_M\|_1 + \mathcal{E}(e), \quad (1)$$

where M is the target sparsity, μ_M denotes the best M -term approximation to μ , obtained by retaining its M largest entries, and $\mathcal{E}(e)$ is an error contribution controlled by the measurement noise. The first term vanishes for exactly M -sparse signals and remains small whenever μ is dominated by few significant interactions, so the guarantee adapts to the instance at hand rather than presupposing exact sparsity. Bounds of the form (1) are known as ℓ_2/ℓ_1 guarantees. Stronger guarantees of the ℓ_2/ℓ_2 type instead bound the reconstruction error by $(1 + O(\varepsilon)) \|\mu - \mu_M\|_2$ up to noise. Guarantees of the latter form are called instance optimal, as the reconstruction is then, up to a factor arbitrarily close to one, as accurate as the best M -term approximation itself.

These features are well aligned with the needs of practical device characterization. The ability to learn interaction terms that are not known in advance is particularly relevant in the context of crosstalk, a persistent issue in complex multi-qubit devices [11–14]. Moreover, Pauli terms of larger weight, say $k \geq 5$, naturally arise in the stabilizer checks of quantum error-correcting codes [15, 16], in fermion-to-qubit encodings [17], and in demanding quantum-simulation applications [18].

Existing compressed-sensing protocols for Hamiltonian learning are, however, efficient only in some of the relevant resources. While they achieve favorable sample complexities, the recovery guarantees known to be robust to noise are of the weaker ℓ_2/ℓ_1 type. More severely, their classical post-processing inherits the $O(n^k)$ scaling of the ambient space, as standard compressed-sensing decoders, whether based on convex programming or on greedy thresholding, maintain and update a dense vector over all candidate k -body Pauli terms. In principle, it is known how to overcome this scaling. This is precisely the subject of sublinear-time compressed sensing [19–22], where both the number of measurements and the decoding time scale with the sparsity and polylogarithmic with the ambient dimension, making them overall sub-linear in the ambient dimension. Such schemes, however, place far more elaborate demands on the measurement ensemble, since the measurement matrix must be extremely sparse and yet structured enough to remain efficiently decodable. The technical difficulty of compressed-sensing Hamiltonian learning hence lies largely in identifying sparse-recovery algorithms whose measurements can be realized efficiently in a quantum experiment.

Our contribution. In this work, we identify a sublinear-time sparse-recovery scheme whose measurements are natively implemented by simple quantum operations. We thereby obtain, to our knowledge, the first protocol for learning an unknown k -body Pauli Hamiltonian with an instance-optimal ℓ_2/ℓ_2 error guarantee that is simultaneously sample and computationally efficient in all resources that enter the protocol: the total evolution time, measurement count, and decoding time all scale sublinearly in the ambient dimension, which is of size $O(n^k)$. evolution under the unknown Hamiltonian, interleaved with single-qubit Pauli operations, together with the preparation and measurement of GHZ-type probe states as used in standard phase-estimation primitives. For an M -term Pauli Hamiltonian and target ℓ_2 reconstruction error ϵ_H , the protocol uses $\tilde{O}(M \log n)$ measurements and total evolution time $\tilde{O}(M \log n / \epsilon_H)$, thus operating at the Heisenberg limit, while the classical decoding time is reduced from the $O(n^k)$ of conventional decoders to $\tilde{O}(M \text{polylog}(n))$. Here, $\tilde{O}(\cdot)$ suppresses further polylogarithmic factors as well as the dependence on the body order k , which is exponential in k but independent of n . Our recovery guarantee is of the $(1 + O(\varepsilon))$ instance-optimal form discussed above. Achieving it for Hamiltonians that are only approximately sparse incurs an additional

overhead exponential in k , which we quantify below. The classical cost per experiment is consequently dominated by the $O(n)$ readout of the n qubits. Across all experiments, the total readout volume amounts to $\tilde{O}(nM \log n)$. Beyond the known weight cutoff k and sparsity, we make no structural assumptions such as geometric locality, and we require no prior knowledge of the interaction terms, which renders our scheme especially attractive for benchmarking and verification.

Our analysis is carried out in a deterministic residual model, in which arbitrary, possibly biased and correlated, errors on the estimated frequencies are charged only by their total energy. This gives robustness to measurement noise without statistical assumptions. Modeling errors, such as omitted higher-weight interaction terms, are treated on the same footing, as they enter the error bound only through the residual energy they induce. Our guarantees hence persist whenever this energy is controlled. Taken together, our results show that assumption-light Hamiltonian learning is possible at essentially the cost of reading out the device.

I. Setting and access model. We consider an unknown n -qubit Hamiltonian

$$H = \sum_{P \in \mathcal{P}_n^{(k)}} \mu_P P, \quad (2)$$

supported on the set $\mathcal{P}_n^{(k)}$ of non-identity Pauli strings of weight at most k , and aim to recover the coefficient vector μ in ℓ_2 norm. Our protocol is organized around commuting sectors. For a basis string $\beta \in \{x, y, z\}^n$, the Pauli strings diagonal in the product basis β form a commuting family $K_\beta = \{P_{\beta,c} : c \in \mathbb{F}_2^n\}$, where $P_{\beta,c}$ acts as σ^{β_i} precisely on the qubits i with $c_i = 1$. Interleaving the time evolution with random Pauli strings from K_β projects H onto its component in K_β . Within a sector, the eigenvalue on the product eigenstate labeled by $b \in \mathbb{F}_2^n$ is the Walsh sum $\lambda_b = \sum_c x_c \chi_b(c)$, where $\chi_b(c) = (-1)^{b \cdot c}$ and the vector x with entries $x_c = \mu_{P_{\beta,c}}$ collects the sector coefficients. The experimental primitive is Heisenberg-limited phase estimation of eigenvalue differences $\Delta_b = \lambda_b - \lambda_0$ with GHZ-type probe states. We work within a single sector and without noise until the error model of Sect. IV. The lift to full k -body Hamiltonians follows in Sect. V.

II. A sparse-recovery measurement primitive. The structural observation at the heart of our protocol is quantum measurement scheme that natively implements the measurements of a sublinear-time sparse-recovery algorithm. Let $h : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^r$ be a random linear map, set $B = 2^r$, and fix a shift $a \in \mathbb{F}_2^r$. Then, for every bucket $v \in \mathbb{F}_2^r \setminus \{0\}$,

$$\frac{1}{B} \sum_{u \in \mathbb{F}_2^r} (-1)^{u \cdot v} \Delta_{a+h^T u} = \sum_{c : hc=v} \chi_a(c) x_c =: Y_{h,a}(v). \quad (3)$$

The right-hand side is a signed hash bucket in the sense of CountSketch [23]. Here, the map h hashes the exponentially many labels into B buckets, while the random signs $\chi_a(c)$ decorrelate colliding coefficients. The left-hand side shows how to obtain it. A single Walsh–Hadamard transform of B estimated eigenvalue differences produces the entire vector of bucket values, at a cost of B frequency queries. Equally important, since $\chi_{a+g}(c) = \chi_a(c)(-1)^{g \cdot c}$, replacing the shift a by $a + g$ flips the sign contributed by the label c according to the linear form $g \cdot c$. In this way, any $\mathbb{F}_2$ -linear function of the unknown label can be imprinted onto the sign of an isolated coefficient, and it is this programmability of the signs that drives the identification step below.

III. Identification and estimation. Our recovery algorithm first identifies a short list of candidate labels and then estimates their coefficients on a fresh, independent sketch, repeating the hashing in each stage with independently drawn maps and shifts. For identification, the shifted signs $a + g_1, \dots, a + g_\ell$ imprint the coordinates $g_j \cdot c$ of a binary linear code of length $\ell = O(k \log n)$ onto the sign pattern of an isolated heavy coefficient, so that decoding the pattern recovers the label even when tail collisions and measurement errors corrupt a constant fraction of the signs. We construct a suitable BCH-type code whose zero-error decoder runs in expected time $\text{poly}(k, \log n)$ and never enumerates the ambient label space. For estimation, each candidate is re-estimated by the median of its signed bucket values over the repetitions in which it collides with no other candidate, so that a corrupted bucket degrades at most one estimate, while the fresh sketch keeps the candidate list independent of the estimation randomness, as made precise in Sect. IV.

IV. Recovery guarantee for one sector. We analyze the algorithm in a deterministic error model, in which the implemented buckets are $\tilde{Y} = Y + e$. Here, Y collects the exact linear measurements of Eq. (3) and the residual e is completely arbitrary, possibly biased and correlated. If the buckets are formed from noisy frequency estimates with error vector ξ , orthogonality of the Walsh–Hadamard transform gives $\|e\|_2 = \|\xi\|_2 / \sqrt{B}$, so the residual energy contracts already at the level of the measurement primitive. We normalize the residual energies per repetition as $\rho_{\text{id}}^2 = \|e^{\text{id}}\|_2^2 / (\ell R_{\text{id}})$ and $\rho_{\text{est}}^2 = \|e^{\text{est}}\|_2^2 / R_{\text{est}}$, where e^{id} and e^{est} collect the residuals of the identification and

estimation stage and R_{id} and R_{est} denote the corresponding numbers of hashing repetitions. The residual may be adversarial but must be nonanticipating, in the sense that the identification residual, and with it the candidate list, is fixed before the fresh estimation sketch is drawn.

Theorem 1 (One-sector recovery). *Let $s \geq 1$ and $\varepsilon, \delta \in (0, 1/2)$. With $B = \tilde{O}(s/\varepsilon)$ buckets, ℓ sign shifts per identification repetition, and $O(\log(s/\delta))$ repetitions per stage, the algorithm outputs an $O(s)$ -sparse vector $\hat{x}$ such that, with probability at least $1 - \delta$,*

$$\|x - \hat{x}\|_2^2 \leq (1 + C\varepsilon) \|x - x_s\|_2^2 + C\rho^2, \quad \rho^2 = \rho_{\text{id}}^2 + \rho_{\text{est}}^2, \quad (4)$$

where x_s denotes the best s -term approximation to x , retaining its s largest entries.

Eq. (4) is the instance-optimal ℓ_2/ℓ_2 form advertised in the introduction. Two comments are in order. First, the guarantee is of the for-each kind, holding with probability over the sketching randomness for each fixed signal and each nonanticipating error process, rather than uniformly in the spirit of the restricted isometry property. This is a feature rather than a limitation. By known lower bounds, uniform ℓ_2/ℓ_2 recovery requires a number of measurements linear in the ambient dimension, so relaxing to a for-each statement is what makes the stronger error bound attainable at all [24]. As a fresh sketch is drawn for every run of the protocol, the for-each guarantee is also the operationally relevant one. Second, the error energy ρ is the sole interface between experiment and analysis: phase-estimation errors, imperfect sector projection, calibrated state-preparation and measurement bias, and out-of-model Hamiltonian terms enter the guarantee exclusively through ρ .

V. From one sector to k -body Hamiltonians. A fixed Pauli term of weight w is contained in a uniformly random sector with probability $3^{-w} \geq 3^{-k}$, so drawing $L = O(3^k \log(s/\delta))$ independent random bases covers the s largest coefficients of μ with high probability. The sector outputs are then merged. Estimates of a term appearing in several sectors are averaged, and a single hard threshold to $O(s)$ terms is applied only at the very end, since thresholding within each sector would accumulate the approximation tail L times.

Theorem 2 (Full Hamiltonian recovery). *Let μ be supported on $\mathcal{P}_n^{(k)}$ and let $\varepsilon_{\text{cs}} \in (0, 1/2)$ be the sparse-recovery parameter of the sector algorithm, taking the role of ε in Theorem 1. With probability at least $1 - \delta$ over basis selection and all sketches, the output $\hat{\mu}$ is $O(s)$ -sparse and satisfies*

$$\|\mu - \hat{\mu}\|_2^2 \leq (1 + CL\varepsilon_{\text{cs}}) \|\mu - \mu_s\|_2^2 + C \sum_{i=1}^L \rho_i^2, \quad (5)$$

where ρ_i is the residual energy of the run in the i -th basis.

The tail factor in Eq. (5) exposes a trade-off between the two sparsity regimes. If μ is exactly s -sparse, the tail vanishes, ε_{cs} may be taken to be an absolute constant, and the protocol carries the expected $\tilde{O}(3^k)$ overhead. To obtain a genuine $(1 + \varepsilon)$ approximation guarantee for merely approximately sparse μ , however, one must choose $\varepsilon_{\text{cs}} = \Theta(\varepsilon/L)$ to compensate for the L -fold accumulation of sector tails, which raises the overhead to $\tilde{O}(9^k)$.

VI. Resource requirements. For an exactly M -sparse Hamiltonian, the sparsity parameter of the sector algorithm is set to $s = M$ and ε_{cs} may be taken to be an absolute constant. The number of eigenvalue-difference estimates (equivalently, of physical measurements up to polylogarithmic shot overhead) and the total evolution time for target ℓ_2 error ϵ_H are then

$$Q = \tilde{O}(3^k M k \log n), \quad T = \tilde{O}\left(L^{3/2} \frac{M k \log n}{\epsilon_H}\right), \quad L = O(3^k \log(M/\delta)). \quad (6)$$

In the approximate-sparse regime, where $\varepsilon_{\text{cs}} = \Theta(\varepsilon/L)$, the query count for target sparsity s becomes $\tilde{O}(9^k s \varepsilon^{-1} k \log n)$. The total evolution time T is obtained by allocating the error budget of the Heisenberg-limited frequency estimation across sectors and exhibits the $1/\epsilon_H$ scaling of the Heisenberg limit. The classical post-processing time is $\tilde{O}(3^k M \text{poly}(k, \log n))$, in expectation over the decoder's internal randomness, and thus sublinear in the ambient dimension $|\mathcal{P}_n^{(k)}| = O(n^k)$ in the relevant sparse regime. These bounds count eigenvalue-difference estimates and algebraic operations on the sketch data. The $O(n)$ cost of emitting and reading out each n -qubit experiment is attributed to the experimental interface, as quoted in the introduction.

VII. Outlook. Our work opens up several directions of further research. It is interesting to ask whether the $L^{3/2}$ overhead of the sector merge can be reduced by a more careful error allocation, and whether structured instead of uniformly random basis families can lower the 3^k coverage cost for physically motivated interaction classes. On a broader level, our results indicate that the toolbox of sublinear-time compressed sensing algorithms is naturally compatible with quantum measurement primitives, and we expect this connection to bear fruit beyond Hamiltonian learning.

-
- [1] A. Anshu, S. Arunachalam, T. Kuwahara, and M. Soleimanifar, *Nature Physics* **17**, 931 (2021).
 - [2] J. Haah, R. Kothari, and E. Tang, in *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, 2022), pp. 135–146.
 - [3] A. Bakshi, A. Liu, A. Moitra, and E. Tang, in *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, 2024), pp. 1037–1050.
 - [4] H.-Y. Huang, Y. Tong, D. Fang, and Y. Su, *Physical Review Letters* **130**, 200403 (2023).
 - [5] D. Hangleiter, I. Roth, J. Fuksa, J. Eisert, and P. Roushan, *Nature Communications* **15**, 9595 (2024).
 - [6] J. Eisert, D. Hangleiter, N. Walk, I. Roth, D. Markham, R. Parekh, U. Chabaud, and E. Kashefi, *Nature Reviews Physics* **2**, 382 (2020).
 - [7] N. Wiebe, C. Granade, C. Ferrie, and D. G. Cory, *Physical review letters* **112**, 190501 (2014).
 - [8] M. Ma, S. T. Flammia, J. Preskill, and Y. Tong, arXiv preprint arXiv:2410.18928 (2024).
 - [9] A. Shabani, M. Mohseni, S. Lloyd, R. L. Kosut, and H. Rabitz, *Physical Review A—Atomic, Molecular, and Optical Physics* **84**, 012107 (2011).
 - [10] K. Rudinger and R. Joynt, *Physical Review A* **92**, 052322 (2015).
 - [11] M. Sarovar, T. Proctor, K. Rudinger, K. Young, E. Nielsen, and R. Blume-Kohout, *Quantum* **4**, 321 (2020).
 - [12] S. Krinner, S. Lazar, A. Remm, C. K. Andersen, N. Lacroix, G. J. Norris, C. Hellings, M. Gabureac, C. Eichler, and A. Wallraff, *Physical Review Applied* **14**, 024042 (2020).
 - [13] P. Murali, D. C. McKay, M. Martonosi, and A. Javadi-Abhari, in *Proceedings of the twenty-fifth international conference on architectural support for programming languages and operating systems* (2020), pp. 1001–1016.
 - [14] K. Rudinger, C. W. Hogle, R. K. Naik, A. Hashim, D. Lobser, D. I. Santiago, M. D. Grace, E. Nielsen, T. Proctor, S. Seritan, et al., *PRX quantum* **2**, 040338 (2021).
 - [15] S. Bravyi, A. W. Cross, J. M. Gambetta, D. Maslov, P. Rall, and T. J. Yoder, *Nature* **627**, 778 (2024).
 - [16] K. Wang, Z. Lu, C. Zhang, G. Liu, J. Chen, Y. Wang, Y. Wu, S. Xu, X. Zhu, F. Jin, et al., *Nature Physics* pp. 1–7 (2026).
 - [17] J. T. Seeley, M. J. Richard, and P. J. Love, *The Journal of chemical physics* **137** (2012).
 - [18] E. Gaz, P. P. Popov, G. Pardo, M. Lewenstein, P. Hauke, and E. Zohar, *Physical Review Research* **7**, 033012 (2025).
 - [19] A. C. Gilbert, Y. Li, E. Porat, and M. J. Strauss, *SIAM Journal on Computing* **41**, 436 (2012).
 - [20] A. C. Gilbert, M. J. Strauss, J. A. Tropp, and R. Vershynin, in *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing* (2007), pp. 237–246.
 - [21] V. Nakos and Z. Song, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing* (2019), pp. 289–297.
 - [22] N. Fischer and V. Nakos, in *2025 IEEE 66th Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, 2025), pp. 1442–1457.
 - [23] M. Charikar, K. Chen, and M. Farach-Colton, in *International colloquium on automata, languages, and programming* (Springer, 2002), pp. 693–703.
 - [24] A. Cohen, W. Dahmen, and R. DeVore, *Journal of the American mathematical society* **22**, 211 (2009).