Imperial College London

Imperial College of Science, Technology and Medicine

Department of Computing

MSc Advanced Computing

Efficient Learning of Quantum Systems at Thermal Equilibrium

Author:
Oliver Kanders

CID:
06019970

Imperial College Supervisor:
Roberto Bondesan

Second Marker:
Dario Paccagnan

September 15, 2025

Abstract

We study the efficient prediction of local observables in quantum lattice systems at zero and finite temperature through a unified locality framework that covers both gapped ground states and Gibbs states. Leveraging Lieb-Robinson bounds together with quasi-local propagation filters (quasi-adiabatic spectral flow at $T{=}0$ and quantum belief propagation at $T{>}0$) we derive parallel derivative identities and show that a local perturbation effectuates an exponentially decaying local response. In the thermal setting, we arrive at an explicit decay rate $\mu_\beta = \min\{\nu/2,\ \mu\,\pi/[8\beta\,(v_{\mathrm{LR}}+\pi/\beta)]\}$, which leads us to a logarithmic radius-accuracy law: $\delta_\beta(\varepsilon) = \mu_\beta^{-1}(\log(K_\beta/\varepsilon) + c_D)$. These results guide us directly to a learning theory: a locality-aware feature map and ℓ_1-constrained linear predictors (LASSO) achieve population risk $(\varepsilon_1+\varepsilon_2)^2+\varepsilon_3$ with sample complexity $\widetilde{O}(N(\varepsilon_1)\varepsilon_3^{-2}\log|S|)$, and symmetry (equivariance) reduces the logarithmic factor to $\log|S/G|$, yielding constant-in-n behavior on translation-invariant rings. For power-law interactions we recover polynomial radii for $\alpha > 2D$ and observe a breakdown of efficient locality when $\alpha \le 2D$. We verify our theory with a purification-TEBD pipeline on disordered Heisenberg and long-range Ising chains across temperatures: (i) open-boundary Heisenberg demonstrates logarithmic growth of required samples; (ii) periodic/equivariant Heisenberg is nearly constant in n; (iii) Ising with $\alpha = 3$ shows near logarithmic growth; and (iv) $\alpha = 1.5$ displays significantly steeper, approximately linear growth. Our framework advances a single, temperature-aware technique from physical locality to finite-sample learning guarantees and practical algorithms for predicting thermal quantum observables.

Acknowledgments

I would like to express my gratitude to Dr. Roberto Bondesan for his wonderful mentorship over the course of this research, guiding me toward challenging questions and compelling directions.

I would also like to extend my gratitude to Štěpán Šmíd for all his generous advice that helped shape this work.

Contents

1 Introduction **5**
- 1.1 The challenge of predicting many-body observables . . . 5
- 1.2 Research aim and contributions . . . 6
- 1.3 Ethical considerations . . . 7
- 1.4 Declaration . . . 8
- 1.5 Code Repository . . . 8

2 Preliminaries **9**
- 2.1 Local Hamiltonians and locality . . . 9
- 2.2 Spectral gaps and gapped phases . . . 10
- 2.3 Lieb-Robinson Bounds (effective light cones) . . . 10
- 2.4 Exponential clustering of correlations . . . 12
  - 2.4.1 Local truncation . . . 13
- 2.5 Thermal states and finite-temperature correlations . . . 14
- 2.6 Quantum Markov properties and conditional independence . . . 16

3 Quantum Information Bound **18**
- 3.1 Introduction . . . 18
- 3.2 Setup and Mathematical Framework . . . 18
  - 3.2.1 Basic definitions and notation . . . 18
  - 3.2.2 Derivative identities and the unified formula . . . 20
- 3.3 The Locality Principle . . . 22
  - 3.3.1 Quasi-locality of the propagation filter Φ_σ . . . 22
  - 3.3.2 Local perturbation leads to local response . . . 25
- 3.4 Consequences for Observable Dynamics . . . 31
  - 3.4.1 Sensitivity to Hamiltonian parameters . . . 31

4 Machine Learning Bound **33**
- 4.1 Setup, hypothesis class, and data . . . 33
- 4.2 Locality reduction: choosing $\delta(\varepsilon_1)$ . . . 35
  - 4.2.1 Short range (exponential) . . . 36
  - 4.2.2 Power law interactions $\alpha > 2D$ . . . 36
- 4.3 Unified LASSO generalization bound . . . 37
- 4.4 Equivariance and sample-complexity reduction . . . 38
- 4.5 Random features and convolutional realizations . . . 38
- 4.6 Complexity summary and limitations . . . 39

5 Experiments **40**
- 5.1 Data . . . 40
  - 5.1.1 Models studied . . . 40
  - 5.1.2 Targets and normalization . . . 40
  - 5.1.3 State preparation (imaginary-time TEBD) . . . 41
  - 5.1.4 TenPy realization (nearest-neighbor) [13] . . . 41
  - 5.1.5 Validation and scope . . . 41
- 5.2 Machine learning setup . . . 41
  - 5.2.1 Random Fourier features [24] . . . 42

5.2.2 Predictor and training 42
5.2.3 Sample-complexity protocol 43
5.3 Results 44
5.3.1 Heisenberg model 44
5.3.2 Long-range Ising model 48

6 Conclusion and Future Work **51**
6.1 Conclusion 51
6.1.1 Summarized Contributions 52
6.2 Future Work 52

1 Introduction

Predicting properties of quantum many-body systems is a central challenge in physics and quantum information. Direct simulation of ground or thermal (Gibbs) states is generally unmanageable because the Hilbert-space dimension grows exponentially with the number of particles; storing an arbitrary n-qubit state requires 2^n complex amplitudes. In worst cases, ground state problems are QMA-hard [15]. Fortunately, physically realistic systems often exhibit structure that renders local predictions as tractable, whether that structure be geometric locality of interactions, spectral gaps, or clustering of correlations. In short-range gapped systems, Lieb-Robinson (LR) bounds impose a finite "light cone" for information propagation. Together with gap assumptions, they can result in exponential clustering of correlations and quasi-local responses [19, 12, 10]. These components make up quasi-adiabatic spectral flow (QASF), which transports ground states within a phase by almost-local unitaries [12, 2]. We are able to understand that distant degrees of freedom have exponentially small influence on local observables, so one may learn or approximate local physics using only nearby data.

1.1 The challenge of predicting many-body observables

We confront the challenge that if given a local Hamiltonian, can we efficiently predict expectation values of *local* observables in their ground state ($T = 0$) or Gibbs state ($T > 0$) to accuracy ε, using resources that scale polylogarithmically (ideally, constantly) in system size.

A naive approach reconstructs the full state (exponential cost). However, a more promising method is through learning: train from measurements or smaller instances and generalize, provided that the inductive bias reflects geometry and locality. A growing literature shows that, within a fixed phase of matter, geometric priors drastically reduce sample complexity. For ground states, recent works show polylogarithmic or even sample complexity independent of size when the predictor is locality aware and trained within one phase, see, e.g., the main theorems in [18, Sec. 1] and [27, Sec. 1]. Furthermore, [27] extend the framework to settings with long-range interactions and symmetries by exploiting locality and equivariance.

Why locality enables learning

For gapped short-range models, LR bounds and QASF imply that local perturbations produce exponentially decaying responses in the ground state [12, 2]. Thus the value of a local observable depends weakly on couplings far from its support, enabling local truncation arguments and neighborhood predictors.

For thermal states, we notably find that the same pattern holds. Gibbs states of short-range Hamiltonians exhibit exponential clustering under standard conditions, and local observables satisfy *locality of temperature*, meaning that their values depend on couplings in a finite neighborhood [16]. Quantum belief propagation (QBP) imparts a valuable mathematical framework for thermal responses. The QBP derivative identity expresses parameter derivatives of thermal expectations through a time filtered Heisenberg evolution with the explicit kernel $\kappa_\beta(t)$. By combining LR with a Δ-split analysis of κ_β [25, Eq. (III.5), Lem. III.2], we are able to establish quasi-locality of the filtered operator (we will go into further detail of all these steps). As a result, thermal responses decay exponentially with a rate set by the Lieb-Robinson data and β. This validates our overall

goal to learn local observables from local information.

In both the ground state ($T = 0$) and thermal ($T > 0$) cases, physical locality fundamentally constrains the statistical complexity of the learning problem. This constraint manifests through three connected mechanisms: first, Lieb-Robinson bounds ensure that information propagates with finite velocity, producing quasi-local response functions. Second, the presence of spectral gaps or exponential clustering of correlations restricts the effective range of quantum correlations. Third, the mathematical frameworks of quasi-adiabatic spectral flow (QASF) for ground states and quantum belief propagation (QBP) for thermal states elicit constructions of quasi-local generators and filters, respectively. Taken together, these three mechanisms suggest that, in order to reach accuracy ε on a local observable, only couplings within a finite radius $\delta(\varepsilon)$ matter (quasi-adiabatic spectral flow at $T = 0$, QBP at $T > 0$). As a result, learning architectures that enforce these locality constraints achieve polylogarithmic or even constant sample complexity within a given phase of matter [18, 27]. These guarantees align with the physical locality intuition developed in [12, 2, 16, 25, 10].

1.2 Research aim and contributions

Aim

Our aim is to develop a unified theoretical and empirical framework for efficient learning of local observables in both gapped ground states (T=0) and thermal (Gibbs) states (T>0) of local quantum lattice systems, by leveraging Lieb-Robinson locality, clustering of correlations, and quasi-local propagation filters (quasi-adiabatic spectral flow for T=0 and quantum belief propagation for T>0) [12, 2, 16, 25]. Within this framework, we translate physical locality into sample-complexity guarantees for learning within a phase, extending ground state results to finite temperature [18, 27].

Contributions

- **Foundations reproduced and standardized**
  - Recast quasi-adiabatic spectral flow (T=0) and thermal quantum belief propagation (T>0) in a common expression, recording the derivative identities and quasi-local generators/filters under a common notation and LR-data convention [12, 2, 16, 25].
  - Collect uniform assumptions and constants (LR parameters, clustering rates, filter tails) needed to make all later bounds system-size independent (QBP identity (III.5), κ_β Δ-*split* (III.3)-(III.6), two-dynamics LR (III.6), and quasi-locality of Φ^β (Lemma III.2) in [25]).
- **New theoretical results (departures from prior work)**
  - *Unified framework (GS + Thermal).* Extend the Šmíd *et al.* local observable learning technique from ground states to *thermal* phases by importing the QBP/clustering machinery of Rouzé *et al.* and placing both regimes under one LR constant convention. We produce the following thermal decay rate

$$\mu_\beta = \min\left\{\frac{\nu}{2},\ \frac{\mu\,\pi}{8\beta\,(v_{\mathrm{LR}}+\pi/\beta)}\right\}.$$

    [12, 2, 16, 25, 27].

– *Downstream bounds for quantum information and machine learning.* From this unified locality module derive: (i) finite-T truncation inequalities and the logarithmic radius accuracy law $\delta_\beta(\varepsilon) = \mu_\beta^{-1}(\log(K_\beta/\varepsilon)+c_D)$ (with a polynomial variant for long-range $\alpha > 2D$ consistent with [27]); (ii) a locality based feature map $\varphi_{I,\delta}$ with bias control $\leq \varepsilon_1\|O_I\|$; (iii) a LASSO generalization bound that includes both T=0 and T>0 with sample complexity $\tilde{O}(N(\varepsilon_1)\varepsilon_3^{-2}\log|S|)$ (extending [27] to T>0); and (iv) equivariance (parameter tying) reductions $\log|S| \to \log|S/G|$ that yield constant-in-n behavior for 1D translation invariant rings.

- **Empirical validation**

  – Implement a purification-TEBD pipeline and validate the theory on Heisenberg (short range) and long-range Ising chains across temperatures and boundary conditions: (i) open-boundary Heisenberg exhibits $\log n$ sample growth; (ii) periodic/equivariant Heisenberg is near constant-in-n; (iii) Ising with α=3 ($\alpha > 2D$) shows near logarithmic growth; (iv) α=1.5 ($\alpha \leq 2D$) indicates linear growth.

1.3 Ethical considerations

This thesis is purely theoretical and computational. All results are gathered from numerical simulations of quantum spin models; no human participants, biological materials, or sensitive/personal data are involved. In line with the Department's guidance, no formal ethics application is required for this work, but we completed and retained an internal checklist and will re-review it if the scope changes (e.g., to add human data collection).

- **Humans / embryos / tissues:** No human participants, no human embryos/foetal tissue, no human cells/tissues. *N/A*.
- **Personal data & tracking:** No collection or processing of personal or sensitive data, no secondary use/merging of datasets, no participant tracking beyond routine institutional system logs. *N/A*.
- **Animals:** No animal use. *N/A*.
- **Developing countries / participant risk:** No fieldwork; no activities that could place individuals at risk. *N/A*.
- **Environment, health & safety:** No hazardous materials/equipment. Computational experiments acknowledge energy use. We minimized runs, profiled code, and reused cached results where possible.
- **Dual use & misuse:** Methods concern condensed-matter simulation and learning bounds. No military or security-sensitive aims, no export controlled items, no cryptanalytic or surveillance functionality. We commit to responsible communication if any dual-use risk is identified.
- **Legal issues (licensing/data protection):** All third-party libraries are open-source and correctly cited. Licenses are compatible with this project's distribution.

- **Research integrity, transparency, reproducibility:** We followed best practices for citation and plagiarism avoidance; code, configuration, and random seeds are documented. We will release source code and scripts under a permissive license to enable independent verification.

All checklist items were considered. All relevant entries are *No* with justification as above. This project follows principles of honesty and responsibility of computational resources.

1.4 Declaration

I acknowledge the use of ChatGPT 4o (OpenAI, https://chat.openai.com/) to check Vancouver reference formatting and for Latex syntax. I confirm that no content generated by AI has been presented as my own work.

1.5 Code Repository

Repository: `github.com/okanders/Efficient-Learning-Thermal-QS`

2 Preliminaries

In this chapter, we introduce several core primitives from a first principles lens: local Hamiltonians, spectral gaps, Lieb-Robinson bounds, exponential clustering of correlations, thermal (Gibbs) states, and quantum Markov properties. For each concept, we provide a formal definition and highlight how it contributes to efficient learning algorithms. These primitives will introduce necessary concepts for the techniques and results developed in subsequent chapters.

2.1 Local Hamiltonians and locality

Physical interactions are often local, where each degree of freedom interacts primarily with those in its immediate vicinity. This motivates the definition of local Hamiltonians, which formalize the notion of locality in quantum systems.

Definition 2.1 (Local Hamiltonian)**.** A Hamiltonian H on a lattice (or graph) of N sites is called *k-local* (or *geometrically local*) if it can be written as a sum of terms, each acting nontrivially on at most k neighboring sites. We express

$$H = \sum_{X \subseteq \Lambda} h_X,$$

where each h_X is an operator supported on a finite region X (of size $|X| \leq k$) and X corresponds to a contiguous cluster of sites (with respect to the lattice geometry). Often one assumes a fixed maximum interaction range R such that $h_X = 0$ whenever the diameter of X exceeds R. We then say H is finite-range (or exponentially decaying) local [6, 15].

Example 2.2 (Nearest-neighbor chain)**.** On a 1D chain,

$$H = \sum_{i=1}^{N-1} h_{i,i+1},$$

with each $h_{i,i+1}$ acting only on sites i and $i{+}1$.

Remark 2.3 (No instantaneous action at a distance). Local Hamiltonians respect an emergent causality. Influence must propagate through a sequence of local interactions (Lieb-Robinson bound) [19].

Local Hamiltonians are the natural starting point for efficient learnability because they impose structure on eigenstates and on Gibbs (thermal) states. If interactions were all-to-all with comparable strength, even a small perturbation could strongly correlate distant regions, hindering learning. However, local interactions typically generate approximately local correlations (in gapped/short range settings they decay exponentially) [12]. Many quantum lattice models of interest—such as spin systems, the Hubbard model, and the quantum Ising family—are few-local. Furthermore, for long-range interactions decaying as $1/r^{\alpha}$, generalized Lieb-Robinson bounds provide linear or sublinear light cones depending on α and the spatial dimension [17, 28].

Locality also plays a role in tensor-network and related machine-learning approaches. When we are restricted to local Hamiltonians, we guarantee that each degree of freedom

is directly influenced by only $\mathcal{O}(1)$ neighbors, which is vitally important for scalable algorithms [22, 26, 32, 29]. In what follows, we combine locality with additional properties (gaps, clustering, Markov structure) to quantify how correlations are constrained in both ground and thermal phases.

2.2 Spectral gaps and gapped phases

While locality limits the actions of interaction terms to nearby sites, it does not by itself guarantee anything about the strength or range of correlations in a quantum state. A complementary concept is the presence of a spectral gap.

Definition 2.4 (Spectral Gap)**.** Consider a Hamiltonian H with eigenvalues $E_0 \leq E_1 \leq E_2 \leq \cdots$ (where E_0 is the ground state energy). The *spectral gap* of H is defined as

$$\Delta \;:=\; E_1 - E_0.$$

We say H is *gapped* if its spectral gap Δ is bounded below by a constant $\Delta_* > 0$ independent of system size (in the thermodynamic limit). It is *gapless* if Δ vanishes or decreases to zero as the system size grows [11, 12, 2].

Remark 2.5. For a *uniformly gapped* family $\{H^{(n)}\}$, the ground-state sector is separated from excitations by an n-independent gap. Such gaps are stable under small local perturbations, and quasi-adiabatic continuation identifies gapped phases [2]. In short-range systems, a nonvanishing gap implies exponential clustering of connected ground-state correlations [12].

In one dimension, a uniform gap further implies an area law for entanglement entropy and thus allows for efficient matrix-product representations of ground states. For learning, restricting to gapped phases provides finite correlation length and short-range entanglement. We find that many efficient guarantees for local-observable prediction assume a gap [14]. At criticality (gapless systems), algebraic correlations typically require more data and more expressive models.

2.3 Lieb-Robinson Bounds (effective light cones)

Underlying both locality and spectral gap phenomena is a fundamental result known as the Lieb-Robinson bound. This theorem provides a quantitative statement of causality in non-relativistic quantum systems with local interactions. It establishes that information and correlations cannot propagate arbitrarily fast through a lattice, even if the Hamiltonian is not relativistic.

Definition 2.6 (Lieb-Robinson bound)**.** Let $H = \sum_X h_X$ be a local Hamiltonian on a lattice Λ, and write $\tau_t(\cdot) = e^{iHt}(\cdot)e^{-iHt}$. There exist constants $C, \mu, v > 0$ (depending only on the interaction, not on system size) such that for any local observables A, B with supports $X = \operatorname{supp} A$ and $Y = \operatorname{supp} B$,

$$\big\|[\,\tau_t(A),\, B\,]\big\| \;\leq\; C\,\|A\|\,\|B\|\,\exp\Big(-\mu\big(d(X,Y) - v|t|\big)\Big).$$

If $|t| < d(X,Y)/v$, the commutator is exponentially small, so $\tau_t(A)$ and B approximately commute [19, 21, 10].

Remark (emergent light cone; quasi-locality). We will be interpreting Def. 2.6 as a *quasi-locality* statement: there is an operator $A_{X_r}(t)$ supported on the r-fattening $X_r := \{z : d(z, X) \leq r\}$ such that

$$\left\|\tau_t(A) - A_{X_r}(t)\right\| \;\leq\; C\,\|A\|\,e^{-\mu\,(r - v_{\mathrm{LR}}|t|)} \qquad (r \geq v_{\mathrm{LR}}|t|),$$

e.g. the Heisenberg evolution of a local observable can be dressed into a strictly local operator up to an exponentially small tail outside the LR cone of radius $v_{\mathrm{LR}}|t|$ [21, Thm. 2.3 and Eq. (2.15)]. We are able to maintain that an influence cannot propagate faster than v_{LR}. If a region Y lies outside the cone of X at time t, then $[\,\tau_t(A), B\,]$ is exponentially suppressed.

Physical intuition and consequences.

The LR bound makes locality quantitative, asserting that a local perturbation influences operators mainly inside a cone of radius $v_{\mathrm{LR}}|t|$, with exponentially small effect outside. Though dynamical, LR has static consequences. We find that in short range systems with a uniform spectral gap, connected ground state correlations decay exponentially with distance [12]. For computation and learning, the cone identifies a finite spacetime neighborhood sufficient to approximate a local observable to accuracy ε, with errors decaying exponentially in the truncation scale.

Power-law interactions and generalized light cones.

Many lattice models do not have strictly short range interactions. Instead, the coupling between two regions X and Y often decays as a *power* of their separation,

$$\|h_{X,Y}\| \;\propto\; d(X, Y)^{-\alpha} \qquad (\text{spatial dimension } d).$$

This is called a *power-law* (or algebraic) decay. In contrast to an exponential decay $\propto e^{-d(X,Y)/\xi}$ (which introduces a finite length scale ξ), an algebraic tail has no single cutoff distance, basically far apart degrees of freedom can still interact weakly. The exponent α quantifies how fast the interaction falls off, and thus how strongly distant sites can influence one another.

Lieb-Robinson bounds continue to constrain information propagation in this setting. However, the shape of the light cone now depends on α: when the decay is sufficiently steep one recovers a linear (light cone) bound, whereas more slowly decaying interactions produce a cone whose radius grows only as a power of time (rather than linearly) [17, 28, 9, 7] .

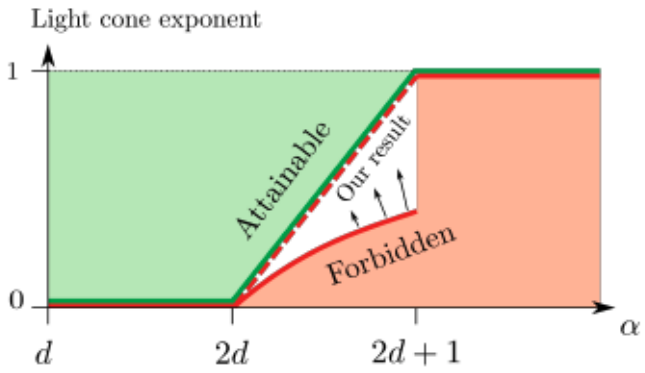


Figure 1: **Generalized Lieb-Robinson light cones for power-law interactions [28].** For interactions decaying as $1/r^\alpha$ in spatial dimension d, the light-cone shape depends on α. The curve $\zeta(\alpha)$ encodes the radius-time relation $r \sim t^{\zeta(\alpha)}$: $\zeta(\alpha) = 1$ for $\alpha > 2d{+}1$ (linear cone), $0 < \zeta(\alpha) < 1$ for $2d < \alpha \leq 2d{+}1$ (polynomial cone), and no finite power exponent for $d < \alpha \leq 2d$ (only logarithmic control).

We summarize the regime as follows:

- **Linear cone ($\alpha > 2d{+}1$).** $r \lesssim v\,t$ (equivalently, $t \gtrsim r/v$), e.g., a linear light cone as in short range systems [17].
- **Polynomial cone ($2d < \alpha \leq 2d{+}1$).** $r \lesssim t^{\,1/(\alpha-2d)+o(1)}$ (equivalently, $t \gtrsim r^{\,\alpha-2d-o(1)}$) [28, Eq. (15)].
- **Logarithmic cone ($d < \alpha \leq 2d$).** $t \gtrsim \log r$ (no finite power exponent governs the front) [28, Eq. (15)].

The same thresholds appear in our truncation bounds: for $\alpha > 2d{+}1$ we recover short range behavior with logarithmic radii, while for $2d < \alpha \leq 2d{+}1$ the truncation radius grows polynomially in $1/\varepsilon$; see Section 2.4.1.

2.4 Exponential clustering of correlations

One of the most significant consequences of locality (Def. 2.1) combined with a spectral gap (Def. 2.4) is the phenomenon of exponential decay of correlations, often simply called exponential clustering. Informally, this means that any two operators acting on distant regions have correlations that fall off (typically exponentially) with the distance separating them. We now define this notion explicitly.

Definition 2.7 (Exponential decay of (connected) correlations)**.** A state ρ on a lattice exhibits *exponential clustering* if there exist constants $C, \mu > 0$ (independent of system size) such that for all disjoint finite regions $X, Y \subset \Lambda$ and all local observables $A \in \mathsf{B}_X$, $B \in \mathsf{B}_Y$,

$$\big|\operatorname{Tr}(\rho\,AB) - \operatorname{Tr}(\rho\,A)\operatorname{Tr}(\rho\,B)\big| \;\leq\; C\,\|A\|\,\|B\|\,e^{-\mu\,d(X,Y)}.$$

We call $\xi := \mu^{-1}$ the *correlation length*. In particular, for $d(X,Y) \gg \xi$ the connected correlation is exponentially small [12, Thm. 2.8].

Hastings and Koma established this property for ground states of uniformly gapped, short range local Hamiltonians. They established that the correlation length follows the estimate

$$\xi \lesssim \frac{v}{\Delta},$$

where v is the Lieb-Robinson velocity and Δ is the (uniform) spectral gap [12]. A larger gap therefore will require a shorter range of correlations.
In practice, exponential clustering legitimizes local reasoning. In order to predict an observable supported on a region X, one may ignore degrees of freedom beyond a buffer of a few correlation lengths around X with only exponentially small error. We now formalize this as a truncation statement.

Remark 2.8 (Pointer to truncation radius). Clustering implies one can truncate couplings outside a finite buffer around I with small error; see Section 2.4.1 for the quantitative $\delta(\varepsilon)$ laws and a formal statement.

2.4.1 Local truncation

Exponential clustering says that degrees of freedom far from a target region I contribute only exponentially small corrections to $\langle O_I \rangle$. Operationally, this means there exists a *truncation radius* $\delta(\varepsilon)$ such that, to achieve error at most ε, it suffices to keep only those interaction terms whose supports lie within graph distance $r \geq \delta(\varepsilon)$ of I. We discarded the rest at an exponentially small cost. We clarify this in the construction below. We gather the nearby couplings into an index set $S_{I,r}$ and define a truncated coupling vector $\chi_{S_{I,r}}(x)$ that zeros out all terms beyond the buffer. Figure 2 illustrates the selection rule, and Proposition 2.10 quantifies how large r must be in short range (gapped) and power-law regimes.

Definition 2.9 (Truncated coupling vector)**.** Fix a target region $I \subseteq \Lambda$ and $r \in \mathbb{N}$. Define

$$S_{I,r} := \{\, j : \ \mathrm{dist}(A_j, I) \leq r \text{ and } \mathrm{diam}(A_j) \leq r \,\},$$

the indices of local terms whose supports lie within distance r of I (and have diameter at most r). The truncated coupling vector is

$$[\chi_{S_{I,r}}(x)]_j = \begin{cases} x_j, & j \in S_{I,r}, \\ 0, & j \notin S_{I,r}. \end{cases}$$

Thus $H(\chi_{S_{I,r}}(x))$ retains only couplings in the r-neighborhood of I [27].

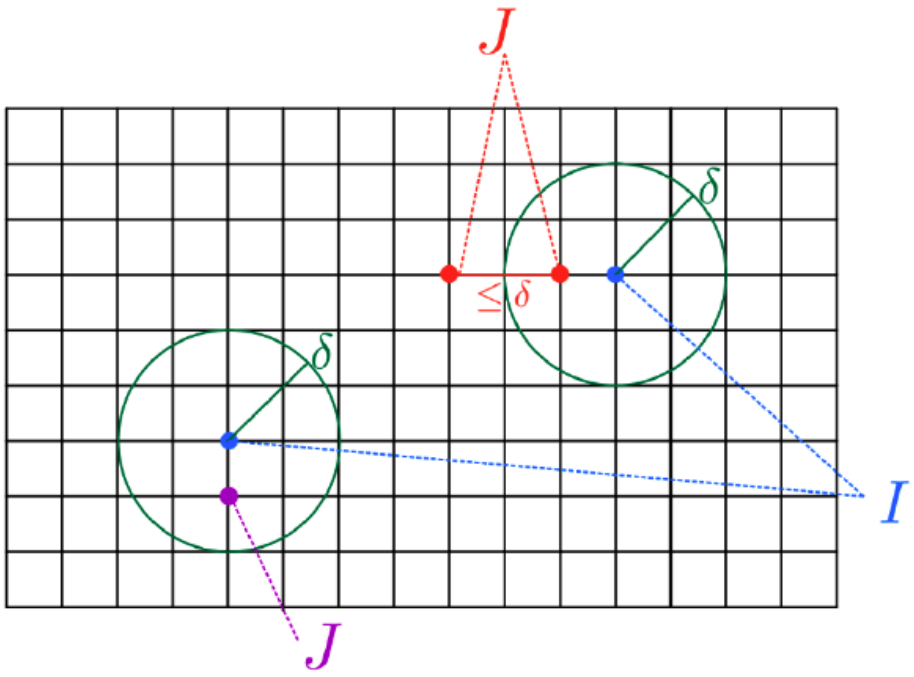


Figure 2: Local truncation around a target region I. The radius-r neighborhood $S_{I,r}$ (circle) determines which couplings are kept (colored) and which are discarded (gray). By clustering, choosing $r \geq \delta(\varepsilon)$ bounds the contribution of discarded couplings to any O_I by at most $\varepsilon\|O_I\|$ [27].

Proposition 2.10 (Ground-state local truncation error [27, Prop. 2.4 & Cor. A.6.1])**.** *Let $H(x)$ be a local Hamiltonian on Λ with a uniform spectral gap along the parameter path under consideration, and let $\rho_0(x) := P(x)/\operatorname{Tr} P(x)$ denote the normalized ground state projector. Then there exist constants $c, \tilde{c} > 0$ (independent of $|\Lambda|$) such that, for any $\varepsilon > 0$, if*

$$r \gtrsim \begin{cases} \dfrac{2}{\mu(\gamma)} \log^2\!\big(c/\varepsilon\big), & \textit{(short range)}, \\ \tilde{c}\, \max\!\Big(\varepsilon^{-1/(\nu-D)},\ \log^{2/\mu(\gamma)}\!\big(c'/\varepsilon\big)\Big), & \textit{(power-law tails, } \alpha > 2D\textit{)}, \end{cases}$$

then for every local observable O_I supported on I,

$$\Big|\operatorname{Tr}\big[O_I\, \rho_0(x)\big] - \operatorname{Tr}\big[O_I\, \rho_0\big(\chi_{S_{I,r}}(x)\big)\big]\Big| \ \leq\ \varepsilon\, \|O_I\|.$$

Equivalently,

$$\big\|\operatorname{Tr}_{\Lambda\setminus I}\big[\rho_0(x) - \rho_0\big(\chi_{S_{I,r}}(x)\big)\big]\big\|_1 \ \leq\ \varepsilon.$$

Remark 2.11 (Implications for learning). Proposition 2.10 shows that to estimate a local observable within error ε, it suffices to restrict to a buffer of radius $r = \delta(\varepsilon)$ around its support, as contributions from outside are exponentially small in r. This justifies locality-based truncations in simulation, finite-depth time-evolution approximations, and bounded support feature maps in our learning.

2.5 Thermal states and finite-temperature correlations

So far, we have focused on ground states of local Hamiltonians in gapped phases. Equally important for both physics and learning are thermal states (also called Gibbs states) at

finite temperature. These states may share locality-induced properties, such as finite correlation lengths, considering the temperature is high enough (or more precisely, away from any phase transition). We first define what a thermal state is.

Definition 2.12 (Gibbs (Thermal) State)**.** Given a Hamiltonian H and an inverse temperature $\beta = 1/(k_B T)$ (set $k_B = 1$), the *thermal state* at temperature T is the density operator

$$\rho_\beta \;=\; \frac{e^{-\beta H}}{Z(\beta)}, \qquad Z(\beta) = \mathrm{Tr}\big(e^{-\beta H}\big).$$

Equivalently, ρ_β is the unique state that maximizes entropy $S(\rho)$ subject to $\mathrm{Tr}(H\rho)$ being fixed by the energy expectation.

Thermal states describe systems in equilibrium at temperature T. At $\beta = 0$ ($T \to \infty$), ρ_β becomes the maximally mixed state. At $\beta \to \infty$ ($T = 0$), ρ_β approaches the ground state projector. For intermediate temperatures, ρ_β interpolates between these extremes, generally becoming more correlated as β increases (temperature lowers).

Definition 2.13 (Exponential decay of thermal correlations)**.** A Gibbs state ρ_β demonstrates *thermal exponential clustering* if there exist $C_\beta, \nu_\beta > 0$ (independent of system size) such that for all disjoint finite regions $X, Y \subset \Lambda$ and all local observables $A \in \mathsf{B}_X$, $B \in \mathsf{B}_Y$,

$$\big|\mathrm{Cov}_{\rho_\beta}(A,B)\big| := \big|\,\mathrm{Tr}(\rho_\beta AB) - \mathrm{Tr}(\rho_\beta A)\,\mathrm{Tr}(\rho_\beta B)\big| \;\le\; C_\beta\, \min\{|X|,|Y|\}\,\|A\|\,\|B\|\, e^{-\nu_\beta\, d(X,Y)}.$$

We call $\xi_\beta := \nu_\beta^{-1}$ the (thermal) correlation length. Note: this is the thermal version of Def. 2.7

A natural question is whether Gibbs states of local Hamiltonians also exhibit exponential clustering (Def. 2.13). In practice the answer is *yes* in the regimes we use. Above any thermal critical point, short range quantum systems retain a finite correlation length, which can be proved by means of Lieb-Robinson/clustering techniques [16, Thm. 1]. At a continuous thermal transition, the correlation length diverges ($\xi_\beta \to \infty$) and correlations decay algebraically.

Remark 2.14. Figure 3 (left) visualizes this: when supports A and B are far apart, the connected covariance $\mathrm{Cov}_{\rho_\beta}(X_A, X_B)$ is exponentially small in their separation.

Learning outlook (thermal).

When ρ_β expresses exponential clustering, local observables can be predicted from a finite buffer around their support, with exponentially small error from the exterior. We can then deploy the locality-based truncation and feature map ideas developed for ground states, which can carry over to thermal states. Recent works leverage this to learn families of Gibbs states within a phase from a small number of samples [25]; under clustering and an approximate Markov property, one can even prepare and estimate local properties with shallow circuits [3].

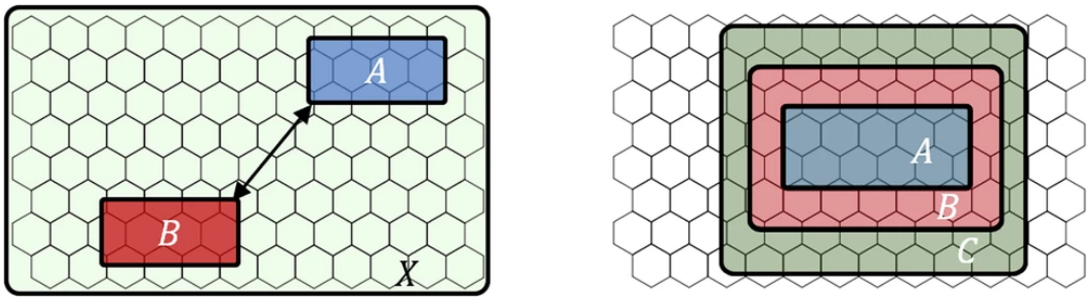


Figure 3: **Illustration of exponential clustering and the Markov property in quantum many-body states [25].** Left: Regions A (blue) and B (red) separated by a large distance in a lattice. When the state has exponential decay of correlations, the covariance between any operator on A and any operator on B falls off sharply as the distance increases. Right: A partition of the system into three regions A (blue), B (red), and C (green), where B is a buffer (or shield) separating A and C. As the buffer grows thicker, conditional mutual information $I(A{:}C \mid B)$ drops, effectively decoupling A and C.

2.6 Quantum Markov properties and conditional independence

The final primitive we discuss is the notion of a quantum Markov property. This concept generalizes the idea of conditional independence (Markovianity) from classical probability distributions to quantum states. In a classical Gibbs distribution on a graph, conditioning on a site's immediate neighbors renders it independent of the rest of the lattice. In terms of quantum mechanics, observables need not share a common eigenbasis, but a related notion of conditional independence can be formulated in terms of density matrices and entropy.

Definition 2.15 (Quantum Markov State)**.** Consider a tripartition of the lattice into regions A, B, and C, such that B separates A and C (e.g. B could be a thick boundary region around A that shields it from C). We say a state ρ_{ABC} is a *quantum Markov state* (or satisfies the Markov property across A–B–C) if the conditional mutual information

$$I(A:C\,|\,B)_\rho \;:=\; S(\rho_{AB}) + S(\rho_{BC}) - S(\rho_B) - S(\rho_{ABC})$$

is zero. Equivalently, ρ has an exact quantum Markov chain structure, if there exists a completely positive recovery map $\mathcal{R}_{B\to BC}$ such that $\rho_{ABC} = (\mathrm{id}_A \otimes \mathcal{R}_{B\to BC})(\rho_{AB})$. In this case, all correlations between A and C are mediated by B [25].

When $I(A:C\,|\,B) = 0$, region B perfectly shields A from C, all correlation between A and C can be accounted for by their mutual correlation with B. This is the quantum generalization of the Markov property in a classical Bayesian network or Markov random field. If B is empty, $I(A:C\,|\,\emptyset) = 0$ would mean A and C are uncorrelated (product state on AC). With a non-empty B, the Markov property allows A and C to be correlated, but only through B.

Remark 2.16 (Approximate Markov Locality; see Fig. 3 (right)). In clustering phases, one expects the conditional mutual information $I(A:C \mid B)_\rho$ to decrease as the width of the buffer region B separating A and C increases. We see that the state ρ behaves similarly to a quantum Markov chain structured as A–B–C over large distances. Exact quantum Markov states constitute a specialized class, such as Gibbs states derived from commuting

Hamiltonians. In contrast, generic Gibbs states with noncommuting terms are not precisely Markov but approximate this property in noncritical conditions. For instance, at elevated temperatures or within one-dimensional short range model, where $I(A:C \mid B)_\rho$ remains small and manageable [16]. Quantitatively, a small conditional mutual information ensures the existence of a recovery map with minimal error, as established by the Fawzi-Renner bound [8].

Relevance for learning.

When a state is regarded as approximately quantum Markov, it becomes possible to infer the properties of a subsystem based solely on its immediate surroundings. By incorporating a fixed buffer B around A, there is no need to model distant regions C to describe the local statistics of A. This perspective substantiates quantum belief propagation for Gibbs states, supports the use of local tensor-network approximations (such as projected entangled pair states in gapped phases), and justifies the application of neural models with bounded receptive fields. In the absence of approximate Markov locality, global patterns dominate, requiring significantly larger sample sizes or model parameters to capture long-range correlations [25].

3 Quantum Information Bound

3.1 Introduction

A notable feature of *gapped quantum matter* is that local perturbations only produce local responses. At zero temperature, the standard framework that formalizes this principle is *quasi-adiabatic spectral flow* [12]: given a smooth path of gapped Hamiltonians $H(s) = H(0) + sV$, $s \in [0, 1]$, it is possible to construct an almost-local unitary $U(s)$ that converts the ground state of $H(0)$ into that of $H(s)$ while maintaining the gap. At finite temperature an apparently different technique of *quantum belief propagation* (QBP) is employed. It supplies a differential equation governing how the Gibbs state

$$\rho_\beta(s) \;=\; \frac{e^{-\beta H(s)}}{\mathrm{Tr}\big(e^{-\beta H(s)}\big)}$$

responds to the same local perturbation V.

At first sight these formalisms appear to inhabit different spaces: spectral flow presumes a spectral gap $\gamma > 0$ and makes no explicit reference to temperature, whereas QBP is born thermal. Its generator carries an inverse temperature β and does not demonstrate dependence on a gap.

However, we reveal that they are two aspects of a *single locality mechanism*, held together by three shared traits:

(i) a Lieb-Robinson (LR) bound on the speed of information,

(ii) rapid decay of a time filter kernel that suppresses long-time propagation, and

(iii) exponential (or at least sufficiently fast) decay of correlations in the reference state.

The aim of this section is an exhibition of two points: (i) to expose the shared algebraic skeleton of spectral flow and QBP, and (ii) to show that QBP can be viewed as a *finite-temperature spectral flow* that smoothly interpolates between the zero-temperature ($\beta \to \infty$) and high-temperature modes. This provides a unified language for stability proofs, continuity bounds, and phase classification across the entire (T, γ) plane.

3.2 Setup and Mathematical Framework

3.2.1 Basic definitions and notation

We work on a D-dimensional hypercubic lattice $\Lambda \subset \mathbb{Z}^D$ of linear size L and cardinality $n = L^D$, equipped with the periodic (torus) graph metric. All distances and diameters below refer to this metric. Throughout, $k, p = \mathcal{O}(1)$ are fixed integers independent of n.

Operator algebras and metrics.

For any finite $X \subset \Lambda$ we write B_X for the algebra of bounded operators on the qudits in X. For $X, Y \subset \Lambda$

$$\mathrm{dist}(X, Y) := \min_{i \in X, j \in Y} d(i, j), \qquad \mathrm{diam}(X) := 1 + \max_{i,j \in X} d(i, j).$$

Norms $\|\cdot\|$ are operator (spectral) norms unless indicated otherwise.

Definition 3.1 (Block neighbourhood)**.** Let $\mathcal{P}_{\le k}(\Lambda) = \{\, J \subset \Lambda \mid |J| \le k \,\}$. For any $I \subset \Lambda$ and radius $\delta \in \mathbb{N}$, define

$$S_{I,\delta} := \{\, J \in \mathcal{P}_{\le k}(\Lambda) \mid \operatorname{dist}(I, J) \le \delta,\ \operatorname{diam}(J) \le \delta\}$$

Definition 3.2 ((ε, δ)-insensitivity)**.** Let $F : [-1, 1]^m \to \mathbb{R}$. Fix a target block $I \in \mathcal{P}_{\le k}(\Lambda)$ and let $S_{I,\delta}$ be as above. We say F is (ε, δ)-*insensitive* (around I) if

$$\left|\, F(x) \;-\; F\big(\chi_{S_{I,\delta}}(x)\big)\right| \;\le\; \varepsilon \quad \text{for all } x \in [-1, 1]^m,$$

where $\chi_{S_{I,\delta}}(x)$ is formed from x by zeroing all coordinates x_J with $J \notin S_{I,\delta}$.

Hamiltonian family. For each parameter vector

$$x = (x_I)_{I \in \mathcal{P}_{\le k}(\Lambda)} \in [-1, 1]^m, \qquad m = \left|\mathcal{P}_{\le k}(\Lambda)\right|,$$

we consider a k-body Hamiltonian [27]:

$$H(x) = \sum_{I \in \mathcal{P}_{\le k}(\Lambda)} h_I(x_I), \qquad \operatorname{supp} h_I \subset I$$

where the *interaction terms* h_I obey one of the following locality hypotheses:

(a) *Exponential tails:* $\|h_I\|, \|\partial_{x_I} h_I\| \le C_0\, \mathrm{e}^{-\nu \operatorname{diam}(I)}$.

(b) *Power-law tails of order* $\alpha > 2D$: $\|h_I\|, \|\partial_{x_I} h_I\| \le C_0 \operatorname{diam}(I)^{-((|I|-2)D+\alpha)}$.

We denote by $\gamma(x)$ the many-body gap above the ground state and assume *uniform stability*: $\inf_x \gamma(x) \ge \gamma_{\min} > 0$.

Reference states. We treat ground and thermal cases in parallel:

$$\sigma_\beta(x) = \begin{cases} P = |\psi_0(x)\rangle\langle\psi_0(x)| & (\beta = \infty,\ \text{a fixed pure ground state}), \\ \rho_\beta(x) := \mathrm{e}^{-\beta H(x)} / \operatorname{Tr}(\mathrm{e}^{-\beta H(x)}) & (0 < \beta < \infty,\ \text{Gibbs state}). \end{cases}$$

Clustering assumption. We assume exponential clustering for the state $\sigma_\beta(x)$: there exist $C, c > 0$ and a correlation length $\xi(\beta)$ such that for observables A_X, B_Y,

$$\left|\operatorname{Cov}_{\sigma_\beta(x)}(A_X, B_Y)\right| \;\le\; C\, \|A_X\|\, \|B_Y\|\, e^{-\operatorname{dist}(X,Y)/\xi(\beta)}.$$

For uniformly gapped, short-range $H(x)$ this holds at $\beta = \infty$ for any *pure* ground state [12]. For Gibbs states we assume exponential clustering whenever needed (e.g. at high temperature or under suitable mixing conditions).

It is convenient to keep in mind the operational length scale

$$\xi_*(\beta) \;:=\; \max\Big\{\xi_{\mathrm{LR}},\ \tfrac{v_{\mathrm{LR}}}{\gamma_{\min}},\ \tfrac{2 v_{\mathrm{LR}} \beta}{\pi}\Big\},$$

which reflects the LR cone and the filter tails appearing in our proofs; implicit constants are absorbed in the $\lesssim$ notation below.

Observables. For a bounded p-body observable $O = \sum_{I \in \mathcal{P}_{\le p}(\Lambda)} O_I$ we abbreviate

$$f_I(x) := \operatorname{Tr}\big[O_I\, \sigma_\beta(x)\big], \quad f(x) = \sum_I f_I(x)$$

The task is to quantify how *local* the map $x \mapsto f_I(x)$ is, and how accurately it can be approximated by discarding parameters that are far away from the support I.

3.2.2 Derivative identities and the unified formula

Consider a local perturbation path $H(s) = H + sV$ on a lattice system with finite Lieb-Robinson speed v.

Gapped ground states ($T = 0$).

Suppose $H(s)$ remains gapped by $\gamma > 0$, and let

$$P(s) = |\psi_0(s)\rangle\langle\psi_0(s)|$$

be its ground state projector. Quasi-adiabatic continuation constructs an almost-local generator

$$D_\gamma(s) \;=\; \Phi^{(\gamma)}_{H(s)}\big(V\big),$$

where

Definition 3.3 (Spectral filter)**.** For any Hamiltonian H and local perturbation V, define

$$\Phi^{(\gamma)}_H(V) := \int_{-\infty}^{\infty} W_\gamma(t)\, e^{itH}\, V\, e^{-itH}\, \mathrm{d}t,$$

with W_γ any odd, integrable kernel band-limited to the gap window $(-\gamma, \gamma)$ as described in [2, 27].

Taking expectations in the ground state gives the spectral-flow derivative identity,

$$\frac{\mathrm{d}}{\mathrm{d}s}\,\langle O\rangle_{\psi_0(s)} = i\,\langle [D_\gamma(s),\, O]\rangle_{\psi_0(s)} \tag{1}$$

Remark 3.4. By combining the Lieb-Robinson bound with the rapid decay of $W_\gamma(t)$, we show that $D_\gamma(s)$ is *quasi-local*: for any observable O_X supported in region X,

$$\big\|[D_\gamma(s),\, O_X]\big\| \;\le\; C\,\|O_X\|\, e^{-\mu\,\mathrm{dist}(\mathrm{supp}\,V, X)}$$

Thermal (Gibbs) states ($T > 0$)

For $\beta < \infty$ the Gibbs state

$$\rho_\beta(s) \;=\; \frac{e^{-\beta H(s)}}{\mathrm{Tr}\; e^{-\beta H(s)}}$$

evolves under the *quantum belief propagation (QBP) flow*

$$\partial_s \rho_\beta(s) \;=\; -\frac{\beta}{2}\,\big[\,\Phi^{(\beta)}_{H(s)}(V),\, \rho_\beta(s)\,\big]$$

Then by trace cyclicity,

$$\begin{aligned}\frac{d}{ds}\,\mathrm{Tr}\big[\rho_\beta(s)\,O\big] &= -\frac{\beta}{2}\,\mathrm{Tr}\Big(\rho_\beta(s)\,[\,\Phi^{(\beta)}_{H(s)}(V),\, O\,]\Big) + \beta\,\langle V\rangle_{\rho_\beta(s)}\,\langle O\rangle_{\rho_\beta(s)} \\ &= -\,\beta\;\mathrm{Cov}_{\rho_\beta(s)}\big(O,\; \Phi^{(\beta)}_{H(s)}(V)\big).\end{aligned} \tag{2}$$

Here we adopt the covariance from [25, Eq. 3] that upholds the above logic:

$$\mathrm{Cov}_\sigma(A, B) := \tfrac{1}{2}\,\mathrm{Tr}\big[\sigma\,\{\,A - \langle A\rangle_\sigma,\; B - \langle B\rangle_\sigma\,\}\big] \tag{3}$$

Definition 3.5 (Thermal filter (QBP))**.** For any Hamiltonian H and operator V, define

$$\Phi_H^{(\beta)}(V) := \int_{\mathbb{R}} \kappa_\beta(t)\, e^{itH}\, V\, e^{-itH}\, \mathrm{d}t,$$

where the (even) kernel has Fourier transform $\widehat{\kappa}_\beta(\omega) = \frac{\tanh(\beta\omega/2)}{\beta\omega/2}$. An explicit expression (for $t \neq 0$) is

$$\kappa_\beta(t) = \frac{2}{\pi\beta}\, \log\frac{e^{\pi|t|/\beta}+1}{e^{\pi|t|/\beta}-1},$$

which exhibits logarithmic behavior at small times, $\kappa_\beta(t) \sim \frac{2}{\pi\beta}\log\frac{1}{|t|}$ as $t \to 0$, and exponential decay at large times via [25, 1, Eq. III.3; Appendix B]

$$\kappa_\beta(t) \;\leq\; \frac{4}{\pi\beta}\,\frac{1}{e^{\pi|t|/\beta}-1}$$

Δ-split (small/large time)[25, Lem. III.2]. Here Δ (a threshold for time) is unrelated to the spatial radius δ used in neighborhoods $S_{I,\delta}$.

Fix

$$\Delta \;\geq\; \Delta_1 := \frac{2\beta}{\pi}\,\mathrm{asinh}\frac{1}{2},$$

then the kernel obeys the two integral bounds

$$\int_{|t|>\Delta} |\kappa_\beta(t)|\,\mathrm{d}t \;\leq\; \frac{16}{\pi^2}\, e^{-\frac{\pi}{2\beta}\Delta}, \qquad \int_{|t|\leq\Delta} |\kappa_\beta(t)|\, e^{v|t|}\,\mathrm{d}t \;\leq\; \frac{4}{\pi^2}\, e^{(v+\pi/\beta)\Delta},$$

which we will use whenever we split time integrals into short/long pieces. These expressions derive from [25, Lem. III.2]

Remark 3.6 (Zero vs finite temperature skeleton). At T=0 (gapped case), the adiabatic/spectral flow gives $\partial_s\langle O\rangle = i\,\langle[\,\Phi^{(\gamma)}_{H(s)}(V),\, O\,]\rangle$. At T>0 (Gibbs case), the variation of expectations is

$$\partial_s\langle O\rangle = -\beta\,\mathrm{Cov}\big(O, \Phi^{(\beta)}_{H(s)}(V)\big),$$

where

$$\mathrm{Cov}_\sigma(A,B) := \tfrac{1}{2}\,\mathrm{Tr}\big[\sigma\,\{\,A-\mathrm{Tr}(\sigma A),\; B-\mathrm{Tr}(\sigma B)\,\}\big]$$

	Spectral flow (T=0)	**QBP** (T>0)
	(*ground projector*)	
reference state σ	$P = \lvert\psi_0\rangle\langle\psi_0\rvert$	Gibbs state ρ_β
prefactor κ_σ	i	$-\beta/2^\dagger$
filter parameter	γ	β
filter kernel f_σ	$W_\gamma(t)$ (BMNS kernel)	$\kappa_\beta(t)$ (QBP kernel)
spatial decay scale	$\xi \sim v_{LR}/\gamma$	$\xi_\beta \sim \frac{4\beta}{\pi}\,\big(v_{\mathrm{LR}} + \pi/\beta\big)$

Table 1: Dictionary between quasi-adiabatic spectral flow and quantum belief propagation. †In covariance form one has $\partial_s\langle O\rangle = -\,\beta\;\mathrm{Cov}\big(O, \Phi^\beta(V)\big)$, e.g. prefactor $-\beta$ and covariance with the *filtered* perturbation.

3.3 The Locality Principle

3.3.1 Quasi-locality of the propagation filter Φ_σ

In a local lattice model, we aim to demonstrate that the dressed operator (filter-evolved observable)

$$\Phi_\sigma(V) \;=\; \int_{-\infty}^{\infty} f_\sigma(t)\,\tau_t(V)\,dt, \qquad \tau_t(\cdot) := e^{iHt}(\cdot)e^{-iHt},$$

is quasi-local. This means that its weight outside the original support of V decays rapidly (exponentially for short-range interactions, algebraically for long-range). This emergent locality arises from two components: (i) the Lieb-Robinson (LR) bound; and (ii) a fast-decaying time filter $f_\sigma(t)$.

Unified notation

We adopt the dictionary

$$(\sigma,\ \kappa_\sigma,\ f_\sigma) \;=\; \begin{cases} (P,\ \gamma,\ W_\gamma) & \text{ground state } (T=0), \\ (\rho_\beta,\ \beta,\ \kappa_\beta) & \text{Gibbs state } (T>0) \end{cases}$$

For the thermal kernel we do *not* assume a global pointwise exponential bound (it is false near t=0). Instead, we always work with the Δ-split above, which captures the small-$|t|$ logarithm and the large-$|t|$ exponential tail uniformly with the stated constants [25, Eqs. III.3-III.6].

For the ground state (quasi-adiabatic) kernel W_γ we use the sub-exponential time decay encoded in [2, Lem. 2.6(iv)]: writing

$$I_\gamma(t) \;:=\; \int_{|s|>t} |W_\gamma(s)|\,ds,$$

there exists a function G such that $I_\gamma(t) \le G(\gamma|t|)$, with G decaying faster than any power (e.g. $G(\xi) \lesssim u_a(\xi)$ for some $a>0$) [27, 2, Lem. 2.2; Lem. 2.6(iv)].

LR and filter data (constant convention).

We fix LR data $(C_{\mathrm{LR}}, \mu, v_{\mathrm{LR}})$ so that for any operators A, B supported a distance d apart,

$$\big\|\,[\tau_t(A), B]\,\big\| \;\le\; C_{\mathrm{LR}}\,\|A\|\,\|B\|\,\exp\!\Big(-\mu\,\big(d - v_{\mathrm{LR}}|t|\big)_+\Big),$$

where $(x)_+ := \max\{x, 0\}$. *(See [27, Appendix A.1, Prop. A.3]: Lemma A.1 for the exp-decay LR bound and Proposition A.3 for the inside-cone estimate used in I_1.)* In either case $\|f_\sigma\|_1 := \int_{\mathbb{R}} |f_\sigma(t)|dt < \infty$.

Lemma 3.7 (Spectral flow generator-locality)**.** *Let $s \mapsto H(s)$ be a C^1 family of Hamiltonians on a finite lattice, $s \in [0,1]$, with a uniform gap $\ge \gamma > 0$ and uniformly bounded $\|H'(s)\|$. Let $P(s)$ be the ground state projector and*

$$D_\gamma(s) \;=\; \int_{\mathbb{R}} W_\gamma(t)\,\tau_t^{H(s)}\big(H'(s)\big)\,dt$$

Then $\frac{d}{ds}P(s) = i\,[D_\gamma(s), P(s)]$, and $D_\gamma(s)$ is quasi-local: if $\operatorname{supp} H'(s) \subset J$, then for any observable X,

$$\big\|\,[D_\gamma(s),\,X]\,\big\| \;\le\; C\,\|H'(s)\|\,\|X\|\,e^{-\mu\,\operatorname{dist}(J,\operatorname{supp} X)} + C_N\,\|H'(s)\|\,\|X\|\,\big(1+\operatorname{dist}(J,\operatorname{supp} X)\big)^{-(N-1)},$$

for every $N \in \mathbb{N}$, with constants independent of system size [27, 2, Lem. 2.1; Cor. 2.8].

Remark 3.8. $D_\gamma(s)$ is self-adjoint (the integrand $\tau_t^{H(s)}(H'(s))$ is Hermitian and W_γ is real), and [2, Cor. 2.8] gives $\partial_s P(s) = i\,[D_\gamma(s), P(s)]$.

What we use below is just the quasi-locality of $D_\gamma(s)$, obtained by combining a Lieb-Robinson bound with the tail control of the quasi-adiabatic kernel, $I_\gamma(t) := \int_{|s|>t} |W_\gamma(s)|\,ds \leq G(\gamma t)$.

Why does locality emerge? (following split time proof of [2])

Let $d := \mathrm{dist}(\mathrm{supp}\, V, \mathrm{supp}\, O)$ and choose the LR cutoff $t^* := d/(2v_{\mathrm{LR}})$. Then

$$[\Phi_\sigma(V), O] \;=\; \int_{\mathbb{R}} f_\sigma(t)\,[\tau_t(V), O]\,dt \;=\; \underbrace{\int_{|t|\leq t^*} f_\sigma(t)\,[\tau_t(V), O]\,dt}_{I_1} \;+\; \underbrace{\int_{|t|>t^*} f_\sigma(t)\,[\tau_t(V), O]\,dt}_{I_2}.$$

(Split at t^* and use the same I_1/I_2 decomposition, with the "LR bound is vacuous" step for $|t| > t^*$.) For $|t| \leq t^*$, $d - v_{\mathrm{LR}}|t| \geq d/2$, thus

$$\|I_1\| \;\leq\; C_{\mathrm{LR}}\|V\|\,\|O\|\,e^{-\mu d/2}\,\|f_\sigma\|_1.$$

For the tail $|t| > t^*$ we use the trivial commutator bound $\|[X,Y]\| \leq 2\|X\|\,\|Y\|$:

$$\|I_2\| \;\leq\; 2\|V\|\,\|O\| \int_{|t|>t^*} |f_\sigma(t)|\,dt.$$

Thermal case ($\sigma = \rho_\beta$). If $t^* \geq \Delta_1 := \frac{2\beta}{\pi}\,\mathrm{asinh}\frac{1}{2}$, then

$$\|I_2\| \;\leq\; 2\|V\|\,\|O\| \int_{|t|>t^*} |\kappa_\beta(t)|\,dt \;\leq\; \frac{32}{\pi^2}\,e^{-\frac{\pi}{2\beta}\,t^*}\,\|V\|\,\|O\|,$$

using the tail bound $\int_{|t|>\Delta} |\kappa_\beta(t)|dt \leq \frac{16}{\pi^2} e^{-\frac{\pi}{2\beta}\Delta}$ for $\Delta \geq \Delta_1$. For general d (thus arbitrary t^*), we bypass this direct I_2 estimate and instead use the quasi-locality of the filtered operator (by Lemma III.2 [25]):

$$\|\Phi_H^{(\beta)}(V) - \Phi_{H_B}^{(\beta)}(V)\| \leq c'\,\|V\|\,e^{-\mu'\,\mathrm{dist}(\mathrm{supp}\, V, B^c)}, \quad \mu' = \mu \min\Big\{\tfrac{1}{2},\; \tfrac{\pi}{4\beta\,(v_{\mathrm{LR}}+\pi/\beta)}\Big\},$$

and then combine with clustering/LR for O to obtain the final exponential response bound.

Ground-state case ($\sigma = P$). Using $f_\sigma = W_\gamma$ and the definition of I_γ,

$$\|I_2\| \;\leq\; 2\|V\|\,\|O\| \int_{|t|>t^*} |W_\gamma(t)|\,dt \;=\; 2\|V\|\,\|O\|\,I_\gamma(t^*) \;\leq\; 2\,G(\gamma t^*)\,\|V\|\,\|O\|$$

(Here $I_\gamma(t) \leq G(\gamma|t|)$ with $G(\xi) = \frac{1}{\gamma}\big[\,K/2 \;\; (0 \leq \xi \leq \xi_*),\;\; 130\,e^{2\xi}/\big(10\,u_{2/7}(\xi)\big) \;\; (\xi > \xi_*)\big]$ and $u_a(\xi) = \exp[-a\,\xi/\log^2 \xi]$; thus sub-exponential decay [27, 2, Lem. 2.2, Eq. 4.3]. By Lemma 2.2 of [27], there exist constants K and $\xi_* > 0$ such that, for t^* obeying $\gamma t^* > \xi_*$, one has the *sub-exponential* control $G(\gamma t^*) \lesssim u_a(\gamma t^*)$, and so for every $N \in \mathbb{N}$ there is C_N'' with

$$G(\gamma t^*) \;\leq\; C_N''\,(t^*)^{-(N-1)}.$$

With $t^* = d/(2v_{\mathrm{LR}})$ this gives

$$\|[\Phi_H^{(\gamma)}(V), O]\| \;\leq\; C_1\,\|V\|\,\|O\|\,e^{-\mu d/2} \;+\; C_N''\,\|V\|\,\|O\|\,\Big(\tfrac{d}{v_{\mathrm{LR}}}\Big)^{-(N-1)}.$$

Now, we are able to see that the short time (inside-cone) piece produces the LR exponential, while the long time piece I_2 is controlled by the quasi-adiabatic filter tail I_γ, which is sub-exponential.

Lemma 3.9 (Quasi-locality of Φ)**.** *Let $H = \sum_{I\subset\Lambda} h_I$ be a local Hamiltonian, fix h_A supported on A, and for $B \supset A$ define $H_B := \sum_{\mathrm{supp}(h_I)\subset B} h_I$. Then:*

(a) ***Thermal case.*** *There exist constants $c', \mu' > 0$ (independent of system size) such that*

$$\left\| \Phi_H^{(\beta)}(h_A) - \Phi_{H_B}^{(\beta)}(h_A) \right\| \;\le\; c'\,|A|\,\|h_A\|\,\exp\!\big(-\mu'\,\mathrm{dist}(A, B^c)\big). \tag{4}$$

(Picking $\Delta = \Delta_1 + \frac{\mu}{2\left(v_{\mathrm{LR}}+\pi/\beta\right)}\,\mathrm{dist}(A,B^c)$ gives $\mu' = \mu\min\!\left\{\frac{1}{2}, \frac{\pi}{4\beta\,(v_{\mathrm{LR}}+\pi/\beta)}\right\}$ [25].)

(b) ***Ground state case.*** *For every $N \in \mathbb{N}$ there exists $C_N > 0$ (independent of system size) such that*

$$\begin{aligned}\left\| \Phi_H^{(\gamma)}(h_A) - \Phi_{H_B}^{(\gamma)}(h_A) \right\| \le\; & C_1\,|A|\,\|h_A\|\,e^{-\mu\,\mathrm{dist}(A,B^c)/2} \\ & +\; C_N\,|A|\,\|h_A\|\left(\frac{\mathrm{dist}(A,B^c)}{v_{\mathrm{LR}}}\right)^{-(N-1)}.\end{aligned} \tag{5}$$

Consequently, in both cases $\Phi_H(h_A)$ is quasi-local around A (exponential at finite T, exponential up to a sub-exponential correction at $T = 0$).

Proof. (Use the two-dynamics LR bound $\|\alpha_t - \alpha_t^B\|$ (Eq. (III.6)) together with Lemma III.2 of [25] for Φ-quasi-locality at finite T.) An LR estimate for dynamics with and without terms outside B yields constants $c, \mu, v_{\mathrm{LR}} > 0$ such that

$$\left\|\tau_t^H(h_A) - \tau_t^{H_B}(h_A)\right\| \;\le\; c\,|A|\,\|h_A\|\,\exp\!\Big(-\mu\left(\mathrm{dist}(A,B^c) - v_{\mathrm{LR}}|t|\right)_+\Big).$$

Set $d := \mathrm{dist}(A, B^c)$ and $t^* := d/(2v_{\mathrm{LR}})$. Then

$$\begin{aligned}\|\Phi_H(h_A) - \Phi_{H_B}(h_A)\| &\le \int_{\mathbb{R}} |f_\sigma(t)|\left\|\tau_t^H(h_A) - \tau_t^{H_B}(h_A)\right\| dt \\ &\le c\,|A|\,\|h_A\|\Big(\|f_\sigma\|_1\,e^{-\mu d/2} + \int_{|t|>t^*} |f_\sigma(t)|\,dt\Big).\end{aligned}$$

Thermal case ($\sigma = \rho_\beta$)**.** Choose

$$\Delta \;=\; \Delta_1 + \frac{\mu}{2\left(v_{\mathrm{LR}} + \pi/\beta\right)}\,d, \qquad \Delta_1 = \frac{2\beta}{\pi}\,\mathrm{asinh}\frac{1}{2},$$

and use the Δ-split:

$$\int_{|t|>\Delta} |\kappa_\beta(t)|\,dt \le \frac{16}{\pi^2} e^{-\frac{\pi}{2\beta}\Delta}, \qquad \int_{|t|\le\Delta} |\kappa_\beta(t)|\,e^{v|t|}\,dt \le \frac{4}{\pi^2} e^{(v+\pi/\beta)\Delta}.$$

Combining with the two-dynamics LR bound gives 4 with $\mu' = \mu\min\!\left\{\frac{1}{2},\ \frac{\pi}{4\beta\,(v_{\mathrm{LR}}+\pi/\beta)}\right\}$.

Ground-state case ($\sigma = P$)**.** With $f_\sigma = W_\gamma$,

$$\int_{|t|>t^*} |W_\gamma(t)|\,dt \;=\; I_\gamma(t^*) \;\le\; G(\gamma t^*),$$

which is sub-exponential and, for any N, satisfies $G(\gamma t^*) \le C'_N (t^*)^{-(N-1)}$. This returns 5. □

Long-range interactions.

For power-law interactions $\|h_{ij}\| \propto \mathrm{dist}(i,j)^{-\alpha}$ with $\alpha > 2D$, locality degrades from exponential to polynomial. The LR light cone becomes algebraic for power-law decays, so truncation errors fall off as a power of the distance rather than exponentially; this is the mechanism used in our bounds. Consequently, for $B \supset A$ one obtains a polynomially decaying truncation estimate of the form

$$\left\| \Phi_H(h_A) - \Phi_{H_B}(h_A) \right\| \;\le\; C\,|A|\,\|h_A\| \left(1 + \tfrac{\mathrm{dist}(A,B^c)}{\xi}\right)^{-p}, \tag{6}$$

for some $p = p(\alpha, D) > 0$ and length scale $\xi > 0$ (the precise power depends on the long-range LR exponents and the summability step). For $\alpha \in (2D, 2D{+}1)$, [27] show that the truncation radius necessary to achieve accuracy ε scales polynomially as $\delta \gtrsim \varepsilon^{-1/(\nu-D)}$, and they extend this to all $\alpha > 2D$ with adjusted constants [27, Sec. 2.2.2-2.2.3]. In contrast, when $\alpha \le 2D$ their outer-region sum diverges ($S_2 = \infty$), we note that we *cannot* make the truncation error small by enlarging B; observables may depend on parameters outside any fixed neighborhood [27, App. A, Prop. A.8].

3.3.2 Local perturbation leads to local response

Having established the quasi-locality of the propagation filter $\Phi_\sigma(V)$, we now formalize the Local Perturbation → Local Response principle, which provides a unified description for both zero-temperature (gapped ground state) and finite-temperature (thermal Gibbs state) settings. The principle asserts that a perturbation of the Hamiltonian localized in some region A can significantly influence expectation values of observables far away in B only through operator spreading within a Lieb-Robinson light cone, with the long-time contributions dampened by the chosen filter. In gapped ground states, we see that this is governed by quasi-adiabatic continuation (Lemma 3.7) together with the notable sub-exponential decaying tail integral I_γ *(spectral-flow locality and sub-exponential tail for I_γ: Lemma 2.6(iv); explicit generator: Cor. 2.8 of [2])*. In thermal states, the analogous role is played by quantum belief propagation (QBP) together with exponential time decay of κ_β and (when assumed, as we have discussed) exponential clustering of correlations in Gibbs states [25]. In both cases the influence is exponentially small in the distance, and up to a sub-exponential correction in the ground state case.

Throughout this discussion, we assume a uniform bound on the interaction strengths:

$$\|h_I\| \;\le\; h_{\max} = O(1)\,, \qquad \forall I \subset \Lambda\,,$$

so that all constants can be taken independent of the system size (we re-iterate). We also assume that the unperturbed state σ (ground or Gibbs) satisfies a uniform Lieb-Robinson bound with parameters $(C_{\mathrm{LR}}, \mu, v_{\mathrm{LR}})$ as fixed above *(standard LR for exponentially decaying/short-range interactions [12, Theorem A.2])*, and (for the thermal statements below) that exponential clustering holds as an explicit condition.

Conventions for Proofs. We repeatedly use:

1. triangle inequality under the integral, $\left\| \int g(t) X_t \, dt \right\| \le \int |g(t)|\,\|X_t\|\,dt$;
2. unitary invariance of the operator norm, $\|\tau_t(X)\| = \|X\|$;
3. the trivial commutator/covariance bounds $\|[A,B]\| \le 2\|A\|\,\|B\|$ and $|\,\mathrm{Cov}_\sigma(A,B)| \lesssim \|A\|\,\|B\|$;
4. Hölder's inequality $\left|\,\mathrm{Tr}(\rho X)\right| \le \|\rho\|_1 \|X\| = \|X\|$ for states ρ

Locality in gapped ground states.

Theorem 3.10 (Locality of Ground State Observables)**.** *Let $H(x) = \sum_J x_J\, h_J$ be a family of local Hamiltonians on a D-dimensional lattice, with a uniform spectral gap $\gamma > 0$ for all parameter values x. Denote by $\langle\cdot\rangle_x$ the expectation in the (possibly degenerate) ground state mixture $\rho(x) = P(x)/\operatorname{Tr} P(x)$. Then for any two disjoint regions $I, J \subset \Lambda$ and any observable O_I supported on I, one has for all $N \in \mathbb{N}$,*

$$\Big|\,\partial_{x_J}\,\langle O_I\rangle_x\,\Big| \;\le\; C\,\|O_I\|\,\|h_J\|\,\Big(e^{-(\mu/2)\,\operatorname{dist}(I,J)} \;+\; C_N\,(1+\operatorname{dist}(I,J))^{-(N-1)}\Big),$$

where $C, \mu > 0$ and $C_N > 0$ are independent of the system size.

Remark 3.11. The first (exponential) term is the short-time contribution controlled by the LR light cone; the second is the sub-exponential tail stemming from the quasi-adiabatic filter via the tail integral $I_\gamma(t)$ (Lemma 3.7) *(Lemma 2.6(iv) of [2]; explicit G in [27, Lem. 2.2])*. If, in addition, the ground state exhibits exponential clustering, one may replace the polynomial tail by an auxiliary (possibly weaker) exponential with adjusted constants when d is sufficiently large *(gapped short-range → exponential clustering; [12])*; otherwise keep the explicit sub-exponential term.

Proof of Theorem 3.10. **Setup.** Fix disjoint regions I, J and write $d := \operatorname{dist}(I, J) > 0$. Our goal is to bound $\big|\partial_{x_J}\langle O_I\rangle_x\big|$.

Spectral-flow representation. We start by expressing the derivative through the spectral-flow generator. By Lemma 3.7, for each coupling x_J there exists a quasi-local operator

$$D_J(x) \;=\; \int_{\mathbb{R}} W_\gamma(t)\,\tau_t^{H(x)}(h_J)\,dt \qquad \text{with} \qquad \partial_{x_J} P(x) \;=\; i\,[D_J(x), P(x)]\,.$$

(Compare the explicit construction and identities $D = \int W_\gamma(t)\, e^{itH} H' e^{-itH} dt$ and $\frac{d}{ds}P = i[D, P]$ in [2, Prop. 2.4, Cor. 2.8].)

From projectors to expectations. Passing from the projector to the ground state density matrix $\rho(x) = P(x)/\operatorname{Tr} P(x)$ and differentiating the expectation, we get

$$\partial_{x_J}\langle O_I\rangle_x = \operatorname{Tr}\big(O_I\,\partial_{x_J}\rho(x)\big) = i\,\operatorname{Tr}\big(O_I\,[D_J(x), \rho(x)]\big) = i\,\operatorname{Tr}\big([O_I, D_J(x)]\,\rho(x)\big),$$

where we used cyclicity of the trace to move the commutator onto O_I. Since $\rho(x) \ge 0$ with $\operatorname{Tr}\rho(x) = 1$, the expectation against $\rho(x)$ is dominated by the operator norm. We see that,

$$\big|\partial_{x_J}\langle O_I\rangle_x\big| \le \big\|[O_I, D_J(x)]\big\|.$$

LR time split. We now bound this commutator using the integral representation of $D_J(x)$. Apply the triangle inequality for integrals:

$$\big\|[O_I, D_J(x)]\big\| \le \int_{\mathbb{R}} |W_\gamma(t)|\,\big\|[O_I, \tau_t(h_J)]\big\|\,dt.$$

To isolate the region where LR information is nontrivial, introduce the LR cutoff

$$t^* \;:=\; \frac{d}{2v_{\mathrm{LR}}}.$$

Inside the cone ($|t| \le t^*$)**.** Inside this window $|t| \le t^*$ the Heisenberg-evolved perturbation $\tau_t(h_J)$ has not reached I by more than half the distance: the LR estimate guarantees $\operatorname{dist}(\operatorname{supp}\tau_t(h_J), I) \ge d - v_{\mathrm{LR}}|t| \ge d/2$. Therefore the LR bound gives

$$\big\|[O_I, \tau_t(h_J)]\big\| \le C_{\mathrm{LR}}\,\|O_I\|\,\|h_J\|\,e^{-\mu(d-v_{\mathrm{LR}}|t|)} \le C_{\mathrm{LR}}\,\|O_I\|\,\|h_J\|\,e^{-\mu d/2},$$

(standard $d - v|t|$ LR form; [12, Thm. A.2]) uniformly for $|t| \le t^*$. Integrating this against $|W_\gamma(t)|$ over the inside-cone window yields

$$\int_{|t|\le t^*} |W_\gamma(t)|\,\big\|[O_I, \tau_t(h_J)]\big\|\,dt \le C_{\mathrm{LR}}\,\|O_I\|\,\|h_J\|\,e^{-\mu d/2}\int_{|t|\le t^*} |W_\gamma(t)|\,dt \le C_1\,\|O_I\|\,\|h_J\|\,e^{-\mu d/2},$$

with $C_1 := C_{\mathrm{LR}}\|W_\gamma\|_1$ (we use $\|W_\gamma\|_1 < \infty$ from the quasi-adiabatic construction [2, Lem. 2.6(iii)]).

Outside the cone ($|t| > t^*$)**.** Outside the light cone the LR bound is vacuous, so we use the trivial commutator estimate $\|[X, Y]\| \le 2\|X\|\,\|Y\|$ together with unitary invariance $\|\tau_t(h_J)\| = \|h_J\|$:

$$\int_{|t|>t^*} |W_\gamma(t)|\,\big\|[O_I, \tau_t(h_J)]\big\|\,dt \le 2\,\|O_I\|\,\|h_J\| \int_{|t|>t^*} |W_\gamma(t)|\,dt.$$

It is convenient to write the tail as the two-sided integral $I_\gamma(t) := \int_{|s|>t} |W_\gamma(s)|\,ds$, so the outside-cone contribution is

$$\le 2\,\|O_I\|\,\|h_J\|\,I_\gamma(t^*).$$

Tail control. By the sub-exponential decay of the quasi-adiabatic tail, there exists a function G with $I_\gamma(t) \le G(\gamma t)$ and $G(\xi)$ decaying faster than any power *([2, Lem. 2.6(iv)]; see also the explicit G in [27, Lem. 2.2])*. Substituting $t^* = d/(2v_{\mathrm{LR}})$ gives

$$I_\gamma(t^*) \;\le\; C'_N \left(\tfrac{d}{2v_{\mathrm{LR}}}\right)^{-(N-1)}.$$

(If γt^* is not large, this tail can be absorbed into the exponential piece by adjusting constants. Either way the final bound holds with constants independent of system size.)

Collecting terms. Collecting the inside-cone and outside-cone contributions and absorbing harmless numerical factors into C and C_N, we arrive at

$$\big\|[O_I, D_J(x)]\big\| \;\le\; C\,\|O_I\|\,\|h_J\|\,\Big(e^{-\mu d/2} + C_N\,(1+d)^{-(N-1)}\Big),$$

and thus, by the reduction at the start,

$$\Big|\,\partial_{x_J}\,\langle O_I\rangle_x\,\Big| \;\le\; C\,\|O_I\|\,\|h_J\|\,\Big(e^{-(\mu/2)\,d} + C_N\,(1+d)^{-(N-1)}\Big).$$

All constants depend only on the LR data, $\|W_\gamma\|_1$, and the interaction bound $\|h_J\| \le h_{\max} = O(1)$, and are uniform in the system size. □

Locality in thermal states (Quantum Belief Propagation)

We assume throughout:

Condition 3.12 (Exponential clustering of correlations)**.** For each inverse temperature $\beta > 0$ and any parameter vector x, the Gibbs state $\sigma_\beta(x) = e^{-\beta H(x)}/\operatorname{Tr}(e^{-\beta H(x)})$ satisfies: there exist $C_{\mathrm{cl}}, \nu > 0$ such that

$$\left|\operatorname{Cov}_{\sigma_\beta(x)}(A,B)\right| \;\le\; C_{\mathrm{cl}}\,\min\{|A|,|B|\}\,\|A\|_\infty\,\|B\|_\infty\,e^{-\nu\,\operatorname{dist}(\operatorname{supp}A,\operatorname{supp}B)}$$

for all observables A, B with disjoint supports. [25, Condition I.1]

Definition 3.13 (Lipschitz seminorm)**.** For an observable L on Λ,

$$\|L\|_{\mathrm{Lip}} := \max_{i\in\Lambda}\ \inf_{M\in\mathcal{B}(\mathcal{H}_{i^c})}\ \big\|\,L - M\otimes\mathbb{I}_i\,\big\|_\infty.$$

In particular, if $L = \sum_{j=1}^{M} O_j$ with each O_j supported on $O(1)$ sites and with bounded overlap, then $\|L\|_{\mathrm{Lip}} = O(1)$ [25, Eq. II.1].

Further assumptions to note for proof of proposition:

QBP derivative identity and thermal filter.

$$\partial_{x_J}\langle L\rangle_{\sigma_\beta(x)} = -\,\beta\ \operatorname{Cov}_{\sigma_\beta(x)}\Big(L,\ \Phi^{(\beta)}_{H(x)}(h_J)\Big), \qquad \Phi^{(\beta)}_H(V) := \int_{\mathbb{R}} \kappa_\beta(t)\,\tau^H_t(V)\,dt. \tag{7}$$

Kernel κ_β: explicit form and Δ-split. For $t \neq 0$,

$$\kappa_\beta(t) = \frac{2}{\pi\beta}\log\frac{e^{\pi|t|/\beta}+1}{e^{\pi|t|/\beta}-1}.$$

Furthermore, for any $\Delta \ge \Delta_1 := \frac{2\beta}{\pi}\operatorname{asinh}\frac{1}{2}$,

$$\int_{|t|>\Delta}|\kappa_\beta(t)|\,dt \le \frac{16}{\pi^2}\,e^{-\frac{\pi}{2\beta}\Delta}, \qquad \int_{|t|\le\Delta}|\kappa_\beta(t)|\,e^{v|t|}\,dt \le \frac{4}{\pi^2}\,e^{(v+\pi/\beta)\Delta}. \tag{8}$$

Lieb-Robinson (two-dynamics) bound [25, Eq. III.6] For $A \subset B$,

$$\|\tau_t(A) - \tau^B_t(A)\| \;\le\; c\,|A|\,\|A\|\,e^{\,v_{\mathrm{LR}}|t| - \mu\,\operatorname{dist}(\operatorname{supp}A,\,B^c)}. \tag{9}$$

Quasi-locality of the thermal filter [25, Lem. III.2] For $A \subset B$, we have

$$\begin{aligned}\big\|\Phi^{(\beta)}_{H(x)}(O_A) - \Phi^{(\beta)}_{H_B(x)}(O_A)\big\| &\le c'\,|A|\,\|O_A\|\,\exp\big(-\mu'\,\operatorname{dist}(A,B^c)\big),\\ \mu' &= \mu\,\min\Big\{\tfrac{1}{2},\ \tfrac{\pi}{4\beta\,(v_{\mathrm{LR}}+\pi/\beta)}\Big\}.\end{aligned} \tag{10}$$

"Remove-and-telescope" covariance step for Lipschitz L. For any operator X_B supported on B and any $r \ge 0$,

$$\left|\operatorname{Cov}_{\sigma_\beta(x)}(L,\,X_B)\right| \;\le\; 2h\,|B(r)|\,\|L\|_{\mathrm{Lip}} \;+\; 2C_{\mathrm{cl}}\,|B|\,h\,\|L\|_\infty\,e^{-\nu r}, \tag{11}$$

where $h := \sup_I \|h_I\|$ and $|B(r)|$ is the r-neighbourhood volume [25, Prop. III.3].

Proposition 3.14 (Locality of Thermal Response — LP-LR form)**.** *Assume Condition 3.12. Fix an interaction index $J \in \mathcal{P}_{\le k}(\Lambda)$ and $\beta > 0$. If L is* local *(supported on a finite region I) with finite $\|L\|_{\mathrm{Lip}}$, then*

$$\Big|\, \partial_{x_J} \langle L \rangle_{\sigma_\beta(x)} \Big| \;\le\; \beta\, C_{\mathrm{th}}\, \|L\|_{\mathrm{Lip}}\, \|\partial_{x_J} h_J\|\, e^{-\mu_\beta\, d}, \qquad d := \mathrm{dist}(\mathrm{supp}\, L, \mathrm{supp}\, h_J)$$

where

$$\mu_\beta \;:=\; \min\Big\{ \tfrac{\nu}{2},\; \mu\, \tfrac{\pi}{8\beta\,(v_{\mathrm{LR}} + \pi/\beta)} \Big\},$$

and $C_{\mathrm{th}} > 0$ depends only on local dimensions/geometry, the LR data $(c, \mu, v_{\mathrm{LR}})$, and (C_{cl}, ν), but not on system size.

Proof. **QBP identity and filtered perturbation.** Using (2) and writing $H_J^{(\beta)}(x) := \Phi_{H(x)}^{(\beta)}(h_J)$,

$$\partial_{x_J} \langle L \rangle_{\sigma_\beta(x)} = -\beta\, \mathrm{Cov}_{\sigma_\beta(x)}\big(L,\, H_J^{(\beta)}(x)\big).$$

LR time split. Let $d := \mathrm{dist}(\mathrm{supp}\, L, \mathrm{supp}\, h_J)$ and split the defining time integral at the LR cutoff $t^* := d/(2v_{\mathrm{LR}})$:

$$\big|\, \mathrm{Cov}(L, H_J^{(\beta)}) \big| \le \int_{|t| \le t^*} |\kappa_\beta(t)|\, \big|\, \mathrm{Cov}\big(L, \tau_t(h_J)\big) \big|\, dt + \int_{|t| > t^*} |\kappa_\beta(t)|\, \big|\, \mathrm{Cov}\big(L, \tau_t(h_J)\big) \big|\, dt =: \mathrm{I}_{\le} + \mathrm{I}_{>}.$$

Inside the cone $|t| \le t^*$. By LR, $\mathrm{dist}(\mathrm{supp}\, L, \mathrm{supp}\, \tau_t(h_J)) \ge d - v_{\mathrm{LR}}|t| \ge d/2$. Using exponential clustering (Condition 3.12) and the remove-and-telescope step (11),

$$\big|\, \mathrm{Cov}\big(L, \tau_t(h_J)\big) \big| \;\le\; C_1\, \|L\|_{\mathrm{Lip}}\, \|\partial_{x_J} h_J\|\, e^{-(\nu/2)\, d},$$

uniformly for $|t| \le t^*$. Since $\|\kappa_\beta\|_1 < \infty$,

$$\mathrm{I}_{\le} \;\le\; C_2\, \|L\|_{\mathrm{Lip}}\, \|\partial_{x_J} h_J\|\, e^{-(\nu/2) d}.$$

Outside the cone $|t| > t^*$. By the triangle inequality,

$$\mathrm{I}_{>} \;=\; \Big| \int_{|t| > t^*} \kappa_\beta(t)\, \mathrm{Cov}(L, \tau_t(h_J))\, dt \Big| \;\le\; \int_{\mathbb{R}} |\kappa_\beta(t)|\, |\, \mathrm{Cov}(L, \tau_t(h_J))|\, dt.$$

By quasi-locality of the thermal filter ([25, Lem. III.2]) and the remove-and-telescope bound (11), for any $R \in [0, d]$ and with B_R the radius-R neighborhood of $\mathrm{supp}\, h_J$,

$$\int_{\mathbb{R}} |\kappa_\beta(t)|\, |\, \mathrm{Cov}(L, \tau_t(h_J))|\, dt \;\le\; C_3\, \|L\|_{\mathrm{Lip}}\, \|\partial_{x_J} h_J\| \Big(e^{-\nu\,(d-R)} + e^{-\mu' R} \Big), \quad \mu' = \mu \min\Big\{ \tfrac{1}{2},\; \tfrac{\pi}{4\beta\,(v_{\mathrm{LR}} + \pi/\beta)} \Big\}.$$

Since $v_{\mathrm{LR}} + \pi/\beta \ge \pi/\beta$, we have $\frac{\pi}{4\beta\,(v_{\mathrm{LR}} + \pi/\beta)} \le \frac{1}{4} \le \frac{1}{2}$, therefore $\mu' = \frac{\mu\,\pi}{4\beta\,(v_{\mathrm{LR}} + \pi/\beta)}$. Choosing $R = d/2$ yields

$$\mathrm{I}_{>} \;\le\; C_3\, \|L\|_{\mathrm{Lip}}\, \|\partial_{x_J} h_J\| \Big(e^{-(\nu/2) d} + e^{-\frac{\mu\,\pi}{8\beta\,(v_{\mathrm{LR}} + \pi/\beta)}\, d} \Big).$$

Combining and conclusion. Together with $\mathrm{I}_{\le} \le C_2\, \|L\|_{\mathrm{Lip}}\, \|\partial_{x_J} h_J\| e^{-(\nu/2) d}$ and $a + b \le 2 \max\{a, b\}$,

$$e^{-(\nu/2) d} + e^{-\frac{\mu\,\pi}{8\beta\,(v_{\mathrm{LR}} + \pi/\beta)}\, d} \le 2\, e^{-\mu_\beta d}, \quad \mu_\beta := \min\Big\{ \tfrac{\nu}{2},\; \tfrac{\mu\,\pi}{8\beta\,(v_{\mathrm{LR}} + \pi/\beta)} \Big\}.$$

Thus,

$$\big|\, \mathrm{Cov}(L, H_J^{(\beta)}) \big| \;\le\; C_4\, \|L\|_{\mathrm{Lip}}\, \|\partial_{x_J} h_J\|\, e^{-\mu_\beta d},$$

and multiplying by β via (2) gives $|\partial_{x_J} \langle L \rangle_{\sigma_\beta(x)}| \le \beta\, C_{\mathrm{th}}\, \|L\|_{\mathrm{Lip}}\, \|\partial_{x_J} h_J\|\, e^{-\mu_\beta d}$, with $C_{\mathrm{th}} := C_4$. □

Remark 3.15. The two exponentials arise from distinct mechanisms: the *state*'s clustering ($\nu/2$) and the *filter+LR tail* ($\mu\,\pi/[8\beta\,(v_{\mathrm{LR}}+\pi/\beta)]$ via the Δ-split).

Remark 3.16 (Extensive L and the polylog(n) prefactor). For additive L (e.g. total magnetization), localize both $H_J^{(\beta)}$ and L to radius R around $\operatorname{supp} h_J$, bound the two truncation errors by [25, Lem. III.2] and Condition 3.12, and choose $R \asymp \log n$ as in [25, Prop. III.3]. This returns a polylog(n) prefactor while the rate $\mu\,\pi/[8\beta\,(v_{\mathrm{LR}}+\pi/\beta)]$ remains unchanged.

Path integration step (common to both regimes)

Fix $x, y \in \mathcal{X}$ and the affine path $x(s) = x + s\,(y-x)$, $s \in [0,1]$, so $\dot{x}(s) = y - x$. For any observable O,

$$\langle O\rangle_y - \langle O\rangle_x \;=\; \int_0^1 \dot{x}(s)\cdot\nabla_x \langle O\rangle_{x(s)}\, ds \;=\; \sum_J (y_J - x_J) \int_0^1 \partial_{x_J}\langle O\rangle_{x(s)}\, ds.$$

Taking absolute values and using $\|\dot{x}(s)\|_1 = \|x-y\|_1$,

$$\big|\langle O\rangle_y - \langle O\rangle_x\big| \;\le\; \sum_J |y_J - x_J| \sup_{s\in[0,1]} \big|\partial_{x_J}\langle O\rangle_{x(s)}\big|. \tag{12}$$

Plugging in locality of the derivative. Assume the Lieb-Robinson data are uniform for $H\big(x(s)\big)$ on $[0,1]$; in the GS case the spectral gap is uniform, and in the thermal case exponential clustering holds uniformly in s. Let $A_J := \operatorname{supp} h_J$ and $d_J := \operatorname{dist}(\operatorname{supp} O, A_J)$. By Theorem 3.10 (gapped ground state) and Proposition 3.14 (thermal), for all $s \in [0,1]$,

$$\big|\partial_{x_J}\langle O\rangle_{x(s)}\big| \;\le\; \begin{cases} C_{\mathrm{GS}}\,\|O\|\,\|h_J\|\,e^{-\frac{\mu}{2}\,d_J} \;+\; C_{\mathrm{GS},N}\,\|O\|\,\|h_J\|\,(1+d_J)^{-(N-1)}, & \text{(GS)},\\ \beta\,C_{\mathrm{th}}\,\|O\|_{\mathrm{Lip}}\,\|\partial_{x_J}h_J\|\,e^{-\mu_\beta d_J}, & \text{(Thermal)}. \end{cases} \tag{13}$$

Here $C_{\mathrm{GS}}, C_{\mathrm{GS},N} > 0$ absorb the (uniform) LR constants and the quasi-adiabatic tail; and $C_{\mathrm{th}} > 0$ absorbs the clustering/filter constants so that $C_1 e^{-(\nu/2)d_J} + C_2 e^{-\mu\,\frac{\pi}{8\beta\,(v_{\mathrm{LR}}+\pi/\beta)}\,d_J} \le C_{\mathrm{th}}\,e^{-\mu_\beta d_J}$. All these constants depend only on LR data and local geometry/dimensions (and on the uniform gap for GS / on β for thermal), but not on the system size or on s.

The decay rates are

$$\mu_{\mathrm{GS}} \;:=\; \frac{\mu}{2}, \qquad \mu_\beta \;:=\; \min\Big\{\tfrac{\nu}{2},\; \mu\,\frac{\pi}{8\,\beta\,\big(v_{\mathrm{LR}}+\pi/\beta\big)}\Big\}. \tag{14}$$

Distance reduction. Let $\gamma_{x,y} := \{J : x_J \neq y_J\}$ and $d := \operatorname{dist}(\operatorname{supp} O, \gamma_{x,y}) = \min_{J\in\gamma_{x,y}} d_J$. For any nonincreasing $\phi \ge 0$,

$$\sum_{J\in\gamma_{x,y}} |y_J - x_J|\,\phi(d_J) \;\le\; \|x-y\|_1\,\phi(d). \tag{15}$$

Unified LP–LR estimate. Let $D := \operatorname{dist}(\operatorname{supp} O, \gamma_{x,y})$. Combining (12)–(15) with

$$\|h_J\| \le h_{\max} := \sup_J \|h_J\| \quad \text{(GS)}, \qquad \|\partial_{x_J}h_J\| \le h^{(\partial)}_{\max} := \sup_J \|\partial_{x_J}h_J\| \quad \text{(Thermal)}$$

gives

$$\begin{aligned} \text{(GS)}\quad |\langle O\rangle_y - \langle O\rangle_x| &\le C_{\mathrm{GS}}\,h_{\max}\,\|O\|\,\|x-y\|_1\,e^{-\mu_{\mathrm{GS}}D} \\ &\quad + C_{\mathrm{GS},N}\,h_{\max}\,\|O\|\,\|x-y\|_1\,(1+D)^{-(N-1)}. \end{aligned} \tag{16}$$

$$\text{(Thermal)}\quad |\langle O\rangle_y - \langle O\rangle_x| \le \beta\,C_{\mathrm{th}}\,h^{(\partial)}_{\max}\,\|O\|_{\mathrm{Lip}}\,\|x-y\|_1\,e^{-\mu_\beta D}. \tag{17}$$

Here $h^{(\partial)}_{\max} := \sup_J \|\partial_{x_J}h_J\|$. In the *linear* case $\partial_{x_J}h_J = h_J$, so $h^{(\partial)}_{\max} = h_{\max}$.

3.4 Consequences for Observable Dynamics

3.4.1 Sensitivity to Hamiltonian parameters

Assumptions & notation.

We write $r := \operatorname{dist}(I, \operatorname{supp} h_J)$, fix LR data $(C_{\mathrm{LR}}, \mu, v_{\mathrm{LR}})$, and (for thermal statements) assume exponential clustering with rate ν. Set

$$\mu_\beta \;:=\; \min\Big\{\tfrac{\nu}{2},\; \tfrac{\mu\,\pi}{8\,\beta\,(v_{\mathrm{LR}}+\pi/\beta)}\Big\}.$$

We use $\|O_I\|_{\mathrm{Lip}}$ for Lipschitz observables; for sums of $O(1)$-local terms with bounded overlap, $\|O_I\|_{\mathrm{Lip}} = O(1)$.

What we inherit from the ground state.

Let $H(x) = \sum_J x_J h_J$ act on a D-dimensional lattice, and let O_I be a p-body observable supported on I. Define $f_I(x) = \operatorname{Tr}\big[O_I\,\rho_0(x)\big]$ for the normalized ground state density matrix $\rho_0(x) := P(x)/\operatorname{Tr} P(x)$. Smíd *et al.* prove that, in a uniformly gapped phase, the directional derivative of f_I with respect to any scalar coupling x_J decays with the distance $r = \operatorname{dist}\big(I, \operatorname{supp} h_J\big)$ [27]:

$$\big|\partial_{x_J} f_I(x)\big| \;\le\; \|\partial_{x_J} h_J\|\; F_0(r), \qquad F_0(r) = \begin{cases} C_1\, e^{-\mu(\gamma)\, r}, & \text{exponential interactions,} \\ C_2\, r^{-p(\alpha,D)}, & \text{power-law interactions,} \end{cases} \tag{18}$$

with constants depending only on microscopic parameters (dimension D, interaction decay rates, and the gap), but independent of system size. Integrating along a straight path in parameter space and using a geometric shell-counting bound for the number N_r of interaction terms with support at distance r from I, they derive the radius accuracy law

$$\delta_0(\varepsilon) = \begin{cases} \dfrac{2}{\mu(\gamma)}\, \log^2\big(c/\varepsilon\big), & \text{exponential interactions,} \\ c' \max\Big\{\varepsilon^{-1/(\nu-D)},\; \log^{2/\mu(\gamma)}\big(c''/\varepsilon\big)\Big\}, & \text{power law } \alpha \in (2D, 2D+1], \end{cases} \tag{19}$$

which guarantees $|f_I(x) - f_I(\chi_{S_{I,\delta}} x)| \le \varepsilon\, \|O_I\|$ for all $\delta \ge \delta_0(\varepsilon)$.

Thermal Gibbs states ($T > 0$).

Let $\sigma_\beta(x) = e^{-\beta H(x)}/\operatorname{Tr}(e^{-\beta H(x)})$ and $f_I(x) = \operatorname{Tr}[O_I\,\sigma_\beta(x)]$. Using the QBP derivative identity (2) together with Φ-locality (Lemma 3.9) and Proposition 3.14, we obtain:

Lemma 3.17 (Thermal gradient bound)**.** *For all I, J and $\beta > 0$,*

$$\big|\partial_{x_J} f_I(x)\big| \;\le\; \beta\, C_{\mathrm{th}}\, \|O_I\|_{\mathrm{Lip}}\, \|\partial_{x_J} h_J\|\, e^{-\mu_\beta r}, \qquad r := \operatorname{dist}(I, \operatorname{supp} h_J),$$

with $\mu_\beta = \min\big\{\frac{\nu}{2},\; \frac{\mu\,\pi}{8\beta\,(v_{\mathrm{LR}}+\pi/\beta)}\big\}$. The constant C_{th} depends only on local dimension/geometry and the (uniform) clustering/LR data, and is independent of system size.

(In the linear parametrization $H(x) = \sum_J x_J h_J$ one may replace $\|\partial_{x_J} h_J\|$ by $\|h_J\|$.)

Local truncation at finite temperature.

Let $S_{I,\delta}$ collect all couplings with support within graph distance δ of I and set $x^{loc} = \chi_{S_{I,\delta}}(x)$. Along the affine path $x(s) = x^{loc} + s\,(x - x^{loc})$, the thermal perturbation identity (path interpolation + generalized covariance; Eq. (5) of [16]) gives

$$f_I(x) - f_I(x^{loc}) = \int_0^1 \sum_{J \notin S_{I,\delta}} (x_J - x_J^{loc})\, \partial_{x_J} f_I\big(x(s)\big)\, ds.$$

By Lemma 3.17, for $r_J := \operatorname{dist}(I, \operatorname{supp} h_J)$,

$$\big|\partial_{x_J} f_I(x(s))\big| \le \beta\, C_{\mathrm{th}}\, \|O_I\|_{\mathrm{Lip}}\, \|\partial_{x_J} h_J\|\, e^{-\mu_\beta r_J}.$$

Thus, with $h_{\max} := \sup_J \|\partial_{x_J} h_J\|$ (and $h_{\max} = \sup_J \|\partial_{x_J} h_J\|$ in the linear case),

$$\big|f_I(x) - f_I(x^{loc})\big| \;\le\; \beta\, C_{\mathrm{th}}\, h_{\max}\, \|O_I\|_{\mathrm{Lip}} \sum_{r>\delta} \mathcal{N}(r)\, e^{-\mu_\beta r}, \tag{20}$$

where $\mathcal{N}(r)$ is the number of terms at distance r from I.

Lemma 3.18 (Shell counting)**.** *There exist $C_D, K_D > 0$ such that $\mathcal{N}(r) \le C_D r^{D-1}$ and, for all $\delta \ge 0$,*

$$\sum_{r=\delta+1}^{\infty} \mathcal{N}(r)\, e^{-\mu_\beta r} \;\le\; K_D\, (1+\delta)^{D-1} e^{-\mu_\beta \delta}.$$

Combining (20) with Lemma 3.18 yields

$$\big|f_I(x) - f_I(x^{loc})\big| \;\le\; K_\beta\, \|O_I\|_{\mathrm{Lip}}\, (1+\delta)^{D-1} e^{-\mu_\beta \delta}, \qquad K_\beta := \beta\, C_{\mathrm{th}}\, h_{\max}\, C_D K_D.$$

Corollary 3.19 (Logarithmic radius-accuracy law)**.** *Given $\varepsilon > 0$, any*

$$\delta_\beta(\varepsilon) \;=\; \frac{1}{\mu_\beta}\Big(\log \tfrac{K_\beta}{\varepsilon} + c_D\Big)$$

(with c_D absorbing the $(1+\delta)^{D-1}$ factor) guarantees $|f_I(x) - f_I(x^{loc})| \le \varepsilon\, \|O_I\|_{\mathrm{Lip}}$. This is the finite-temperature analogue of the exponential truncation bound and its inversion in [16, Cor. 2, Eq. 13, Eq. 20], with the decay rate μ_β provided by QBP+LR by means of Lemma 3.17 and Proposition 3.14.

Long-range interactions (thermal analogue).

Replacing Φ-locality by its long-range form (6) (see *Long-range interactions* above) and repeating the same reduction yields:

Lemma 3.20 (Thermal gradient under power-law interactions)**.** *If $\|h_{ij}\| \propto \operatorname{dist}(i,j)^{-\alpha}$ with $\alpha > 2D$, then there exist $\widetilde{C}_{\mathrm{th}} > 0$, $\xi_\beta = O(1)$ and $p = p(\alpha, D) > 0$ such that*

$$\big|\partial_{x_J} f_I(x)\big| \;\le\; \beta\, \widetilde{C}_{\mathrm{th}}\, \|O_I\|_{\mathrm{Lip}}\, \|\partial_{x_J} h_J\| \left(1 + \tfrac{r}{\xi_\beta}\right)^{-p}, \qquad r = \operatorname{dist}(I, \operatorname{supp} h_J).$$

For $\alpha \le 2D$, no uniform truncation in r is possible (outer-shell sum diverges).

Corollary 3.21 (Polynomial radius-accuracy law (power-law case))**.** *With $p = p(\alpha, D)$ as above,*

$$\delta_\beta(\varepsilon) \;\lesssim\; \varepsilon^{-1/\kappa(\alpha,D)}, \qquad \kappa(\alpha, D) = p(\alpha, D) - D,$$

with constants independent of $|\Lambda|$, as in the long-range LPPL/LI analysis in [5].

4 Machine Learning Bound

We now derive learning bounds for predicting local observables in quantum phases. The concept that we demonstrate is that physical locality constrains the effective complexity of the learning task. By leveraging the finite range of correlations and the Lieb-Robinson (LR) bounds in both settings, we show that a simple linear model (LASSO predictor) can learn the expectation value of a sum of local observables with sample complexity that grows only moderately with system size and desired accuracy. In the following, we set up the class of predictors and observables, explain the locality based feature map and truncation strategy, and then state the generalization bound (our LASSO theorem) that equally treats both ground and thermal cases.

We study predictors of $f_O(x) := \mathrm{Tr}\big[O\,\sigma(x)\big]$ that use local information concerning the Hamiltonian parameters x. As in Šmíd *et al.*, the learning architecture is LASSO over a locality aware feature map. The physics enters exclusively through the truncation radius $\delta(\varepsilon_1)$, which fixes both the approximation bias and the size of the dictionary $m_\varphi \le |S|\,N(\varepsilon_1)$, and therefore also the ℓ_1-budget $B = r(O)\,N(\varepsilon_1)$ used in the capacity bound. Once $\delta(\varepsilon_1)$ is set (by the locality theory), our experimental work (Sec. 5) will ultimately trace the ground state analysis in [27] and apply it to the thermal setting.

4.1 Setup, hypothesis class, and data

Let $\Lambda \subset \mathbb{Z}^D$ with $|\Lambda| = n$. Consider a k-local family $H(x) = \sum_{J\in\mathcal{P}_{\le k}(\Lambda)} h_J(x_J)$ with $k = O(1)$. Assume (standing hypotheses): (i) a uniform spectral gap in the GS setting; (ii) a Lieb-Robinson (LR) bound with constants (μ, v_{LR}); (iii) exponential clustering in the thermal setting (with rate $\nu > 0$); and (iv) decay of $\|h_J\|$ and $\|\partial h_J\|$ with $\mathrm{diam}(J)$ (exponential or power law). All these assumptions we had established in Section 3, and are recalled and invoked. Let

$$O = \sum_{I\in S} O_I, \qquad S \subseteq \mathcal{P}_{\le p}(\Lambda),\ \ p = O(1), \qquad r(O) := \sum_{I\in S} \|O_I\|.$$

We either work with the renormalized $O' := O/r(O)$ (so $r(O') = 1$) or assume $r(O) = O(1)$, which yields the dependence $O(\log n)$ referenced below.

We are given a dataset $\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_{i=1}^N$ with bounded label noise $|y^{(i)} - \mathrm{Tr}(O\,\sigma(x^{(i)}))| \le \varepsilon_2$ [27, Eq. 5]. Each $x^{(i)}$ is a full parameter vector of couplings, and $y^{(i)}$ is a possibly noisy estimate of the target expectation. The noise bound can reflect finite precision simulation or measurement. We train a linear predictor under an ℓ_1 constraint using empirical MSE (standard LASSO [20]).

Local neighborhoods and (discretization) features

For $I \subset \Lambda$ and $\delta \in \mathbb{N}$,

$$S_{I,\delta} := \{\, J \in \mathcal{P}_{\le k}(\Lambda) :\ \mathrm{dist}(I, J) \le \delta,\ \mathrm{diam}(J) \le \delta \,\},$$

as in [27, Eq. 1]. Let $X_{I,\delta}$ be a finite grid on the coordinates x_J with $J \in S_{I,\delta}$, with mesh δ_2 chosen as in [27, Eq. 20] (by means of the gradient bound, Lem A.9) so that the per-cell discretization error is $\le \varepsilon_1$. For $x' \in X_{I,\delta}$ define the cell

$$T_{x',I} := \Big\{\, x :\ \|x_J - x'_J\|_\infty < \tfrac{\delta_2}{2} \text{ for all } J \in S_{I,\delta} \Big\},$$

following [27, Eqs. 21-23]. The feature map is the indicator partition

$$\varphi(x)_{I,x'} := \mathbf{1}\big(x \in T_{x',I}\big), \qquad h_w(x) = \sum_{I \in S} \sum_{x' \in X_{I,\delta}} w_{I,x'}\, \varphi(x)_{I,x'},$$

scaled so that $\|\varphi(x)\|_\infty \le 1$. This produces $m_\varphi \le |S|\, N(\varepsilon_1)$ and the norm budget $\|w\|_1 \le B := r(O)\, N(\varepsilon_1)$ [27, Lemma A.11].

Remark 4.1. The neighborhood $S_{I,\delta}$ selects the coordinates near I, and the grid $X_{I,\delta}$ tiles this local parameter subspace by ℓ_∞-cubes of mesh δ_2 chosen through the gradient bound so that $\sup_{x \in T_{x',I}} |f_I(x) - f_I(x')| \le \varepsilon_1 \|O_I\|$. Each cell $T_{x',I}$ contributes a one-hot indicator, so the feature map is a blockwise partition with $\|\varphi(x)\|_\infty \le 1$ and h_w acts as a table lookup proxy for $\sum_I f_I(x|_{S_{I,\delta}})$. From the step construction, we get $m_\varphi \le |S|\, N(\varepsilon_1)$ and $\|w\|_1 \le r(O)\, N(\varepsilon_1)$. The mesh–capacity trade-off mirrors bias–variance. A shrinking δ_2 lowers approximation bias but increases $N(\varepsilon_1)$. Because $\delta(\varepsilon_1)$ depends only on locality (and temperature) constants, capacity is controlled by $N(\varepsilon_1)$, with only a mild $\log |S|$ term in the generalization bound.

Example 4.2 (Toy illustration: Ising chain)**.** Consider a 1D transverse field Ising model

$$H(x) = \sum_i J_i\, Z_i Z_{i+1} + h_i\, X_i$$

with local couplings $x = (J_0, J_1, \ldots, h_0, h_1, \ldots)$. Let $O_I = Z_0 Z_1$ act on sites $I = \{0, 1\}$. For neighborhood radius $\delta = 1$, the relevant set is

$$S_{I,1} = \{J_0, J_1, h_0, h_1, h_2\},$$

e.g. the bond on I itself, the adjacent bond, and the nearby fields. By locality, only these five parameters affect $\langle Z_0 Z_1 \rangle$ up to exponentially (or polynomially) small tails.

Next, discretize each parameter to a mesh δ_2 chosen by the gradient bound; for instance, with $\delta_2 = 0.5$ on $[0, 2]$, each variable has grid $\{0, 0.5, 1.0, 1.5, 2.0\}$. The full grid $X_{I,1}$ thus has $N(\varepsilon_1) = 5^5 = 3125$ points. Each $x' \in X_{I,1}$ defines a cube

$$T_{x',I} = \Big\{ x : \ |x_J - x'_J| < \tfrac{\delta_2}{2} \ \ \forall\, J \in S_{I,1} \Big\},$$

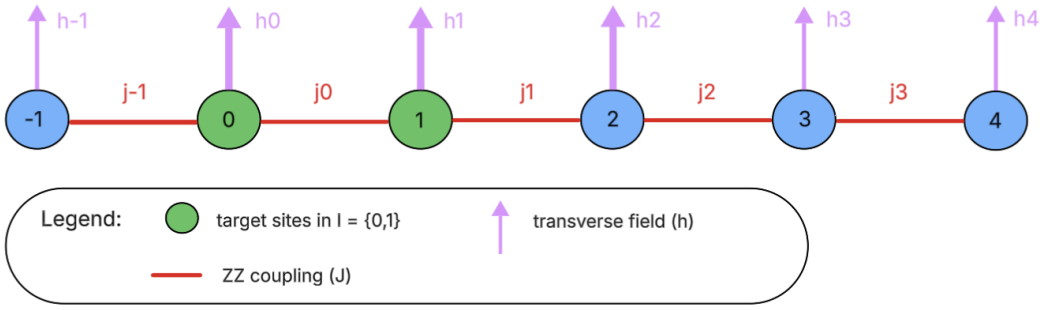
and the associated indicator $\varphi(x)_{I,x'} = \mathbf{1}[x \in T_{x',I}]$. For a given x, exactly one indicator is active, so the feature map is a one-hot partition of the local parameter space.

The surrogate predictor

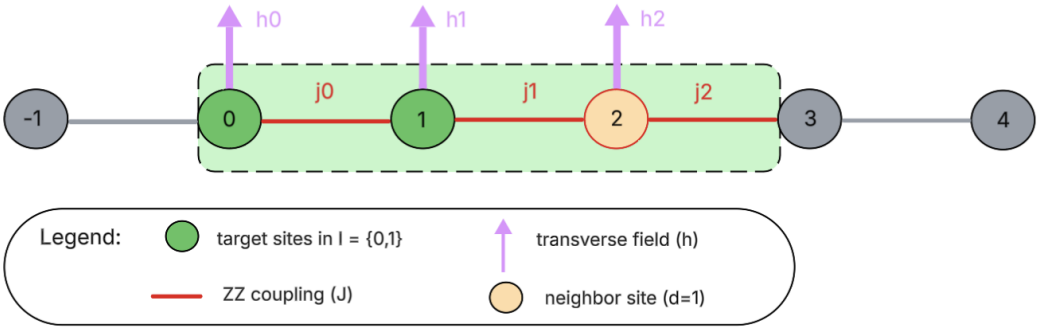
$$h_w(x) = \sum_{x' \in X_{I,1}} w_{I,x'}\, \varphi(x)_{I,x'}$$

is a table lookup approximation to $f_I(x) = \mathrm{Tr}[Z_0 Z_1\, \sigma(x)]$, with per-cell error bounded by $\varepsilon_1 \|Z_0 Z_1\|$. Summing over all $I \in S$ gives the blockwise structure described above. Therefore, $m_\varphi \le |S|\, N(\varepsilon_1)$ and $\|w\|_1 \le r(O)\, N(\varepsilon_1)$ follow exactly as in the general construction.

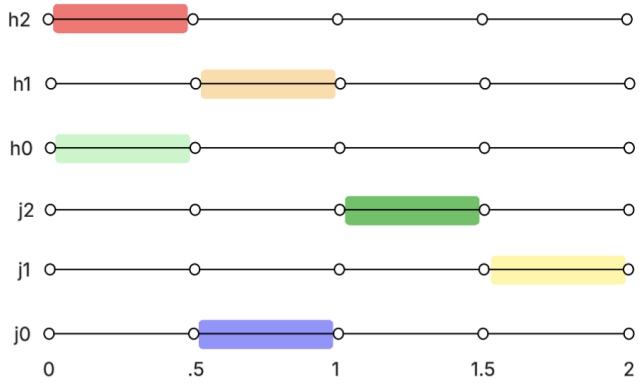
This example shows how the abstract definitions work. Locality fixes the neighborhood $S_{I,\delta}$, discretization via δ_2 ensures bias $\le \varepsilon_1$, and the indicator features φ give a finite dictionary.



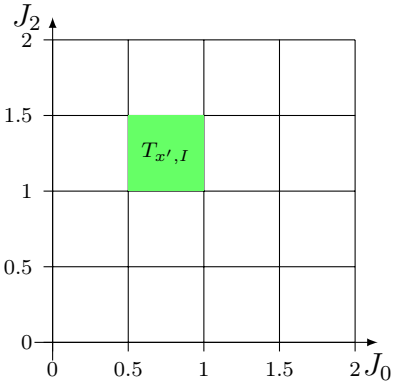
(a) Full chain with observable $O_I = Z_0 Z_1$ on sites $I = \{0, 1\}$ (green).



(b) Radius-$\delta = 1$ neighborhood $S_{I,1}$ (green+orange) containing the relevant couplings $\{J_0, J_1, J_2, h_0, h_1, h_2\}$.



(c) Discretization with mesh δ_2.



(d) Indicator partition (2D slice; axes in steps of $\delta_2 = 0.5$).

Figure 4: Toy Ising illustration of the locality-discretization pipeline. (a) Select observable support I in the full chain. (b) Restrict to the δ-neighborhood $S_{I,\delta}$ of relevant couplings. (c) Discretize the local parameter space with mesh δ_2 (highlighted intervals define $T_{x',I}$). (d) Represent each cube $T_{x',I}$ by a one-hot indicator feature $\varphi(x)_{I,x'}$.

4.2 Locality reduction: choosing $\delta(\varepsilon_1)$

We convert a global prediction task into a sum of local ones by proving that each target term $\mathrm{Tr}(O_I\, \sigma(x))$ can be uniformly approximated (to accuracy $\varepsilon_1 \|O_I\|$) by a functional depending only on the couplings within a radius-δ neighborhood of I. Remember that the choice of $\delta = \delta(\varepsilon_1)$ is only determined by locality physics. At $T = 0$ we use quasi-adiabatic spectral flow combined with LR bounds; at $T > 0$ we use quantum belief propagation plus

thermal clustering. In both cases, distant couplings contribute only with exponentially (or algebraically, for long-range) decaying weight, so truncating them brings about a controlled bias *independent of* $|\Lambda|$. This is the bridge from physics to statistics. Once $\delta(\varepsilon_1)$ is fixed, the dictionary size $N(\varepsilon_1)$ and the ℓ_1-budget $r(O)N(\varepsilon_1)$ in the LASSO bound are determined by local shell counting rather than by system size.

For each $I \in S$ there exists a local functional f_I acting only on the truncated parameter vector $x|_{S_{I,\delta}}$ such that, for $0 < \varepsilon_1 \le e^{-1}$,

$$\left|\mathrm{Tr}(O_I\,\sigma(x)) - f_I\big(x|_{S_{I,\delta}}\big)\right| \;\le\; \varepsilon_1\,\|O_I\|, \tag{21}$$

and therefore

$$\left|\mathrm{Tr}(O\,\sigma(x)) - \sum_{I\in S} f_I\big(x|_{S_{I,\delta}}\big)\right| \;\le\; \varepsilon_1\, r(O). \tag{22}$$

All constants below are independent of $|\Lambda|$. The radius $\delta = \delta(\varepsilon_1)$ depends on the interaction decay and on the regime (gapped $T{=}0$ vs. Gibbs $T{>}0$) as follows.

4.2.1 Short range (exponential)

$$\textbf{Ground state } (T{=}0): \quad \delta_0(\varepsilon_1) \;=\; \frac{2}{\mu_{\mathrm{GS}}}\,\log^2\!\Big(\frac{c}{\varepsilon_1}\Big), \quad \mu_{\mathrm{GS}} := \min\!\left\{\mu, \frac{\gamma}{2v_{\mathrm{LR}}}\right\},$$

$$\textbf{Thermal } (T{>}0): \quad \delta_\beta(\varepsilon_1) \;=\; \frac{1}{\mu_\beta}\Big(\log\frac{c}{\varepsilon_1} + c_D\Big), \quad \mu_\beta := \min\!\left\{\frac{\nu}{2}, \frac{\mu\,\pi}{8\beta\,(v_{\mathrm{LR}} + \pi/\beta)}\right\}$$

Here the GS form follows from quasi-adiabatic spectral flow and LR (Prop. A.3 / Cor. A.3.1 of [27]). The thermal form $(\delta_\beta, \mu_\beta)$ is imported from our QBP/clustering locality section, Sec 3.4.

4.2.2 Power law interactions $\alpha > 2D$

For long-range tails the light cone is algebraic, and δ must control both the LR part and the filter tail. In the GS analysis one obtains

$$\delta_0(\varepsilon_1) \;\ge\; c\,\varepsilon_1^{-1/(\nu-D)} \quad \text{up to polylog factors}$$

(Cor. A.6.1 of [27]); equivalently, dictionary growth matches the shell-counting exponent below. In the thermal case, the same shell counting yields a *polynomial* truncation radius (no extra $\log^2$ factor from the quasi-adiabatic tail), by our QBP derivation.

Basis size per neighborhood.

Let $N(\varepsilon_1)$ be the number of discretization cells on $S_{I,\delta}$ needed to approximate each f_I within ε_1 (indicator-partition features). With the above $\delta(\varepsilon_1)$,

$$N(\varepsilon_1) = \begin{cases} 2^{O(\mathrm{polylog}(1/\varepsilon_1))}, & \text{exponential decay,} \\ 2^{O\left(\varepsilon_1^{-\omega}\log(1/\varepsilon_1)\right)}, \quad \omega = \dfrac{kD}{\nu - D}, & \alpha > 2D, \end{cases}$$

so that $m_\varphi \le |S|\,N(\varepsilon_1)$ and the true ℓ_1 budget satisfies $\|w\|_1 \le r(O)\,N(\varepsilon_1)$ [27, Lem. A.11]. These are exactly the quantities entering the LASSO generalization bound (Thm. 3.1 of [27]).

Remark 4.3. Think of $N(\varepsilon_1)$ as the number of cells per neighborhood needed to resolve the local map $x|_{S_{I,\delta}} \mapsto f_I$ to accuracy ε_1. Greater accuracy forces a finer mesh δ_2. Through $\delta(\varepsilon_1)$, we are able to increase the local dimension $d_I \sim |S_{I,\delta}|$. With exponential decay, the required cells grow exponentially. With power-law tails and $\alpha > 2D$,they grow sub-exponentially. The radius $\delta(\varepsilon_1)$ depends only on locality constants and not on n. $N(\varepsilon_1)$, and therefore m_φ and the ℓ_1 budget, are system-size independent up to a mild $\log|S|$ factor in the bound.

4.3 Unified LASSO generalization bound

Theorem 4.4 (Unified ML generalization bound (LASSO)). *Assume the locality hypotheses of the previous subsection (short range with exponential clustering, or power-law with $\alpha > 2D$), and fix a truncation radius $\delta = \delta(\varepsilon_1)$ accordingly.*

Let $O = \sum_{I\in S} O_I$ and use the indicator-partition features on $S_{I,\delta}$ so that

$$m_\varphi \le |S|\, N(\varepsilon_1), \qquad \|\varphi(x)\|_\infty \le 1, \qquad \|w\|_1 \le B := r(O)\, N(\varepsilon_1).$$

Given a dataset $\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_{i=1}^N$ with $|y^{(i)} - \mathrm{Tr}(O\,\sigma(x^{(i)}))| \le \varepsilon_2$, let

$$\hat{w} \in \arg\min_{\|w\|_1 \le B} R_{\mathcal{D}}(h_w), \qquad R_{\mathcal{D}}(h) := \frac{1}{N}\sum_{i=1}^{N} \big(h(x^{(i)}) - y^{(i)}\big)^2,$$

and assume the near-minimizer condition

$$R_{\mathcal{D}}(h_{\hat{w}}) \;\le\; \frac{\varepsilon_3}{2} \;+\; \min_{\|w\|_1 \le B} R_{\mathcal{D}}(h_w).$$

Then for any $\varepsilon_1, \varepsilon_2, \varepsilon_3 \in (0,1)$, with probability at least $1-\gamma$,

$$\mathbb{E}\big[\, |h_{\hat{w}}(x) - \mathrm{Tr}(O\,\sigma(x))|^2 \,\big] \;\le\; (\varepsilon_1 + \varepsilon_2)^2 \;+\; \varepsilon_3, \tag{23}$$

provided the sample size satisfies

$$N \;\ge\; C\, r(O)^4\, \varepsilon_3^{-2}\, N(\varepsilon_1)\, \ln\!\Big(\tfrac{|S|}{\gamma}\Big), \tag{24}$$

for a universal constant $C > 0$. In particular, if $r(O) = O(1)$ and $|S| = O(n^k)$ with fixed $k = O(1)$, then $N = O(\log n)$ at fixed accuracies.

Proof: we summarize steps from the proof of [27].

(i) Local approximation. By locality, for each I there exists f_I on $S_{I,\delta}$ with $|\,\mathrm{Tr}(O_I\,\sigma(x)) - f_I(x|_{S_{I,\delta}})| \le \varepsilon_1 \|O_I\|$, providing us

$$\Big|\mathrm{Tr}(O\,\sigma(x)) - \sum_{I\in S} f_I(x|_{S_{I,\delta}})\Big| \;\le\; \varepsilon_1\, r(O).$$

(ii) Linearization over a bounded dictionary. Each f_I is ε_1-approximated by $N(\varepsilon_1)$ indicator cells on $S_{I,\delta}$, producing a linear predictor $h_{w^\dagger}$ with $\|w^\dagger\|_1 \le r(O)\, N(\varepsilon_1)$ and training MSE $\le (\varepsilon_1 + \varepsilon_2)^2$.

(iii) Empirical → population. By [27, Lem. 3.2] with r_∞=1, A=m_φ, B=$r(O)N(\varepsilon_1)$, and [27, Lem. 3.3], the population risk of $\hat{w}$ is bounded by the training risk plus a complexity term of order $B^2\sqrt{\ln m_\varphi + \ln(1/\gamma)}/\sqrt{N}$. Choosing N as in (24) (e.g., [27, Eq. (7)]) gives us (23).

□

Remark 4.5 (Equality of GS and thermal bounds). Temperature enters only through the locality radius $\delta(\varepsilon_1)$ and the associated rates used to build the same feature dictionary and the same ℓ_1 budget $B = r(O)N(\varepsilon_1)$. Thus, the learning bound (23)-(24) is identical in both regimes.

Corollary 4.6 (Equivariance reduction)**.** *If the interaction hypergraph $\mathcal{I}$ has automorphism group $G = \mathrm{Aut}(\mathcal{I})$ and the model is made G-equivariant (tying weights along orbits), then $|S|$ in (24) is replaced by $|S/G|$. On a translation-invariant 1D ring and $O = \sum_i O_i$ with O_i translates of a fixed local term, $|S/G| = 1$ and the logarithmic factor becomes $O(1)$.*

4.4 Equivariance and sample-complexity reduction

Let $\mathcal{I} = (\Lambda, E, h)$ be the interaction hypergraph and $G = \mathrm{Aut}(\mathcal{I})$ its automorphism group (permutations of sites that permute interaction terms). The group acts on subsets and parameters by $gI := \{gi : i \in I\}$ and $(g{\cdot}x)_J := x_{gJ}$. In this notation the target map

$$f(O_I, x) \;=\; \mathrm{Tr}\big[O_I\, \sigma(x)\big]$$

is G-equivariant:

$$f(O_{gI}, x) \;=\; f(O_I, g{\cdot}x) \qquad \forall\, g \in G,$$

as a direct consequence of the covariance of $H(x)$ and $\sigma(x)$ under G [27, Lem. 3.4]. Additionally, the locality neighborhoods and discretization grids transform compatibly, $S_{gI,\delta} = gS_{I,\delta}$ and $X_{gI,\delta} = g^{-1}\cdot X_{I,\delta}$, so the features $\varphi(x)_{I,x'} = \mathbf{1}[x \in T_{x',I}]$ admit a G-equivariant parametrization. Specifically, choosing weights to satisfy

$$w_{gI,\,x'} \;=\; w_{I,\,g^{-1}\cdot x'} \qquad (\text{all } g \in G)$$

enforces equivariance at the model level, $h_{O_{gI}}(x) = h_{O_I}(g{\cdot}x)$ [27, Prop. 3.5]. Under this tying, one free parameter is retained per G-orbit in S, so the number of independent coefficients drops from $|S|$ to $|S/G|$, and the logarithmic factor in the generalization bound ([27, Thm 3.1]) correspondingly decrease:

$$N \;=\; \frac{r(O)^4}{\varepsilon_3^2}\, N(\varepsilon_1)\, O\big(\log(|S/G|/\gamma)\big)\,,$$

e.g., replace $\log|S|$ by $\log|S/G|$ [27, (Cor. 3.5.1)]. For a 1D ring with translations and $O = \sum_i O_i$ where all O_i are translates of a fixed local term, $|S/G| = 1$ and the log factor is $O(1)$, yielding constant-in-n sample complexity. All statements here follow the equivariant construction and bounds in [27].

Remark 4.7. The single observable reduction does not require distributional symmetry. G-invariance of the sampling distribution matters only for multi-observable or shadow extensions. Approximate equivariance, such as with open boundaries, still improves conditioning in practice.

4.5 Random features and convolutional realizations

For a fixed locality radius δ, the indicator-partition features may be replaced (in practice) by *random features* on the local block $Z_I(x) := (x_J)_{J\in S_{I,\delta}}$:

$$\Phi_{\mathrm{RF}}(x)_{I,r} := \sigma\big(b_{I,r}^{\top} Z_I(x)\big), \qquad F_{O_I}(x) = \sum_{r=1}^{R} a_{I,r}\, \Phi_{\mathrm{RF}}(x)_{I,r},$$

with i.i.d. $b_{I,r}$ and a bounded, 1-Lipschitz activation σ (e.g. tanh), scaled so that $\|\Phi_{\mathrm{RF}}(x)\|_\infty \le 1$. In translation-invariant (more generally, G-equivariant) settings, tying parameters along orbits produces a two layer convolutional realization and the same orbit reduction $|S| \to |S/G|$ [27, Prop. 3.6, Sec. 3.2.3, Sec. 4.1].

Proposition 4.8 (Random Fourier Feature bound)**.** *Assume an RF model with R features per block is used and normalized so that $\|\Phi_{\mathrm{RF}}(x)\|_\infty \le 1$, and (optionally) weights are tied along G-orbits so that only $|S/G|$ blocks remain free (Sec. 4.4). If there exist coefficients with $\|w\|_1 \le B_{\mathrm{RF}}$ whose training MSE is at most $(\varepsilon_1+\varepsilon_2)^2$ (here ε_1 is the locality/truncation error from Sec. 4), then the same linear ℓ_1 generalization machinery used in Theorem 4.4 applies, yielding*

$$\mathbb{E}\big[|h_{\hat{w}}(x) - \mathrm{Tr}(O\,\sigma(x))|^2\big] \;\le\; (\varepsilon_1+\varepsilon_2)^2 + \varepsilon_3$$

whenever

$$N \;\ge\; C\,B_{\mathrm{RF}}^2\,\varepsilon_3^{-2}\,\ln\!\Big(\tfrac{M}{\gamma}\Big), \qquad M = \begin{cases} R\,|S|, & \text{no tying,} \\ R\,|S/G|, & \text{with tying,} \end{cases}$$

for a universal constant $C > 0$.

Including temperature as an input

Fixed β. If all data are at a single β_0, appending β to Z_I has no effect (it is a constant coordinate). The feature linear bound and Prop. 4.8 remain unchanged.

4.6 Complexity summary and limitations

Sample complexity and risk

Under the locality setup of Sec. 4 (short range with exponential clustering, or power-law with $\alpha > 2D$), using the locality dictionary with $m_\varphi \le |S|\,N(\varepsilon_1)$, $\|\varphi(x)\|_\infty \le 1$, and $\|w\|_1 \le B := r(O)\,N(\varepsilon_1)$, the LASSO estimator $\hat{w}$ satisfies, with probability at least $1-\gamma$,

$$\mathbb{E}\big[|h_{\hat{w}}(x) - \mathrm{Tr}(O\,\sigma(x))|^2\big] \;\le\; (\varepsilon_1+\varepsilon_2)^2 + \varepsilon_3,$$

provided

$$N \;\ge\; C\,r(O)^4\,N(\varepsilon_1)\,\varepsilon_3^{-2}\,\ln\!\Big(\tfrac{|S|}{\gamma}\Big),$$

for a universal constant $C > 0$.

Growth in n and ε_1

If $r(O) = O(1)$ and $|S| = O(n^k)$ with fixed $k = O(1)$, then

$$N = \tilde{O}\big(N(\varepsilon_1)\,\varepsilon_3^{-2}\,\log n\big), \qquad N(\varepsilon_1) = \begin{cases} 2^{O(\mathrm{polylog}(1/\varepsilon_1))}, & \text{short range,} \\ 2^{O\left(\varepsilon_1^{-\omega}\log(1/\varepsilon_1)\right)}, \quad \omega = \dfrac{kD}{\nu - D}, & \alpha > 2D. \end{cases}$$

Equivariance adjustment

If weights are tied along G-orbits (Sec. 4.4), replace $\log|S|$ by $\log|S/G|$. For a 1D translation-invariant ring with $O = \sum_i O_i$ (translates of a fixed local term), $|S/G| = 1$ and the log factor is $O(1)$.

Once locality fixes $\delta(\varepsilon_1)$ (GS with QASF, thermal with QBP), the *same* LASSO analysis applies in both regimes; thermalization changes only the locality rate through μ_β, not the learning form of the bound.

5 Experiments

In this section, we evaluate the performance of the locality-based LASSO model in predicting thermal quantum observables, following the framework of Smíd *et al.* [27]. We consider one-dimensional Heisenberg and long-range Ising chains at various inverse temperatures ($\beta = 1, 3, 5, 7, 10$). For each system, we generate datasets of coupling parameters and corresponding thermal expectation values of chosen observables, and we examine how the number of training samples required for a fixed accuracy scales with system size. The results are compared against theoretical predictions (e.g., logarithmic scaling in system size), with attention to any deviations in regimes where the theory's conditions may break down.

5.1 Data

Notation

We write L for chain length and use a fixed temperature grid $\beta \in \{1, 3, 5, 7, 10\}$. Couplings are i.i.d. draws (specified below). Observables are normalized so their variance is $O(1)$ across L.

5.1.1 Models studied

Definition 5.1 (Heisenberg chain)**.** A spin-$\frac{1}{2}$ nearest-neighbor Heisenberg model on a 1D chain (open or periodic),

$$H \;=\; \sum_{\langle ij \rangle} J_{ij} \left(X_i X_j + Y_i Y_j + Z_i Z_j \right),$$

with i.i.d. bond couplings $J_{ij} \sim \mathrm{Unif}[0, 2]$ (as in prior work).

Definition 5.2 (Long-range transverse field Ising chain)**.** A power-law Ising model with transverse field,

$$H \;=\; \sum_{i<j} \frac{J_i J_j}{|i-j|^{\alpha}} Z_i Z_j \;+\; \sum_i h_i X_i,$$

with site couplings $J_i \sim \mathrm{Unif}[0, 2]$ and a uniform field $h_i \equiv h$. We consider $\alpha = 3$ (fast decay, $\alpha > 2D$ in 1D) and $\alpha = 1.5$ (slower decay, $\alpha < 2D$).

Remark 5.3. The Heisenberg model tests short range locality, while the long-range Ising family probes how weaker locality (α smaller) impacts learning difficulty and required truncation radii.

5.1.2 Targets and normalization

Definition 5.4 (Measured observables)**.** For Heisenberg we use the two-site correlator and the energy. To compare fairly across L, we report a standardized extensive observable

$$C_{i,i+1} \;=\; \tfrac{1}{3}\left(X_i X_{i+1} + Y_i Y_{i+1} + Z_i Z_{i+1} \right), \qquad \widetilde{E}_L \;:=\; \frac{1}{\sqrt{L}} \sum_{\langle ij \rangle} J_{ij} \left(X_i X_j + Y_i Y_j + Z_i Z_j \right)$$

and analogously standardized sums for Ising. This keeps label variance $O(1)$ as L grows.

Remark 5.5 (Scale choice from [27]). Standardizing sums by $\sqrt{L}$ avoids trivial shrinkage of variance with L and makes a fixed RMSE threshold comparable across sizes.

5.1.3 State preparation (imaginary-time TEBD)

Definition 5.6 (Purification via imaginary-time TEBD)**.** We prepare thermal states by evolving the infinite-temperature purification $|\psi(0)\rangle$ on system⊗ancilla to $|\psi(\beta/2)\rangle = e^{-\beta H/2}|\psi(0)\rangle$ and set

$$\rho_\beta \;\propto\; \mathrm{Tr}_{\mathrm{anc}}\big[\,|\psi(\beta/2)\rangle\langle\psi(\beta/2)|\,\big],$$

e.g., imaginary time runs only to $\beta/2$ because H acts on the physical system while the ancilla is traced out [26, Eqs. 237-238]. Time evolution is performed with second-order (Strang) Trotter-Suzuki TEBD: split $H = \sum_b h_b$ into even/odd layers and apply two-site gates $e^{-\Delta\tau\, h_b}$ alternately, followed by SVD-truncate after each gate [30, 4]. The purification/ancilla approach to finite-temperature DMRG is standard [26, 4].

Remark 5.7. Purification halves the imaginary-time budget (to $\beta/2$) and lets us reuse standard two-site TEBD. For gapped short range models, Lieb-Robinson locality keeps required bond dimensions moderate.

5.1.4 TenPy realization (nearest-neighbor) [13]

1. *Initialization.* Build the MPO H (e.g. `HeisenbergChainOpenBC`) and create the infinite-temperature purification $|\psi(0)\rangle$ with `PurificationMPS.from_infiniteT`.
2. *Engine.* Use `tenpy.algorithms.purification.PurificationTEBD` with second-order Trotter-Suzuki, step size $\Delta\tau$ (code variable `dt`), and truncation parameters (max bond $\chi_{\max}$, SVD cutoff). Each Trotter layer applies two-site gates then SVD-truncate.
3. *Scheduling.* For a monotone grid $\beta_1 < \cdots < \beta_K$, call `eng.run_imaginary((beta_k - beta_{k-1})/2)` so that the cumulative imaginary time equals $\beta_k/2$ after step k.
4. *Measurements.* After each step, contract MPO/MPS to obtain the standardized energy (we report $\widetilde{E}_L$ for ML) and the two-site correlator $C_{i,i+1} = \frac{1}{3}(X_iX_{i+1} + Y_iY_{i+1} + Z_iZ_{i+1})$. We monitor discarded weight to track truncation error.

Long-range specifics ($\alpha = 1.5, 3$)

We approximate $1/r^\alpha$ by a short sum of exponentials on $[1, L]$ to form a tractable MPO; evolution uses either TEBD (when the layout permits) or `PurificationApplyMPO` with the same SVD-truncate controls. These runs are costlier due to larger MPO bond dimension.

5.1.5 Validation and scope

For $L \lesssim 20$ we cross-check against exact diagonalization. For larger L we use the purification-TEBD pipeline with tight SVD cutoff and sufficiently large $\chi_{\max}$. The dataset comprises input and output pairs $(x, \beta) \mapsto \langle O\rangle_{\rho_\beta(x)}$ for $L \in \{8, \ldots, 128\}$ and $\beta \in \{1, 3, 5, 7, 10\}$.

5.2 Machine learning setup

Objects and locality

As in Sec. 4, we use the radius-δ neighborhood $S_{I,\delta}$ and its local grid $X_{I,\delta}$ to restrict predictors to nearby couplings (see Eq. 21 for the truncation guarantee).

Let $O = \sum_{I \in S} O_I$ with $\mathrm{supp}(O_I) = I \subset \Lambda$ and $|I| \le p = O(1)$. Write the Hamiltonian parameters as $x = (x_J)_{J \in \mathcal{P}_{\le k}(\Lambda)}$. For $\delta \in \mathbb{N}$, define

$$S_{I,\delta} := \{\, J \in \mathcal{P}_{\le k}(\Lambda) : \ \mathrm{dist}(I, J) \le \delta, \ \mathrm{diam}(J) \le \delta \,\}.$$

Definition 5.8 (Local parameter block)**.** For each $I \in S$, set $Z_I^{(\delta)}(x) := (x_J)_{J \in S_{I,\delta}} \in \mathbb{R}^{\ell_I(\delta)}$ and augment with the inverse temperature: $z_I^{(\delta)}(x, \beta) := (Z_I^{(\delta)}(x),\ \beta) \in \mathbb{R}^{\ell_I(\delta)+1}$.

Remark 5.9. $Z_I^{(\delta)}$ is the local view of x available to O_I; locality bounds imply sites farther than δ contribute only exponentially/polynomially small tails, so we truncate them.

5.2.1 Random Fourier features [24]

These randomized smooth features are a practical replacement for the indicator-partition features used in the proof (Sec. 4.5), preserving the same locality factorization.

Let $R \in \mathbb{N}$ and draw i.i.d. $\omega_1, \dots, \omega_R \sim \mathcal{N}(0, \gamma I)$ with bandwidth $\gamma > 0$.

Definition 5.10 (RFF map)**.** For $z \in \mathbb{R}^{\ell_I(\delta)+1}$ define

$$\phi_{\mathrm{RFF}}(z) := \big(\cos(\omega_1^\top z),\ \sin(\omega_1^\top z), \dots, \cos(\omega_R^\top z),\ \sin(\omega_R^\top z)\big)^\top \in \mathbb{R}^{2R}.$$

The global feature vector is the concatenation

$$\Phi^{(\delta)}(x, \beta) := \big[\, \phi_{\mathrm{RFF}}(z_I^{(\delta)}(x, \beta)) \,\big]_{I \in S} \in \mathbb{R}^{m_\Phi}, \qquad m_\Phi \le 2R\,|S|.$$

Remark 5.11. RFFs give a smooth, bounded basis for functions of the local block $z_I^{(\delta)}$, acting as a differentiable proxy for the indicator locality features (that we had mentioned in Sec 4) while preserving spatial decoupling.

5.2.2 Predictor and training

Definition 5.12 (Linear predictor with ℓ_1 budget)**.** We learn

$$h_w(x, \beta) := \langle w, \Phi^{(\delta)}(x, \beta) \rangle, \qquad \hat{w} \in \arg\min_{\|w\|_1 \le B} \ \frac{1}{N} \sum_{i=1}^{N} \big(h_w(x^{(i)}, \beta^{(i)}) - y^{(i)}\big)^2.$$

Remark 5.13. The ℓ_1 constraint encourages using only a few local features per neighborhood. Effective capacity is governed by (m_Φ, B); δ (from locality theory) controls bias, while (R, γ, B) are tuned by cross-validation.

Equivariance

Remark 5.14. If the interaction hypergraph $\mathcal{I} = (\Lambda, E, h)$ demonstrates a nontrivial automorphism group $G = \mathrm{Aut}(\mathcal{I})$, we tie parameters along orbits: for $g \in G$, set $w_{gI} := w_I$ after transporting inputs accordingly. This reduces free parameters from $|S|$ to $|S/G|$ and tightens the log-factor in the sample complexity. Symmetry tells the model that equivalent neighborhoods should be treated the same, so a single learned coefficient generalizes across all symmetry related sites.

LASSO model and training

Our LASSO objective matches the ℓ_1-constrained ERM analyzed in Sec. 4. The sample-size requirement and population risk are given by Thm. 4.4.

Definition 5.15 (Predictor and objective)**.** Given the feature map $\Phi^{(\delta)}(x, \beta)$, we learn an affine predictor

$$\hat{y}(x, \beta) = y_0 + \langle w, \Phi^{(\delta)}(x, \beta) \rangle,$$

by solving the LASSO problem

$$\min_{y_0, w} \frac{1}{2N} \sum_{i=1}^{N} \Big(y^{(i)} - y_0 - \langle w, \Phi^{(\delta)}(x^{(i)}, \beta^{(i)}) \rangle \Big)^2 + \lambda \|w\|_1,$$

where $\lambda > 0$ is the regularization strength.

Remark 5.16. The ℓ_1 penalty encourages using only a few local features per neighborhood, controlling variance and improving interpretability. Its effect is complementary to the truncation radius δ (bias) and the RFF budget R (approximation quality).

Hyperparameters and calibration.

Definition 5.17 (Chosen grids and reuse of features)**.** We fix the locality radius to $\delta = 4$ (sufficient under exponential clustering) and tune the remaining hyperparameters by 5-fold cross-validation over

$$R \in \{5, 10, 20, 40\}, \qquad \gamma \in \{0.4, 0.5, 0.6, 0.65, 0.7, 0.75\}, \qquad \lambda \in \{2^{-8}, 2^{-7}, 2^{-6}, 2^{-5}\}.$$

Once selected, these hyperparameters are held fixed across all system sizes and β values for the given model. A single draw of RFF frequencies $\{\omega_r\}_{r=1}^{R}$ is used to construct *all* training and test features. We do not re-sample Ω between splits.

Remark 5.18 (Why fix Ω). Keeping the same random features defines a single kernel space for the entire study, ensuring that changes in performance reflect physics (size/temperature) rather than feature re-sampling noise.

5.2.3 Sample-complexity protocol

Definition 5.19 (Procedure for $N_{\text{req}}(n)$)**.** For each model, system size n, and inverse temperature β:

1. Draw a held-out test set of $N_{\text{test}} = 40$ fresh disorder instances.
2. Train on an increasing number N of i.i.d. training samples from the same distribution, starting small and incrementing N.
3. After each increment, compute the test RMSE and record the smallest N achieving the target accuracy ϵ.
4. Repeat with several random seeds (we report mean $\pm$ one standard deviation of N_{req}).

Targets follow the literature: for Heisenberg, $\epsilon = 0.55$; for long-range Ising, $\epsilon = 0.30$ at $\alpha = 3$ and $\epsilon = 0.15$ at $\alpha = 1.5$. The resulting $N_{\text{req}}(n)$ traces the empirical sample-complexity scaling.

Implementation note

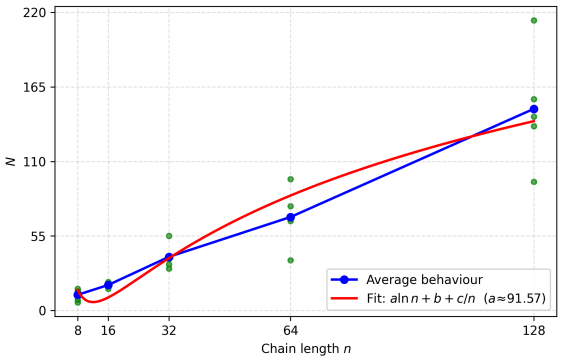
All models use `scikit-learn`'s LASSO (coordinate descent, with intercept) [23]. Results are stable across random seeds for the training order, cross-validation folds, and the fixed RFF matrix Ω.

5.3 Results

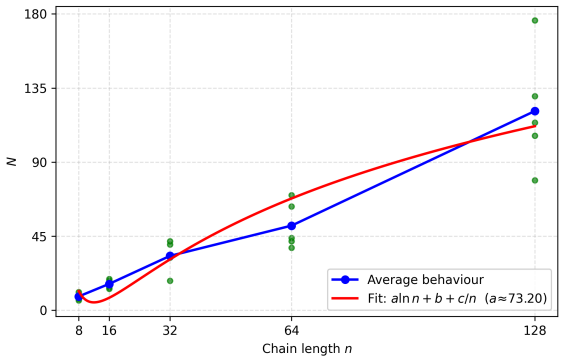
5.3.1 Heisenberg model

Heisenberg (open): logarithmic sample complexity.

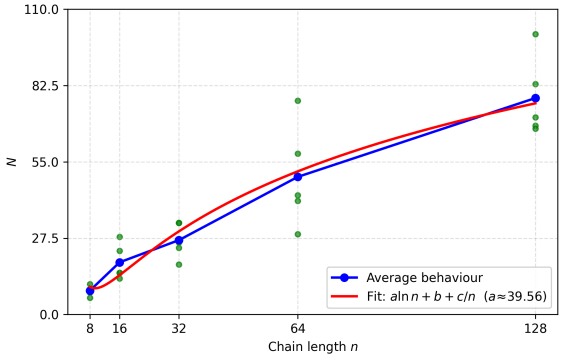
Across all temperatures the required samples grow only logarithmically with system size, in line with the locality-based prediction for short range gapped systems. Temperature mainly changes the prefactor of the $\ln n$ fit. The mid-temperature case $\beta \approx 5$ is easiest, since its fitted coefficient of the log term is smallest (panel values are $a_{\beta=1} \approx 92$, $a_3 \approx 73$, $a_5 \approx 40$, $a_7 \approx 64$, $a_{10} \approx 71$), and the data adhere most closely to the $a \ln n + b + c/n$ curve. This matches the physics intuition. At low β (high T) thermal averaging weakens the local signal, while at high β the correlation length increases and the thermal state approaches the ground state. With a fixed locality radius ($\delta = 4$) the optimal truncation radius $\delta^*(\beta)$ likely grows, so under-truncation introduces bias that offsets any expected gain from being closer to the GS. As a result, we observe a shallow increase in samples at large β relative to the mid-β case. Residual scatter around the fit stems from disorder realizations and finite test sets, as well as machine learning variability from cross-validated hyperparameters and finite RFF features, and is largest for small n where the c/n finite-size term dominates.



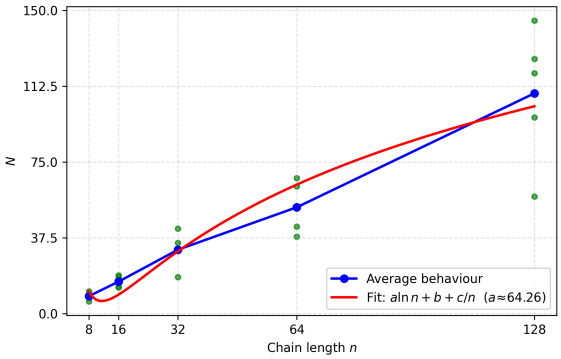
(a) $\beta = 1$



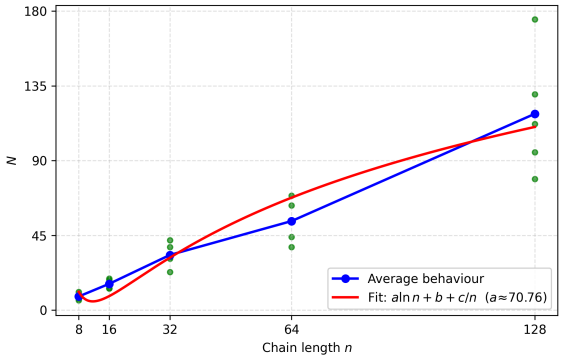
(b) $\beta = 3$



(c) $\beta = 5$



(d) $\beta = 7$



(e) $\beta = 10$

Figure 5: Required training samples $N_{\mathrm{req}}(n)$ to reach $\epsilon = 0.55$ as a function of system size n for $\beta \in \{1, 3, 5, 7, 10\}$. Points are individual trials; solid curves show the fit $N = a \ln n + b + c/n$.

Temperature sweep at fixed size.

At fixed chain length ($n = 32$) and locality window ($\delta = 4$), Fig. 6 plots the training samples needed to reach $\epsilon = 0.55$ versus β (mean $\pm 1\sigma$ over five repeats; faint points are per-repeat counts). The dependence is a shallow U-shape: N_{tr} is smallest around $\beta \approx 4$–5 (mid-20s on average), with mild increases and larger dispersion near $\beta \approx 2$ and $\beta \approx 6$. For

$\beta \gtrsim 8$ the mean returns to the same plateau. Our interpretation of this is twofold: at low β (high T) thermal averaging flattens the local signal, requiring more data, while at higher β the correlation length grows so a fixed radius $\delta = 4$ under-captures longer-range influence, slightly increasing bias/variance and thus N_{tr}.

Temperature dependence at fixed size (open boundaries)

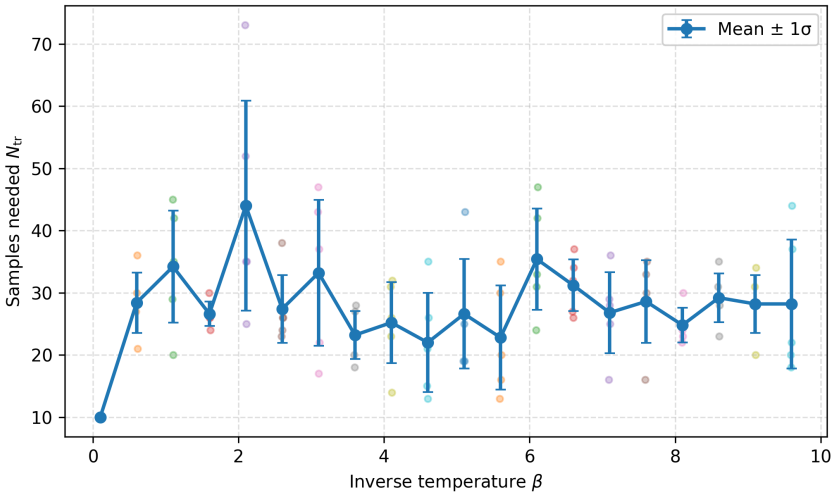
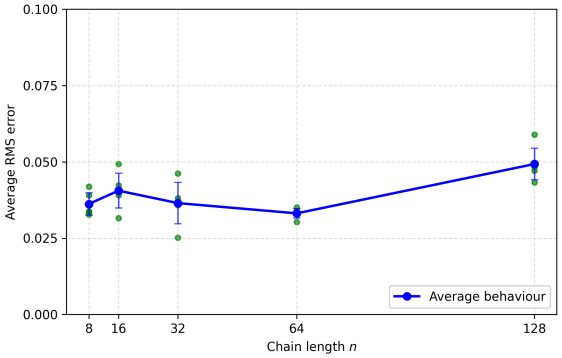


Figure 6: Required training samples N_{tr} to reach $\epsilon = 0.55$ versus inverse temperature β for a fixed system size (here $n = 32$) and locality window $\delta = 4$. Error bars show mean $\pm$ 1σ over 5 repeats; faint points are the repeat trials. Random Fourier features and LASSO hyperparameters were selected by cross-validation.

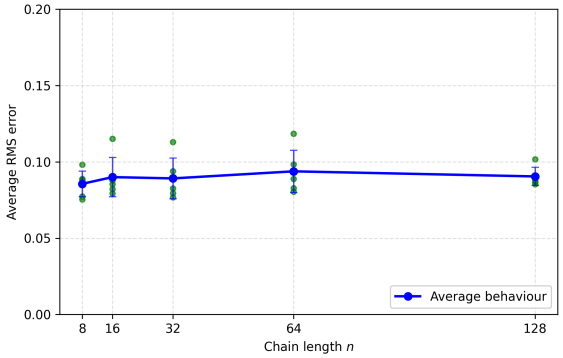
Heisenberg (periodic): constant-in-n via equivariance.

With periodic boundaries the predictor is tied across translation orbits, so the number of free coefficients is independent of n (one orbit of nearest-neighbor bonds). Locality then implies that the distribution of δ-neighborhoods is stationary in n, so the generalization error at a fixed training budget should be almost constant. We observe this behavior for each $\beta \in \{1, 3, 5, 7, 10\}$, where the RMSE curve is flat within error bars as n ranges from 8 to 128. The β sweep reflects physics rather than size. Errors are smallest at low or intermediate temperatures (e.g. $\beta \approx 1$ to 5) and slightly higher at large β, consistent with a growing correlation length that a fixed radius $\delta = 4$ undercaptures. Residual fluctuations at very small n arise from finite-size effects and disorder averaging, not from a systematic trend with n. Overall the figure supports the symmetry prediction. Once equivariance reduces the parameter count from $|S|$ to $|S/G| = 1$ the sample complexity does not grow with system size, in contrast with the logarithmic growth under open boundaries.

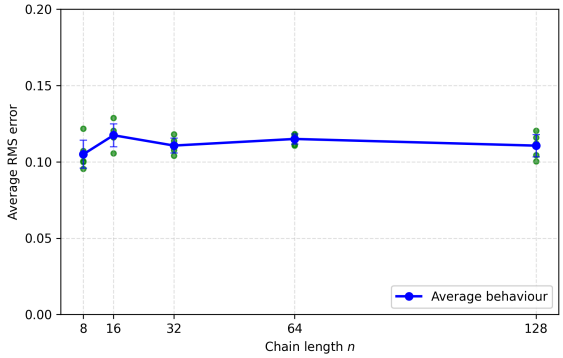
Prediction error for the disordered Heisenberg chain (periodic boundaries)



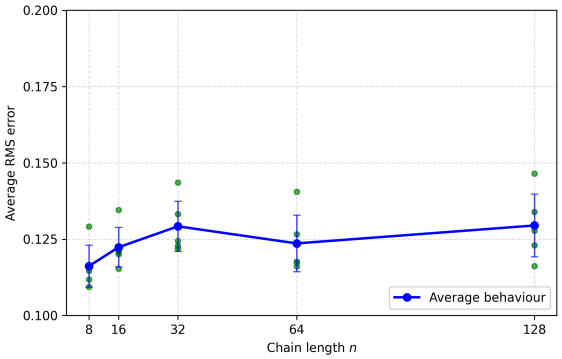
(a) $\beta = 1$



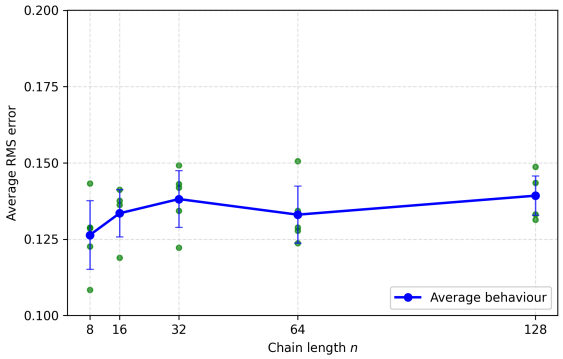
(b) $\beta = 3$



(c) $\beta = 5$



(d) $\beta = 7$



(e) $\beta = 10$

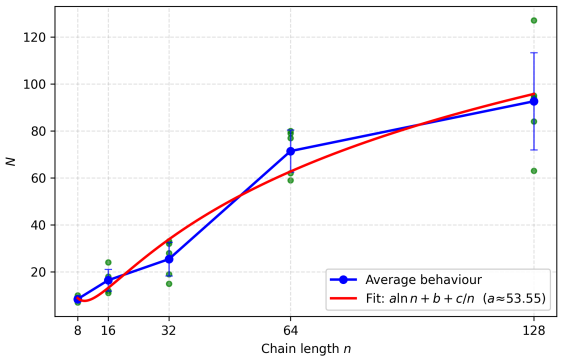
Figure 7: Average RMSE for the energy predictor as a function of system size n at fixed training budget, for inverse temperatures $\beta \in \{1, 3, 5, 7, 10\}$ (panels (a)-(e)). Green dots show individual trials; the blue curve shows the mean across trials with error bars. The error remains approximately constant with n, indicating that the number of training samples required to maintain a fixed accuracy does not grow with system size in the translationally invariant setting.

5.3.2 Long-range Ising model

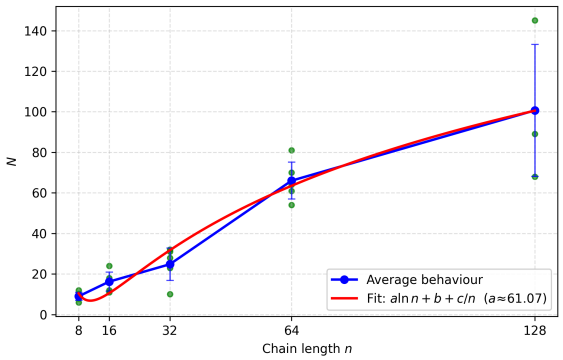
Long-range Ising model ($\alpha = 3$): logarithmic scaling in the fast-decay regime

We examine the long-range Ising chain with power-law interactions $1/r^3$, which lies in the regime $\alpha > 2D$ where efficient learning is theoretically expected. We predict the normalized energy (a sum of two-body Z_iZ_j terms plus transverse field terms) at several temperatures. For a fixed accuracy $\epsilon = 0.3$, the required training set size grows only logarithmically with n: for $n \geq 8$ the data fit well to $N_{\text{req}}(n) = a \ln n + b + c/n$. The leading $\ln n$ behavior agrees with the finite-temperature learning applied to $\alpha > 2D$, reflecting that correlations decay sufficiently fast to keep the effective model complexity in check (see Fig. 8).

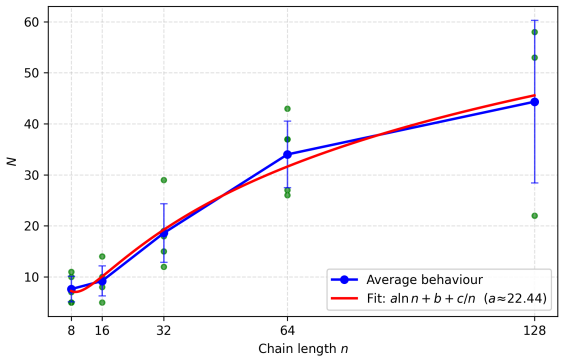
Sample complexity for the long-range Ising chain ($\alpha = 3$)



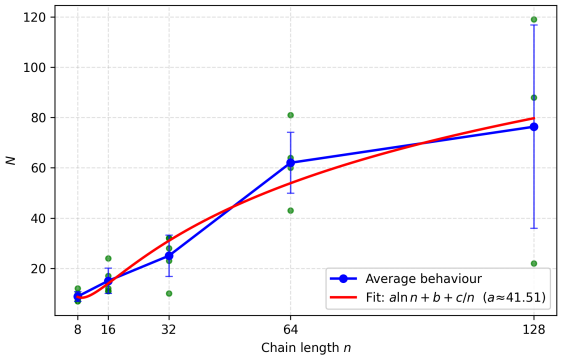
(a) $\beta = 1$



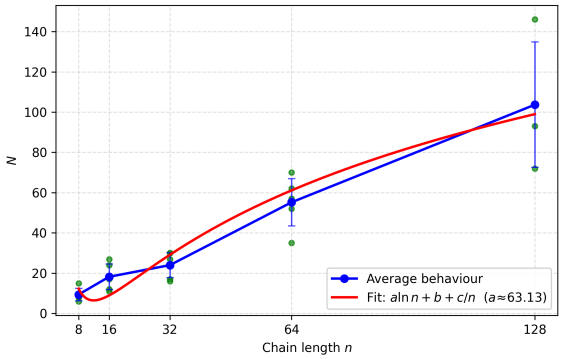
(b) $\beta = 3$



(c) $\beta = 5$



(d) $\beta = 7$



(e) $\beta = 10$

Figure 8: Required training samples $N_{\text{req}}(n)$ to reach $\epsilon = 0.3$ as a function of system size n for $\beta \in \{1, 3, 5, 7, 10\}$. Points are individual trials; solid curves show the fit $N = a \ln n + b + c/n$. The near logarithmic growth across temperatures is consistent with the $\alpha > 2D$ theory.

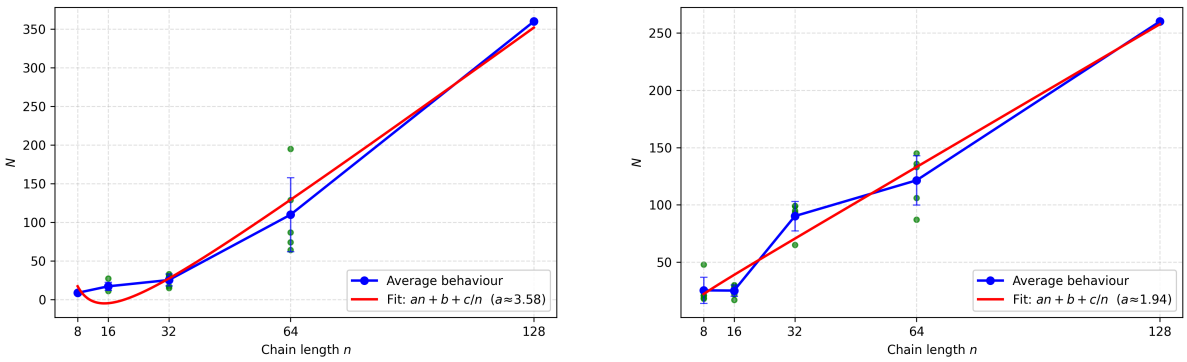
Long-range Ising model ($\alpha = 1.5$): deviation from efficient scaling in the slow-decay regime

We finally consider the more demanding $\alpha = 1.5$ case, which violates the $\alpha > 2D$ condition (with $2D = 2$ in 1D). In this regime the interactions are sufficiently long-ranged that

the locality-based theory does not guarantee efficient logarithmic learnability. Consistent with this, our experiments show clear linear scaling. To reach a comparable accuracy (we use $\epsilon = 0.15$), the required training set size grows much faster than logarithmically and is roughly linear over the sizes we could simulate. A fit of the form $N \approx a\,n + b + c/n$ captures the data well.

Computational note. Generating data at $\alpha = 1.5$ is notably more expensive than for $\alpha = 3$. Our TenPy-based simulator represents the $1/r^{\alpha}$ tail by a sum of exponentials fitted on $[1, L]$, which increases the MPO width and, together with imaginary-time purification TEBD, drives the bond dimension adaptively (up to a cap). Each β point requires evolving the purification to $\beta/2$ with adaptive time steps and repeated recompressions; see the code excerpt above (`_fit_power_law` and `_simulate_beta`)) for the exact routine (sum-of-exponentials fit, adaptive χ, stopping criteria). Because of this computational cost we report two representative temperatures, $\beta = 1$ and $\beta = 5$ (Fig. 9). We caution that these two panels do not necessarily capture the asymptotic convergence of the energy at those β values (the TEBD stopping rule balances accuracy and runtime), and thus the displayed $N_{\mathrm{req}}(n)$ should be viewed as indicative rather than definitive. Nonetheless, we do reveal a qualitative trend: the $\alpha = 1.5$ chain demands many more samples as n grows than the $\alpha = 3$ chain, emphasizing that slow decaying interactions degrade the effectiveness of locality driven learning. This aligns with theoretical expectations that when interactions are too long-ranged (or correlation lengths too large), the underlying locality assumptions fail to curb the model complexity, and the sample complexity may grow polynomially in n rather than logarithmically.

Sample complexity for the long-range Ising chain ($\alpha = 1.5$)



(a) $\beta = 1$ (b) $\beta = 5$

Figure 9: Required training samples $N_{\mathrm{req}}(n)$ to reach $\epsilon = 0.15$ as a function of system size n for the inverse temperatures shown (panels (a)-(b)). Points are individual trials; solid red curves show the fit $N = a\,n + b + c/n$. The growth is significantly steeper than $\ln n$, consistent with the breakdown of the efficient learning regime when $\alpha \leq 2D$.

6 Conclusion and Future Work

6.1 Conclusion

This dissertation develops a unified locality framework for learning observables of quantum lattice systems at both zero and finite temperature by placing quasi-adiabatic spectral flow (QASF) and quantum belief propagation (QBP) on the same standing and deriving machine learning guarantees. On the physics side, we recognize common derivative identities (Eqs. 1-2) and prove that the propagation filters are quasi-local under Lieb-Robinson (LR) dynamics. In the thermal setting this produces the explicit decay rate

$$\mu_\beta = \min\left\{\frac{\nu}{2},\ \frac{\mu\pi}{8\beta\,(v_{\mathrm{LR}} + \pi/\beta)}\right\},$$

and the logarithmic radius–accuracy law

$$\delta_\beta(\varepsilon) = \frac{1}{\mu_\beta}\left(\log\frac{K_\beta}{\varepsilon} + c_D\right),$$

while at $T{=}0$ we import the ground-state law $\delta_0(\varepsilon) = \frac{2}{\mu_{\mathrm{GS}}}\log^2\!\big(c/\varepsilon\big)$ with $\mu_{\mathrm{GS}} = \mu/2$ (Sec. 3.3). These locality inputs can be placed directly into a unified LASSO generalization theorem with sample complexity $\tilde{O}\big(N(\varepsilon_1)\,\varepsilon_3^{-2}\log|S|\big)$ and an equivariance reduction $|S| \to |S/G|$ producing constant-in-n behavior on translation invariant rings (Thm. 4.4, Cor. 4.6).

Short- vs long-range locality bounds.

- **Short range / exponential interactions.** Quasi-locality of the filters holds with exponential tails:

  $$\big\|\Phi_H^{(\beta)}(h_A) - \Phi_{H_B}^{(\beta)}(h_A)\big\| \le c'\,|A|\,\|h_A\|\,e^{-\mu'\,\mathrm{dist}(A,B^c)}, \quad \mu' = \mu\min\Big\{\tfrac{1}{2},\ \tfrac{\pi}{4\beta\,(v_{\mathrm{LR}}+\pi/\beta)}\Big\},$$

  and, at $T{=}0$,

  $$\big\|\Phi_H^{(\gamma)}(h_A) - \Phi_{H_B}^{(\gamma)}(h_A)\big\| \le C_1|A|\|h_A\|e^{-\mu\,\mathrm{dist}(A,B^c)/2} + C_N|A|\|h_A\|\Big(\tfrac{\mathrm{dist}(A,B^c)}{v_{\mathrm{LR}}}\Big)^{-(N-1)}.$$

  These suggest the radius laws stated above: $\delta_0(\varepsilon) = \frac{2}{\mu_{\mathrm{GS}}}\log^2(c/\varepsilon)$ and $\delta_\beta(\varepsilon) = \mu_\beta^{-1}\big(\log(K_\beta/\varepsilon) + c_D\big)$ (Secs. 3.3.1, 3.4.1).

- **Long range / power-law interactions ($\alpha > 2D$).** Quasi-locality degrades to a polynomial tail:

  $$\big\|\Phi_H(h_A) - \Phi_{H_B}(h_A)\big\| \le C\,|A|\,\|h_A\|\Big(1 + \tfrac{\mathrm{dist}(A,B^c)}{\xi}\Big)^{-p}, \quad p = p(\alpha,D) > 0,$$

  leading to polynomial radius laws. Specifically rendered as

  $$\delta_0(\varepsilon) \gtrsim \varepsilon^{-1/(\nu-D)}\ \text{(up to polylog factors)}, \qquad \delta_\beta(\varepsilon) \lesssim \varepsilon^{-1/\kappa(\alpha,D)},\ \kappa(\alpha,D) = p(\alpha,D) - D,$$

  while for $\alpha \le 2D$ no uniform truncation is possible (Sec. 3.3.1, Sec. 4.2.2).

Empirically, we validate these predictions on (i) short-range Heisenberg chains, observing $N_{\mathrm{req}}(n) \sim a \ln n + b$ under open boundaries and size-independent error under periodic equivariant training (Figs. 5-8); and (ii) long-range Ising chains, where the $\alpha = 3 > 2D$ case (fast decay) exhibits near-logarithmic growth (Fig. 8), while $\alpha = 1.5 \leq 2D$ (slow decay) departs from the efficient regime and shows roughly linear growth over accessible sizes (Fig. 9). These results align with the theory. Exponential (or sufficiently fast) clustering supports small truncation radii, whereas $\alpha \leq 2D$ removes uniform truncation guarantees and inflates sample complexity (Secs. 3.3, 5.3).

6.1.1 Summarized Contributions

- *Foundations reproduced and standardized.* QASF (T=0) and QBP (T>0) recast in a common notation; derivative identities and quasi-local generators/filters stated under a unified LR-data convention; uniform assumptions collected (QBP identity, κ_β Δ-split, two-dynamics LR, and quasi-locality of $\Phi^{(\beta)}$) so all constants are independent of the system size.

- *New theoretical results.* Unified locality for GS and thermal phases; explicit rates μ_β and radius laws $\delta_\beta(\varepsilon)$ (short range) and their polynomial analogues (long range); T=0 law $\delta_0(\varepsilon)$; and a unified LASSO bound with equivariance reduction $|S| \to |S/G|$.

- *Downstream ML guarantees.* Local feature map $\varphi_{I,\delta}$ with bias control $\leq \varepsilon_1 \|O_I\|$; sample complexity $\tilde{O}\big(N(\varepsilon_1)\varepsilon_3^{-2} \log |S|\big)$; constant-in-$n$ behavior on translation invariant rings with parameter tying.

- *Empirical validation.* Purification-TEBD pipeline confirms the theory across temperatures and boundary conditions: open-boundary Heisenberg $\sim \log n$; periodic/equivariant Heisenberg nearly constant-in-n; Ising with α=3 $\sim \log n$; α=1.5 roughly linear.

6.2 Future Work

F1. Complete the $\alpha = 1.5$ study across all β.

Goal. Systematically map $N_{\mathrm{req}}(n)$ for $\alpha = 1.5$ over the full temperature grid $\beta \in \{1, 3, 5, 7, 10\}$ (and, time permitting, a finer grid), extending the indicative results in Fig. 9.

Protocol.

- Fix observables and training target ε as in Sec. 5.2-5.3.

- For each β, sweep n and fit $N_{\mathrm{req}}(n)$ to $an + b + c/n$.

Rationale. For $\alpha \leq 2D$ the LR analysis implies only logarithmic-cone control (no uniform exponential clustering), so truncation cannot be made small uniformly in r and sample complexity can grow polynomially with n (Sec. 3.3.1, Long-range). Hopefully, this experiment can quantify that polynomial behavior and its dependence on β.

F2. β–sweeps at different locality radii $\delta \in \{2, 6\}$.

Goal. Make the $\beta \mapsto \delta^*(\beta)$ tradeoff more exact by repeating the Heisenberg and $\alpha = 3$ Ising experiments with fixed radii $\delta = 2$ and $\delta = 6$, tuning the ML hyperparameters for each (β, δ), and plotting test error and N_{req} vs. β.

Hypothesis. The theoretical prediction $\delta_\beta(\varepsilon) = \frac{1}{\mu_\beta}\left(\log \frac{K_\beta}{\varepsilon} + c_D\right)$ (Cor. 3.19) says that as β increases (lower T), μ_β may decreases because of the β in the denominator, so the optimal radius grows. Therefore, we expect $\delta = 2$ to under-truncate at large β (increased bias), while $\delta = 6$ may be over-conservative at small β (risking overfitting and increased variance). Mapping these cases will empirically verify the manipulation of μ_β derived from QBP+LR (Prop. 3.14, Lem. 3.17, Cor. 3.19).

Deliverables. For each model: (i) curves of error and N_{req} vs β at $\delta = 2, 6$; (ii) an empirical $\hat{\delta}^*(\beta)$ from cross-validated performance.

F3. Sampling β from a distribution.

Goal. Treat temperature as a random input and study multi-temperature generalization: How does training with $\beta \sim P(\beta)$ affect N_{req} and error at held-out β values?

Protocol.

- Use the same locality radius δ and feature map as in Sec. 5.2, keeping β as an explicit feature.
- Consider $P(\beta)$ choice at first from a uniform on $[1, 10]$. We can experiment with other distributions as well.
- Compare to single β baselines and evaluate worst-case and average test error across β.

F4. Local neural network (GS & finite T) per [31].

Goal. Reproduce the constant-in-n ground-state result of [31], then apply the same local network to finite temperature. Compare against our LASSO/RFF baseline.

Model (basic, local MLP). For each symmetry orbit $[I]$ of local blocks, use a small two-layer tanh MLP f_θ shared across the orbit and sum the outputs:

$$\text{GS:} \quad \hat{f}(x) = \sum_{[I]} w_{[I]}\, f_\theta\big(Z_I^{(\delta)}(x)\big), \qquad \text{Thermal:} \quad \hat{f}(x, \beta) = \sum_{[I]} w_{[I]}\, f_\theta\big(Z_I^{(\delta)}(x),\, \beta\big).$$

Train with squared error.

Protocol.

(i) *GS replication.* Heisenberg benchmark. Sompare Sobol vs. uniform sampling. Report train/test RMSE and simple norms (last-layer $\|w\|_1$, max weight).

(ii) *Finite-T extension.* Use the theory based radius $\delta = \delta_\beta(\varepsilon)$ (Cor. 3.19). Measure $N_{\text{req}}(n)$ and RMSE for Heisenberg (short range) under periodic (equivariant) boundaries.

Outputs. Curves of N vs. RMSE and $N_{\text{req}}(n)$ fits (log vs. poly), gains from tying ($|S| \to |S/G|$).

References

[1] Anurag Anshu, Srinivasan Arunachalam, Tomotaka Kuwahara, and Mehdi Soleimanifar. Sample-efficient learning of interacting quantum systems. *Nature Physics*, 17(8):931–935, May 2021.

[2] Sven Bachmann, Spyridon Michalakis, Bruno Nachtergaele, and Robert Sims. Automorphic equivalence within gapped phases of quantum lattice systems. *Communications in Mathematical Physics*, 309(3):835–871, November 2011.

[3] Fernando G. S. L. Brandao and Michael J. Kastoryano. Finite correlation length implies efficient preparation of quantum thermal states, 2019.

[4] Jacob C Bridgeman and Christopher T Chubb. Hand-waving and interpretive dance: an introductory course on tensor networks. *Journal of Physics A: Mathematical and Theoretical*, 50(22):223001, May 2017.

[5] Ángela Capel, Massimo Moscolari, Stefan Teufel, and Tom Wessel. From decay of correlations to locality and stability of the gibbs state. *Communications in Mathematical Physics*, 406(2), January 2025.

[6] Alexander M. Dalzell, Sam McArdle, Mario Berta, Przemyslaw Bienias, Chi-Fang Chen, András Gilyén, Connor T. Hann, Michael J. Kastoryano, Emil T. Khabiboulline, Aleksander Kubica, Grant Salton, Samson Wang, and Fernando G. S. L. Brandão. *Quantum Algorithms: A Survey of Applications and End-to-end Complexities*. Cambridge University Press, April 2025.

[7] Nicolò Defenu, Tobias Donner, Tommaso Macrì, Guido Pagano, Stefano Ruffo, and Andrea Trombettoni. Long-range interacting quantum systems. *Reviews of Modern Physics*, 95(3), August 2023.

[8] Omar Fawzi and Renato Renner. Quantum conditional mutual information and approximate markov chains. *Communications in Mathematical Physics*, 340(2):575–611, September 2015.

[9] Michael Foss-Feig, Zhe-Xuan Gong, Charles W. Clark, and Alexey V. Gorshkov. Nearly linear light cones in long-range interacting quantum systems. *Physical Review Letters*, 114(15), April 2015.

[10] M. B. Hastings. Locality in quantum systems, 2010.

[11] M. B. Hastings and Xiao-Gang Wen. Quasiadiabatic continuation of quantum states: The stability of topological ground-state degeneracy and emergent gauge invariance. *Physical Review B*, 72(4), July 2005.

[12] Matthew B. Hastings and Tohru Koma. Spectral gap and exponential decay of correlations. *Communications in Mathematical Physics*, 265(3):781–804, April 2006.

[13] Johannes Hauschild, Jakob Unfried, Sajant Anand, Bartholomew Andrews, Marcus Bintz, Umberto Borla, Stefan Divic, Markus Drescher, Jan Geiger, Martin Hefel, Kévin Hémery, Wilhelm Kadow, Jack Kemp, Nico Kirchner, Vincent S. Liu, Gunnar Möller, Daniel Parker, Michael Rader, Anton Romen, Samuel Scalet, Leon Schoonderwoerd,

Maximilian Schulz, Tomohiro Soejima, Philipp Thoma, Yantao Wu, Philip Zechmann, Ludwig Zweng, Roger S. K. Mong, Michael P. Zaletel, and Frank Pollmann. Tensor network Python (TeNPy) version 1. *SciPost Phys. Codebases*, page 41, 2024.

[14] 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), September 2022.

[15] Julia Kempe, Alexei Kitaev, and Oded Regev. The complexity of the local hamiltonian problem, 2004.

[16] M. Kliesch, C. Gogolin, M.J. Kastoryano, A. Riera, and J. Eisert. Locality of temperature. *Physical Review X*, 4(3), July 2014.

[17] Tomotaka Kuwahara and Keiji Saito. Strictly linear light cones in long-range interacting systems of arbitrary dimensions. *Physical Review X*, 10(3), July 2020.

[18] Laura Lewis, Hsin-Yuan Huang, Viet T. Tran, Sebastian Lehner, Richard Kueng, and John Preskill. Improved machine learning algorithm for predicting ground state properties. *Nature Communications*, 15(1), January 2024.

[19] Elliott H. Lieb and Derek W. Robinson. The finite group velocity of quantum spin systems. *Communications in Mathematical Physics*, 28(3):251–257, September 1972.

[20] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, 2 edition, 2018.

[21] Bruno Nachtergaele and Robert Sims. Lieb-robinson bounds in quantum many-body physics, 2010.

[22] Román Orús. A practical introduction to tensor networks: Matrix product states and projected entangled pair states. *Annals of Physics*, 349:117–158, October 2014.

[23] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Andreas Müller, Joel Nothman, Gilles Louppe, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in python, 2018.

[24] Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems*, volume 20. Curran Associates, Inc., 2007.

[25] Cambyse Rouzé, Daniel Stilck França, Emilio Onorati, and James D. Watson. Efficient learning of ground and thermal states within phases of matter. *Nature Communications*, 15(1):7755, September 2024.

[26] Ulrich Schollwöck. The density-matrix renormalization group in the age of matrix product states. *Annals of Physics*, 326(1):96–192, January 2011.

[27] Štěpán Šmíd and Roberto Bondesan. Efficient Learning of Long-Range and Equivariant Quantum Systems. *Quantum*, 9:1597, January 2025.

[28] Minh C. Tran, Andrew Y. Guo, Christopher L. Baldwin, Adam Ehrenberg, Alexey V. Gorshkov, and Andrew Lucas. Lieb-robinson light cone for power-law interactions. *Physical Review Letters*, 127(16), October 2021.

[29] F. Verstraete and J. I. Cirac. Renormalization algorithms for quantum-many body systems in two and higher dimensions, 2004.

[30] Guifré Vidal. Efficient simulation of one-dimensional quantum many-body systems. *Physical Review Letters*, 93(4), July 2004.

[31] Marc Wanner, Laura Lewis, Chiranjib Bhattacharyya, Devdatt Dubhashi, and Alexandru Gheorghiu. Predicting ground state properties: Constant sample complexity and deep learning algorithms, 2024.

[32] Steven R. White. Density matrix formulation for quantum renormalization groups. *Phys. Rev. Lett.*, 69:2863–2866, Nov 1992.