FEEDBACK-BASED LEARNING OF GROUND STATE PROPERTIES USING TENSOR CROSS INTERPOLATION

Florian J. Kiwit [1,2,*], **Bernhard Jobst** [3,4], **Andre Luckow** [1,2] & **Carlos A. Riofrío** [2]

[1]Ludwig Maximilian University, Munich, Germany
[2]BMW Group, Munich, Germany
[3]Technical University of Munich, Garching, Germany
[4]Munich Center for Quantum Science and Technology (MCQST), Munich, Germany

ABSTRACT

The integration of machine learning with quantum computing is a vital pathway for accelerating scientific discovery in many-body physics. However, mapping out the properties of quantum materials over extended parameter regimes, such as required for identifying phase transitions, remains computationally expensive because individual evaluations, e.g., on a quantum computer, are expensive. We propose an active learning loop to explore parameter spaces of quantum models in a more sample-efficient way. We treat the quantum processor as a high-cost oracle and use tensor cross interpolation (TCI) as a classical agent to control it. TCI efficiently reconstructs a surrogate model by adaptively querying the oracle at informative parameter points and interpolating the remaining parameter space. We benchmark this approach on the transverse-field XY model, focusing on learning the energy gap across its phase diagram. Our results show that TCI autonomously concentrates its sampling budget near critical phase boundaries, where the structure of ground state properties is most complex. In these regions, our framework achieves higher accuracy than standard regression baselines with the same sampling budget, providing a scalable strategy for AI-driven materials discovery in the era of early fault-tolerant quantum computing.

1 INTRODUCTION

Quantum computers allow the computation of certain quantum-material properties that may be beyond the computational reach of classical hardware. While early milestones for demonstrating quantum advantage primarily focused on sampling tasks (Arute et al., 2019), and were subsequently challenged by refined tensor-network methods (Pan & Zhang, 2022), recent advances shifted to the computation of physical expectation values that are hard to obtain classically. For instance, Kim et al. (2023) presented evidence for quantum utility in the pre-fault-tolerant era by computing time-dependent observables in the $2D$ Ising model, though their results quickly faced scrutiny from classical approaches using sophisticated tensor network techniques (Tindall et al., 2024; Begušić et al., 2024). More recently, Google Quantum AI and Collaborators (2025) measured out-of-time-ordered correlators that are sensitive to detailed features of the underlying dynamics, and currently remain beyond the reach of classical methods.

While this advocates that quantum computers can offer an advantage in computing properties for a given material configuration, one often faces the problem of needing to evaluate a property over a whole range of parameters to identify regions of interest. This is not directly addressed by quantum simulation algorithms, and mapping an entire parameter space remains a bottleneck. This motivates the integration of quantum hardware with machine learning frameworks to predict properties across broader parameter landscapes. Yet, many quantum processes exhibit a learning barrier where exponentially many samples would be required for any classical algorithm to learn them, such as the period-finding routines in Shor's algorithm (Shor, 1994). For these processes, the quantum advantage cannot be bypassed by data-driven classical approximations (Huang et al., 2021a;b). Fortu-

*florian.kiwit@ifi.lmu.de

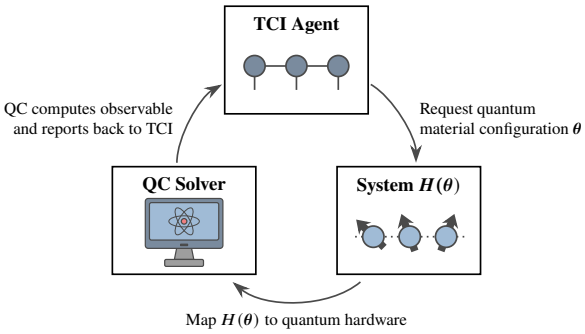


Figure 1: The TCI agent learns a surrogate model of the function by identifying informative parameter coordinates $\boldsymbol{\theta}$ where the current surrogate model is least accurate. It then requests a quantum material configuration. The QC Solver computes the physical observables and reports the values back to the agent. This process iteratively updates the surrogate model.

nately, this barrier does not extend to all physical systems. Huang et al. (2020; 2022) demonstrated that for gapped Hamiltonians, the function $f : \boldsymbol{\theta} \mapsto p$ mapping the parameters of a Hamiltonian $H(\boldsymbol{\theta})$ to its ground state properties p can be learned efficiently. This implies that a classical learner can construct an accurate surrogate model from a limited number of quantum samples, thereby avoiding the high cost of exhaustive parameter sweeps.

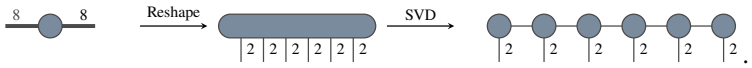
While Huang et al. (2020; 2022) established the theoretical learnability of these systems, the acquisition of the dataset relied on random sampling. This raises the question: how can we select specific parameter values that improve the model most effectively? In this work, we address this by introducing a feedback-based learning routine. We treat the quantum processor as a high-cost oracle within a feedback-based learning strategy based on tensor cross interpolation (TCI) (Oseledets & Tyrtyshnikov, 2010; Savostyanov, 2014). Our framework enables a classical agent to dynamically target informative regions of the Hamiltonian parameter space. Beyond serving as an active learning strategy, the function class TCI can learn is particularly well-suited for representing smooth functions with local nonanalyticities, such as cusps and discontinuities (Ali & Nouy, 2023; 2025; Jobst et al., 2024). This is crucial because the most interesting physics often concentrates near phase boundaries where the ground state properties can change nonanalytically.

In this work, we use the one-dimensional transverse-field XY model as a benchmark and focus on learning the map from Hamiltonian parameters to the spectral gap. We construct this surrogate using TCI in a closed-loop, query-efficient manner, treating the solver as a costly oracle that is evaluated only at adaptively selected parameter points. An overview of this workflow is provided in Figure 1. As a widely used baseline, we compare against Gaussian process regression (GPR) under the same sampling budget. We find that TCI concentrates evaluations in the physically relevant high-curvature regions—most notably near the critical line—and achieves substantially lower error there than active GPR modeling, highlighting the promise of feedback-based tensor-network methods for data-efficient phase-diagram exploration, achieving accurate reconstruction with only 1.6% of the full parameter grid.

2 METHOD

Our aim is to learn a multivariate function $f : \boldsymbol{\theta} \to \mathbb{R}$ that maps a vector of n Hamiltonian parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n) \in \mathbb{R}^n$ to a physical ground state property p, such as the ground state energy or correlation functions. TCI learns this function on a finely discretized grid in the form of a matrix product state (MPS) (Khoromskij, 2011)—also called a tensor train in the literature (Oseledets, 2011). In the two-dimensional case, if we exponentially finely discretize each variable to 2^L values, we may think of a $2^L \times 2^L$ matrix $A = [f(x_i, y_j)]_{i,j=1}^{2^L}$ as a lookup table for a function $f(x, y)$. We can reshape this array into a tensor $f_{\sigma_1,\ldots,\sigma_{\mathcal{L}}}$ with $\mathcal{L} = n\,L$ indices. To bring this tensor into the form of an MPS, one usually applies successive SVDs (Schollwöck, 2011). Diagrammatically,

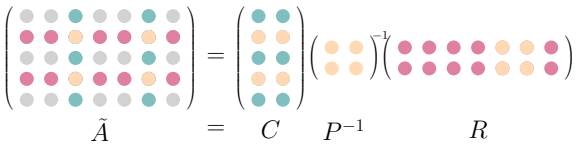
discretizing the function to a 8×8 matrix, the two steps look as follows:



(1)

A representation as an MPS can be very memory-efficient, if one requires high resolution on small length scales while simultaneously defining the function over a large domain. Storing all $2^{\mathcal{L}}$ function values on this grid may be intractable, but if the function f possesses structure, the resulting bond dimension χ of the MPS often remains quite small and one only needs to store $\mathcal{O}(\mathcal{L}\chi^2)$ parameters (Ali & Nouy, 2023; 2025; Jobst et al., 2024).

While the standard decomposition of a tensor into an MPS via successive SVDs is close to optimal, it requires access to all tensor entries. This is practically infeasible because populating the entire grid requires an exhaustive number of oracle evaluations, which is exactly the bottleneck we aim to avoid. Instead, there are alternative techniques based on interpolative matrix decompositions, such as the matrix cross interpolation (MCI). The MCI is given by



(2)

where C is a subset of columns, R is a subset of rows, and P is the matrix of pivots containing the intersection of the chosen rows and columns. For a matrix A with rank r, the decomposition is exact if we select r linearly independent rows and columns. For more general matrices that are only approximately low-rank, the interpolation quality depends heavily on the selection of suitable pivots. While finding the optimal set, defined by the maximum-volume submatrix (Goreinov et al., 1997a;b), is NP-hard (Bartholdi, 1982), there exist practical heuristic algorithms to find good pivots (Cortinovis et al., 2020).

The TCI generalizes the MCI from two-dimensional tensors (i.e., matrices) to multi-dimensional tensors by making successive use of the MCI. Notably, the MCI decompositions only ever need to be performed on slices of the high-dimensional tensor, instead of the full, exponentially large tensor. In this way, it can reconstruct a tensor with $2^{\mathcal{L}}$ entries by evaluating only $\mathcal{O}(m_{\text{sweep}}\chi^2\mathcal{L})$ entries of the tensor, where m_{sweep} is a small integer typically around 10 which counts the number of iterations in the algorithm until convergence. For a detailed discussion of TCI, see the recent review by Núñez Fernández et al. (2025).

In our setup, the TCI acts as an active learning agent that samples the function $f(\boldsymbol{\theta})$ at specific points $\boldsymbol{\theta}$ to iteratively construct an MPS representation. This creates a feedback loop where the classical learner controls the quantum computer: (i) The TCI algorithm identifies *interesting* points in the parameter space where the current approximation has the largest error. (ii) The quantum oracle $f(\boldsymbol{\theta})$ is queried only at these specific coordinates. (iii) The MPS is updated, and the bond dimension grows dynamically to capture the necessary complexity of the property surface. Once this representation is constructed, we gain access to the extensive MPS toolbox, which includes efficient methods for integration, optimization, and further analysis.

For comparison with our TCI-based approach, we also employ active GPR as a baseline method, with implementation details provided in App. A.1.

3 Results

We evaluate the proposed feedback-based learning framework on the task of learning the spectral gap of parameterized quantum spin Hamiltonians. As a representative benchmark, we consider the transverse-field XY model on a one-dimensional chain in the thermodynamic limit, defined by the Hamiltonian $H = -\sum_{j\in\mathbb{Z}} \left[\frac{1+\gamma}{2}\sigma_j^x\sigma_{j+1}^x + \frac{1-\gamma}{2}\sigma_j^y\sigma_{j+1}^y + h\sigma_j^z\right]$. The Hamiltonian is parameterized by the transverse field $h \in [0, 2]$ and the anisotropy parameter $\gamma \in [0, 2]$. This model is exactly solvable via a Jordan-Wigner transformation followed by a Bogoliubov transformation (Lieb et al., 1961), mapping the spin system to free fermions with dispersion relation $\epsilon_k = 2\sqrt{(h - \cos k)^2 + (\gamma \sin k)^2}$. We choose the spectral gap $\Delta(h, \gamma) = \min_k \epsilon_k$ as the target property $f(\boldsymbol{\theta})$, with $\boldsymbol{\theta} = (h, \gamma)$. A more detailed review of the Hamiltonian is provided in App. A.2.

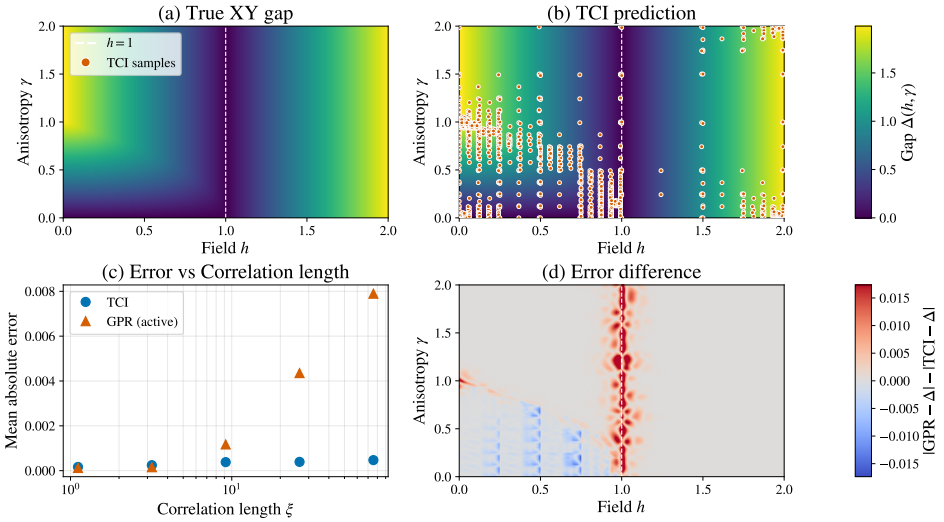


Figure 2: (a) The exact spectral gap $\Delta(h, \gamma)$ for the transverse-field XY model. The critical line at $h = 1$ is highlighted with a white dashed line. Another critical line for $\gamma = 0$ and $h \in [0, 1]$ lies at the boundary of the parameter regime, but is not highlighted. (b) TCI reconstruction of the gap function. Red markers show the TCI 1060 sampling locations. (c) Mean absolute error of TCI and active GPR predictors as a function of correlation length ξ. (d) Spatial distribution of approximation error difference $|\Delta_{\text{GPR}} - \Delta| - |\Delta_{\text{TCI}} - \Delta|$. The color scale is symmetrically clipped at the 99th percentile to avoid domination by extreme outliers. Red regions indicate where TCI outperforms GPR, with improvements concentrated near the phase transition.

We evaluate our feedback-based learning framework on the task of constructing a surrogate model for the spectral gap over a two-parameter Hamiltonian landscape under a fixed budget of expensive evaluations. Concretely, we treat the solver that returns the target quantity $f(h, \gamma)$, here, the XY-model spectral gap $\Delta(h, \gamma)$, as a high-cost oracle. While in our benchmark this oracle is implemented using the exact analytical solution, in practice this oracle could represent any expensive computation, such as a quantum computer evaluation. We use TCI as described in Sec. 2 to learn an efficient MPS representation from adaptively selected oracle queries (h, γ). In this closed-loop setting, TCI iteratively identifies informative regions of the parameter space and requests new oracle evaluations only where needed. We discretize the parameter space using a quantics grid with $L = 8$ resulting in a resolution of 256×256 for the $n = 2$ parameters. We limit the budget to 1000 evaluations, with TCI completing full sweeps until this budget is reached or exceeded, resulting in 1060 total queries. Details on TCI's convergence behavior are provided in App. A.3. As baseline, we also fit an active GPR model with a Matérn kernel (Genton, 2001) ($\nu = 1.5$) using the same sampling budget of 1060 evaluations.

Figure 2 (a,b) shows the exact gap landscape $\Delta(h, \gamma)$ together with the reconstruction obtained from TCI using 1060 oracle calls—only 1.6% of the total 256×256 grid points. While the critical line $h = 1$ is captured efficiently with relatively few queries, sampling concentrates most heavily along the nonanalytic boundary $h = 1 - \gamma^2$. This targeted acquisition enables the MPS surrogate to reproduce both the global structure of $\Delta(h, \gamma)$ and the sharp features associated with criticality. Figure 2(c) quantifies this behavior by relating predictive error to the correlation length ξ. An increasing correlation length indicates closeness to a phase transition, where it diverges and ground state properties can change most rapidly. As the correlation length increases, the active GPR baseline exhibits a pronounced degradation in accuracy, consistent with the smoothness imposed by its kernel. While both methods perform comparably at short correlation lengths, TCI maintains substantially lower error at high correlation lengths, demonstrating superior performance precisely in the regime with most rapidly changing features. Figure 2(d) localizes where these improvements occur by visualizing the signed error difference $|\Delta_{\text{GPR}} - \Delta| - |\Delta_{\text{TCI}} - \Delta|$. Positive (red) regions identify parameter values where TCI outperforms the GPR baseline, revealing most improvements in approximation quality are concentrated in a window around the phase transition. Notably, no

extended blue regions are observed, indicating that TCI does not underperform GPR anywhere in the vicinity of the transition. A light blue region is visible in the bottom-left area, bounded by the line $h = 1 - \gamma^2$ along which the remaining light red region concentrates. Along this line the dispersion goes from having two degenerate minima to having a single minimum, which leads to another nonanalytic feature but not a phase transition—see App. A.2. Separate depictions of the TCI and GPR errors are provided in App. A.4.

4 Discussion and Conclusion

Our results on the transverse-field XY model demonstrate that TCI achieves superior accuracy compared to active GPR using the same sampling budget. This performance required evaluating only 1.6% of the full parameter grid. Apart from providing a more accurate surrogate model, TCI is also computationally cheaper. Active GPR is limited by an $\mathcal{O}(N^4)$ cost with respect to the number of samples N, which comes from the $\mathcal{O}(N^3)$ kernel matrix inversion after drawing each new sample. The computational cost of the TCI scales as $\mathcal{O}(m_{\mathrm{sweep}} n L \chi^3)$, where n is the number of variables with discretization level 2^L and χ being the bond dimension. Since the TCI requests $N = \mathcal{O}(m_{\mathrm{sweep}} n L \chi^2)$ samples and the bond dimension typically does not need to increase with more samples, the computational cost is roughly linear in the number of samples. We observe this computational advantage significantly in practice. In our experimental setup (see App. A.5), performed on a standard laptop, the TCI learning completes in just 0.1 seconds. In contrast, the passive GPR requires 14 seconds, while the active GPR takes approximately 42 minutes to complete the same sampling budget.

Another key distinction between TCI and GPR lies in how they select sample points. Active GPR maximizes predictive uncertainty through posterior variance, which assumes the target function is sampled from a Gaussian probability distribution and requires well-matched kernels with consistent smoothness assumptions. Near quantum critical points, however, standard stationary kernels may become overconfident despite large local errors. In contrast, TCI employs error-driven adaptivity and new samples are selected where the current MPS representation fails to interpolate existing data.

As quantum hardware transitions toward fault tolerance, automated parameter space exploration becomes increasingly important for materials discovery. The TCI-based approach can address systems where exhaustive sampling is computationally prohibitive. Future work will extend this approach to other properties such as conductivities and correlation functions, and address interacting quantum systems where analytical solutions are unavailable. Another interesting direction would be investigating the framework's performance in higher-dimensional parameter spaces.

Acknowledgments

The authors thank Frank Pollmann for reviewing the manuscript and providing valuable comments, and Giovanni Concheri for insightful discussions and for sharing his TCI implementation guiding the early stages of this project.

References

Mazen Ali and Anthony Nouy. Approximation theory of tree tensor networks: Tensorized univariate functions. *Constructive Approximation*, 58:463–544, October 2023. doi: 10.1007/s00365-023-09620-w.

Mazen Ali and Anthony Nouy. Approximation theory of tree tensor networks: Tensorized multivariate functions. *arXiv:2101.11932*, December 2025. doi: 10.48550/arXiv.2101.11932.

Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando G. S. L. Brandao, David A. Buell, et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 2019.

John J. Bartholdi. A good submatrix is hard to find. *Operations Research Letters*, 1(5):190–193, November 1982. ISSN 0167-6377. doi: https://doi.org/10.1016/0167-6377(82)90038-4.

Tomislav Begušić, Johnnie Gray, and Garnet Kin-Lic Chan. Fast and converged classical simulations of evidence for the utility of quantum computing before fault tolerance. *Science Advances*, 10(3): eadk4321, 2024.

Alice Cortinovis, Daniel Kressner, and Stefano Massei. On maximum volume submatrices and cross approximation for symmetric semidefinite and diagonally dominant matrices. *Linear Algebra and its Applications*, 593:251–268, 2020.

Fabio Franchini, Alexander R. Its, and Vladimir E. Korepin. Renyi entropy of the XY spin chain. *Journal of Physics A: Mathematical and Theoretical*, 41(2):025302, December 2007. doi: 10.1088/1751-8113/41/2/025302.

Marc G Genton. Classes of kernels for machine learning: a statistics perspective. *Journal of machine learning research*, 2(Dec):299–312, 2001.

Google Quantum AI and Collaborators. Observation of constructive interference at the edge of quantum ergodicity. *Nature*, 646(8086):825–830, Oct 2025. ISSN 1476-4687. doi: 10.1038/s41586-025-09526-6. URL https://doi.org/10.1038/s41586-025-09526-6.

Sergei A. Goreinov, Eugene E. Tyrtyshnikov, and Nickolai L. Zamarashkin. A theory of pseudoskeleton approximations. *Linear Algebra and its Applications*, 261(1):1–21, August 1997a. ISSN 0024-3795. doi: https://doi.org/10.1016/S0024-3795(96)00301-1.

Sergei A. Goreinov, Nickolai L. Zamarashkin, and Eugene E. Tyrtyshnikov. Pseudo-skeleton approximations by matrices of maximum volume. *Mathematical Notes*, 62(4):515–519, October 1997b. ISSN 1573-8876. doi: 10.1007/BF02358985.

James Hensman, Nicolo Fusi, and Neil D. Lawrence. Gaussian processes for big data. *arXiv preprint arXiv:1309.6835*, 2013.

Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, 2020.

Hsin-Yuan Huang, Michael Broughton, Masoud Mohseni, Ryan Babbush, Sergio Boixo, Hartmut Neven, and Jarrod R McClean. Power of data in quantum machine learning. *Nature communications*, 12(1):2631, 2021a.

Hsin-Yuan Huang, Richard Kueng, and John Preskill. Information-theoretic bounds on quantum advantage in machine learning. *Physical Review Letters*, 126(19):190505, 2021b.

Hsin-Yuan Huang, Richard Kueng, Giacomo Torlai, Victor V. Albert, and John Preskill. Provably efficient machine learning for quantum many-body problems. *Science*, 377(6613):eabk3333, 2022.

Bernhard Jobst, Kevin Shen, Carlos A. Riofrío, Elvira Shishenina, and Frank Pollmann. Efficient MPS representations and quantum circuits from the Fourier modes of classical image data. *Quantum*, 8:1544, December 2024. ISSN 2521-327X. doi: 10.22331/q-2024-12-03-1544.

Nick G. Jones and Ruben Verresen. Asymptotic correlations in gapped and critical topological phases of 1d quantum systems. *Journal of Statistical Physics*, 175(6):1164–1213, April 2019. ISSN 1572-9613. doi: 10.1007/s10955-019-02257-9.

Boris N. Khoromskij. $O(d \log N)$-quantics approximation of N-d tensors in high-dimensional numerical modeling. *Constructive Approximation*, 34(2):257–280, 2011.

Youngseok Kim, Andrew Eddins, Sajant Anand, Ken Xuan Wei, Ewout Van Den Berg, Sami Rosenblatt, Hasan Nayfeh, Yantao Wu, Michael Zaletel, Kristan Temme, et al. Evidence for the utility of quantum computing before fault tolerance. *Nature*, 618(7965):500–505, 2023.

Elliott Lieb, Theodore Schultz, and Daniel Mattis. Two soluble models of an antiferromagnetic chain. *Annals of Physics*, 16(3):407–466, 1961.

Gerhard Müller and Robert E. Shrock. Implications of direct-product ground states in the one-dimensional quantum XYZ and XY spin chains. *Phys. Rev. B*, 32:5845–5850, November 1985. doi: 10.1103/PhysRevB.32.5845.

Yuriel Núñez Fernández, Marc K. Ritter, Matthieu Jeannin, Jheng-Wei Li, Thomas Kloss, Thibaud Louvet, Satoshi Terasaki, Olivier Parcollet, Jan von Delft, Hiroshi Shinaoka, et al. Learning tensor networks with tensor cross interpolation: new algorithms and libraries. *SciPost Physics*, 18(3): 104, 2025.

Ivan Oseledets and Eugene Tyrtyshnikov. TT-cross approximation for multidimensional arrays. *Linear Algebra and its Applications*, 432(1):70–88, January 2010. ISSN 0024-3795. doi: 10.1016/j.laa.2009.07.024.

Ivan V. Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011.

Feng Pan and Pan Zhang. Simulation of quantum circuits using the big-batch tensor network method. *Physical Review Letters*, 128(3):030501, 2022.

Carl Edward Rasmussen. Gaussian processes in machine learning. In *Summer school on machine learning*, pp. 63–71. Springer, 2003.

Dmitry V. Savostyanov. Quasioptimality of maximum-volume cross interpolation of tensors. *Linear Algebra and its Applications*, 458:217–244, October 2014. ISSN 0024-3795. doi: 10.1016/j.laa.2014.06.006.

Ulrich Schollwöck. The density-matrix renormalization group in the age of matrix product states. *Annals of Physics*, 326(1):96–192, January 2011. doi: 10.1016/j.aop.2010.09.012.

Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando De Freitas. Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104(1): 148–175, 2015.

Peter W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms. In *Proceedings of the 35th Annual Symposium on Foundations of Computer Science*, volume 124, pp. 124–134, 1994.

Joseph Tindall, Matthew Fishman, E. Miles Stoudenmire, and Dries Sels. Efficient tensor network simulation of IBM's Eagle kicked Ising experiment. *PRX Quantum*, 5(1):010308, 2024.

Michalis Titsias. Variational learning of inducing variables in sparse Gaussian processes. In *Artificial intelligence and statistics*, pp. 567–574. PMLR, 2009.

Ruben Verresen, Nick G. Jones, and Frank Pollmann. Topology and edge modes in quantum critical chains. *Phys. Rev. Lett.*, 120:057001, January 2018. doi: 10.1103/PhysRevLett.120.057001.

A Appendix

A.1 Gaussian process regression.

Gaussian Processes (GPs) are a standard tool for Bayesian active learning (Rasmussen, 2003; Shahriari et al., 2015). In GP regression, the predictive distribution at a test point $\boldsymbol{x}$ given training data $\mathcal{D}$ follows

$$f(\boldsymbol{x}) \mid \mathcal{D} \sim \mathcal{N}(\mu(\boldsymbol{x}), \sigma^2(\boldsymbol{x})), \tag{3}$$

where $\mu(\boldsymbol{x})$ is the posterior mean and $\sigma^2(\boldsymbol{x})$ is the posterior variance. In uncertainty sampling active learning, the next evaluation point is chosen as

$$\boldsymbol{x}^* = \arg\max_{\boldsymbol{x}} \sigma^2(\boldsymbol{x}), \tag{4}$$

selecting locations where the model is most uncertain. This principled uncertainty quantification enables acquisition strategies that balance exploration and exploitation. However, exact GP regression scales cubically with the number of observations, $\mathcal{O}(N^3)$, and requires $\mathcal{O}(N^2)$ memory. Active learning compounds this cost, as each of the N acquisition steps requires a full GP update, resulting in overall $\mathcal{O}(N^4)$ scaling. While sparse and variational GP approximations (Titsias, 2009; Hensman et al., 2013) mitigate this issue, these methods introduce their own approximation errors and require careful parameter tuning that depends on the complexity of the target function.

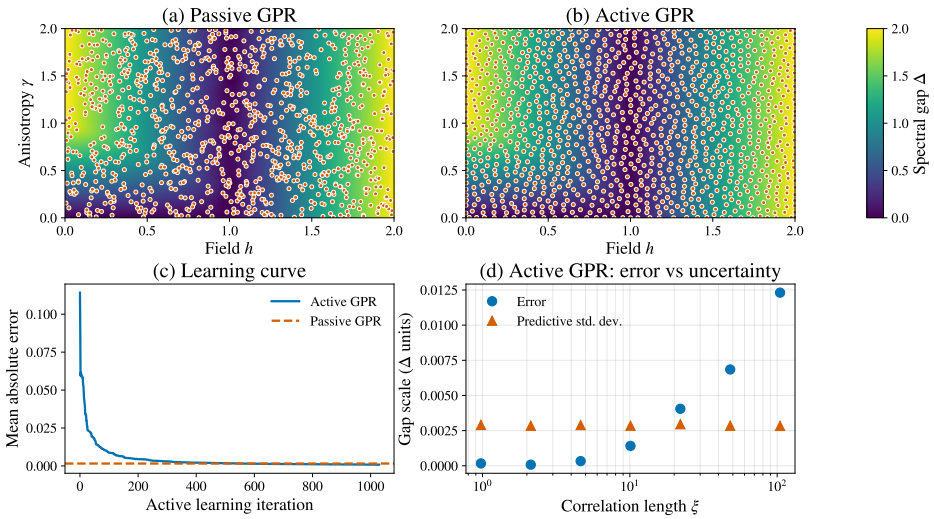


Figure 3: (a) Passive GPR sampling of the parameter space (h, γ) using a uniform or low-discrepancy design. (b) Active GPR sampling driven by variance-based acquisition. (c) Mean absolute error (MAE) as a function of active learning iterations. The active GPR (solid blue line) achieves lower error than the passive GPR baseline (dashed red line) for the same total sampling budget. (d) Evaluation of the GPR surrogate performance across different correlation lengths ξ. The plot compares the actual interpolation error against the predictive standard deviation.

In our experimental setup, we implement active GP regression using a Matérn kernel (Genton, 2001) with $\nu = 1.5$. The Matérn kernel controls the smoothness of function samples: functions drawn from a GP with Matérn kernel are $\lfloor \nu \rfloor$ times differentiable, where $\lfloor \cdot \rfloor$ denotes the floor function. Thus, $\nu = 1.5$ produces once-differentiable functions, providing a balance between the overly smooth RBF kernel (which corresponds to $\nu \to \infty$) and the non-differentiable exponential kernel ($\nu = 0.5$). We sample a candidate pool of 5000 points at each iteration and select only the single point with highest predictive variance. This conservative batch size ensures each acquisition is maximally informative but results in slower exploration compared to TCI's adaptive tensor decomposition. In the context of learning the spectral gap over discretized parameter grids, GP-based approaches face two practical limitations. First, the cost of updating the surrogate model increases rapidly as more samples are acquired, making dense sampling regimes prohibitively expensive. Second, standard kernels impose smoothness assumptions that may be violated near phase boundaries or critical regions, where sharp variations occur. While specialized kernels can partially address this issue, their design typically requires problem-specific insight.

As shown in Fig. 3, active GPR (b) distributes samples more spread out than passive random sampling (a), achieving lower prediction error with the same sampling budget (Fig. 3(c)). However, the variance-based acquisition strategy tends to spread samples across regions of high uncertainty rather than concentrating on regions where the surrogate model makes the largest error, which tend to be the most physically relevant, critical parameter regimes.

A.2 The transverse-field XY model

The transverse-field XY model can be solved by a Jordan-Wigner transformation to noninteracting fermions. For each site j, we can introduce two Majorana operators $\gamma_{2j}, \gamma_{2j+1}$ that square to identity $\gamma_k^2 = 1$ and satisfy the anti-commutation relation $\{\gamma_k, \gamma_\ell\} = 2\delta_{k,\ell}$ for all $k, \ell \in \mathbb{Z}$. After the Jordan-Wigner transform, the Hamiltonian then becomes

$$H = \sum_{j \in \mathbb{Z}} \tfrac{1+\gamma}{2}\, i\gamma_{2j+1}\gamma_{2j+2} + \tfrac{1-\gamma}{2}\, i\gamma_{2j+1}\gamma_{2j-2} - h\, i\gamma_{2j+1}\gamma_{2j}, \tag{5}$$

which describes a one-dimensional superconducting chain. This fermionic Hamiltonian can be diagonalized with a Bogoliubov transformation.

In our case it is more convenient to realize that this Hamiltonian belongs to a special class of free-fermion Hamiltonians (the BDI class, i.e., spinless fermions with time-reversal symmetry) for which a number of exact results have been derived (Verresen et al., 2018; Jones & Verresen, 2019). Any Hamiltonian in this class can be constructed as a linear combination

$$H = \frac{1}{2}\sum_{\alpha\in\mathbb{Z}} t_\alpha H_\alpha \tag{6}$$

from terms of the form

$$H_\alpha = \sum_{j\in\mathbb{Z}} i\gamma_{2j+1}\gamma_{2j+2\alpha}. \tag{7}$$

To get a local Hamiltonian with a finite coupling range all t_α with α above or below some threshold are zero. The information about the coupling strengths in the Hamiltonian can be equivalently encoded in a Laurent polynomial

$$f(z) = \sum_{\alpha\in\mathbb{Z}} t_\alpha z^\alpha. \tag{8}$$

It has been shown that much information about the physical properties can be obtained from the function $f(z)$. The single-particle excitation spectrum is given by $\epsilon_k = |f(e^{ik})|$, i.e., the magnitude on the unit circle. The ground state is given by the absence of single-particle excitations, and thus the energy gap is simply the minimum of the single-particle spectrum $\Delta = \min_k \epsilon_k$. If we denote the roots of $f(z)$ by ζ_i, the correlation length of the state is given by $\xi = \max_i 1/|\ln(|\zeta_i|)|$. Note that $f(z)$ has a pole at $z = 0$ if there are nonzero couplings with negative α, with the order of the pole N_p given by the most negative α with a nonzero coupling t_α, but if we factor out the contribution $1/z^{N_p}$ from the pole, we are left with finding the roots of a polynomial.

In our case for the transverse-field XY model, the corresponding Laurent polynomial is given by

$$f(z) = (1+\gamma)\,z + (1-\gamma)\,\frac{1}{z} - 2h = \frac{1}{z}\big((1+\gamma)\,z^2 - 2h\,z + (1-\gamma)\big). \tag{9}$$

Thus, for the single-particle excitation spectrum we find

$$\begin{aligned}\epsilon_k &= |f(e^{ik})| = |(1+\gamma)\,e^{ik} + (1-\gamma)\,e^{-ik} - 2h| \\ &= \sqrt{[(1+\gamma)\cos k + (1-\gamma)\cos k - 2h]^2 + [(1+\gamma)\sin k - (1-\gamma)\sin k]^2} \\ &= \sqrt{(2\cos k - 2h)^2 + (2\gamma\sin k)^2} = 2\sqrt{(h-\cos k)^2 + (\gamma\sin k)^2},\end{aligned} \tag{10}$$

which is the dispersion given in the main text. The quantum phase transitions occur at those points where there are zero-energy excitations, which we can see happens either at zero momentum if $h = 1$, which is the line shown in Fig. 2 in the main text, at π momentum if $h = -1$, or if $\gamma = 0$ and $h \in [-1, 1]$ at momentum $k = \pm\arccos(h)$. Away from the phase transitions, we can obtain the spectral gap Δ by finding the minimum of the excitation spectrum. To find the extrema, we consider its derivative

$$\partial_k\epsilon_k = \frac{2}{\epsilon_k}\left[2(h-\cos k)\sin k + 2\gamma^2\sin k\cos k\right] = \frac{4}{\epsilon_k}\left[h\sin k + (\gamma^2-1)\sin k\cos k\right] \tag{11}$$

and set it to zero. Since we consider the gapped phases where $\epsilon_k > 0$, we need to solve

$$h\sin k + (\gamma^2-1)\sin k\cos k = 0, \tag{12}$$

which has solutions $k = 0$, $k = \pi$ and—if $|h| < |1-\gamma^2|$—two additional solutions at

$$k = k^*_\pm \equiv \pm\arccos\frac{h}{1-\gamma^2} \tag{13}$$

that coincide with either $k = 0$ or $k = \pi$ on the line $|h| = |1-\gamma^2|$. This gives the extremal values $\epsilon_0 = 2|h-1|$ at $k = 0$, $\epsilon_\pi = 2|h+1|$ at $k = \pi$, and

$$\begin{aligned}\epsilon_{k^*_\pm}/2 &= \sqrt{\left[h - \left(\frac{h}{1-\gamma^2}\right)\right]^2 + \gamma^2\left[1 - \left(\frac{h}{1-\gamma^2}\right)^2\right]} = \sqrt{\left(\frac{h\gamma^2}{1-\gamma^2}\right)^2 + \gamma^2 - \left(\frac{h\gamma}{1-\gamma^2}\right)^2} \\ &= \sqrt{\left(\frac{h\gamma}{1-\gamma^2}\right)^2(\gamma^2-1) + \gamma^2} = \sqrt{\gamma^2 - \frac{(h\gamma)^2}{1-\gamma^2}} = |\gamma|\sqrt{1 - \frac{h^2}{1-\gamma^2}}\end{aligned} \tag{14}$$

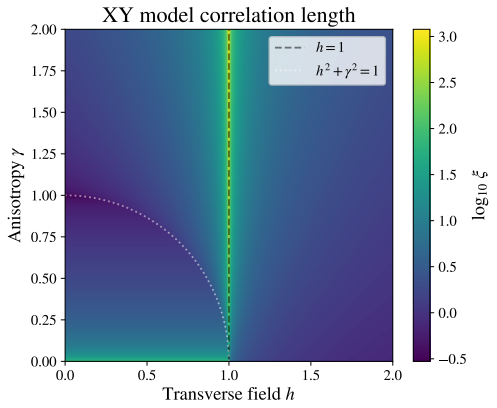


Figure 4: Correlation length $\xi(h, \gamma)$ of the transverse-field XY model across the two-dimensional parameter space spanned by the transverse field h and the anisotropy γ. The correlation length diverges along the critical line $h = 1$, signaling the quantum phase transition between the paramagnetic and ferromagnetic phases.

at $k = k^*_\pm$. If we focus on the parameter regime $h, \gamma \geq 0$ considered in the main text, then we have $\epsilon_0 \leq \epsilon_\pi$. In the regime $h \geq |1 - \gamma^2|$, there are no other extrema and this the minimum. For $h < |1-\gamma^2|$, we compare ϵ_0 with $\epsilon_{k^*_\pm}$ to find when the minimum is at $k^*_\pm$. Defining $s = \mathrm{sign}(1-\gamma^2)$, we have

$$\begin{aligned} \epsilon^2_{k^*_\pm} = 4\gamma^2\left(1 - \tfrac{h^2}{1-\gamma^2}\right) &< 4(h-1)^2 = \epsilon_0^2 \\ \implies \quad s\gamma^2\left(1-\gamma^2-h^2\right) &< s(1-\gamma^2)(h^2-2h+1) \\ \implies \quad 0 &< s\left(h^2 - 2(1-\gamma^2)h + (1-\gamma^2)^2\right) \\ \implies \quad 0 &< s\left(h-(1-\gamma^2)\right)^2. \end{aligned} \tag{15}$$

This is only fulfilled if the sign s is positive, i.e., if $1 - \gamma^2 > 0$ or $\gamma < 1$.

Taken together, the gap function for $h, \gamma \geq 0$ is thus

$$\Delta(h,\gamma) = \begin{cases} 2|\gamma|\sqrt{1 - \frac{h^2}{1-\gamma^2}} & \text{if } \gamma < 1 \text{ and } h < 1-\gamma^2, \\ 2|h-1| & \text{otherwise.} \end{cases} \tag{16}$$

Note that the parabola $h = 1-\gamma^2$ where the two function definitions meet is the second line in Fig. 2 where the TCI outperforms the GPR apart from the phase transition. Note that in both cases the gap function becomes nonanalytic. Around the critical line $h = 1$ the gap function is $\sim |h - 1|$ which is nondifferentiable. To see that also along $h = 1 - \gamma^2$ the gap function is nonanalytic, consider the h-derivative. Approaching the line from below, the h-derivative is

$$\partial_h \Delta(h,\gamma)\Big|_{h \nearrow 1-\gamma^2} = \frac{2|\gamma|h}{\sqrt{1-\frac{h^2}{1-\gamma^2}}}\Bigg|_{h \nearrow 1-\gamma^2} = \frac{-2|\gamma|(1-\gamma^2)}{\sqrt{1-\frac{(1-\gamma^2)^2}{1-\gamma^2}}} = -2(1-\gamma^2), \tag{17}$$

whereas when approaching from above it is

$$\partial_h \Delta(h,\gamma)\Big|_{h \searrow 1-\gamma^2} = -2. \tag{18}$$

These two results agree only for $\gamma = 0$, so the gap function is nondifferentiable along $h = 1 - \gamma^2$ for $0 < \gamma < 1$.

Finally, we can also compute the correlation length $\xi(h, \gamma)$ as a function of the Hamiltonian parameters. For this, we need to find the roots of the function in Equation 9. We obtain

$$\zeta_\pm = \frac{2h \pm \sqrt{4h^2 - 4(1+\gamma)(1-\gamma)}}{2(1+\gamma)} = \frac{h \pm \sqrt{h^2+\gamma^2-1}}{1+\gamma}. \tag{19}$$

The correlation length can then be computed as $\xi = \max_{i=\pm} 1/|\ln(|\zeta_i|)|$, which is shown as a color plot in Figure 4. Note that along the critical line $h = 1$ the roots are $\zeta_\pm = (1\pm|\gamma|)/(1+\gamma)$, and along the critical line $\gamma = 0$ with $0 \leq h \leq 1$ the roots are $\zeta_\pm = h \pm i\sqrt{1-h^2}$. In both cases at least one of the absolute values satisfies $|\zeta_i| = 1$ and the correlation length diverges $\xi = 1/|\ln(|\zeta_i|)| = \infty$, as expected from a critical point and visible in the plot. A second feature is visible in the plot along the circle $h^2 + \gamma^2 = 1$, where the roots change from being real to complex. Within the circle, the roots are the complex conjugate pair

$$\zeta_\pm = \frac{h \pm i\sqrt{1-h^2-\gamma^2}}{1+\gamma} \tag{20}$$

which have the absolute value

$$|\zeta_\pm| = \frac{\sqrt{h^2+1-h^2-\gamma^2}}{1+\gamma} = \frac{\sqrt{1-\gamma^2}}{1+\gamma} = \sqrt{\frac{1-\gamma}{1+\gamma}}. \tag{21}$$

Notably, the absolute value of the roots is independent of h, and therefore so is the correlation length in that regime. Conversely, outside the circle the roots are real and depend on h. Thus, the circle $h^2 + \gamma^2 = 1$ defines a line where the correlation length is nonanalytic but finite.

As a final comment, note that the free-fermion picture only allows to treat states that do not spontaneously break the $\mathbb{Z}_2$ symmetry, which is what would happen in the spin-model in the thermodynamic limit. If one allows superpositions that break the $\mathbb{Z}_2$ symmetry of the two degenerate symmetric ground states, another peculiarity exists along the line $h^2 + \gamma^2 = 1$. There, one can find a suitable superposition such that the ground state is a product state with correlation length $\xi = 0$ (Müller & Shrock, 1985; Franchini et al., 2007).

A.3 TCI CONVERGENCE AND ADAPTIVE COMPLEXITY

Figure 5(a) shows the decay of the relative in-sample error as a function of the TCI rank for three grid resolutions[1]. In all cases the error drops rapidly and reaches the target tolerance of 10^{-3} at a rank of approximately 13–15, largely independent of the grid resolution. This resolution-independence indicates that the rank required by TCI is governed by the intrinsic complexity of the spectral gap function rather than by the size of the discretization grid. Figure 5(b) shows the internal structure of the learned model by plotting the virtual bond dimensions D_ℓ against the tensor site index ℓ. Rather than growing monotonically, the bond dimension peaks at intermediate scales before stabilizing or even decreasing at finer resolutions. This adaptive behavior demonstrates that TCI automatically discovers the intrinsic complexity of the underlying function—higher resolutions require larger bond dimensions to capture fine-scale features, but the algorithm avoids over-parameterization by adapting χ to the actual structure of the spectral gap landscape rather than the worst-case scenario of the discretization grid.

A.4 SPATIAL ERROR DISTRIBUTION

Fig 6 reveals differences in how approximation errors are distributed across the parameter space for TCI and GPR methods. Figure 6(a) shows that TCI concentrates errors in narrow regions in the parabolic boundary $h = 1 - \gamma^2$. In contrast, Fig. 6(b,c) demonstrates that GPR develops correlated error ridges that propagate across extended regions of the parameter space, creating interpolation corridors that connect the sampled training points (red markers). These filamentary patterns arise from the kernel's smoothness constraints, which create systematic interpolation bias when the underlying function violates the GP's regularity assumptions. The logarithmic color scale reveals that TCI achieves several orders of magnitude better accuracy in smooth regions, while both methods struggle near critical points where the physics is inherently nonsmooth. The logarithmic color scale reveals that TCI achieves several orders of magnitude better accuracy in smooth regions. Notably, TCI accurately resolves the quantum critical line ($h = 1$). The remaining localized errors (seen in the bottom left) are concentrated along the parabolic boundary $h = 1 - \gamma^2$. As presented in App. A.2, the gap function is nonanalytic in this area.

[1]The steep decline at the final rank is an artifact of the error tracking mechanism in `xfac`. While the algorithm records the maximum error encountered across all iterations for earlier ranks, it terminates immediately upon reaching the target tolerance. As a result, this highest rank is evaluated over far fewer sample points than the previous ranks, which may lead to an underestimated global error.

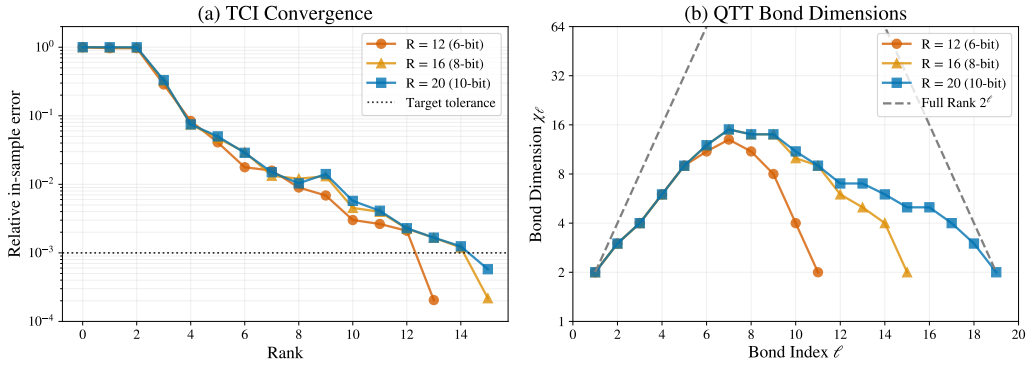


Figure 5: (a) Convergence of the relative in-sample error as a function of the TCI rank for three grid resolutions. (b) Profile of the virtual bond dimensions D_ℓ across the tensor train sites $\ell = 1, \ldots, nL$ for the final converged model. The sites ℓ correspond to decreasing length scales $2^{-\ell}$ in the parameter space.

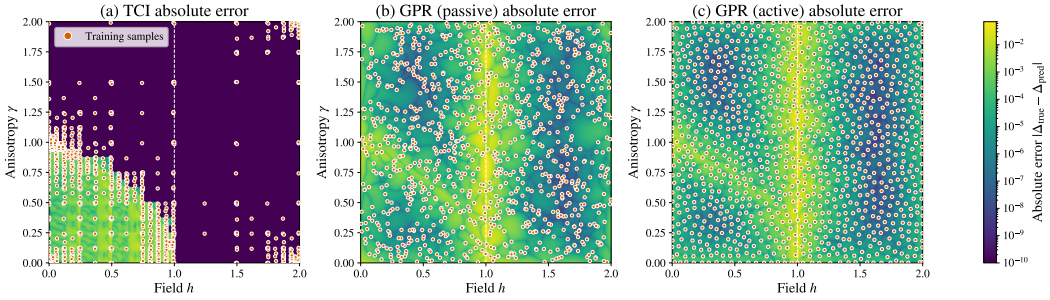


Figure 6: Spatial distribution of the absolute approximation error for the XY-model spectral gap $\Delta(h, \gamma)$ using (a) TCI and (b) passive and (c) active GPR under the same fixed sampling budget. Errors are shown on a logarithmic color scale to resolve variations across several orders of magnitude. To ensure numerical stability for the logarithmic visualization a small constant with value 10^{-10} was added to the absolute error values before plotting. Red markers indicate the locations of oracle evaluations used to construct the respective surrogate models. While TCI exhibits highly localized error concentrated near regions of nonanalytic behavior, the GPR baseline develops extended filamentary error structures across the parameter space. These correlation ridges arise from kernel-induced interpolation bias and reflect the global smoothness assumptions imposed by the GP prior, which are violated near critical regions of the phase diagram.

A.5 Experimental setup

TCI was performed using the Python bindings of the `xfac` library, which provides efficient implementations of Quantics tensor trains. For the experiments involving GPR, we employed the `GaussianProcessRegressor` from the `sklearn` library.